Twin Research for Everyone

From Biology to Health, Epigenetics, and Psychology

Edited by

Adam D. Tarnoki
Medical Imaging Centre, Semmelweis University, Hungarian Twin Registry, Budapest, Hungary

David L. Tarnoki
Medical Imaging Centre, Semmelweis University, Hungarian Twin Registry, Budapest, Hungary

Jennifer R. Harris
Centre for Fertility and Health, The Norwegian Institute of Public Health, Oslo, Norway

Nancy L. Segal
Department of Psychology, California State University, Fullerton, CA, United States

ACADEMIC PRESS
An imprint of Elsevier
elsevier.com/books-and-journals

CHAPTER

Twins and omics: the role of twin studies in multi-omics

32

Fiona A. Hagenbeek[a,b], Jenny van Dongen[a,b,c], René Pool[a,b], Dorret I. Boomsma[a,b,c]

[a]*Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

[b]*Amsterdam Public Health Research Institute, Amsterdam, The Netherlands*

[c]*Amsterdam Reproduction and Development Research Institute, Amsterdam, The Netherlands*

32.1 Introduction

The "-omics" suffix denotes a discipline in biology, while the related suffix "-ome" signifies the object of study in this field.[1] Genomics, transcriptomics, proteomics, and metabolomics, referring to the study of the genome (DNA), transcriptome (RNA), proteome (proteins), and small molecules involved in metabolism, respectively,[2] cover the core molecules in the central dogma of biology.[3] The central dogma of biology describes how proteins are formed by the transcription and translation of genetic information (genomics → transcriptomics → proteomics).[3] Metabolomics, the study of the metabolites, that is, all small-molecules in an organism,[2] and the central dogma together describe the omics cascade from genes to metabolites (Fig. 32.1).[4] In addition to the linear, unidirectional oriented connections in the omics cascade, more complex relationships exist between and within the different omics layers, including feedback loops among omics levels.[3] Increasingly, other omics layers, such as the epigenome, microbiome, glycome, phosphoproteome, lipidome, fluxome, or exposome, are added to the omics cascade.[5] Many of these, such as the glycome or phosphoproteome, reflect regulatory and modulatory processes,[5] others, such as the exposome, reflect exposures to the environment.[6]

Large-scale omics studies are often carried out in cohorts of unrelated individuals. This is, in part, because many statistical models originally designed to study omics data rely on standard techniques for association and regression. In the field of genomics, particularly for genome-wide association (GWA) studies, it was quickly recognized that leveraging the information contained within the many twin registries around the world would result in many advantages, if we properly account for the clustering of observations.[7] This recognition spurred efforts to apply approaches, such as mixed models and generalized estimating equations, to account for relatedness among participants in twin and family studies.[8] Approaches that allow for the inclusion of related individuals led to the inclusion of large numbers of samples of

Twin Research for Everyone. DOI: https://doi.org/10.1016/B978-0-12-821514-2.00029-5

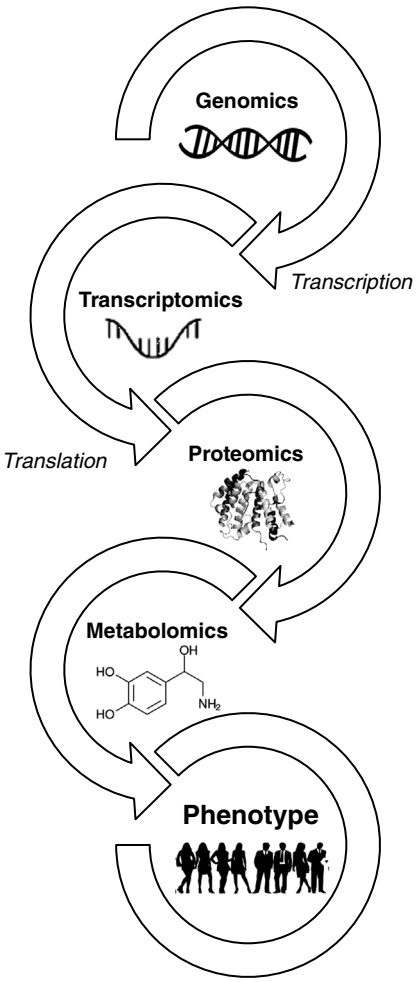

FIG. 32.1 The omics cascade—the omics cascade describes the cascade from genotype to phenotype.

well-phenotyped participants from twin registries in e.g., GWA studies of migraine, major depression, educational attainment.[9–11] However, twin designs themselves are powerful analytical tools for omics data beyond contributing to association studies.[12] In this chapter, we will first introduce some often studied omics domains: genomics, epigenomics, transcriptomics, and metabolomics. Next, for each of these domains, we outline the contributions made by twin studies and consider the added value of twin research in omics. We illustrate some designs such as the discordant twin design, in some detail and consider a combination of the classical twin design with genome-wide genotype data.

32.2 Genomics

32.2.1 What is genomics and how do we measure the genome?

Deoxyribonucleic acid (DNA) polymer molecules contain the hereditary information of the organism. DNA consists of two polynucleotide chains ("strands") that form a double-helical structure that is stabilized by hydrogen bonding between the nucleotides of both strands. These hydrogen bonds are formed between complementary nucleotides. There are four nucleotide types, where adenine (A) pairs with thymine (T) and guanine (G) pairs with cytosine (C). Segments of DNA contain genes, that consist of a few hundred to more than two million base pairs.[13] Genes consist of multiple long noncoding regions called introns and shorter coding regions called exons.[14] By coding, we mean coding for a function in the next omics layer(s). Originally it was believed that all genes contain the instructions to encode proteins, however, we now know that many genes are not protein coding. Almost all DNA molecules are contained in the nucleus of each cell. The cell nucleus is approximately 5–8 μm in diameter. By contrast, unfolded human DNA is approximately 2 m in length. To fit DNA in the cell nucleus, DNA is packed into highly condensed structures called chromosomes, each of which comes in two copies (one inherited from each parent). Humans have 23 pairs of chromosomes: 22 autosomal chromosome pairs and a sex chromosome pair.[13]

With genome we refer to the complete set of hereditary information, where the word "genome" is a conjunction "gene" and "chromosome." Therefore, genomics has been coined to refer to the study of the structure, function, and mapping of genomes.[15] In this chapter, we focus on genomic studies characterizing the DNA sequence variants between individuals. We can distinguish various types of sequence variants, spanning from a single nucleotide to dozens of base pairs and even entire chromosomes. The single nucleotide variants (SNVs), also called point mutations, are variations (substitutions, insertions, or deletions) in a single base pair. When these occur in more than 1% of individuals we refer to them as single nucleotide polymorphisms (SNPs). Small insertions or deletions that affect several[2–50] base pairs are called indels and substitutions of several base pairs are called block substitutions.[16] Differences in copy number (deletions, insertions, duplications), orientation (inversions; i.e. stretches of flipped DNA sequence), or location (translocations, i.e., stretches of DNA that have migrated within the genome) between individuals that span more than 50 base pairs are called structural variants (SVs).[17] The largest SVs can affect whole chromosomes, as such, they are also referred to as chromosomal aberrations.

To characterize nucleotide sequence variations in DNA, two techniques are commonly used: DNA microarrays and sequencing. Microarrays typically measure up to 1 million SNPs, while whole-genome sequencing yields nearly 100% of the (structural) information of the genome. Due to the correlation structure of the DNA sequence, the genomic information in DNA microarray data often suffices when studying the relation between the genome and biological (dys)function. DNA microarrays use a technology comprised of a collection of single-stranded oligonucleotide

probes covalently linked to a flat surface, often times on a medium analogous to a microscope slide. For these probes, their locations in the genome are known. Synthetic oligonucleotide probes interact with highly specific genomic sequences via complementary base pairing (hydrogen bonding between the probe and target DNA sequences) in a process termed hybridization. The probes are typically designed to hybridize to the target sequence immediately upstream of the polymorphic nucleotide. Following hybridization, fluorescently labeled nucleotides are utilized in an extension or ligation reaction to discriminate between the different alleles known to occur at that locus and are subsequently imaged utilizing a laser-powered scanner. After the raw intensity data for samples processed on the DNA microarray are generated; next steps involve genotype calling, quality control of genotypes, including tests of Hardy–Weinberg (HE) equilibrium of alleles, of Mendelian transmission (in family data), and comparison of allele frequencies to reference sets.[24] DNA microarrays can be designed to target SNPs, either in small numbers for dedicated purposes such as arrays targeting (rare) exonic variants,[20] or SNPs of interest for particular traits[21] or contain genome-wide common genetic variants, such as present on the global screening array or the Axiom UK Biobank Array.[22,23]

DNA sequencing technologies allow for the measurement of most variants in the genome. DNA sequencing was first developed by Sanger in 1975, and this technique is now referred to as Sanger sequencing. Sanger sequencing has high accuracy, low throughput (it only produces a single DNA fragment at a time), the maximum sequence length is 1000 base pairs, is relatively expensive, and is not suitable for large-scale sequencing projects.[25] Because of its high accuracy, it is often used as a follow-up of findings that result from other sequencing techniques. Several technological advances have contributed to the development of high-throughput sequencing. One advance was the development of polymerase chain reaction (PCR), which allows for massive amplification of small DNA samples, a development that improved the scalability of sequencing as this could be applied in multiwell plates.[26] These and other developments led to the newer next-generation sequencing (NGS) techniques. A complete overview of all types of NGS techniques is outside the scope of this chapter, the reader is referred to, for example, a review by Goodwin, McPherson, and McCombie.[27]

32.2.2 Sequence differences between monozygotic twins

Because monozygotic (MZ) twins arise from one fertilized oocyte they are taken to be genetically identical; a key assumption in the classic twin design.[28] While MZ twins are genetically identical at conception, somatic mutations can arise during cell division (mitosis).[29] Such somatic mutations cause differences in the DNA sequence across different cells of the body. An individual with different populations of cells with different DNA sequences originating from the same zygote is called a mosaic[29] and mosaic mutations can differ between MZ twins from the same pair. Mutations can also arise in germ cells (germline mutations), and be transmitted to the offspring resulting in a constitutional mutation that is present in all cells.[30] Germline mutations,

or pretwinning *de novo* mutations, are therefore shared between MZ twins, but not between the twins and their parents. By contrast, somatic mutations, or post-twinning *de novo* mutations, are present in only one MZ twin or even only in some of the cells of one twin (mosaicism). The genetic (dis)similarity of MZ twins, therefore, depends on the moment in life at which mutations occur. Multiple genetically different cell lineages within one person can also originate from different zygotes. This is referred to as chimerism and can arise for example if dizygotic (DZ) twin zygotes merge early in development.[29] In contrast to gross chimerism, which is present in the majority of the cells in the total population of cells of a particular cell type, microchimerism is present in less than 1% of the total cell population. It occurs frequently, for example as a result of the passage of blood between mother and child during pregnancy, with twin chimerism as a special case and can be a source of discordance in MZ twin pairs.[31]

DNA sequencing studies suggest that the *de novo* SNV mutation rate in somatic cells is approximately 0.82×10^{-8} to 1.70×10^{-8} mutations per base per generation.[32–34] Study designs utilizing MZ twins allow for distinguishing between prezygotic (present in both twins of a pair) and postzygotic (present in only one twin of a pair) *de novo* mutations and to estimate the postzygotic mutation rate. A whole-genome sequencing study of a healthy MZ twin pair and their parents obtained a rate of 0.97×10^{-8} per base per generation for *de novo* SNVs shared by the twin pair. For twin-specific *de novo* SNVs, rate of 0.34×10^{-8} base pair per generation was calculated for one twin and 0.04×10^{-8} base pair per generation for the other twin,[34] that is, an overall *de novo* SNV rate of 1.01 and 1.31×10^{-8}. A comparison of whole-genome DNA sequence data of two monozygotic twin pairs, 40 and 100 years old, was carried out to detect somatic mosaicism and identified 1720 putative postzygotic mutations in blood cells from the 40-year-old MZ twin pair and 1739 in the 100-year-old pair.[35] The identified postzygotic mutations were nonrandomly distributed across the genome, with enrichment for regulatory elements such as coding exons or genes involved in GTPase activity.

Discordances in MZ twin pairs have also been reported for chromosomal abnormalities, particularly for aneuploidy, where one or more chromosomes are missing or present in an extra copy,[29] such as monosomy X (missing sex chromosome; e.g., Turner Syndrome), or trisomy 21 (gain of extra chromosome 21, e.g., Down Syndrome).[36] Postzygotic *de novo* CNVs have been observed in for example a sample of 1097 unselected MZ twin pairs. One hundred fifty-three putative *de novo* CNVs were detected in peripheral blood and buccal epithelium cells, of which 58.8% were located in the same 15q11.2 region.[37] Replication of 20 candidate CNVs with qPCR validated two CNVs in the same 13-year-old MZ twin pair. The twins had no large phenotypic discordances. The twin with three copies of both CNVs outperformed its cotwin (with 1 and 2 copies, respectively, for each of the CNVs) on school achievement. This study also compared CNVs derived from peripheral blood or buccal epithelium cells in the complete group of 1097 MZ twin pairs. While more CNVs were found in DNA from blood, buccal epithelium DNA CNVs had higher concordance rates per twin pair.

As *de novo* postzygotic mutations may arise at each cell division, it is believed that somatic mutations accumulate with age and that aging might even be a consequence of the accelerated accumulation of somatic mutations.[38] A study of twins and singletons investigated CNV accumulation with age.[39] In a healthy cohort of 159 MZ twin pairs and 296 singletons, CNVs were compared in a younger (≤55) and older (≥60) age group. In contrast to the younger group, where no large CNVs were detected, 3.4% of subjects in the older age group had large CNVs, indicating a relationship between age and CNV occurrence in peripheral blood DNA. In addition, for 18 MZ twin pairs (50.7–72.6 years of age at baseline), data on small CNVs were available longitudinally, measured ten years apart. The longitudinal data showed that both increases and decreases in the number of CNVs can be observed. Thus, CNVs appear to accumulate with age, but the populations of peripheral blood cells with CNVs are not stable.

The discordant MZ design also is a tool to identify trait- or disorder-associated genetic variants. An early study of whole-genome sequencing in MZ discordant twins was published in 2010.[40] In addition to whole-genome DNA sequencing, this study also included data on mRNA sequencing, genome-wide SNP microarrays, and DNA methylation profiles with the objective to identify genetic, transcriptomic, and epigenetic differences between $CD4^+$ T cells of three pairs of MZ twins discordant for multiple sclerosis (MS). Differences in SNPs, indels, CNVs, viral genome sequences, gene expression levels and CpG methylation levels could not be reproducibly detected in $CD4^+$ T cells to explain MS discordance. While this early study on MS showed no clear differences within the MZ discordant pairs, this design has been applied with clearer results for CNVs. For example, comparison of CNVs in peripheral blood in a sample of 19 adult MZ twin pairs, of which 9 pairs were discordant for neurodegenerative disorders and 10 pairs were phenotypically unselected, found a larger number of CNVs in the disease discordant than in the other MZ twin pairs.[41] While some of the CNVs reported in the discordant MZ twins might be pathogenic for the neurodegenerative disorders, the authors stressed that replication in larger samples across multiple (relevant) tissues is necessary. As the last example, a study investigating the contribution of the number and the size of CNVs in attention problems identified 8 pre- and 18 post-twinning CNVs in 50 MZ twin pairs. In this group, for 25 MZ pairs both parents were genotyped so that pretwinning *de novo* CNV events could be detected.[42] Of the three possible pretwinning *de novo* CNVs that were included in a qPCR replication study, one pretwinning *de novo* CNV mutation was confirmed, where both MZ twins had a duplication on chromosome 15q11.2. This region contains the gene *HERC2P3*, which is expressed in the human brain. However, both twins scored in the normal range for attention problems.

32.2.3 Sequence differences between dizygotic twins

Classical twin models assume that MZ twins are genetically identical and that dizygotic (DZ) twin pairs and full siblings share on average 50% of their DNA sequence.[28] This last assumption can be tested empirically by estimating the amount

of genetic material that DZ twins or full siblings have inherited identical-by-descent (IBD). DNA segments are IBD if they are inherited from a common ancestor without recombination. This is in contrast to identity-by-state sharing, where DNA segments are identical between pairs of individuals, but do not need to derive from a common ancestor.[43] Genome-wide microsatellite markers data and SNP data indicated that the proportion of IBD sharing between DZ twins and full siblings ranges from 42% to 58%, and confirmed that the average is indeed close to 50%.[44,45]

32.3 Epigenomics

32.3.1 What is epigenomics and how do we measure the epigenome?

With the exception of *de novo* somatic mutations, all cells in the body have the same DNA sequence (except for red blood cells that do not contain DNA), and differences between cell functions are mainly due to differences in which parts of the DNA sequence are expressed in different cells. Gene expression also is modified in response to developmental and environmental cues[46] and is under tight control through multiple regulating mechanisms.[47] Gene expression occurs in regions of the DNA where the chromatin permits transcription.[48] Chromatin is the macromolecular complex that is responsible for condensing DNA into smaller packages of chromosomes and is built up of nucleosomes; a segment of DNA wound around eight histone proteins.[13] Approximately 99% of a cell's genome is located in so-called heterochromatin, a highly compact state where the DNA is not accessible for transcription.[48] At present, 15 distinct chromatin states have been characterized.[49]

Epigenomics is the comprehensive study of the mechanisms that control gene expression by influencing the accessibility of the genome for transcription and/or the ability of the transcription machinery to adhere to accessible DNA segments.[48] Multiple systems cooperate in epigenetic control: DNA methylation (addition of a methyl group to DNA), histone modification (e.g., methylation or acetylation of histone proteins), nucleosome remodeling (change the position of the DNA wrapped around the nucleosomes), and noncoding RNAs (ncRNAs; which are functional RNA molecules that are transcribed from DNA but not translated into proteins and which can influence DNA methylation and histone modifications).[46] Here our focus is mainly on DNA methylation, which is the best-studied epigenomic mechanism in human studies including twin studies and is currently the only one that is suited for assessment in large-scale human epidemiological studies. The relationship between DNA methylation and transcription depends on the genomic context: whereas DNA methylation at gene promoters is usually associated with transcriptional repression, gene body methylation is a feature of actively transcribed genes. Methylation occurs at the C5 position of the aromatic rings of cytosines (5-methylcytosine). This can occur at any cytosine, but in humans, DNA methylation happens almost exclusively at regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide (CpGs). CpG sites tend to cluster in so-called CpG-islands, regions of at least 200 base pairs consisting of 55% or more CG sites.[50]

Several methods for the analysis of epigenomics are available, of which microarrays and sequencing are the main ones. The most frequently used technologies make use of a bisulfite treatment step of the DNA. Unmethylated cytosines are converted to uracil by sodium bisulfite treatment.[51] In PCR amplification uracil is recognized as thymine, as methylated cytosines are immune to the bisulfate conversion they remain cytosines, therefore methylated cytosines can be distinguished from unmethylated cytosines.[52] The bisulfite-treated DNA is then introduced to a methylation microarray which typically includes several hundreds of thousands of probes. The most commonly used Illumina microarrays return, for each interrogated site, the methylation level (proportion of methylated alleles).[53,54] In DNA that is derived from a mixture of cells, such as found in whole blood, the methylation level represents a continuous variable with values that may range between zero and one. For example, a methylation level of 1 means that all DNA strands had a methyl group attached at this position and a value of 0.5 that 50% of all DNA strands had a methyl group attached at this position. Intermediate values arise when a position is methylated in a fraction of cells or on one of the two chromosomes.

32.3.2 Causes of epigenetic variation

The epigenome is often discussed in the context of environmental explanations for diseases, but the epigenome is also shaped by genetic influences. In fact, the epigenome may be a key mediator of the effects of common genetic variants on complex traits and disease, because these variants usually reside in regulatory regions (rather than protein-coding regions) of the genome.[55] Disease-associated SNPs are often associated with expression levels of transcription factors, which in turn drive variation in the DNA methylation level of distal binding sites.[56] Large-scale methylation Quantitative Trait Loci (mQTL) analyses can map associations between genetic variants (typically, SNPs) and DNA methylation levels across the genome.[57] As MZ twins share their genomes, such mQTLs contribute to their epigenetic similarity. In 49 MZ twin pairs from the Netherlands Twin Register,[58] DNA methylation was measured at ~850.000 sites in the genome with the Illumina EPIC array in buccal samples, which consist for about 80% of epithelial cells and about 20% of white blood cells. After adjusting for cellular composition, the methylation levels of MZ twins were more similar at CpG sites whose methylation level was strongly influenced by SNPs than at CpG sites for which no significant mQTLs were detected.

DNA methylation profiles can be seen as complex traits, or phenotypes, and differences between individuals may be analyzed by the classical twin design to estimate heritability. Data from a large cohort of twins and family members from the Netherlands Twin Register were analyzed to estimate the overall heritability for DNA methylation levels at multiple sites. As the participants were genotyped, the variance explained by genome-wide SNPs could also be estimated. In follow-up analyses, interactions of genetic and environmental influences with age and sex were examined.[59] All results are described in a catalog (http://bbmri.researchlumc.nl/atlas/). In 2603,

genotyped adult individuals (mean age 37.2, sd = 13.3, 66% females), DNA methylation was measured at ~450.000 sites in the genome with the Illumina 450 k array. Based on the twin data, the total heritability was 19% on average across the genome. On average 7% (s.d. = 12%) of the variance of DNA methylation was explained by common genetic variants in the genome (h^2_{SNPs}). Thus, the proportion of the total heritability that can be explained by SNPs, i.e., h^2_{SNPs}/h^2 was 0.37 (s.d. = 0.40).

Epigenetic differences between MZ twins are observed in tissues collected at birth,[60] but may also emerge postnatally: results from both cross-sectional studies and longitudinal studies of adult twins suggest that the epigenomes of MZ twins diverge as they age.[59,61,62] This means that the differences between individuals in a population become larger as a function of age; older individuals show more variation in DNA methylation level at these loci. With data from MZ and DZ twins, the causes of age-related changes in variance (genetic and environmental) can be examined by adding a moderator variable to the classical twin model[63] to test the interaction between age and the genetic and environmental effects. Such models have found that age-interaction effects were widespread: 10.4% of all measured sites showed a significant interaction effect of age and genetic or environmental effects on DNA methylation level.[59] At 82% of sites, the unique environmental variance changed with age. These sites typically showed an increase in the unique environmental variance and total variance with age, and a decrease of the heritability. At 90% of sites with significant age interaction, the heritability was lower at age 50 than at age 25, although the difference in heritability between younger and older people was usually modest.

The average heritability of DNA methylation in blood is almost the same in males (mean $h^2 = 0.199$) and females (mean $h^2 = 0.198$), but a small percentage (0.7% of all measured sites) showed a significant interaction effect of sex and genetic or environmental effects on DNA methylation level. At 59% of these sites, the heritability was lower in women. At 76% of all sites with significant sex interaction, the unique environmental variance (rather than the additive genetic variance) differed between the sexes. At sites with a lower heritability in females, the variance of DNA methylation due to environmental influences was usually larger in females. Such methylation sites with sex-specific variation in epigenetic regulation can be studied in future epigenetic studies of diseases with a sex-specific etiology.

32.3.3 MZ discordant design applied to epigenomics studies

Differences in DNA methylation and histone modifications within MZ twin pairs have been reported for multiple tissues and cell types, including blood cells, buccal cells, and fat.[64] The distinct methylomes of MZ twins are even studied by forensic scientists to develop tools to distinguish MZ twins in forensic settings.[65–70] Epigenetic differences between MZ twins can arise from stochastic (random) events, different environmental exposures of cotwins, and genetic mutations. Here, we highlight a few studies investigating epigenetic differences in (MZ) twins, for more detail, we refer the reader to review articles on this topic.[64,71,72]

Stochastic variation can result from the imperfect molecular control of gene expression. For example, the maintenance of DNA methylation in dividing cells by DNA methyltransferases (DNA MTase, DNMT) enzymes is not 100% accurate. Differences in exposures and lifestyle, such as smoking behavior impact of on the epigenome of circulating cells. MZ twin pairs who are discordant for smoking show DNA methylation differences at several loci in white blood cells.[73] This study of 20 MZ pairs of which one twin smoked regularly and the cotwin never smoked or had stopped smoking more than 10 years ago confirmed several loci identified previously in epigenome-wide association studies (EWAS) that compared unrelated smokers to nonsmokers. Note that a key strength of the MZ twin design is that many alternative explanations are ruled out, because MZ twins are genetically identical. For example, one of the most strongly associated genetic variants for nicotine dependence is located in the DNA methyltransferase gene DNMT3B,[74] which might lead to differences in genome-wide DNA methylation between people with different genotypes at this gene, regardless of their smoking behavior. This would be an example of a pleiotropic genetic effect, where a genotype influences genome-wide DNA methylation as well as smoking behavior. Because MZ twins carry the same genetic predisposition for nicotine dependence, potential pleiotropic effects of genetic variants that influence multiple traits independently are not an issue in MZ twin studies.

Epigenetic differences between MZ twins may cause different usage of the identical DNA code. This can lead to extreme phenotypic differences,[36] as illustrated by the study of one MZ pair, in which one twin had a severe congenital caudal duplication malformation and the other did not.[75] There was a very strong candidate gene for the disorder, for which no DNA sequence differences were found. However, this gene showed strong epigenetic differences between the two girls.

Not all epigenetic differences that are observed in monozygotic twin pairs lead to phenotypic discordance. If the two twins are measured on, for example, different days, on different arrays, technical variation can lead to dissimilarity. Differences between twins in the cellular composition of blood samples can also contribute to differences in DNA methylation between MZ twins. Epigenetic differences can of course arise as a result of a disease of one twin, can represent a marker of a disease-causing event, or can be caused by medication use of the affected twin. This opens up possibilities for the identification of dynamic epigenetic biomarkers (those that indicate the emergence and progression of a disease or that indicate current exposure to a risk factor) and persistent epigenetic biomarkers of environmental exposures in the past.[76] A study of 45 MZ twin pairs discordant for MS measured genome-wide DNA methylation in peripheral blood mononuclear cells and identified disease-associated methylation sites, loci where differences between twins in methylation level reflect whether a person is currently receiving interferon-beta treatment, and a locus whose methylation level reflected prior glucocorticoid treatment.[77] Epigenomic studies in MZ twins can have more power than traditional case-control EWA studies[78] and can contribute to our understanding of the underlying pathways and consequences of disease and to the identification of biomarkers.

32.4 Transcriptomics

32.4.1 What is transcriptomics and how do we measure the transcriptome?

The mechanism by which cells copy DNA information into ribonucleic acid (RNA) is called transcription. In contrast to DNA, RNA is single-stranded and it contains uracil (U) bases instead of the thymine (T) bases found in DNA.[14] During transcription one of the two DNA strands acts as a template for RNA synthesis. The sequence of the RNA is synthesized complementary to the nucleotides of the antisense DNA strand and is therefore a copy of the sense strand (with exception for the substitution of U for T). The entire length of a gene, both introns and exons, is transcribed. Next, RNA splicing removes the introns and combines the exons. Not all exons of a gene need be included in the final RNA transcript. Through alternative splicing, different combinations of exons allow for the production of different proteins from the same gene.[14] Such protein-encoding RNA transcripts are referred to as messenger RNA (mRNA). Other types of RNA include ribosomal (rRNA; which forms the core of ribosomes where mRNA is translated to proteins), transfer RNA (tRNA; which is involved in the process of translation of mRNA into proteins by connecting amino acids for incorporation into the protein), and microRNA (miRNA; which is involved in regulation of gene expression).[14,79] The analysis of the complete set of transcripts (RNAs) in a cell or the study of RNA or RNA variants is known as transcriptomics, often also referred to as gene expression studies. Similar to other omics, two techniques are common to study the transcriptome: microarrays and RNA sequencing (RNA-Seq).[19]

32.4.2 Causes of variation in gene expression levels

Like DNA methylation profiles, transcriptome profiles can be regarded as complex phenotypes, and differences between individuals may be analyzed by the classical twin design to decompose variation into genetic and nongenetic variance components. These analyses provide heritability estimates of gene expression which gives an indication of the extent to which the DNA sequence regulates its own expression. Below we give two illustrations of how twin studies shed light on the causes of variation in gene expression. Both studies derive from the Netherlands Twin Register. Wright et al. (2014) employed several methods to analyze variation in RNA microarray data. These included a classical twin design with MZ and DZ twin pairs, and a design with genotyped DZ twin pairs to obtain SNP-heritability estimates.[80] Gene expression of 18,392 genes was assessed in peripheral blood samples obtained from 2752 twins, including 690 complete MZ and 618 complete DZ twin pairs. Twin-based heritability across all RNA probes was 0.10 (sd = 0.14). To assess the contribution of heritability attributable to local genetic variation, SNPs were selected that were located 1 mega base upstream of a transcription start site and 1 mega base downstream of a transcription end site. Estimates of IBD sharing in DZ-twin pairs for these SNPs were used to estimate the ratio of $h^2_{\text{local IBD}}$ (that is, the variance in

gene expression level explained by all local genetic variants; both common and rare) to overall narrow-sense heritability of gene expression levels. The mean and median for the $h^2_{\text{local IBD}}/h^2$ ratio were 0.11 and 0.30, respectively, across all RNA probes. Second, local $r^2_{\text{Socal SNP}}$ (that is, the variance in gene expression level explained by most significant local SNP within 1 Mb) was estimated in unrelated participants using the GCTA software.[82] The ratio of $r^2_{\text{local SNP}}$ to h^2 had a mean = 0.04 and median = 0.09. These 2 sets of estimates are consistent with a higher explained variation from the total local contribution of a region.

The second study by Ouwens et al. (2020) focused on RNA sequence data and included a subsample of these same twin pairs. Classical twin and GRM- (Genetic Relatedness Matrices, based on SNP data) approaches were used to decompose transcriptome variation from RNA sequence data into genetic and nongenetic variance components.[81] Peripheral blood gene expression was obtained for 52844 genes in 1497 twins, including 459 complete MZ and 150 complete DZ twin pairs.[81] Heritability of gene expression profiles based the classic twin design, was 0.20 on average. The mean contribution of the shared environment was 0.05 and the mean contribution of the unique (unshared) environment was 0.75. Next, this total (twin-based) heritability was compared to the heritability which could be attributed to genome-wide SNP data. This was accomplished by creating two GRMs: one GRM containing all SNPs in a 250 kb window of a gene (referred to as *cis*), and one GRM including all autosomal SNPs for closely-related individuals in the dataset. Because of the large number of related individuals this last GRM captures genetic variance tagged by substantial IBD sharing, with the sum of the two effects being roughly equal to the total heritability, which contains the genetic variation in the *cis*-window a gene (h^2_{cis}) and the residual heritability (h^2_{res}). With this approach, an average total heritability of 0.26 was found, which correlated 0.98 with the h^2 estimate obtained from twin modeling. The mean *cis*-heritability ($h^2_{\text{cis;}}$ that is, the variance in gene expression level explained by local SNPs) was 0.06, and a mean residual heritability of 0.20.

Both these studies were conducted in peripheral blood samples; however, gene expression can be tissue specific.[83] For example, a study in 856 female twins (154 complete MZ and 232 complete DZ twin pairs) investigated the heritability of expressed transcripts and performed *cis*- and *trans*-eQTL analysis of adipose and skin tissue and lymphoblastoid cell lines (LCL).[84] Average heritability for these three tissue-types was 0.26 for adipose, 0.16 for skin, and 0.21 for LCL-based gene expression. The study also reported 3529, 2796, and 4625 adipose, skin and LCL *cis*-eQTLs, and 639, 609, and 557 adipose, skin and LCL *trans*-eQTLs, respectively.

Multivariate extensions of the classic twin design are valuable to characterize the genetic and environmental correlations between multiple outcome traits, for example, between expression levels of different genes, or between gene expression levels and complex traits or diseases. A significant genetic correlation between multiple outcome traits indicates that the observed phenotypic correlation between those traits is to a significant extent caused by overlapping genetic influences. An array-based transcriptome-wide analysis of blood pressure in peripheral leukocytes for 391 twins (193 complete same-sex pairs) identified that expression of the *MOK* gene was

significantly associated with systolic blood pressure and this finding was replicated in an independent population cohort.[85] Additionally, out of 40 genes whose expression levels were previously associated with blood pressure, this study replicated the effects of 12 genes. Heritability for the expression levels of these 12 genes ranged from 6% to 65%. Bivariate models estimated the contribution of genes and environment to the association of blood pressure and gene expression levels. The association of blood pressure with *CD97*, *TIPARP*, and *TPP3* expression levels was determined completely by shared genetic factors. By contrast, the association with *LMNA*, *SLC31A2*, *TSC22D3*, and *TAGLN2* expression levels was determined completely by the environment. The association of *CRIP1*, *F12*, *S100A10*, *TAGAP*, and *MOK* expression levels with blood pressure were determined by both genetic and environmental factors.

After the successes of GWA studies for complex traits and disorders, it became clear that common genetic variants often did not fully account for the heritability of these traits as observed in twin-family studies.[86,87] Gene finding for omics phenotypes have been very successful, but for these traits we also observe a gap between the variance explained by omics QTLs and twin-based heritability estimates. Several explanations for the "missing heritability" problem have been proposed.[88] Most omics QTL studies have focused on common SNPs, while it is likely that rare genetic variants also contribute to the heritability of omics phenotypes. Gene–gene (GxG, or epistasis) and gene–environment (GxE) interactions have also been listed as possible reasons for "missing heritability."[88] With twin designs, both GxG and GxE effects have been identified for gene expression levels.[89] For example, gene-by-body mass index (GxBMI) interactions on gene expression regulation were identified in a cohort of 856 female twin individuals with multitissue RNA sequencing data.[90] In adipose tissue, this study found 16 *cis* and 53 *trans* GxBMI interactions. However, recent findings now strongly suggest that the "still missing heritability" of complex phenotypes is accounted for by rare variants, in particular those in regions of the genome of low linkage disequilibrium.[91]

32.4.3 MZ discordant design applied to transcriptomics studies

The MZ discordant design has been frequently used to identify differentially expressed genes for various traits and disorders. Such differentially expressed genes may provide insight in the underlying biology of traits and disorders and could shed light on disease mechanisms. Below, we give two examples to illustrate the strength of the MZ discordant design for transcriptomics studies. Examples of other traits and diseases for which gene expression has been investigated in discordant MZ twin pairs include Type I Diabetes,[92,93] Rheumatoid Arthritis,[94] treatment of Childhood Primary Myelofibrosis,[95] hormone replacement therapy,[96] Parkinson's Disease,[97] Schizophrenia and Schizophrenia treatment,[98,99] Bipolar Disorder,[100] sleep duration,[101] and neurodevelopmental disorders due to trisomy's, such as Down Syndrome.[102,103]

Our first example concerns obesity. A study of mitochondrial DNA gene expression in subcutaneous fat and peripheral leukocytes in 14 obesity-discordant

MZ twin pairs detected upregulation of genes involved in inflammatory pathways and downregulation of genes in mitochondrial branched-chain amino acid catabolism in obese twins as compared to their lean cotwins.[104] Additional evidence that obesity is associated with dysregulation of cellular metabolism and mitochondrial function comes from a BMI-discordant MZ study on the role of sirtuin (SIRT) and NAD^+ biosynthesis gene expression pathways in obesity.[105] The NAD^+/SIRT pathway is involved in sensing energy levels within cells, with the SIRT proteins involved in, for example, mitochondrial oxidation, lipid oxidation, lipolysis, and adipogenesis. This study found that, compared to their leaner cotwins, heavier MZ twins had reduced expression of genes involved in mitochondrial unfolded protein responses and SIRT and NAD^+ biosynthesis and increased poly-ADP ribose polymerase (PARP) activity in subcutaneous adipose tissue (SAT). Transcriptomics studies in obesity-discordant MZ twins also identified obesity subtypes based on transcriptomic profiles and correlations with clinical characteristics. A study in 26 BMI-discordant MZ twin pairs revealed three distinct subgroups based on their molecular profiles and showed that for subgroup one the transcriptional differences between the heavy and leaner cotwins were benign, transcriptional differences between the MZ twins in subgroup two appeared to be characterized by downregulation of mitochondrial function in the heavy twins, and subgroup three showed a clear inflammation pattern in addition to the downregulated mitochondrial function in the heavy twins.[106]

The second example of the MZ discordant design involves multiple phenotypes. In order to identify differentially expressed genes for multiple phenotypes and integrate mean expression differences across phenotypes, Tangirala and Patel (2018) performed a meta-analysis of MZ discordant studies for seven phenotypes, based on studies from public repositories including ten or more MZ twin pairs.[107] These studies focused on ulcerative colitis, chronic fatigue syndrome, physical activity, intelligence quotient (IQ), intermittent allergic rhinitis, major depressive disorder (MDD), and obesity, with gene expression data measured in different tissues, including peripheral blood, lymphoblastoid cell lines, adipose tissue, muscle tissue, and colon tissue. For each of the seven phenotypes, differential gene expression analysis was performed and results were meta-analyzed per phenotype at the gene level. In total, 5% of the genes in the datasets were significantly differentially expressed between discordant MZ twins across all phenotypes. Little overlap in the differentially expressed genes was observed among the phenotypes, with an average overlap of 0.009%. Meta-analysis of each gene across the seven phenotypes identified no genes that were both overall significant and significant for the individual phenotypes. Differential gene expression for most genes was not heterogeneous across the multiple phenotypes. Overall, this study found a small common gene expression signature across the seven phenotypes, where 0.08% of the full list of differentially expressed genes (across all seven phenotypes) were in fact differentially expressed across all seven phenotypes in discordant MZ twins. The study concluded that the majority of differentially expressed genes are phenotype specific.

32.4.4 Other applications of twin research in transcriptomics studies

The discordant MZ design is often expanded to include discordant DZ twin pairs or case-control groups of unrelated individuals. Effects in this last group represent associations at the population level. A comparison between the unaffected MZ twins from discordant pairs with healthy unrelated controls provides information regarding whether these two groups have comparable transcription levels, or whether unaffected twins exhibit a disease-related profile that is more similar (although perhaps milder) to that of their affected cotwin. Gene expression studies in peripheral blood samples for systemic autoimmune diseases, such as rheumatoid arthritis or systemic lupus erythematosus, reported 92–537 differentially expressed genes between probands and unrelated matched controls.[108,109] They also reported that both human and viral gene expression levels of the unaffected twins were intermediate between the expression levels of their affected cotwin and the healthy unrelated controls. Therefore, they concluded that the unaffected MZ twins may be in a transitional or intermediate state of immune regulation.[108]

MZ twin pairs concordant for a disorder may still present discordant phenotypes with unique transcriptomics profiles. For example, miRNA expression of placenta samples in mono-chorionic twin pairs with ($N = 17$) and without ($N = 16$) selective intrauterine growth restriction (sIUGR) identified seven upregulated and seven downregulated miRNAs among the larger sIUGR twins as compared to their smaller cotwins.[110] This study showed that pathogenesis of sIUGR is associated with miRNA pathways involved in organ size, cell differentiation, cell proliferation, and cell migration. Longitudinal designs can be strengthened by inclusion of MZ twin pairs, as these designs are robust for changes in gene expression profiles due to genetic liabilities. A longitudinal MZ design in 235 MZ twin pairs was used to assess the transcriptional changes in the blood associated with cognitive ability differences over a 10-year interval.[111] While this study found no significant transcripts associated with cognitive level or cognitive change over time, it reported two suggestive transcripts; *POU6F1* was negatively associated with cognitive level and *MAD2L1* was positively associated with cognitive change. In addition, gene set enrichment analyses indicated that genes involved in protein metabolism, translation, RNA metabolism, the immune system, and infectious diseases were correlated with lower cognitive levels and cognitive decline. Similar results had previously been observed in individuals with cognitive impairments, indicating these pathways could play a role in aging and cognitive aging in general.

The discordant MZ twin pair design is a valuable tool to examine causality,[112] as illustrated by an example study that aimed to identify gene expression profiles for smoking behavior and to elucidate whether such gene expression profiles are cause or consequent of smoking.[113] In two Dutch population-based cohorts peripheral gene expression microarray data were available for 743 current smokers, 1686 never smokers, and 890 former smokers (age range: 18–88 years). The study identified 220 gene expression probes (of 132 genes) differentially expressed between current and never smokers, that were enriched for immune system, natural killer cells, blood coagulation, and cancer pathways. The expression levels of the 132 smoking-related

genes were compared between current and former smokers and between former and never smokers, as this comparison informs on the reversibility of gene expression levels. Six out of 132 smoking-related genes smoking had irreversible effects on gene expression levels, 31 out of 132 genes were slowly reversible (expression patterns differ between current and former smokers and between former and never smokers) and 94 out of 132 were reversible. Comparisons of gene expression levels of the 132 smoking-related genes in MZ twin pairs discordant for smoking behavior ($N = 56$ pairs) identified 6 differentially expressed genes, indicating these expression levels changed as a consequence of smoking behavior. Successful look-up of *cis*-eQTLs of the smoking-related genes in a GWA for number of cigarettes smoked per day suggested that *GPR56* and *RARRES3* expression are causative for smoking behavior. Thus, the majority of gene expression differences in smoking behavior are a consequence rather than a cause of smoking, which can be largely reversed after cessation of smoking.

32.5 Metabolomics

32.5.1 What is metabolomics and how do we measure the metabolome?

Metabolites are the small molecules, with low molecular weight (<1 kDa), that are involved in cellular metabolism.[114] In the human body, metabolites have numerous functions, including structure formation, signaling, and energy storage.[115] Metabolites can be endogenous (i.e., originate from within an organism) or exogenous (i.e., originate from outside of an organism, e.g., toxins, drugs, and nutrients)[116] and are a highly diverse set of molecules that include amino acids, keto acids, sugars, and lipids.[117] The metabolome is the complete set of metabolites that can be measured within a specific biofluid (e.g., serum, plasma, urine, cerebrospinal fluid, or saliva) or tissue sample.[118] Metabolomics is the study of the metabolome of a biological system, for example, a tissue, cell, or entire organism.[119] As the field of metabolomics includes a broad spectrum of molecular species of different (physical) chemical nature, many metabolomics subtypes focusing on specific molecule types have arisen. One can think of subtypes that are aimed at exogenous molecules taken up by the organism (drugs, nutrients), or molecules involved in specific biological pathways or systems (hormones, lipids). Among the most studied metabolomics subtypes is lipidomics, the study of lipids.[120] Metabolomics strategies focusing on known metabolites, often of similar chemical structures, are called targeted metabolomics and are common in hypothesis testing. Nontargeted metabolomics aims for global detection of a wide range of metabolites and are commonly used to identify changes in metabolites between conditions without *a priori* knowledge of relevant biological pathways.[121] The number and variety of measured metabolites for targeted and nontargeted strategies depend on the sensitivity of the chosen analytical chemical technology.

Different combinations of separation and detection methods are applied in metabolomics.[116] Nuclear magnetic resonance (NMR) spectroscopy, liquid-chromatography mass spectrometry (LC-MS), and gas-chromatography mass spectrometry (GC-MS) are the most widespread platforms.[122] Most NMR metabolomics studies focus on proton (^{1}H) NMR spectroscopy, because it has higher sensitivity then carbon (^{13}C) NMR spectroscopy due to the low natural abundance of carbon (~1.1%).[123] The identification of metabolites with ^{1}H-NMR is based on the so-called "chemical shifts" of the signals and the relative intensity of these signals. The chemical shift in NMR is the variation in resonance frequencies of protons due to different compositions of the surrounding molecules, with respect to a reference frequency or sample.[124] Like in nuclear magnetic imaging (MRI), an NMR signal is produced by aligning the spin states of all protons via a strong magnetic field. Next, an electromagnetic pulse in the radio frequency range is applied to the sample, causing the proton spin states to resonate. The energy emitted from the protons as they relax from the excited spin state to the one before the pulse is measured.[125]

MS determines the molecular weight of metabolites by measuring the mass to charge ratio (m/z).[126] Prior to MS, separation is important to separate analytes with identical m/z values, to prevent high-abundance metabolites to dominate the MS spectrum, or to select which metabolites may pass into the mass spectrometer. GC- and LC-MS are most commonly applied in metabolomics studies. In GC-MS, metabolites injected into the chromatographic device are heated to approximately 300°C to convert them to a gaseous state. Separation of the metabolites depends on their volatility, as more easily evaporated metabolites are driven through the chromatographic column, and subsequently to the detector, faster than less volatile metabolites.[127] LC-MS setups can be distinguished by separation on hydrophobicity or polarity. In reversed-phase chromotography, dissolved metabolites bind to the column (the stationary phase) based on their hydrophobic interactions with the hydrophilic liquid (the mobile phase) in the column. By making the mobile–mobile phase more hydrophobic, the metabolites are eluded from the column, toward the entrance of the mass spectrometer, by use of a strong hydrophobic solvent.[128,129] Normal phase LC-MS is based on the polarity of the metabolites rather than their hydrophobicity.[130] After separation metabolites are destructed into charged fragments. The fragment composition after destruction serves as a fingerprint for the molecule type and hence enables identification of a given metabolite. The gas-phase ionic fragments are generated by the mass spectrometer at its ionization source where molecules are charged by the removal of electrons. After ionization, the ions enter the mass analyzer through which the ions travel based on its m/z ratio. The ionized sample hits the detector, where the number of separated ions with particular m/z values is recorded (mass spectrum).[128]

32.5.2 Causes of variation in metabolite levels

Differences in metabolite levels among individuals reflect individual differences in genetic make-up, physiology, lifestyle, and behavior or responses to environmental factors.[131] Similarities in genetic and environmental backgrounds between individuals

result in more similar lipid profiles, as shown through hierarchical clustering of plasma lipids (LC-MS) in young adult twins and nontwin siblings.[132,133] Twin-family studies estimated the heritability of metabolite levels from approximately 0% to 80%.[134–139] The average heritability observed for metabolite levels differs among metabolites classes. For example, one study estimated the total and SNP-based heritability of 1097 metabolites (UPLC-MS/MS) in plasma for 1111 individuals, and reported that the median total heritability for lipids was 37% and for amino acids 40%.[140] This is in contrast to heritability estimates derived from a study in 221 MZ and 340 DZ twin pairs, that found higher heritability estimates for NMR-measured lipids (range: 0.48–0.62) and lipoproteins (range: 0.50–0.76) than for amino acids and other small molecules (range: 0.23–0.55).[141] A higher heritability for LC-MS measured amino acids than lipids was also seen in a family cohort.[142] The same study reported higher heritability levels for essential amino acids than for nonessential amino acids. Heritability differences among lipid species were also found in twin and family studies of lipidomics data that reported that sphingolipids and glycerolipids tended to have higher heritability estimates than phospholipids.[140,143,144]

The influence of genetic factors on metabolites levels has also been substantiated through genetic association studies that successfully identified metabolite QTLs.[145] For example, in serum samples from 79 MZ twin pairs, 215 DZ twin pairs, and 413 unrelated individuals, the genetic influence on metabolite levels as obtained from two metabolomics platforms were compared,[146] with 160 metabolites measured on a targeted platform (FIA-MS/MS) and 488 metabolites on a nontargeted platform (combination UHPLC-MS and GC-MS), with 43 metabolites measured on both platforms. The mean correlation between these 43 overlapping metabolites was 0.44, and 29 of these 43 metabolites were heritable on both platforms, with heritability estimates ranging from 0.29 to 0.72. For all metabolites on both platforms, GWA identified 61 significant metabolite-SNP associations at 26 independent loci. Of these 26 loci, 19 loci were associated with metabolites measured on one platform, and 7 loci were associated with six metabolites measured on both platforms. This study observed moderate heritability ($h^2 > 0.26$) and correlation ($r > 0.38$) among five of the metabolites associated with the seven loci. Here, the main message is that genetic influences on metabolite concentrations can be observed from data generated by different platforms, possibly utilizing different techniques (NMR vs MS). Even when concentrations of the same metabolite measured by different platforms correlate only moderately (due to e.g. experimental differences) and have only moderate heritability, the interaction with genetic variants may remain detectable. This enables combining/extending studies based on different platforms.

Metabolite QTL information can be used to obtain additional insights into the genetic architecture of metabolite classes. A recent study investigated the heritability of 361 metabolites, in a cohort of 5117 twin-family members (mean age: 42.1), with an extended GRM-based approach.[147] Four GRMs were obtained based on twin and SNP information: two GRMs defined the total (h^2_{total}) and SNP-based heritability (h^2_{SNP}), and two GRMs defined the contribution of metabolite QTLs of the same or of different metabolite classes. These last two GRMs included all loci from GWA and

(exome-) sequencing studies published between November 2008 and October 2018, which identified >800 loci associated with metabolite levels. In this study, the 361 metabolites could be classified as 309 lipids and 52 organic acids and were measured on four different metabolomics platforms (NMR and MS). The mean and median h^2_{total} for lipids both were 0.47. For the organic acids mean and median heritability were 0.41 and 0.40. The median heritability captured by all metabolite QTLS ($h^2_{\text{metabolite-hits}}$) was 0.06 for lipids and 0.01 for organic acids and was mainly attributable to with class-specific hits. Differences in heritability estimates among subclasses of organic acids, lipids, and among lipid species were investigated with mixed-effect meta-regression models. These analyses demonstrated that subclasses of lipids and organic acids differed significantly in $h^2_{\text{metabolite-hits}}$ and that higher degrees of unsaturation in phosphatidylcholines is associated with higher estimates of $h^2_{\text{metabolite-hits}}$.

Unlike the influence of genetic factors on metabolite levels, contributions of the environment shared by family members has been less well characterized and here the classical twin design is of substantial value. An NMR metabolomics twin study in 221 MZ and 340 DZ twin pairs (aged 22–25 years) for 216 metabolites reported that a model including shared environment was the best one for only 31 metabolites (variance explained by shared environment ranged between 15% and 38%).[141] For 6 of these 31 metabolites shared environment explained all familial resemblance. Thus, shared environment influences metabolite levels for a minority of metabolites in a young adult population. In contrast, a family-based FIA-MS/MS metabolomics study in 48 individuals from 16 families (12 parents [mean age = 42] and 26 children aged 8–18 years) reported shared environmental influences for 55 out of 147 measured metabolites.[148] A study from the Netherlands Twin Register estimated the contribution of genetic and shared environmental influences on 237 metabolite levels measured on three platforms (NMR, FIA-MS/MS, and LC-MS) in 886 MZ and 601 DZ adult twin pairs (mean age = 35).[149] A significant contribution of shared environment was reported for 6 out of 237 metabolites (25% explained variance, range 17%–43%) only. Together these studies indicate that the common environment does not play a large role in adult metabolite levels and that substantial effects are mostly found in studies that include younger participants or small sample sizes.

The value of multivariate extensions of the classic twin design for multiple metabolites was highlighted in a study of 221 MZ and 340 DZ young adult twin pairs that explored the association of serum n-6 and n-3 polyunsaturated (PUFAs), monosaturated (MUFAs), and saturated (SFAs) fatty acids with NMR-measured lipoprotein particle concentrations.[150] Bivariate models were applied to those metabolites with a phenotypic correlation of ≥0.3. The bivariate analysis of total n-6 PUFAs and Linoleic Acid (LA) with triglyceride and VLDL particles showed that approximately half (44%–56%) of the phenotypic covariance between the metabolites pairs was due to genetic factors. For MUFAs genetic factors explained more than half of the phenotypic variance between the metabolites, with bivariate heritability estimates of ~80% of MUFAs and HDL-related metabolites and of 58% to 66% for MUFAs and triglyceride and VLDL subclasses. Thus, shared genetic factors play a large role in explaining the associations of PUFAs and MUFAs with lipoprotein particle concentrations.

32.5.3 MZ discordant design applied to metabolomics studies

In contrast to epigenomics or transcriptomics studies, in metabolomics studies the MZ discordant design is less frequently applied. One example concerns an application to schizophrenia. An ^{1}H-NMR metabolomics study in plasma samples of 21 schizophrenia discordant MZ pairs and 8 pairs of matched unaffected MZ pairs showed that signals for VLDL and LDL lipoproteins and aromatic metabolites were the most important to differentiate affected, unaffected and control twins.[151] The differentiation between affected and unaffected twins was more pronounced for female twin pairs. In discordant pairs, MZ twins with schizophrenia had a 23% increase in plasma VLDL signals and a 14% reduction in plasma aromatic metabolites as compared to their unaffected cotwin.

While the MZ discordant design has not often been applied as the main analysis in metabolomics studies, a design with discordant MZ twin pairs to test for replication has gained popularity. Examples include blood metabolomics profiles of food preference and nutrition[136,152–154] and a recent study of urinary metabolites and neurotransmitter ratios, as measured with LC-MS and GC-MS, and childhood aggression.[155] The discovery sample in the aggression study had 783 MZ and DZ twins, the replication sample 189 MZ twin pairs discordant for aggression, and had an additional validation sample of 183 unrelated children who had been referred to a child psychiatry clinic. Positive associations were reported for two metabolites and childhood aggression in the discovery phase. The study did not replicate or validate its findings, but provided suggestive evidence linking childhood aggression to metabolic dysregulation in energy metabolism, oxidative stress, and neurotransmission pathways.

32.5.4 Other application of twin research in metabolomics studies

Discordant DZ twin pairs control for shared environmental factors and partially for genetic factors. While such a design weakens the ability to control for genetic factors, inclusion of DZ pairs would increase statistical power as discordant MZ twin pairs, particularly longitudinally discordant MZ twin pairs, are relatively scarce. A study investigating the long-term effect of physical activity on the serum NMR metabolome selected 16 same-sex twin pairs (7 MZ and 9 DZ pairs; age range: 50–74 years) longitudinally discordant (32 years) for leisure-time physical activity in addition to three independent population cohorts with longitudinally (>5 years) active and inactive participants (N = 1037, mean ages: 31–52 years).[156] Compared to persistently inactive individuals, the serum metabolome of persistently active individuals was characterized by lower concentrations of very-low-density lipoprotein particles, α1-acid glycoprotein, glucose, isoleucine, and polyunsaturated fatty acids and by higher concentrations of large and very large high-density lipoprotein particles and saturated fatty acids.

A discordant MZ twin pair design is suited for dichotomous traits such as presence or absence of disorders. For continuous traits, paired differences between MZ twins also inform about associations of omics profiles with such traits, adjusted for shared genetic, and environmental factors. A recent paper incorporated this strategy

to elucidate plasma metabolite profiles for metabolic risk factors.[157] For 40 MZ twin pairs (mean age 30.7 years) 111 plasma UPLC-MS metabolites were measured as well as blood lipids, fasting glucose, fasting insulin, C-reactive protein (CRP), adiposity measures and homeostasis model assessment (HOMA). First, the 93 metabolites that survived quality control were regressed against the adiposity and blood biochemistry measures, while accounting for twin relatedness. After correction for multiple testing, 18 metabolites were significantly associated with adiposity measures (BMI, percentage of body fat, abdominal visceral adipose tissue, and liver fat) and 24 with blood biochemistry measures (HOMA, CRP, triglycerides, and high-density lipoprotein cholesterol [HDL-C]). Next, follow-up with within-twin pair moderated *t*-tests (this type of *t*-tests uses the square root of the moderated variance as the SD instead of the sample variance) showed that the associations of 9 metabolites with adipose measures and of 10 with blood biochemistry measures (only HDL-C) were independent of confounding factors shared by twins.

32.6 Twin studies in other omics domains

We have considered in some detail the value of twin studies in genomics, epigenomics, transcriptomics, and metabolomics, but other omics domains also benefited from twin research. Proteomics is the large-scale study of the entire range of proteins, the vital molecules that have direct involvement in cellular function,[158] in a cell type (proteome).[159] Protein synthesis is accomplished by converting the information contained in the mRNA sequence to amino acids, a process called translation. Decoding of mRNA is done by the ribosomes where mRNA travels through the ribosome to translate one codon (block of three mRNA nucleotides) at a time to an amino acid, in this process, tRNA is responsible for forming the covalent peptide bonds between the amino acids.[14] As proteins are three-dimensional structures, folding forms the final protein structure. Some proteins fold spontaneously while they are released from the ribosome, while most others require molecular chaperones to help them fold correctly.[160] Large-scale high-throughput proteomics studies predominantly employ two types of analytical strategies. The first uses analytical protein microarrays that rely on antigen-antibody pairing.[161] While protein microarrays have good sensitivity and reproducibility,[161] they are limited in the number of proteins, and the specific group of proteins or molecular pathways they can assess. Therefore, MS-based proteomics provides a more versatile analytical strategy.

Regardless of the analytical strategy, sample preparation for proteomics experiments are labor-intensive, often involving multiple steps such as purification, enzymatic digestion, cell lysis, and solid-phase extraction.[162] The challenges in sample preparation, combined with those in protein and peptide identification, means that large-scale proteomics studies remain relatively expensive and proteomics has not been as extensively studied in twins. The discordant MZ design has been applied to characterize proteomic profiles for BMI,[163] ischemic stroke,[164] bipolar disorder,[165] fatigue,[166,167] hormone replacement therapy,[168] strabismus,[169] and multiple

autoimmune disorders.[170] Twin studies using various other designs have also been applied to proteomics studies. For example, in 15 pairs of opposite-sex DZ twins, sex-specific differences in LC-MS proteins of human endothelial cells were investigated.[171] This study reported small (average fold difference of 1.1–1.2) sex-specific differences in protein levels for approximately 10% of the measured proteins.

Another omics type that has benefitted from twin studies is the microbiome, which is the total ecological community of microorganism such as bacteria, fungi, and viruses that live on and inside our body.[172] Techniques to examine the human microbiome assess both structure and function of the microbiome. The most common application is structural, aimed at cataloging which microbes are present and what their relative abundance is.[173] This can be done by sequencing the gene that encodes the RNA component of the small ribosomal subunit (16S rRNA), followed by taxonomy of the 16S rRNA sequences.[174] Twin studies suggest a greater similarity for measures of relative abundance in MZ than in DZ twins.[175–178] Environmental factors, ranging from pre- and perinatal conditions to household sharing, may be important contributors to the microbiome composition.[178] Twin studies confirm that cohabiting MZ twin pairs have more similar microbiota communities than noncohabiting MZ twin pairs,[179] and that cohabitation can make microbial strains more similar between twins.[180] Rare SNVs in a fecal metagenomes sequencing study were assessed in a cohort of family members, including some twin families.[181] Strain persistence and within-family strain transmissions were analyzed from birth into adulthood. Strong evidence of transmission of maternal strains was seen for vaginally born infants. Later in childhood there was replacement by strains from the environment, including those from family members, with fathers appearing to be more frequently donors of novel strains to other family members. Twins generally did not have more similar rare SNV profiles than nontwin siblings, consistent with findings from abundance studies.

Other omics domains can often be considered subtypes of the traditional omics domains. Subtypes of proteomics include for example glycomics (i.e., the study of glycosylation, or the attachment of glycans or carbohydrates to proteins),[182] or phosphoproteomics (i.e., the study of proteins containing a phosphate group as a post-translational modification).[183] Fluxomics (i.e., the study of the rate of metabolite conversion or transportation in biochemical reaction networks),[184] can be seen as a subtype of metabolomics. Many of these subtypes currently are not optimized for application on a large scale, and twin studies are scarce.

Finally, the exposome has been defined as the totality of exposure individuals experience over their lives.[185] The exposome "summarizes" all environmental influences and is the accumulation of a person's environmental exposures from conception onward. It characterizes the environmental exposures in space and time on omics and on other phenotypes or phenotypic development. The exposome comprises of three domains: (1) internal, (2) specific external, and (3) general external.[6] The internal exposome refers to processes within the body, for example, body morphology or physical activity, but also encompasses the other omics layers such as the interactions between host and (gut) microflora (i.e., the microbiome). Specific external exposures

are the target of classic epidemiology studies and include exposure to environmental pollutants, diet, or lifestyle. General external exposures may include more general economic or social influences. An overview of twin studies in this research domain would go beyond the scope of the current chapter, but we note that twin studies indicate that exposures that are commonly labeled "environment" may show substantial heritability.[186,187]

32.7 Discussion

We have considered and reviewed the value of multiple twin analytical designs in omics research, from the classical twin design which relies on the comparison of resemblance in mono- and dizygotic twin pairs to the discordant twin design. The classic twin design is still invaluable to determine the contribution of genetic and environmental factors on variation in omics levels, with one of its strengths being the possibility to distinguish shared and unique environment. The classic twin design can be extended in multiple ways. A particular strength is combining the twin design with genome-wide SNP data. A recent example of such a combined analysis investigated the heritability of blood metabolites.[147] Based on the twin and SNP information, four genetic relatedness matrices (GRMs) among participants were obtained. Two GRMs defined the total and the SNP heritability. With the addition of two extra GRMs a distinction was made in the contribution of metabolite SNPs of the same or of different metabolite classes. Thus, this method relies on four GRMs: (1) a GRM including all autosomal SNPs for all closely-related individuals in the pedigree (h^2_{ped}); (2) a GRM including all autosomal SNPs (excluding all metabolite QTLs ± 50 kb) for all individuals in the dataset (h^2_{g}); (3) a GRM including the metabolite QTLs of a specific metabolite class for all individuals in the dataset ($h^2_{\text{class-hits}}$); and (4) a GRM including all metabolite QTLs (excluding all QTLs ± 50 kb as included in the third GRM) for all other metabolite classes for all individuals in the dataset ($h^2_{\text{notclass-hits}}$). In this model, the total heritability (h^2_{total}) is obtained by summing across all four heritabilities, SNP-based heritability is obtained by summing across the variance components obtained from the other 3 GRMs and the variance explained by all metabolite QTLs ($h^2_{\text{metabolite-hits}}$) can be obtained by summing $h^2_{\text{class-hits}}$ and $h^2_{\text{notclass-hits}}$. By specifying separate variance components for $h^2_{\text{class-hits}}$ and $h^2_{\text{notclass-hits}}$ metabolite QTLs of the same metabolite class were found to have higher heritability than metabolite QTLs of all other metabolite classes. The study reported nonzero median $h^2_{\text{notclass-hits}}$ estimates, suggesting that metabolite QTLs of other metabolite classes contribute to variance in metabolite levels. This may mean that more powerful GWA or sequencing studies will find associations of these QTLs for the relevant metabolites or this could be a reflection of metabolic networks which can span across distinct metabolite classes. This example and similar studies demonstrate the versatility of combining twin data with genome-wide SNP data. Thus, joining new omics analytical strategies with twin data will be of great benefit to omics research.

Multiple popular analytical strategies in omics research may benefit from including twin data. First, GWA studies have demonstrated that most complex traits and disorders have a highly polygenic nature. To capture polygenic signatures at the individual level, polygenic scores can be constructed.[188] Polygenic scores are calculated by computing the sum of the risk alleles an individual carries at a particular locus, weighted by the locus effect size, as obtained from a GWA. Similar scores can now be constructed from other omics data, for example, DNA methylation scores,[76,189] the epigenetic equivalent of polygenic scores. DNA methylation scores have been explored for traits such as BMI[190] and smoking.[191] DNA methylation scores hold promise as disease biomarkers that, in contrast to polygenic scores, can capture the cumulative and long-term effects of lifetime environmental exposures and the disease process itself. The MZ twin design offers a unique opportunity to examine if prediction of disease risk can be improved by combining polygenic scores with epigenetic scores. MZ twins have identical polygenic scores, yet their discordance rate for many diseases is high, illustrating that the accuracy of polygenic scores will never be perfect. Future studies can examine if epigenetic scores can aid further stratification of disease risk in individuals with identical polygenic scores.

Second, omics data can be used to construct predictors of biological aging and mortality. Well-established predictors rely on epigenetic markers to create the so-called epigenetic clocks.[192] Epigenetic clocks have also been investigated in twins. These studies indicated that the rate of epigenetic aging of MZ cotwins age tends to be similar but is often not identical and these differences in epigenetic aging between MZ cotwins have been associated with traits such as the cerebroplacental ratio (reflects fetal adaptation to hypoxic conditions),[193] and grip strength.[194] No differences in epigenetic aging between MZ cotwins were reported for studies investigating, for example, the association with leisure-time physical activity,[195] depression symptomatology in elderly twins,[196] or cognitive functioning.[197] While epigenetic clocks are frequently used to determine biological aging, clocks based on data from other omics domains are also being developed. For example, with microarray gene expression of T cells in a sample of 27 MZ twins (age range: 22–98) a transcriptomic signature of 125 genes could be constructed to estimate chronological age.[198] This gene expression clock could be replicated in gene expression datasets of T cells, but had poor performance when calculating it using gene expression data of human muscle, indicating that the gene expression clock is likely tissue-specific. Similarly, a metabolomics predictor for chronological age has been constructed using 56 ^{1}H-NMR blood metabolites as measured in 22 cohorts (N = 18,716).[199] A large, positive, difference between an individual's metabolomic and chronological age (ΔmetaboAge) indicates that, for a given chronological age, this individual has a relatively "old" blood metabolome. This has been associated with poor cardio-metabolic health in Dutch BBMRI (Biobanking and BioMolecular Resources Research Infrastructure) cohorts, and with an increased risk for future cardiovascular disease, higher mortality and lower functionality in independent cohorts of older individuals.

Third, in order to establish causal relationships randomized controlled trials often are the preferred method. However, for many research questions RTCs are not

feasible or ethical. Twin models, such as the discordant MZ twin design or methods investigating intra-pair differences, may serve as alternatives to assess causality.[112] Yet, the MZ discordant design does have a caveat, as *de novo* sequence differences between MZ twin pairs can occur. Furthermore, differences between MZ twins could be inflated by measurement error, as this introduces random divergence within twin pairs.

Based on cross-sectional data from MZ and DZ pairs the direction of causation between two traits can be assessed (Direction of Causation model) if the pattern of heritability and shared environmental influences is not too similar for the two traits.[200,201] Mendelian randomization (MR) employs genetic variants as instrumental variables to detect a causal effect of a risk factor on a complex trait or disease.[202] MR requires strong instrumental variables, and as most genetic variants have small effect sizes it has been proposed to combine them into polygenic scores. However, many genetic variants are pleiotropic, and polygenic scores may violate the "no pleiotropy" assumption (instrumental variables may not have direct effects on the outcome) of MR. Several methods are available to include multiple genetic variants that are robust for the "no pleiotropy" assumption.[203] When integrating MR with the Direction of Causation twin model (MR-DoC), the "no pleiotropy" assumption can be relaxed and polygenic scores can serve as instrumental variables.[204]

Twin studies are also valuable in providing information on the reliability of omics traits and profiles, as illustrated by a study of DNA methylation profiles.[205] Reliable methylation probes, defined as probes with a large correlation between replicate measures of the same DNA, have a higher heritability. In general, unreliable traits cannot be highly correlated in monozygotic twin pairs, and therefore the MZ correlation offers a lower bound for the reliability of a trait.

The majority of the twin omics studies described here tended to focus on a single omics domain. However, while each of the different omics layers provides us with a unique picture of the underlying biology of complex traits and disorders, this is an incomplete picture.[206] Because the multiple omics domains are interrelated and interact, we need to study the omics domains collectively to fully understand biological processes.[207] Studies combining multiple omics domains are becoming more frequent, often including multiple omics layers with the purpose of providing biological or functional interpretation of the results for the first omics domain through study of a second (or more) omics domain. Such a strategy is applied in many GWA or EWA studies, were follow-up analyses investigate colocalization of the top SNPs/CpGs with eQTLs. This type of multiomics integration is called sequential integration, when simultaneously analyzing multiple omics domains this is called parallel integration.[208] Many methods for parallel integration of multiomics data have been developed in order to aid in disease classification or subtyping, biomarker prediction, or obtaining insight into disease biology. Most of the studies in twin samples to date have focused on sequential integration of multiomics data. We anticipate that combining twin designs with parallel multiomics integration strategies will be of benefit in disease classification or subtyping and biomarker prediction.

32.8 Conclusion

We have described the value of twin studies in genomics, epigenomics, transcriptomics, and metabolomics. We have discussed the application of the classical twin design and highlighted the benefits of the MZ discordant twin design for identifying omics profiles for complex traits and disorders and to inform on the causal role of omics domains. Much of the twin research has focused on elucidating the causes of variation in omics data, demonstrating the strength of the classical twin design. We also provided a brief overview of other omics domains that can benefit from more twin research in the future and have suggested analytical designs for omics studies that may benefit from the inclusion of twin data. Due to the wide availability of omics data and the methodological advances in multiomics analyses, twin studies with multiomics designs will likely see substantial growth in the coming years.

Acknowledgments

This work was performed within the framework of BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, nos. 184.021.007 and 184.033.111); NWO-funded X-omics project (184.034.019); and the European Union Seventh Framework Program (FP7/2007–2013) ACTION Consortium (Aggression in Children: Unraveling gene–environment interplay to inform Treatment and InterventiON strategies; Grant number 602768). We thank drs Erik Ehli, Jeff Beck, and Jennifer Harris for their critical review and help with the manuscript.

References

1. Zhang XD. Precision medicine, personalized medicine, omics and big data: concepts and relationships. *J Pharmacogenomics Pharmacoproteomics*. 2015;06:1–2.
2. Kim M, Tagkopoulos I. Data integration and predictive modeling methods for multi-omics datasets. *Mol Omi*. 2018;14:8–25.
3. Buescher JM, Driggers EM. Integration of omics: More than the sum of its parts. *Cancer Metab [Internet]*. 2016;4:1–8. https://doi.org/10.1186/s40170-016-0143-y.
4. Franklin S, Vondriska TM. Genomes, proteomes, and the Central Dogma. *Circ Cardiovasc Genet [Internet]*. 2011;4:576. https://www.ahajournals.org/doi/10.1161/CIRCGENETICS.110.957795.
5. Giera M, Wuhrer M. Recent developments in clinical omics. *Chromatographia*. 2015;78:305–306.
6. Wild CP. The exposome: From concept to utility. *International Journal of Epidemiology*. 2012;41:24–32.
7. Visscher PM, Andrew T, Nyholt DR. Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. *European Journal of Human Genetics*. 2008;16:387–390.
8. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding factors and mixed models. *Plos Genetics*. 2018;14:1–22.

9. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*. 2018;50:1112–1121.
10. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet [Internet]*. 2018;50:668–681. https://linkinghub.elsevier.com/retrieve/pii/S0924977X17304960.
11. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nature Genetics*. 2016;48:856–866.
12. van Dongen J, Slagboom PE, Draisma HHM, et al. The continuing value of twin studies in the omics era. *Nat Rev Genet [Internet]*. 2012;13:640–653. http://www.ncbi.nlm.nih.gov/pubmed/22847273.
13. Alberts B, Bray D, Hopkin K, et al.DNA and chromosomes *Essential Cell Biology*. Third ed. New York, NY: Garland Science; 2010:171–195.
14. Alberts B, Bray D, Hopkins K, et al. From DNA to proteins: how cells read the genome*Essential Cell Biology*. Third ed. New York, NY: Garland Science; 2010:231–268.
15. Singh DD, Datta M. Genomics*Omics Approaches, Technologies And Applications [Internet]*. Singapore: Springer Singapore; 2018:11–38. http://link.springer.com/10.1007/978-981-13-2925-8_2.
16. Haraksingh RR, Snyder MP. Impacts of variation in the human genome on gene regulation. *J Mol Biol [Internet]*. 2013;425:3970–3977. https://doi.org/10.1016/j.jmb.2013.07.015.
17. Escaramís G, Docampo E, Rabionet R. A decade of structural variants: description, history and methods to detect structural variation. *Brief Funct Genomics*. 2015;14:305–314.
18. Bumgarner R. Overview of DNA microarrays: types, applications, and their future*Current Protocols in Molecular Biology [Internet]*. Hoboken, NJ: John Wiley & Sons, Inc.; 2013:1–11. http://doi.wiley.com/10.1002/0471142727.mb2201s101.
19. Rajawat J. Transcriptomics*Omics Approaches, Technologies And Applications [Internet]*. Singapore: Springer Singapore; 2018:39–56. http://link.springer.com/10.1007/978-981-13-2925-8_3.
20. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. *Nature Protocols*. 2014;9:2643–2662.
21. Voight BF, Kang HM, Ding J, et al. The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. In: Gibson G, ed.; 2012. *PLoS Genet [Internet]*. 8:e1002793. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3410907&tool=pmcentrez&rendertype=abstract.
22. Ehli EA, Abdellaoui A, Fedko IO, et al. A method to customize population-specific arrays for genome-wide association testing. *Eur J Hum Genet [Internet]*. 2017;25:267–270. http://dx.doi.org/10.1038/ejhg.2016.152.
23. Beck JJ, Hottenga J-J, Mbarek H, et al. Genetic similarity assessment of twin-family populations by custom-designed genotyping array. *Twin Res Hum Genet [Internet]*. 2019;22:210–219. https://www.cambridge.org/core/product/identifier/S1832427419000410/type/journal_article.
24. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*. 2018;27:1–10.
25. Petersen B-S, Fredrich B, Hoeppner MP, et al. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet [Internet]*. 2017;18:14. http://bmcgenet.biomedcentral.com/articles/10.1186/s12863-017-0479-5.

26. Mardis ER. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*. 2013;6:287–303.
27. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 2016;17:333–351.
28. Boomsma DI, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet [Internet]*. 2002;3:872–882. http://www.nature.com/doifinder/10.1038/nrg932.
29. Strachan T, Read A. Chromosome structure and function*Human Molecular Genetics*. 4th edition. New York, NY, USA: Garland Science; 2011:29–60.
30. Griffiths AJF, Miller JH, Suzuki DT. Somatic versus germinal mutation. In: Freeman WH, (ed.). An Introduction to Genetic Analysis [Internet]. 7th edition. New York, NY; 2000. https://www.ncbi.nlm.nih.gov/books/NBK21894/
31. Johnson BN, Ehli EA, Davies GE, et al. Chimerism in health and potential implications on behavior: a systematic review. *Am J Med Genet Part A*. 2020;182(6):1513–1529.
32. Campbell CD, Chong JX, Malig M, et al. Estimating the human mutation rate using autozygosity in a founder population. *Nature Genetics*. 2012;44:1277–1281.
33. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 2012;488:471–475. http://dx.doi.org/10.1038/nature11396.
34. Dal GM, Ergüner B, Sağiroğlu MS, et al. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *Journal of Medical Genetics*. 2014;51:455–459.
35. Ouwens KG, Jansen R, Tolhuis B, et al. A characterization of postzygotic mutations identified in monozygotic twins. *Human Mutation*. 2018;39:1393–1401.
36. Zwijnenburg PJG, Meijers-Heijboer H, Boomsma DI. Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet Part B Neuropsychiatr Genet*. 2010;153:1134–1149.
37. Abdellaoui A, Ehli EA, Hottenga JJ, et al. CNV concordance in 1,097 MZ twin pairs. *Twin Res Hum Genet*. 2015;18:1–12.
38. Melzer D, Pilling LC, Ferrucci L. The genetics of human ageing. *Nat Rev Genet [Internet]*. 2020;21:88–101. http://dx.doi.org/10.1038/s41576-019-0183-6.
39. Forsberg LA, Rasi C, Razzaghian HR, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet [Internet]*. 2012;90:217–228. http://dx.doi.org/10.1016/j.ajhg.2011.12.009.
40. Baranzini SE, Mudge J, Van Velkinburgh JC, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature*. 2010;464:1351–1356.
41. Bruder CEG, Piotrowski A, Gijsbers A, et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics*. 2008;82:763–771.
42. Ehli EA, Abdellaoui A, Hu Y, et al. De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. *European Journal of Human Genetics*. 2012;20:1037–1043.
43. Strachan T, Read A. Mapping genes conferring susceptibility to complex diseases. *Human molecular genetics*. 4th edition. New York, NY, USA: Garland Science; 2011:467–496.
44. Visscher PM, Macgregor S, Benyamin B, et al. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American Journal of Human Genetics*. 2007;81:1104–1110.
45. van Dongen J, Jansen R, Smit D, et al. The contribution of the functional IL6R polymorphism rs2228145, eQTLs and other genome-wide SNPs to the heritability of plasma sIL-6R levels. *Behav Genet [Internet]*. 2014;44:368–382. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4283105&tool=pmcentrez&rendertype=abstract.

46. Jones PA, Liang G. The human epigenome*Epigenetic Epidemiology [Internet]*. Dordrecht: Springer; 2012:5–20. http://link.springer.com/10.1007/978-94-007-2495-2_2.
47. Hawkins LJ, Al-attar R, Storey KB. Transcriptional regulation of metabolism in disease: from transcription factors to epigenetics. *PeerJ*. 2018;6:e5062. doi: 10.7717/peerj.5062.
48. Carlberg C, Molnár F. What is epigenomics?*Human Epigenomics [Internet]*. Singapore: Springer; 2018:3–18. http://link.springer.com/10.1007/978-981-10-7614-5_1.
49. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature [Internet]*. 2015;518:317–330. http://www.nature.com/articles/nature14248.
50. Carlberg C, Molnár F. DNA methylation*Human Epigenomics [Internet]*. Singapore: Springer; 2018:57–73. http://link.springer.com/10.1007/978-981-10-7614-5_4.
51. Carlberg C, Molnár F. Methods and applications of epigenomics*Human Epigenomics [Internet]*. Singapore: Springer; 2018:19–38. http://link.springer.com/10.1007/978-981-10-7614-5_2.
52. Li Y, Tollefsbol TO. DNA methylation detection: bisulfite genomic sequencing analysis*Methods in Molecular Biology (Clifton, NJ) [Internet]*; 2011:11–21. http://www.ncbi.nlm.nih.gov/pubmed/21913070.
53. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics [Internet]*. 2011;98:288–295. http://dx.doi.org/10.1016/j.ygeno.2011.07.007.
54. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol [Internet]*. 2016;17:1–17. http://dx.doi.org/10.1186/s13059-016-1066-1.
55. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–1195. doi:10.1126/science.
56. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49:131–138.
57. Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 2021;53:1311–1321. https://doi.org/10.1038/s41588-021-00923-x.
58. Van Dongen J, Ehli EA, Jansen R, et al. Genome-wide analysis of DNA methylation in buccal cells: a study of monozygotic twins and mQTLs. *Epigenetics Chromatin [Internet]*. 2018;11:1–14. https://doi.org/10.1186/s13072-018-0225-x.
59. van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun [Internet]*. 2016;7:11115. http://www.nature.com/doifinder/10.1038/ncomms11115.
60. Martino D, Loke YJ, Gordon L, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome biol.*. 2013;14(5):R42. doi: 10.1186/gb-2013-14-5-r42.
61. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U.S.A*. 2005.
62. Talens RP, Christensen K, Putter H, et al. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell*. 2012;11:694–703.
63. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Res*. 2002;5(6):554–571. doi:10.1375/136905202762342026.

64. Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med*. 2014;6:60.
65. Vidaki A, Daniel B, Court DS. Forensic DNA methylation profiling—potential opportunities and challenges. *Forensic Science International: Genetics*. 2013;7:499–507.
66. Vidaki A, Kayser M, Nothnagel M. Unsupported claim of significant discrimination between monozygotic twins from multiple pairs based on three age-related DNA methylation markers. *Forensic Science International: Genetics*. 2019;39:e1–e2.
67. Li C, Zhang S, Que T, Li L, Zhao S. Identical but not the same: the value of DNA methylation profiling in forensic discrimination within monozygotic twins. *Forensic Sci Int Genet Suppl Ser*. 2011;3:e337–e338.
68. Xu J, Fu G, Yan L, et al. LINE-1 DNA methylation: a potential forensic marker for discriminating monozygotic twins. *Forensic Science International: Genetics*. 2015;19:136–145.
69. Stewart L, Evans N, Bexon KJ, Van Der Meer DJ, Williams GA. Differentiating between monozygotic twins through DNA methylation-specific high-resolution melt curve analysis. *Analytical Biochemistry*. 2015;476:36–39.
70. Du Q, Zhu G, Fu G, et al. A Genome-Wide Scan of DNA Methylation Markers for Distinguishing Monozygotic Twins. *Twin Res Hum Genet*. 2015;18:670–679.
71. Tan Q. The epigenome of twins as a perfect laboratory for studying behavioural traits. *Neuroscience Biobehavioral Rev*. 2019;107:192–195.
72. Palma-Gudiel H, Córdova-Palomera A, Navarro V, et al. Twin study designs as a tool to identify new candidate genes for depression: a systematic review of DNA methylation studies. *Neurosci Biobehavioral Rev*. 2020;112:345–352.
73. Allione A, Marcon F, Fiorito G, et al. Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. *PLoS One*. 2015;10:e0128265.
74. Hancock DB, Guo Y, Reginsson GW, et al. Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol Psychiatry*. 2017;23(9):1911–1919. doi:10.1038/mp.2017.193.
75. Oates NA, Van Vliet J, Duffy DL, et al. Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am J Hum Genet*. 2006;79(1):155–162. doi:10.1086/505031.
76. Nwanaji-Enwerem JC, Colicino E. DNA methylation–based biomarkers of environmental exposures for human population studies. *Curr Environ Health Rep*. 2020;7(2):121–128. doi:10.1007/s40572-020-00269-2.
77. Souren NY, Gerdes LA, Lutsik P, et al. DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. *Nat Commun*. 2019;10:2094. https://doi.org/10.1038/s41467-019-09984-3.
78. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *International Journal of Epidemiology*. 2015;44:1429–1441.
79. Hu Y, Ehli EA, Boomsma DI. MicroRNAs as biomarkers for psychiatric disorders with a focus on autism spectrum disorder: current progress in genetic association studies, expression profiling, and translational research. *Autism Res*. 2017;10:1184–1203.
80. Wright FA, Sullivan PF, Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*. 2014;46:430–437.
81. Ouwens KG, Jansen R, Nivard MG, et al. A characterization of cis- and trans-heritability of RNA-Seq-based gene expression. *Eur J Hum Genet [Internet]*. 2019;10. http://www.nature.com/articles/s41431-019-0511-5.

82. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet [Internet]*. 2011;88:76–82. http://dx.doi.org/10.1016/j.ajhg.2010.11.011.
83. Aguet F, Ardlie KG. Tissue specificity of gene expression. *Curr Genet Med Rep*. 2016;4:163–169.
84. Grundberg E, Small KS, Hedman ÅK, et al. Mapping cis-and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*. 2012;44:1084–1089.
85. Huang Y, Ollikainen M, Sipil P, et al. Genetic and environmental effects on gene expression signatures of blood pressure: a transcriptome-wide twin study. *Hypertension*. 2018;71:457–464.
86. Maher B. Personal genomes: the case of the missing heritability. *Nature [Internet]*. 2008;456:18–21. http://www.nature.com/doifinder/10.1038/456018a.
87. A MT, FS C, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature [Internet]*. 2009;461:747–753. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2831613&tool=pmcentrez&rendertype=abstract.
88. Génin E. Missing heritability of complex diseases: case solved? *Hum Genet [Internet]*. 2020;139:103–113. https://doi.org/10.1007/s00439-019-02034-4.
89. Buil A, Brown AA, Lappalainen T, et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet [Internet]*. 2015;47:88–91. http://www.nature.com/articles/ng.3162.
90. Glastonbury CAA, Viñuela A, Buil A, et al. Adiposity-dependent regulatory effects on multi-tissue transcriptomes. *American Journal of Human Genetics*. 2016;99:567–579.
91. Wainschtein P, Jain DP, Yengo L, et al. Recovery of trait heritability from whole genome sequence data. *bioRxiv*. 2019.
92. Caramori ML, Kim Y, Moore JH, et al. Gene expression differences in skin fibroblasts in identical twins discordant for type 1 diabetes. *Diabetes*. 2012;61:739–744.
93. Beyan H, Drexhage RC, Nieuwenhuijsen L, et al. Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. *Diabetes*. 2010;59:1751–1755.
94. Haas CS, Creighton CJ, Pi X, et al. Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins. *Arthritis and Rheumatism*. 2006;54:2047–2060.
95. Ding N, Zhang Z, Yang W, et al. Transcriptome analysis of monozygotic twin brothers with childhood primary myelofibrosis. *Genomics, Proteomics Bioinforma [Internet]*. 2017;15:37–48. http://dx.doi.org/10.1016/j.gpb.2016.12.002.
96. Ronkainen PHA, Pöllänen E, Alén M, et al. Global gene expression profiles in skeletal muscle of monozygotic female twins discordant for hormone replacement therapy. *Aging Cell*. 2010;9:1098–1110.
97. Alieva AK, Rudenok MM, Novosadova EV, et al. Whole-transcriptome analysis of dermal fibroblasts, derived from three pairs of monozygotic twins, discordant for Parkinson's Disease. *Journal of Molecular Neuroscience*. 2020;70:284–293.
98. Kakiuchi C, Ishiwata M, Nanko S, et al. Up-regulation of ADM and SEPX1 in the lymphoblastoid cells of patients in monozygotic twins discordant for schizophrenia. *Am J Med Genet Part B Neuropsychiatr Genet*. 2008;147:557–564.
99. Nakazawa T, Kikuchi M, Ishikawa M, et al. Differential gene expression profiles in neurons generated from lymphoblastoid B-cell line-derived iPS cells from monozygotic twin cases with treatment-resistant schizophrenia and discordant responses to clozapine. *Schizophr Res [Internet]*. 2017;181:75–82. http://dx.doi.org/10.1016/j.schres.2016.10.012.

100. Matigian N, Windus L, Smith H, et al. Expression profiling in monozygotic twins discordant for bipolar disorder reveals dysregulation of the WNT signalling pathway. *Molecular Psychiatry*. 2007;12:815–825.
101. Watson NF, Buchwald D, Delrow JJ, et al. Transcriptional signatures of sleep duration discordance in monozygotic twins. *Sleep*. 2017;40(1):zsw019. doi:10.1093/sleep/zsw019.
102. Stamoulis G, Garieri M, Makrythanasis P, et al. Single cell transcriptome in aneuploidies reveals mechanisms of gene dosage imbalance. *Nature Communication*. 2019;10:1–11.
103. You S-H, Lee Y-S, Chang Y-J, et al. Gene expression profiling of amniotic fluid mesenchymal stem cells of monozygotic twins discordant for trisomy 21. *Gene [Internet]*. 2020;738:144461. https://doi.org/10.1016/j.gene.2020.144461.
104. Pietiläinen KH, Naukkarinen J, Rissanen A, et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: Pathways behind acquired obesity. *Plos Medicine*. 2008;5:0472–0483.
105. Jukarainen S, Heinonen S, Rämö JT, et al. Obesity is associated with low nad+/sirt pathway expression in adipose tissue of BMI-discordant monozygotic twins. *Journal of Clinical Endocrinology and Metabolism*. 2016;101:275–283.
106. Muniandy M, Heinonen S, Yki-Järvinen H, et al. Gene expression profile of subcutaneous adipose tissue in BMI-discordant monozygotic twin pairs unravels molecular and clinical changes associated with sub-types of obesity. *International Journal of Obesity*. 2017;41:1176–1184.
107. Tangirala S, Patel CJ. Integrated analysis of gene expression differences in twins discordant for disease and binary phenotypes. *Sci Rep [Internet]*. 2018;8:1–9. http://dx.doi.org/10.1038/s41598-017-18585-3.
108. O'Hanlon TP, Rider LG, Gan L, et al. Gene expression profiles from discordant monozygotic twins suggest that molecular pathways are shared among multiple systemic autoimmune diseases. *Arthritis Res Ther [Internet]*. 2011;13:R69. http://arthritis-research.biomedcentral.com/articles/10.1186/ar3506.
109. Gan L, O'Hanlon TP, Lai Z, et al. Gene expression profiles from disease discordant twins suggest shared antiviral pathways and viral exposures among multiple systemic autoimmune diseases. *PLoS One*. 2015;10:1–15.
110. Wen H, Chen L, He J, Lin J. MicroRNA expression profiles and networks in placentas complicated with selective intrauterine growth restriction. *Mol Med Rep*. 2017;16:6650–6673.
111. Nygaard M, Larsen MJ, Thomassen M, et al. Global expression profiling of cognitive level and decline in middle-aged monozygotic twins. *Neurobiol Aging [Internet]*. 2019;84:141–147. https://doi.org/10.1016/j.neurobiolaging.2019.08.019.
112. Vitaro F, Brendgen M, Arseneault L. The discordant MZ-twin method: one step closer to the holy grail of causality. *Int J Behav Dev [Internet]*. 2009;33:376–382. http://jbd.sagepub.com/cgi/doi/10.1177/0165025409340805.
113. Vink JM, Jansen R, Brooks A, et al. Differential gene expression patterns between smokers and non-smokers: cause or consequence?. *Addiction Biology*. 2017;22:550–560.
114. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*. 2012;13:263–269.
115. Dunn WB, Broadhurst DI, Atherton HJ, et al. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev [Internet]*. 2011;40:387–426. http://www.ncbi.nlm.nih.gov/pubmed/20717559.
116. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, et al. Untargeted metabolomics strategies—challenges and emerging directions. *J Am Soc Mass Spectrom [Internet]*. 2016;27:1897–1905. https://linkinghub.elsevier.com/retrieve/pii/S0031938416312148.

117. Adamski J, Suhre K. Metabolomics platforms for genome wide association studies-linking the genome to the metabolome. *Current Opinion in Biotechnology*. 2013;24:39–47.
118. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. *TrAC Trends Anal Chem [Internet]*. 2005;24:285–294. http://linkinghub.elsevier.com/retrieve/pii/S0165993605000348.
119. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Molecular Biology*. 2002;48:155–171.
120. Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov*. 2005;4:594–610.
121. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*. 2011;40:387–426.
122. Li B, He X, Jia W, Li H. Novel applications of metabolomics in personalized medicine: a mini-review. *Molecules [Internet]*. 2017;22:1173. http://www.mdpi.com/1420-3049/22/7/1173.
123. Clendinen CS, Lee-McMullen B, Williams CM, et al. 13 C NMR metabolomics: applications at natural abundance. *Anal Chem [Internet]*. 2014;86:9242–9250. https://pubs.acs.org/doi/10.1021/ac502346h.
124. Lenz EM, Wilson ID. Analytical strategies in metabonomics. *Journal of Proteome Research*. 2007;6:443–458.
125. Berger A. How does it work? Magnetic resonance imaging. *Br Med J*. 2002;324:35.
126. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev [Internet]*. 2007;26:51–78. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1904337&tool=pmcentrez&rendertype=abstract.
127. Issaq HJ, Abbott E, Veenstra TD. Utility of separation science in metabolomic studies. *J Sep Sci [Internet]*. 2008;31:1936–1947. http://www.ncbi.nlm.nih.gov/pubmed/18348322.
128. Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem Soc Rev*. 2009;38:1882–1896.
129. Vaughan AA, Dunn WB, Allwood JW, et al. Liquid chromatography-mass spectrometry calibration transfer and metabolomics data fusion. *Anal Chem [Internet]*. 2012;84:9848–9857. http://www.ncbi.nlm.nih.gov/pubmed/23072438.
130. Trivedi DK, Iles RK. Do not just do it, do it right: urinary metabolomics -establishing clinically relevant baselines. *Biomed Chromatogr [Internet]*. 2014;28:1491–1501. http://www.ncbi.nlm.nih.gov/pubmed/24788800.
131. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol [Internet]*. 2002;48:155–171. http://www.ncbi.nlm.nih.gov/pubmed/11860207.
132. Draisma HHM, Reijmers TH, Bobeldijk-Pastorova I, et al. Similarities and differences in lipidomics profiles among healthy monozygotic twin pairs. *OMICS [Internet]*. 2008;12:17–31. http://www.ncbi.nlm.nih.gov/pubmed/18266560.
133. Draisma HHM, Reijmers TH, Meulman JJ, van der Greef J, Hankemeier T, Boomsma DI. Hierarchical clustering analysis of blood plasma lipidomics profiles from mono- and dizygotic twin families. *Eur J Hum Genet [Internet]*. 2013;21:95–101. http://www.nature.com/articles/ejhg2012110.
134. Nicholson G, Rantalainen M, Maher AD, et al. Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol Syst Biol [Internet]*. 2011;7:525. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3202796&tool=pmcentrez&rendertype=abstract.

135. Alul FY, Cook DE, Shchelochkov OA, et al. The heritability of metabolic profiles in newborn twins. *Heredity (Edinb) [Internet]*. 2013;110:253–258. http://dx.doi.org/10.1038/hdy.2012.75.
136. Menni C, Zhai G, MacGregor A, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics [Internet]*. 2013;9:506–514. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3608890&tool=pmcentrez&rendertype=abstract.
137. van 't Erve TJ, Doskey CM, Wagner BA, et al. The heritability of glutathione and related metabolites in, stored red blood cells. *Free Radic Biol Med [Internet]*. 2014;76:107–113. http://www.sciencedirect.com/science/article/pii/S0891584914003578.
138. van 't Erve TJ, Wagner BA, Martin SM, et al. The heritability of metabolite concentrations in stored human red blood cells. *Transfusion [Internet]*. 2014;54:2055–2063. http://www.ncbi.nlm.nih.gov/pubmed/24601981.
139. Shin S-Y, Fauman EB, Petersen A-K, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet [Internet]*. 2014;46:543–550. http://www.ncbi.nlm.nih.gov/pubmed/24816252.
140. Darst BF, Koscik RL, Hogan KJ, et al. Longitudinal plasma metabolomics of aging and sex. *Aging (Albany NY) [Internet]*. 2019;11:1262–1282. http://www.aging-us.com/article/101837/text.
141. Kettunen J, Tukiainen T, Sarin A-P, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet [Internet]*. 2012;44:269–276. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3605033&tool=pmcentrez&rendertype=abstract.
142. Rhee EP, Ho JE, Chen M-H, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab [Internet]*. 2013;18:130–143. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3973158&tool=pmcentrez&rendertype=abstract.
143. Bellis C, Kulkarni H, Mamtani M, et al. Human plasma lipidome is pleiotropically associated with cardiovascular risk factors and death. *Circ Cardiovasc Genet [Internet]*. 2014;7:854–863. https://www.ahajournals.org/doi/10.1161/CIRCGENETICS.114.000600.
144. Frahnow T, Osterhoff MA, Hornemann S, et al. Heritability and responses to high fat diet of plasma lipidomics in a twin study. *Science Reports*. 2017;7:1–11.
145. Kastenmüller G, Raffler J, Gieger C, et al. Genetics of human metabolism: an update. *Hum Mol Genet [Internet]*. 2015;24:R93–101. http://www.hmg.oxfordjournals.org/lookup/doi/10.1093/hmg/ddv263.
146. Yet I, Menni C, Shin SY, et al. Genetic influences on metabolite levels: a comparison across metabolomic platforms. *PLoS One [Internet]*. 2016;11(4). http://dx.doi.org/10.1371/journal.pone.0153672.
147. Hagenbeek FA, Pool R, van Dongen J, et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun [Internet]*. 2020;11:39. http://www.nature.com/articles/s41467-019-13770-6.
148. Tremblay BL, Guénard F, Lamarche B, et al. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. *Nutr Res [Internet]*. 2019;61:22–30. https://doi.org/10.1016/j.nutres.2018.10.003.
149. Pool R, Hagenbeek FA, Hendriks A, et al. Genetics and not shared environment explains familial resemblance in adult metabolomics data. *Twin Res Hum Genet*. 2020;23:145–155.
150. Jelenkovic A, Bogl LH, Rose RJ, et al. Association between serum fatty acids and lipoprotein subclass profile in healthy young adults: exploring common genetic and

environmental factors. *Atherosclerosis [Internet]*. 2014;233:394–402. http://www.ncbi.nlm.nih.gov/pubmed/24530769.
151. Tsang TM, Huang JTJ, Holmes E, et al. Metabolic profiling of plasma from discordant schizophrenia twins: Correlation between lipid signals and global functioning in female schizophrenia patients. *Journal of Proteome Research*. 2006;5:756–760.
152. Pallister T, Sharafi M, Lachance G, et al. Food Preference Patterns in a UK Twin Cohort. *Twin Res Hum Genet*. 2015;18:793–805.
153. Pallister T, Jennings A, Mohney RP, et al. Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. *PLoS One*. 2016;11:1–16.
154. Pallister T, Jackson MA, Martin TC, et al. Untangling the relationship between diet and visceral fat mass through blood metabolomics and gut microbiome profiling. *Int J Obes [Internet]*. 2017;41:1106–1113. http://dx.doi.org/10.1038/ijo.2017.70.
155. Hagenbeek FA, Roetman PJ, Pool R, et al. Urinary amine and organic acid metabolites evaluated as markers for childhood aggression: the ACTION biomarker study. *Front Psychiatry [Internet]*. 2020;11(165). https://www.frontiersin.org/article/10.3389/fpsyt.2020.00165/full.
156. Kujala UM, Mäkinen V-P, Heinonen I, et al. Long-term Leisure-time Physical Activity and Serum Metabolome. *Circulation [Internet]*. 2013;127:340–348. https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.112.105551.
157. Muniandy M, Velagapudi V, Hakkarainen A, et al. Plasma metabolites reveal distinct profiles associating with different metabolic risk factors in monozygotic twin pairs. *Int J Obes [Internet]*. 2019;43:487–502. http://dx.doi.org/10.1038/s41366-018-0132-z.
158. Vaz C, Tanavde V. Proteomics*Omics Approaches, Technologies And Applications [Internet]*. Singapore: Springer; 2018:57–73. http://link.springer.com/10.1007/978-981-13-2925-8_4.
159. Aslam B, Basit M, Nisar MA, et al. Proteomics: Technologies and their applications. *Journal of Chromatographic Science*. 2017;55:182–196.
160. Alberts B, Bray D, Hopkin K, et al. Protein structure and function. *Essential Cell Biology*. 2010:119–170.
161. Sutandy FXR, Qian J, Chen C-S, et al. Overview of Protein Microarrays. *Curr Protoc Protein Sci [Internet]*. 2013;72. 27.1.1-27.1.16. http://doi.wiley.com/10.1002/0471140864.ps2701s72.
162. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature [Internet]*. 2003;422:198–207. http://www.nature.com/articles/nature01511%0Ahttp://www.nature.com/nature/journal/v422/n6928/abs/nature01511.html.
163. Sahebekhtiari N, Saraswat M, Joenväärä S, et al. Plasma proteomics analysis reveals dysregulation of complement proteins and inflammation in acquired obesity—a study on rare BMI-discordant monozygotic twin pairs. *Proteomics Clin Appl*. 2019;13(4):e1800173. doi:10.1002/prca.201800173.
164. Vadgama N, Lamont D, Hardy J, et al. Distinct proteomic profiles in monozygotic twins discordant for ischaemic stroke. *Mol Cell Biochem [Internet]*. 2019;456:157–165. http://dx.doi.org/10.1007/s11010-019-03501-2.
165. Kazuno A-a, Ohtawa K, Otsuki K, Usui M, Sugawara H, Okazaki Y, et al. Proteomic analysis of lymphoblastoid cells derived from monozygotic twins discordant for bipolar disorder: a preliminary study. *PLoS One*. 2013;8(2):e53855. https://doi.org/10.1371/journal.pone.0053855.
166. Ciregia F, Giusti L, Da Valle Y, et al. A multidisciplinary approach to study a couple of monozygotic twins discordant for the chronic fatigue syndrome: A focus on potential

salivary biomarkers. *J Transl Med*. 2013;11:243. https://doi.org/10.1186/1479-5876-11-243.
167. Ciregia F, Kollipara L, Giusti L, et al. Bottom-up proteomics suggests an association between differential expression of mitochondrial proteins and chronic fatigue syndrome. *Transl Psychiatry*. 2016;6(9):e904..
168. Laakkonen EK, Soliymani R, Karvinen S, et al. Estrogenic regulation of skeletal muscle proteome: a study of premenopausal women and postmenopausal MZ cotwins discordant for hormonal therapy. *Aging Cell*. 2017;16:1276–1287.
169. Liu G, Bai H, Yan Z, et al. Differential expression of proteins in monozygotic twins with discordance of infantile esotropic phenotypes. *Molecular Vision*. 2011;17:1618–1623.
170. O'Hanlon TP, Li Z, Gan L, et al. Plasma proteomic profiles from disease-discordant monozygotic twins suggest that molecular pathways are shared in multiple systemic autoimmune diseases. *Arthritis Res Ther*. 2011;13(2):R69. doi:10.1186/ar3330.
171. Witt E, Lorenz M, Völker U, et al. Sex-specific differences in the intracellular proteome of human endothelial cells from dizygotic twins. *J Proteomics [Internet]*. 2019;201:48–56. https://doi.org/10.1016/j.jprot.2019.03.016.
172. Peterson J, Garges S, Giovanni M, et al. The NIH Human Microbiome Project. *Genome Res [Internet]*. 2009;19:2317–2323. http://genome.cshlp.org/cgi/doi/10.1101/gr.096651.109.
173. Young VB. The role of the microbiome in human health and disease: An introduction for clinicians. *BMJ*. 2017;356:831. doi:10.1136/bmj.j831.
174. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev [Internet]*. 2004;68:669–685. http://www.ncbi.nlm.nih.gov/pubmed/15590779.
175. Stewart JA, Chadwick VS, Murray A. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J Med Microbiol [Internet]*. 2005;54:1239–1242. https://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.46189-0.
176. Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell [Internet]*. 2014;159:789–799. http://dx.doi.org/10.1016/j.cell.2014.09.053.
177. Goodrich JK, Davenport ER, Beaumont M, et al. Genetic determinants of the gut microbiome in UK Twins. *Cell Host Microbe [Internet]*. 2016;19:731–743. http://dx.doi.org/10.1016/j.chom.2016.04.017.
178. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555:210–215.
179. Finnicum CT, Beck JJ, Dolan CV, et al. Cohabitation is associated with a greater resemblance in gut microbiota which can impact cardiometabolic and inflammatory risk. *Bmc Microbiology*. 2019;19:1–10.
180. Koo H, Hakim JA, Crossman DK, et al. Sharing of gut microbial strains between selected individual sets of twins cohabitating for decades. *PLoS One*. 2019;14:1–13.
181. Korpela K, Costea P, Coelho LP, et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res [Internet]*. 2018;28:561–568. http://genome.cshlp.org/lookup/doi/10.1101/gr.233940.117.
182. Raman R, Raguram S, Venkataraman G, et al. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods [Internet]*. 2005;2:817–824. http://www.nature.com/articles/nmeth807.
183. Cohen P. The origins of protein phosphorylation. *Nat Cell Biol*. 2002;4(5):E127–30. doi:10.1038/ncb0502-e127.

184. Cortassa S, Caceres V, Bell LN, et al. From metabolomics to fluxomics: a computational procedure to translate metabolite profiles into metabolic fluxes. *Biophys J [Internet]*. 2015;108:163–172. http://dx.doi.org/10.1016/j.bpj.2014.11.1857.
185. Wild CP. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology and Prevention Biomarkers*. 2005;14:1847–1850.
186. Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychol Med [Internet]*. 2007;37:615. http://www.journals.cambridge.org/abstract_S0033291706009524.
187. Vinkhuyzen AAE, van der Sluis S, de Geus EJC, et al. Genetic influences on 'environmental' factors. *Genes, Brain Behav [Internet]*. 2010;9:276–287. http://doi.wiley.com/10.1111/j.1601-183X.2009.00554.x.
188. Wray NR, Lee SH, Mehta D, et al. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry Allied Discip*. 2014;55:1068–1087.
189. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores. *Epigenetics [Internet]*. 2020;15:1–11. https://doi.org/10.1080/15592294.2019.1644879.
190. Shah S, Bonder MJ, Marioni RE, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet*. 2015;97(1):75–85. doi:10.1016/j.ajhg.2015.05.014. PMID: 26119815; PMCID: PMC4572498.
191. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenet*. 2014;6:4. https://doi.org/10.1186/1868-7083-6-4.
192. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371–384. doi:10.1038/s41576-018-0004-3.
193. Palma-Gudiel H, Eixarch E, Crispi F, et al. Prenatal adverse environment is associated with epigenetic age deceleration at birth and hypomethylation at the hypoxia-responsive EP300 gene. *Clin Epigenetics*. 2019;11:1–10.
194. Sillanpaä E, Laakkonen EK, Vaara E, et al. Biological clocks and physical functioning in monozygotic female twins. *BMC Geriatr*. 2018;18:1–7.
195. Sillanpää E, Ollikainen M, Kaprio J, et al. Leisure-time physical activity and DNA methylation age—a twin study. *Clin Epigenetics*. 2019;11:1–8.
196. Starnawska A, Tan Q, Soerensen M, et al. Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Transl Psychiatry [Internet]*. 2019;9. http://doi.org/10.1038/s41398-019-0548-9.
197. Starnawska A, Tan Q, Lenart A, et al. Blood DNA methylation age is not associated with cognitive functioning in middle-aged monozygotic twins. *Neurobiol Aging [Internet]*. 2017;50:60–63. http://dx.doi.org/10.1016/j.neurobiolaging.2016.10.025.
198. Remondini D, Intrator N, Sala C, et al. Identification of a T cell gene expression clock obtained by exploiting a MZ twin design. *Science Reports*. 2017;7:1–8.
199. Akker van EB, Trompet S, Wolf J, et al. Predicting biological age based on the BBMRI-NL 1H-NMR metabolomics repository. *bioRxiv*. 2019.
200. Heath AC, Kessler RC, Neale MC, et al. Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics*. 1993;23:29–50.
201. Duffy DL, Martin NG. Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. *Genetic Epidemiology*. 1994;11:483–502.

202. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet [Internet]*. 2014;23:R89–R98. https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddu328.
203. Burgess S, Bowden J, Fall T, et al. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology (Cambridge, Mass.)*. 2017;28:30–42.
204. Minică CC, Dolan CV, Boomsma DI, et al. Extending causality tests with genetic instruments: an integration of mendelian randomization with the classical twin design. *Behav Genet [Internet]*. 2018;48:337–349. http://dx.doi.org/10.1007/s10519-018-9904-4.
205. Sugden K, Hannon EJ, Arseneault L, et al. Patterns of reliability: assessing the reproducibility and integrity of DNA methylation measurement. Patterns [Internet]. 2020;1:100014. https://doi.org/10.1016/j.patter.2020.100014
206. Olivier M, Asmis R, Hawkins GA, et al. The need for multi-omics biomarker signatures in precision medicine. *International Journal of Molecular Sciences*. 2019;20(19).
207. Hillmer RA. Systems Biology for Biologists. In: True-Krob HL, ed.; 2015. *PLOS Pathog [Internet]*. 11:e1004786. https://dx.plos.org/10.1371/journal.ppat.1004786.
208. Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights [Internet]*. 2020;14:117793221989905. https://doi.org/10.1177/1177932219899051.