Genetic factors affecting pregnancy duration in humans

Jonas Bačelis

Department of Obstetrics and Gynecology
Institute of Clinical Sciences
Sahlgrenska Academy
University of Gothenburg
Gothenburg, Sweden, 2018

UNIVERSITY OF
GOTHENBURG

Cover illustration by Mykolas Jocys
Associative depiction of genetic factors in action

Genetic factors affecting pregnancy duration in humans

jonas.bacelis@gu.se

ISBN 978-91-7833-175-8 (PRINT)
ISBN 978-91-7833-176-5 (PDF)

Printed in Gothenburg, Sweden 2018
BrandFactory

This is rather as if you imagine a puddle waking up one morning and thinking, "This is an interesting world I find myself in — an interesting hole I find myself in — fits me rather neatly, doesn't it? In fact it fits me staggeringly well, must have been made to have me in it!" This is such a powerful idea that as the sun rises in the sky and the air heats up and as, gradually, the puddle gets smaller and smaller, frantically hanging on to the notion that everything's going to be alright, because this world was meant to have him in it, was built to have him in it; so the moment he disappears catches him rather by surprise. I think this may be something we need to be on the watch out for.

Douglas Adams. The Salmon of Doubt

Abstract

This thesis investigates the mechanisms behind human pregnancy duration. Too short gestation is a direct cause of perinatal, neonatal, and infant mortality. Deviation from normal pregnancy length is also associated with a child's morbidity, even in the adulthood. The mechanisms determining pregnancy duration are not understood well enough to design an effective preterm birth prevention method, nor a method that would prevent preterm birth sequelae. The three included studies use genomic and epidemiological methods to contribute to our understanding of causal factors triggering birth.

Study I is a hypothesis-free genome-wide search for genetic variants affecting gestational age at birth. The study uses genotyped mothers (n=1921) and children (n=1199) from a Norwegian cohort MoBa. While finding no statistically significant associations, the study empirically shows that the top implicated loci are enriched in genes biologically relevant to the field of obstetrics and gynecology, and that the enrichment is mainly caused by infection/inflammation-related genes.

Study II explores whether a well-known association between maternal height and duration of pregnancy could be causally linked. It utilizes a novel adaptation of Mendelian randomization, which is based on the non-transmitted maternal haplotype and its polygenic risk score for human height. With the help of genomic data from 3485 mother-child pairs from Nordic countries, the study confirms the causal relationship.

Study III follows up on the findings from the Mendelian randomization study, this time using non-genetic epidemiological data to explain the mechanism behind the causal relationship. A uterine distention hypothesis is formulated and tested by comparing the expected and observed patterns of interaction between fetal growth rate, maternal height and the child's gestational age at birth. The twin (n=2846) and singleton (n=527 868) data is obtained from the Swedish Medical Birth Register. Since the observed and expected interaction patterns agree with each other, the study concludes that uterine distention is likely to be one of the causal mechanisms regulating pregnancy duration.

Keywords: gestational age at birth, preterm delivery, preterm birth, genome-wide association study, GWAS, enrichment, Mendelian randomization, causality, uterine distention, interaction.

Sammanfattning på svenska

Denna avhandling undersöker mekanismerna bakom graviditetens längd hos människa. Förtidsbörd är den främsta orsaken till perinatala och neonatala komplikationer och dödligheten hos barn upp till 5 år. Avvikelse från normal graviditetslängd är också associerad med barnets sjuklighet, även upp i vuxen ålder. De mekanismer som bestämmer graviditetslängd hos människa förstås inte i tillräcklig omfattning för att man skall kunna utforma en effektiv strategi för att förebygga förtidsbörd eller dess följder. De tre inkluderade studierna i den här avhandlingen använder genomiska och epidemiologiska metoder, för att bidra till ökad förståelse av orsakssamband till varför förlossningen hos människa startar vid en viss tidpunkt.

Studie I är en hypotesfri undersökning av hur olika genetiska varianter påverkar graviditetslängden vid förlossning. Studien använder sig av genotypade mammor (n=1921) och barn (n=1199) från en norsk kohort (Den norska mor-barn studien, MoBa). Trots att inga statistiskt signifikanta associationer hittades, visar studien ändå att det är främst loci i gener som är biologiskt relevanta inom området för obstetrik och gynekologi, och att de är huvudsakligen anhopade i infektionsrelaterade gener.

Studie II undersöker huruvida en välkänd koppling mellan mammans längd och graviditetens varaktighet kan ha ett orsakssamband och inte bara en epidemiologisk association. Den utnyttjar en ny variant av så kallad Mendelsk randomisering, som är baserad på den icke-transmitterade maternella haplotypen och dess beräknade genetiska risk för mammans längd. Med hjälp av genomisk data från 3485 mor-barn par från nordiska länder bekräftar studien orsakssambandet.

Studie III följer upp resultaten från den Mendelska randomiseringsstudien. Denna gång med icke-genetiska epidemiologiska data för att förklara den bakomliggande mekanismen till orsakssambandet. En hypotes testas om att det är livmoderns utspänning som är en av mekanismerna. Genom att jämföra de förväntade och observerade mönstren av interaktion mellan fetal tillväxt, mammans längd och graviditetens längd vid förlossningen finner man ett sådant samband. Information om enkelbörder (n=527 868) och tvillinggraviditeter (n=2846) erhölls från det svenska medicinska födelseregistret. Eftersom de observerade och förväntade interaktionsmönstren överensstämmer med varandra, drar studien slutsatsen att livmoders utspänning sannolikt kommer att vara en av de kausala mekanismerna som reglerar graviditetens varaktighet hos människa.

List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Bacelis J, Juodakis J, Sengpiel V, Zhang G, Myhre R, Muglia LJ, Nilsson S, Jacobsson B. **Literature-informed analysis of a genome-wide association study of gestational age in Norwegian women and children suggests involvement of inflammatory pathways**.
PLOS One, 2016. 11(8): e0160335. [doi:10.1371/journal.pone.0160335]

II. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, Helgeland Ø, Johansson S, Myhre R, Sengpiel V, Njølstad PR, Jacobsson B, Muglia L. **Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis**.
PLOS Medicine, 2015. 12(8): e1001865. [doi:10.1371/journal.pmed.1001865]

III. Bacelis J, Juodakis J, Adams Waldorf KM, Sengpiel V, Muglia LJ, Zhang G, Jacobsson B. **Uterine distention as a factor in birth timing: retrospective nationwide cohort study in Sweden**.
BMJ Open, 2018. 0:e022929 [doi:10.1136/bmjopen-2018-022929]

Content

01 ABBREVIATIONS

1. Abbreviations

DNBC	Danish National Birth Cohort
FIN	Finnish cohort
GA	gestational age at birth
GWAS	genome-wide association study
IVF	in vitro fertilization
LGA	large for gestational age
LMP	last menstrual period (also a method for GA dating)
MAF	minor-allele frequency
MoBa	Norwegian Mother and Child Cohort
MR	Mendelian randomization
PROM	prelabor rupture of fetal membranes
PTD	preterm delivery (also, preterm birth)
QC	genotyping-data quality control
SGA	small for gestational age
SNP	single-nucleotide polymorphism
UL	ultrasound method for GA dating (from Swedish "**u**ltra**l**jud")

02 TO THE READER

2. To the Reader

Biomedical science is volatile. Its methods evolve ever so quickly. In ten years from now, it might be difficult to imagine what it was like to do science in 2018. The professor, who introduced me to the field of genomics back in 2010, once told a joke about how all experiments that took her an entire PhD to perform, today could be repeated overnight. I feel that a reflection on a current context could make this thesis more inviting. So, to those who read this in the future, I would like to describe the today of my present.

The "dark ages of DNA" are now over, symbolically marked by a recent death of its last doge - Luigi Luca Cavalli-Sforza[*]. His was an epoch of blood-group population genetics, single-marker association studies, and Mendelian phenotypes. Now we find ourselves in the Renaissance of the Genetic Era[1]. In it, reading a full human genome using sequencing technology is fast and affordable, although it still costs around 70 hourly wages[†]. Full-genome association studies are published at a rate of 4000 per year, but the largest genotyped cohorts still have fewer than 1 million humans. Today, an typical genomics researcher has an access to computational power equivalent to 200 billions instructions per second; quantum computing is in its fetal stage and not practical yet. Just recently, as part of ancient-DNA revolution we learned about geographical as well as intimate adventures of our ancient ancestors and their cousins Neanderthals and Denisovans - only by sequencing their 50,000-year-old bones[1]. Genome editing, too, has made a huge leap with CRISPR/Cas9 technology - early clinical trials on humans are already taking place. Today we are wondering what will come first - a widely spread therapeutic gene editing or a wild spread of (currently) extinct woolly mammoth[2]. Last year, *de novo* genome synthesis and assembly has also reached a milestone of 1MB[3]. An average person today is already worried about the climate change, some have concerns about the development of artificial intelligence, but most are still oblivious of potential dangers of engineered gene drive systems[4] or terrorists using synthetic pathogens[5]. The cost of developing a prescription drug that gains market approval is equivalent to a budget of 42 flights to the geosynchronous transfer orbit with 8 tons of payload each[‡].

As a contrast, the field of obstetrics has been stagnant for at least a decade. The non-invasive prenatal tests using cell-free DNA to screen for trisomies are now commercially available but have not replaced amniocentesis. In some regions, magnesium neuroprotection and progesterone[6] treatments have been introduced into clinical routine, as well as fibronectin test[7]. Nonetheless, one in ten babies on the planet are still born too early. For those born very early, the mortality rate gap between high- and low-income countries is currently 10% vs 90%.

---

[*] *Dr. Cavalli-Sforza [25 January 1922 - 31 August 2018] died two weeks after I decided to honour him in a metaphor.*

[†] *with current median income in Sweden.*

[‡] *SpaceX's Falcon-9 launch prices for 2018.*

03 THE PHENOTYPE

3. The Phenotype

3.1. Gestational Age

3.1.1. Meaning, synonyms and units

During conversations with people outside the field of medicine, I have noticed that the most convenient way to describe the object of my research is "the time a baby spends inside the mother's womb". Despite being perfectly inaccurate, this definition is where we should start the journey. There are other technical names to call it: "a child's gestational age at birth", "gestational duration", "gestational length", "pregnancy duration", "pregnancy length", and "timing of birth". As there is no strict consensus on the preferred term, I will use them interchangeably, with a slight bias towards my favourite - "gestational age" (GA, with an implicit note that this age is evaluated at birth).

Gestational age can be evaluated in time units - months, weeks or days. That depends on a context. While it is convenient to refer to "months" in casual conversations, obstetricians usually refer to weeks and academic scholars use the smallest unit of measurement practically available - days. As this might relay an impression of perfect precision, I would like to stress that this number is rarely correct. In fact, it is almost never correct. The first day of an organism's existence is the day a sperm fertilizes an egg. In most practical settings involving humans, the fertilization time is not known[*], thus the time difference between conception and birth can only be guessed, guesstimated, estimated - but not measured.

3.1.2. Methods of estimation

Two time points are required to determine GA. The date of birth is always known to the accuracy of minutes. Determining the date of fertilization, on the other hand, is tricky. All currently applied GA evaluation methods try to estimate the day of conception using extraneous signs, as fertilization event itself is ethically undetectable in natural human pregnancies. The estimations are based on various assumptions that are not guaranteed to hold.

The last menstrual period (LMP) method assumes that ovulation occurs on the 14th day (or mid-cycle) after the LMP. This is rarely the case due to personal variation and population variation in menstrual cycle length (short, long, irregular). Another assumption is that the fertilization day is on the same day as ovulation (although it is pretty accurate). This method also relies heavily on a recall accuracy of the self-reported LMP date.

[*] *An exception could be IVF pregnancies.*

The ultrasonographic (UL) measurement method uses fetal growth curves previously derived by a combination of UL and LMP methods in women with very regular menstrual cycles. UL method assumes that, while in early developmental stages (e.g., first trimester), all fetuses of the same true age have no variation in fetal size (e.g., crown-rump length, biparietal diameter). It also relies on the reference population, equipment accuracy and personnel skills (intra-operator and inter-operator variability).

Other methods rely on detection of the ovulation event (by monitoring basal body temperature or changes in the hormone levels) or the implantation event (monitoring human chorionic gonadotropin levels). These methods also assume that fertilization date can be reliably inferred from ovulation or implantation dates[8]. But most importantly, they require a meticulous personal longitudinal record keeping, thus are not a part of a standard medical practice.

A very important caveat remains to be stated. Even if the true date of fertilization is known, GA is not recorded as the true time difference between fertilization and birth event (as would be more than reasonable to do). A conventional "correction" of 14 days would be added to create a compatibility with historical records that only registered uncorrected LMP date as the starting point of pregnancy.

3.1.3. Which method of estimation is the best?

When compared among each other, the four methods of estimating gestational length show the following order of accuracy: ovulation > implantation > UL > LMP[9]. However, due to practical reasons, only LMP and UL can be considered in standard medical practice, as the other two methods would require daily monitoring and would dependent on the skills* and determination of the woman herself.

UL is currently the most common GA dating method in Sweden; over the last four decades it has gradually pushed the LMP method **(Figure 1**) to a relative obscurity. In the three studies covered by this thesis, the GA data was mostly or exclusively generated by UL method. The second-trimester UL scan measures fetal head circumference via the biparietal diameter and the occipital-frontal diameter. This method is less accurate than the first-trimester UL measurement of the crown-rump length.

When compared to the true known GA in IVF pregnancies, the UL dating was found to differ by +/- 8 days (range; n=1268)[10]. The data available to us from Studies I and II (MoBa) show similar inaccuracy: 95% of the differences between the true GA in IVF pregnancies and UL method were between -7 and +6 days (**Figure 2**).

* *The cost-effective method involves meticulously registering menstrual cycles, sexual intercourses, and regular monitoring of basal body temperature and consistency of cervical mucus.*

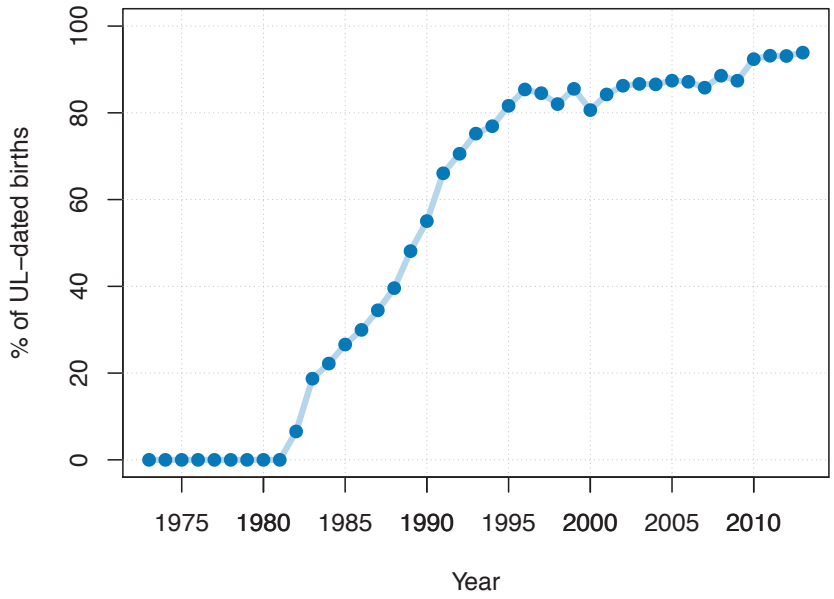


Figure 1. *The percentage of births in Sweden dated using purely ultrasonography increased over time due to a combination of desirable features of this method: accuracy and practicality. Currently, ultrasonography is by far the most common method used to estimate gestational age in the developed world. In Sweden, the UL-based dating was introduced in 1982. In 1990, more than 50% pregnancies were dated using this method and the trend has increased ever since.*

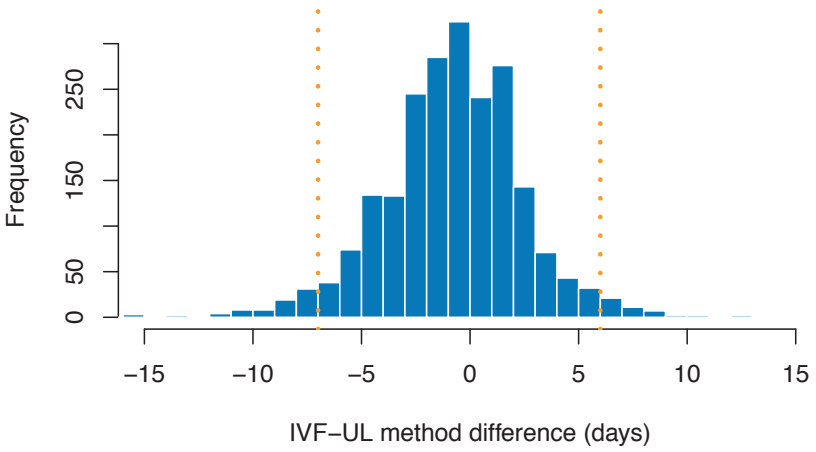


Figure 2. *Comparison of known gestational age and UL-estimated gestational age from IVF-conceived pregnancies. Orange lines indicate the range of differences that contains 95% of all observations. Data from the Norwegian Mother and Child cohort, N=2169. As expected, IVF GA was on average 16 days shorter than UL GA, thus differences were centred to zero.*

3.1.4. When do women deliver?

In Sweden, 50% of all pregnant women give birth in a two-week window ranging from 273 to 286 days (39 to 41 weeks) of gestation, the mean GA is 278 days and the median is 280 days (**Figure 3**). The distribution of GA is left-skewed, which means that there are more early births than there are late births. Both extremes of gestational age (preterm and post-term delivery) increase the risks for the mother's and baby's health and life.

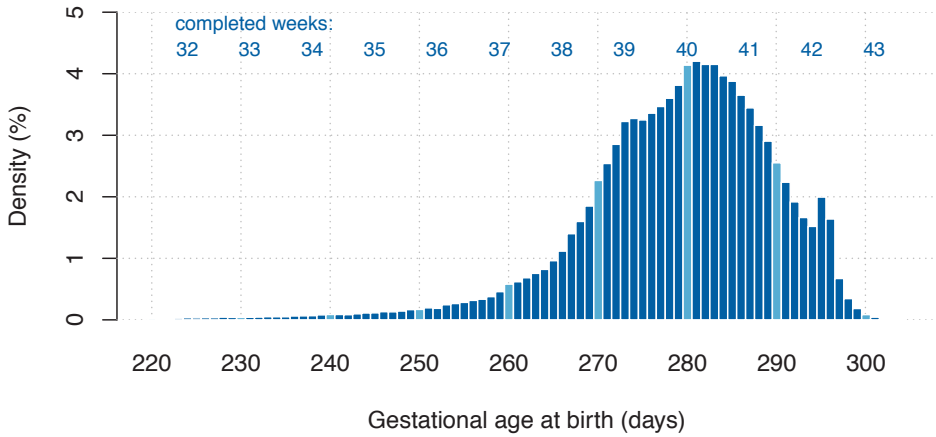


Figure 3. Distribution of child's gestational age at birth. Data from the Swedish Medical Birth Register (2010-2013). Spontaneous and medically induced deliveries, only one observation per pregnancy (e.g., only one of two twins were counted). Gestational age was evaluated using UL method. N=293912.

3.1.5. Why do we care about this number?

In fetal development, maturation monotonically increases with gestational age. Preterm born infants are poorly adapted to the extra-uterine life. The earlier the birth - the lower the probability of survival[11]. Extremely preterm babies are born with severely underdeveloped brains, digestive systems, and lungs. Moreover, somewhere in the 22nd week of gestation there is the so-called "limit of viability", which refers to the minimum gestational age at which a baby currently can survive outside the womb.

It is tempting to rush to the conclusion that the best time to be born is as late as possible. However, a post-term birth also bears death-related risks, not only to a child, but to a mother too - via complicated birth.

The major part of obstetrics is about defining the best time of delivery for both mother and child, aiming to either prolong gestation or induce delivery for some medical reason.

3.2. Preterm Birth

3.2.1. Definition

Preterm birth, or preterm delivery (PTD), is historically defined as a childbirth occurring at less than 37 completed weeks (259 days) of gestation. There are other more nuanced classifications of gestational duration (**Table 1**); however, they will not be used in this thesis.

Table 1. Extended classification of gestational age.

	Gestational age at birth	
Categories	Completed weeks	Gestational day
Post-term birth	$42^{0/7}$ -	294 -
Late term	$41^{0/7}$ - $41^{6/7}$	287 - 293
Full term	$39^{0/7}$ - $40^{6/7}$	273 - 286
Early term	$37^{0/7}$ - $38^{6/7}$	259 - 272
Moderate or late preterm birth	$32^{0/7}$ - $36^{6/7}$	224 - 258
Very preterm birth	$28^{0/7}$ - $31^{6/7}$	196 - 223
Extremely preterm birth	- $27^{6/7}$	- 195

Based on[12,13]. The nomenclature of GA is typically discussed in terms of the number of "completed weeks", but in statistical analyses we have uses gestational days.

The current PTD definition only provides a standardized language but lacks medical or biological meaning. The chosen threshold for gestational age is arbitrary, as the earlier the separation line, the grimmer the birth outcomes. The major transition in terms of needing special care occurs between 34 and 37 weeks[14]. Some also suggest that the current threshold does not serve a useful purpose, because it does not coincide with functional maturity, thus should be shifted to 39 weeks[15].

In general, "preterm" should be distinguished from "premature", which describes a lack of completed fetal development[16]. To exemplify the importance of this distinction: preterm born Black and Asian infants (compared to white European infants) have higher fetal maturity, even though PTD rates in these ethnicities are higher[17]. Unfortunately, approximating maturity by gestational days is much more scalable (simple, cheap, familiar, universal) than quantifying the maturity.

The rate of PTD is estimated as all live births before 37 completed weeks (whether singleton, twin, or higher order multiples) divided by all live births in the population.

The best estimate of global PTD rate is 11.1%, although country-wise rates range from 5% to 18%[18].

The major classification of PTD includes two groups: (1) spontaneous preterm delivery and (2) provider-initiated preterm delivery (defined as induction of labor or elective Caesarean section before 37 completed weeks of gestation for maternal or fetal indications or other non-medical reasons). The second group used to be called "iatrogenic". Since provider-initiated preterm births are regionally and temporally dependent on public policies and developmental level of medical care, we often exclude these types of PTD from analyses. In all three studies covered by this thesis similar action was taken.

It is useful to mention the subclassification for the first group based on how the delivery starts: (1a) spontaneous labor with intact membranes, (1b) preterm prelabor rupture of the membranes.

3.2.2. Consequences

Mortality

Preterm birth is the leading cause of child deaths worldwide: according to the latest global estimate, 15.4% of all the deaths before age 5 were a direct cause of preterm birth[19]. From the year 2000 to 2013, global child mortality dropped; however, the rate of reduction attributable to PTD was one of the smallest out of 17 death causes[19].

Most of these lives are lost during the challenging neonatal period (28 first days of extrauterine life). The common reasons of neonatal death in preterm-born babies are respiratory distress syndrome (breathing difficulty caused by deficiency of surfactant, also known as hyaline membrane disease), bronchopulmonary dysplasia (due to prolonged mechanical ventilation and supplemental oxygen), necrotising enterocolitis (seen almost exclusively in preterm infants), intracranial non-traumatic hemorrhage (with no history of birth or post delivery trauma)[20]. In Level 1 income countries[21], the list expands to neonatal infections, hypothermia, and malnutrition.

Over the last five decades, due to advancements in medical care, in high-income countries neonatal mortality rate has significantly dropped in every strata of gestational age[22]. But this has also widened the survival gap between high- and low-income countries, which is currently 90% vs 10%[23].

During the neonatal period, preterm-born babies can experience retinopathy of prematurity caused by oxygen toxicity (supplementary oxygen received at neonatal intensive care unit), which leads to hypoxia and abnormal blood vessel development in the retina. Even though not deadly, this condition will lead to blindness or severe myopia, thus contributing to morbidity later in life.

Premature infants have a very high readmission rate in the three months after discharge[24]. Readmission is often related to jaundice[25], also to respiratory infection[26].

Morbidity

Preterm-born children often have a lifetime of significant disability. On a global level, of those who survive beyond the first month, 2.7% are estimated to have moderate or severe neurodevelopmental impairment, and additional 4.4% to have mild neurodevelopmental impairment[20]. An estimated 31% of preterm-born children have at least one of the problems: cognition impairments, general developmental delay or learning difficulties; cerebral palsy; impaired vision or blindness, gross motor and coordination impairments; deafness or hearing loss; epilepsy; behavioural problems (sorted by decreasing frequency)[27]. Cognitive and neurologic impairments were still evident at starting school age[28].

In Sweden, 36.1% of extremely preterm children had no disability, 30.4% had mild disability, 20.2% had moderate disability, and 13.4% had severe disability (evaluated at 6.5 years of age; includes cerebral palsy, vision, hearing, and cognitive disability)[29].

At age 11, preterm-born kids significantly more often had functional limitations, compensatory dependency needs, and services above those routinely required by children[30]. A lower mean intelligence quotient[31] and a decline in mean intelligence quotient over time in childhood is documented[32]. In the adulthood, of those who are born preterm, significantly fewer complete the high school or university, more of them receive Social Security benefits and have medical disabilities severely affecting working capacity, less have a high job-related income, less get married or have a partner and less become biological parents[22].

3.2.3. Obstetric care

First of all, it is worth mentioning the existing **preventive measures**. These include generic common-sense recommendations to women, such as leading a healthy life style, good nutrition, physical activity, vitamins and supplements, emotional health[33]. In practice, it would be naive to expect a high effectiveness of such guidelines. They are hard to adhere to. But if we were to imagine a "platonic pregnancy cohort" in which every woman follows the World Health Organisation recommendations[33], we would be likely to find much lower PTD rates than in the real world[34].

Besides prevention, there is **prediction**. It could be useful to have a tool that identifies women at risk. This would give doctors more time to act, would assure that women are monitored and do not deliver at home by accident. Currently, the best predictors of preterm birth are personal history of PTD, cervical length[35], and fibronectin[36]. The composite predictive model is far from perfect[37] and does not help to prevent PTD nor improve perinatal outcome[7].

Lastly, there is **treatment**. Or rather, there is no good treatment. In preterm birth, two complementary directions could be mentioned: (1) reduction of adverse consequences to the fetus, and (2) prolongation of pregnancy. In (1), there are corticosteroids that accelerate maturation of fetal lungs (essential for fetal viability) and brain, also neuroprotective effects of magnesium sulphate (decreased incidence and severity of cerebral palsy, neonatal intraventricular hemorrhage and periventricular leukomalacia)[38]. In (2), there are tocolytics (labor suppressants) that postpone delivery to some extent, allowing the corticosteroid treatment.

The effectiveness of tocolysis is arguable[39,40]. Antibiotics may too prolong the pregnancy for a couple of days but their use is associated with higher rate of neonatal necrotizing enterocolitis[41], almost double risk of cerebral palsy[42], and the benefit of antibiotics used prophylactically in a general population or therapeutically in preterm labor with intact membranes is neither proven nor recommended[42].

It is generally considered that progesterone administration reduces the rate of preterm birth and improves neonatal outcome. However, progesterone has only been shown to improve child outcome in risk pregnancies: women with short cervical length (less than 2% of pregnant women), or for women with previous PTD (less than 10% of pregnant women). The remaining majority is not recommended progesterone and thus remains at risk[43]. Latest studies show that progesterone is generally ineffective[6].

To summarize, modern obstetrics does not have a solution on how to regulate pregnancy duration. At least, not to the extent which would allow the prevention of preterm birth. Due to advancements in obstetrics and neonatology, more children that would otherwise be born dead due to extreme prematurity are now born alive (**Figure 4**), although only a small fraction of survivors are expected to live a regularly healthy life. One positive note: human clinical trials might soon be approved to test the extra-uterine system recreating the intrauterine environment (artificial womb, or "baby bag"). This system was recently shown to improve the condition of preterm-born lambs[44,45].

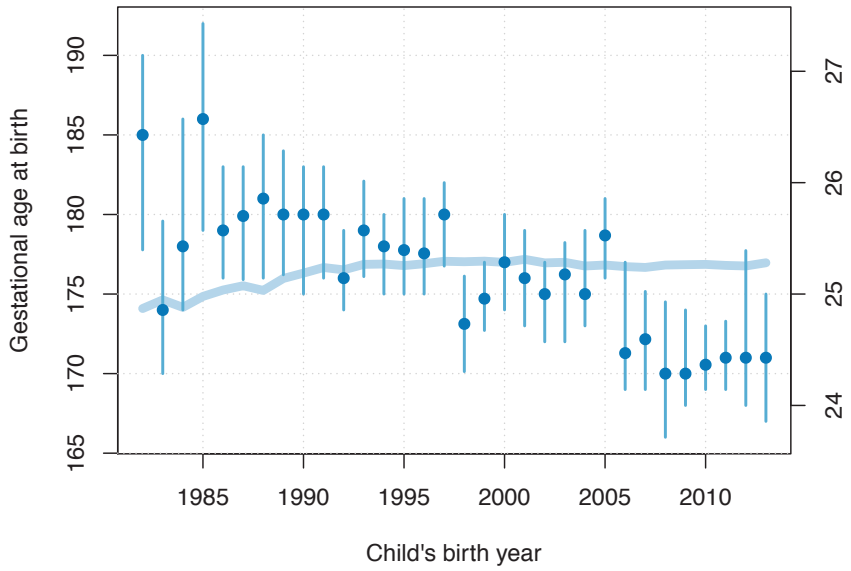


Figure 4. *The decreasing trend of gestational age of live-born children in the 0.05th percentile of gestational age. Swedish Medical Birth Register (1982-2013), spontaneous and medically induced pregnancies with at least one live-born child and UL-based GA dating. The numbers on the right indicate gestational age in completed weeks. Dark dots are the 0.05th percentile of gestational age, and vertical lines are the 95% confidence intervals estimated using bootstrap method. For comparison, the thick curve in the background represents the mean gestational age of the population lowered by 101.6 days to match the mean of 0.05th percentile data.*

3.3. Post-term birth

Post-term delivery is defined as delivery after $41^{6/7}$ completed weeks of gestation (294 days and further). During the post-term period the fetus is at a higher risk for intrauterine death, hypoxia and subsequent meconium aspiration syndrome. With larger fetus, delivery might be complicated by obstructive labour, shoulder dystocia, plexus injuries in the baby and pelvic floor injuries in the mother. The post-term delivery is arguably a lesser problem than the preterm birth, as it is easier to induce delivery than to make gestation last longer.

3.4. Evolutionary context

Allometric scaling studies suggest that human gestation is shorter relative to other primates and that 18–21 months would be required for humans to be born at neurological and cognitive developmental stage equivalent to that achieved by a chimpanzee neonate[46]. This is thought to be caused by two phenotypic shifts, both favoured by natural selection: a shift from tree-climbing to bipedal locomotion, and an increase in the brain size and cranial volume. Considered separately, both shifts are advantageous; however, acting together they impose a new threat to the evolutionary fitness, because there is a physical limit to which the outlet size of maternal pelvic bones and fetal head size can vary without causing complicated childbirth. The "Obstetric Dilemma" hypothesis suggests that mutations shortening the time of gestation were favoured by the natural selection in order to avoid physical constraints during childbirth[47]. As a consequence, human neonates are born in a completely parent-dependent state (altriciality), the birth involves complicated head/shoulder rotations and unique occipitoanterior birth position, which demands assistance during a delivery. It might be that selective pressure towards shorter pregnancy did not push all involved mutations to fixation. In other words, some DNA positions that have an effect on pregnancy length might still contain variation in the human population: some individuals having the ancestral "long gestation" alleles and others having new "short gestation" alleles. If that is the case, we should be able to identify these genetic variants using genotype-phenotype association analysis (Study I).

There are also other evolutionary forces at play. Long after the ancestral branches of humans and other primates have split, stabilizing natural selection must still be active. In prehistoric "natural" conditions, mothers delivering at far tails of gestational age distribution would have lower evolutionary fitness: without medical assistance, preterm-born children would rarely survive; post-term birth would be a serious mortality risk to the mother and (because of altriciality) to the newborn. Mutations and common genetic variants that determine extreme gestational age must have been selected against. Stabilizing natural selection must be favoring alleles that increase the likelihood of delivery at gestational age, which we with hindsight call "term". In modern times, due to obstetrics and neonatology, stabilizing natural selection has very limited power to swipe out the risk alleles. With every generation, the number of old and new (*de novo*) mutations accumulates and the population becomes more *genetically* susceptible to preterm and post-term delivery.

3.5. Environmental factors

Even though human gestational length is still an unsolved mystery[48], over the years numerous observations have been made about conditions and circumstances, under which pregnancy duration tends to be longer or shorter than normal. I will first describe the group of factors that could be called environmental and in the next subchapter I will expand on the genetic ones.

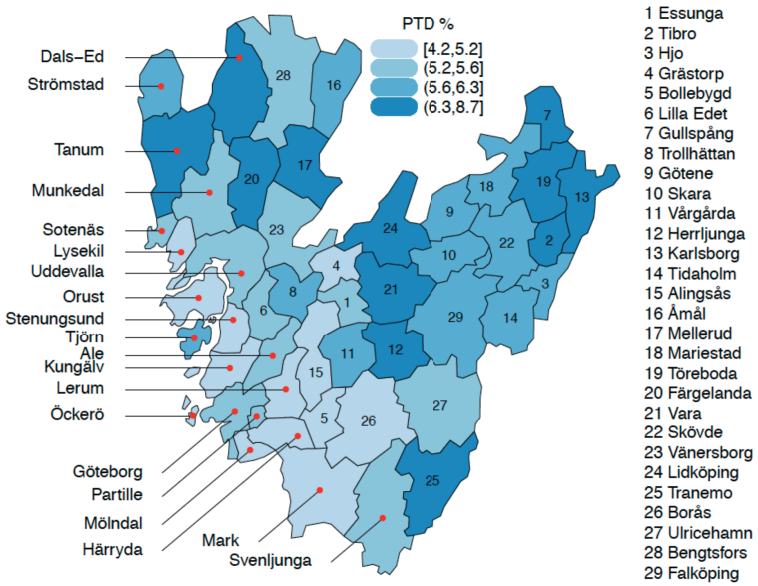


Figure 5. *Preterm delivery rates in Västra Götaland County. Only 1998-2013 births of Sweden-born mothers with Swedish nationality were used. Gestational age was evaluated only using UL method. N=190892, ANOVA p<2.2e-16. Results from an on-going study. The area with the highest PTD rate was Gullspång (39/450 = 8.6%), and the area with the lowest rate was Bollebygd (45/1078 = 4.2%).*

A good example illustrating the existence of non-genetic component contributing to the variation in gestational age is a regional map. In **Figure 5** (as well as our published work[49]) the map is colored by preterm birth incidence rate. Since only the pregnancies of Sweden-born Swedish nationals were used, the population is homogeneous and depleted from genetic factors that tend to segregate geographically (e.g., race). While genetic profile could be assumed uniform, each community has a specific environmental profile: some areas are next to the ocean, many are serviced by different water-cleaning stations, each has a different level of air pollution, micro-climate etc. Since the number of environmental differences is immense and genetic homogeneity is strong, the significant differences in PTD rates should in large part be explained by variation in the environmental exposures between these geographical areas. In other words, some factors are environmental.

There is a number of known environmental factors associated with shorter gestational age: physical traumas[50], physical exertion[51-53], malnutrition[54], infection[55,56], mental stress[57], smoking[58], also their proxy - low economical status and

low education[59,60]. Multiple pregnancy, IVF[61], adolescent pregnancy or advanced parental age[62], short inter-pregnancy interval[63] - all shorten child's gestational age at birth and could be considered as environmental factors.

When compared to the population average, mothers who use supplements (e.g., folate[64]) and eat healthy will deliver slightly later.

Importantly, for many of the aforementioned risks, the cause-and-effect relationship (causality) has only been suggested and not proven. A thorough causal inference is a very tricky task when there is no ethical possibility to conduct a randomized controlled trial (i.e., an experiment). Studies II and III in this thesis are dedicated to the causality question.

As a side note, many of the factors listed above could also be classified as genetic factors. For example, one must recognize that tobacco and alcohol consumption are dependent on genetics[65,66], educational attainment is also partly a genetic trait[67], as well as infection (due to genetic susceptibility[68]) and maternal stress (via genetic propensity for anxiety[69]).

3.6. Genetic factors

Gestational age does have a familial (genetic) nature. A palette of creative methods* has been used to demonstrate this.

The first observation is that the gestational age in maternal relatives is strongly correlated: pregnancies of the same mother[70], pregnancies of a mother and her daughter[70], pregnancies of monozygotic twin-sisters[71], pregnancies of dizygotic twin-sisters[71], pregnancies of full-sisters[70], pregnancies of maternal half-sisters[70]. Such phenotypic correlation between related individuals, especially if they do not share a common environment, is an indication that genes are involved (**Figure 6)**.

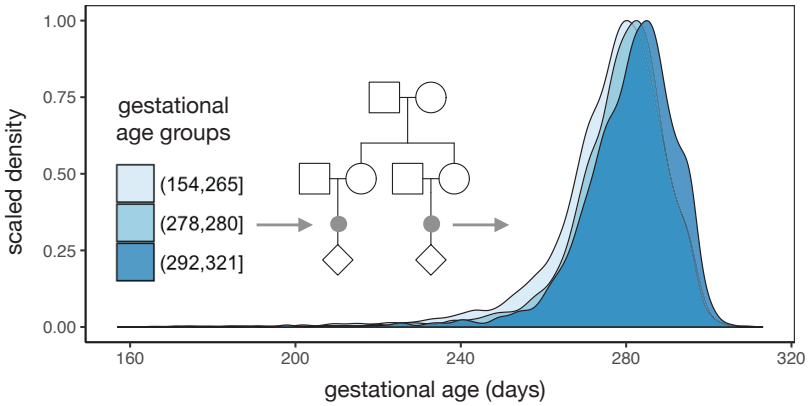


Figure 6. *Evidence for heritability in gestational age. Three equally-sized groups with 15788 pairs of maternal cousins. The group assignment is based on gestational age of the first cousin in a pair. The three overlaid density plots represent gestational age distribution of the second cousins, with original grouping preserved. Grey dots denote pregnancies. Data from the Swedish Medical Birth Register.*

* *These methods do not rely on genetic data, only on the pedigree information and the phenotype.*

To be clear, genes are always involved in everything (after all, all our bodies are built using genetic information), but what we implicitly mean by "genes are involved" is actually "the variation in genetic information can partially explain variation in the phenotype".

The second observation is that maternal and paternal genetic contributions to heritability of gestational age differ. Maternal genetic effects are much stronger than paternal[70], maybe not surprisingly, as the mother's genome can affect the pregnancy via her uterine environment, thus her genotype has more "expressive freedom" to impact the pregnancy duration as compared to the father. Since one half of the fetal genome is inherited from the father, the paternal genome can have an effect on pregnancy duration via the fetus. The paternal genetic effect is very small[72,73]. An indirect indication that fetal genes influence pregnancy length is that boys are born preterm more often than girls[20].

Thirdly, it must be mentioned that gestational age is not a Mendelian trait. No known Mendelian disorder manifests an abnormal gestational length as its primary clinical feature[74]. In other words, these phenotypes do not have a clear inheritance pattern. However, a very small fraction of PTD-affected families exhibit explicit evidence of phenotype aggregation among relatives (**Figure 7**, also[75,76]). These families are rare exceptions, implying a presence of low-frequency mutations with large effects, which are able to impair a normal progress of gestation. Despite the large penetrance, such rare genetic abnormalities do not explain any significant fraction of variance in gestational age in the population.

Gestational age, differently than Mendelian traits, falls into a category of "complex traits". This means that there are many genetic variants with small effects, rather than one single mutation with a very strong effect. Depending on the population, the heritability of gestational age is approximately 30%[71].

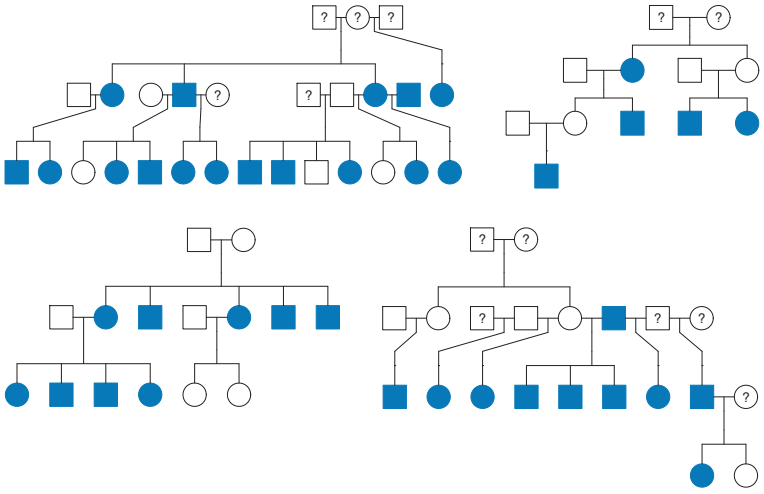


Figure 7. Examples of rare Mendelian-like patterns of PTD familial aggregation in the Swedish Medical Birth Register. Individuals born preterm are highlighted in colour. Question marks denote missing phenotypic data. In order to protect the identity of these families, the pedigrees have been slightly modified without changing the general message.

04 ESSAY ON GENETICS

Jonas Bačelis

4. Essay on Genetics

The field of genetics is broad and fascinating. Even a thick textbook might do little justice to it. This chapter presents merely a couple of basic concepts and ideas that might be helpful to the reader without a background in genetics. To anyone outside this field, I would recommend exploring it deeper. The basic principle behind the Darwin's Theory of Evolution, when honestly and fully understood, stuns the subject with profound realisations about the meaning of life. I would argue that genetics can well compete with astronomy in its potency to change a person's life by providing a grander perspective.

4.1. Human genome

The human genome is a recipe for making humans. While being a poor technical definition, this gives an opportunity to introduce and address a rather common misconception that genomes define who we are. There is a subtle but important distinction between a “blueprint” and a “recipe”. The former implies a fully deterministic, factory-like process with a guaranteed outcome and very little variation. This is not what genomes are. The ”recipe”, however, is a guideline. Its success is much dependent on, to continue the metaphor, the interpretation of “pinch of salt” or “simmer until the desired consistency”. There are no blueprints in nature, but there are recipes.

A human genome is also a recipe for disaster when trying to define it or explain it without the help of a drawing board or fancy Youtube videos. A strong warning is warranted: there are fascinating exceptions to almost every following statement.

A genome is a biological storage medium for information. It is encrypted in a code that uses an alphabet of four letters. These letters symbolize four nucleotides - monomeric chemical fragments that form a long polymeric DNA molecule. At 1% of the genome where the genes are, DNA code is decrypted into protein-building instructions with a 20-letter alphabet, where each letter represents a distinct amino acid type. The other parts of the genome are more mysterious. In general, every cell of an individual has exactly the same copy of the genome. To fully describe a person’s genome means to name approximately 3.2 billion letters residing in one cell, and then to name 3.2 billion more, because human genome consists of two halves.

One half of a person's genome is inherited from his mother while the other half - from his father. Each half could be called a "haploid" genome. Both halves are identical in their architecture, but not identical in their contents. Let us clarify this important idea. An analogy could be two identical houses, each with the same number of rooms and identical floor plan; however, furbished differently - in Modernistic and Victorian styles. Importantly, there is a direct functional correspondence between any location in house 1 and house 2, e.g., the lowest compartment always has the function of a basement. Similarly, any location on paternally inherited genome contains the same type of information as maternally inherited genome in the same location;

however, the flavours can differ. For example, the fragment affecting the eye colour will be exactly at the same location in both halves of the genome, despite encoding different colours. This perfect correspondence allows nature to swap fragments between maternal and paternal halves during the process called meiosis, thus creating new genomes.

To avoid a common misunderstanding: the existence of maternal and paternal halves of person's genome should not be confused with the concept of "double helix". Each half, being a string of ‘letters’, has an identical twin, chemically attached to it, side by side. It is identical in a sense that it contains exactly the same information, only the information is written in an inverse alphabet: each A is swapped for T, each C is swapped for G, and vice versa. For the purposes of this thesis, we will ignore the twin strands and only refer to two halves of person's genome as maternal and paternal.

The next level of complexity is chromosomes. Maternal and paternal halves of the genome are physically split into 23 chunks which we now can visualize as uninterrupted strings of letters. The chunks are of various sizes, but, as mentioned earlier, in terms of size and structure they are exactly the same in the maternal and paternal halves. One exception is the 23rd chromosome. It comes in two different architectures, i.e., contains different genes and, consequently, have different functions. One version of 23rd chromosome is called X. A person carrying two copies of X has a female sex. Another version of 23rd chromosome is called Y. It is much smaller than the X chromosome. An embryo carrying Y without X would not survive due to incomplete genomic recipe. The male sex humans carry one copy of X and one of Y.

So far, we have ignored a small fraction of the genome located in maternally inherited organelles called mitochondria, and we will continue to ignore it since this does not directly relate to this thesis.

4.2. Genomic variation

Many physiological features in which we differ are due to differences in our genomes. Even though the human genome is large, at most of its positions we do not differ among each other at all. For a genetic epidemiologist, the only interesting parts of the genome are those where there is variation between individuals.

Some rare forms of variation are fragment insertions, deletions, rearrangements, and even chromosomal duplications or omissions. But mostly, genetic differences between us are defined by single-position mutations, where some genomes have one particular letter, and some - another.

It is common to consider mutations as being something bad. That is far from the truth. The core driver of evolution is the constant supply of new random mistakes in the genome that get judged in the fitness competition. The bad ones do not survive. The beneficial or neutral ones might get passed to the next generation.

In fact, it is much better to think about mutations as "old" and "new"[77]. The old ones have spent a lot of time in human population without being eliminated by natural selection. Maybe because they were beneficial. Or perhaps because they had no detrimental effect on the reproductive fitness, and it was purely by chance, due to

genetic drift, that they became common. The new mutations are more likely to do harm. Those, which do, are likely to be rare. In all other species, any harmful mutation is already on its way to extinction. A caveat must be mentioned here: with a help of modern medicine we have placed ourselves in a position where natural selection can no longer see us as the players of survival game. Thus, it is less able to regulate the frequencies of genetic variants based on their effect on our health.

An accidental mistake in copying the genome that turns out to be a beneficial mutation will likely come to a fixation. In other words, in many generations it will fully replace the ancestral version of itself. If there are no archaic version copies left, then there is no variation and we can no longer call it a mutation or a genetic variant. Only if fixation is not reached yet and two or more versions at the same genomic position are floating in the gene pool, we can call it a genetic variant or a mutation.

Let us narrow down the genome even further, to only those variants that are discussed in this thesis. Firstly, we will exclude all rare mutations: if only 1 person in 1000 has a particular mutated genomic site, then our analyses will be statistically underpowered. The remaining part is called single-nucleotide polymorphisms (SNPs). We will also ignore those SNPs that have three or four possible letters, the so-called multiallelic mutations, which are rarely used due to analytic complexity. We are left with millions of biallelic genetic markers - those that have only two alleles. One allele is ancestral and one is relatively new.

At a SNP with two possible alleles, a person can have three genotypes. Let us see why. There are two halves of the genome. At each half, there can be only one of the two alleles, e.g., C or A. If a mother and a father did not choose each other because of their preference for a certain allele, both halves are independent. Thus, there are four possible combinations of alleles at this SNP per full (diploid) human genome: AA, AC, CA, CC; where the left-side letter indicates maternal origin, while the right-side letter - paternal. In genetic epidemiology we rarely care whether person's genotype is AC or CA[*]. If allele A is harmful, then both AC and CA individuals should have the same risk for a disease. Moreover, the current bead-chip genotyping technology is unable to differentiate between the two heterozygotes. Thus, AC and CA being "identical", we are left with three genotypes per one biallelic SNP. The matter could be much more complicated in other species (e.g., a triallelic SNP in a tetraploid organism would have 15 possible genotypes).

In general, when we talk about a SNP being associated with a phenotype or a risk for a certain disease, we simply mean that across a population, individuals in the three genotypic groups at that particular genomic site have different phenotypic means or different disease prevalence.

[*] *The distinction between parental origin of alleles will be of paramount importance in Study II of this thesis.*

4.3. Genes in action

A true causal SNP can exert its effect on a phenotype in various ways. It is convenient to classify these modes of action into three categories: additive, dominant and recessive effects. Assume that in a biallelic SNP with alleles T and G, the G allele is less common in the gene pool, thus we call it the minor allele. We often refer to the minor allele when talking about the three modes of action.

An additive genetic effect is when the effect size is *linearly* dependent on the dosage of the minor allele (how many copies of the minor allele are there: 0, 1, or 2). An example of an additive genetic effect could be when the ranking of the three genotypic groups based on their risk for a disease is: TT < TG < GG (**Figure 8**).

A recessive effect is when only individuals with two copies of the minor allele tend to get sick (or have a higher phenotypic mean). Thus, there would be no statistical difference between individuals in genotypic groups TT and TG, and only individuals in the GG genotypic group would have a differing phenotype: TT = TG < GG.

Similarly, a dominant effect is when there is no difference in how many copies of the minor allele individual has, the only important thing is that there is at least one copy: TT < TG = GG.

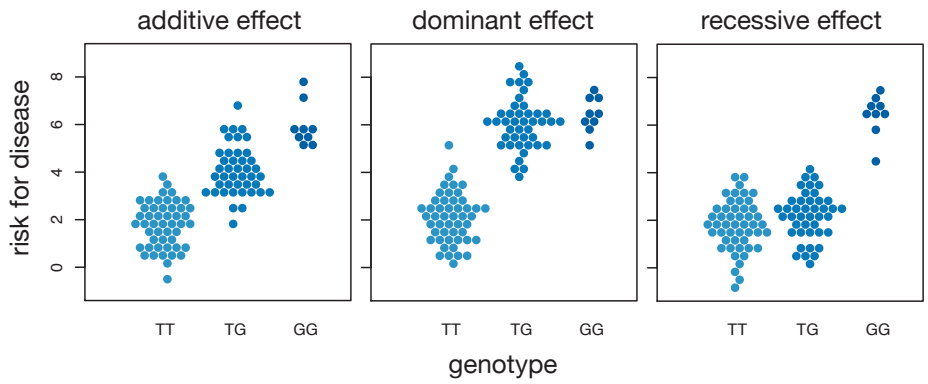


Figure 8. *Modes of genetic action with respect to the minor allele (G allele frequency is 30%). Observations are individuals (n=100). The risk of disease increases along the y-axis but units are arbitrary.*

There is also a fourth mode of action, where the highest risk is experienced by heterozygotes (TG). This scenario is considered to be rare and will not be discussed in this thesis.

A common misconception is to think that one mutation will have an effect on only one trait. This is rarely (and, arguably, never) the case. A genome is not a product of "intelligent design". It was grown by accidental rearrangements and natural selection, which reused and recycled genetic code fragments with one function to other fragments with slightly different functions. Due to such semi-stochastic design, all functions are interconnected; many genes cannot function without a proper function of other genes. Differently than in intelligent design, e.g., car manufacturing, knocking off one gene can have effects on numerous phenotypes, as compared to a single malfunctioning taillight. Similarly, a single phenotype, e.g., human height, can be affected by thousands of genetic variants, all contributing to the phenotypic variation.

These scenarios together with their technical concepts - pleiotropy and polygenicity - are explained in **Figure 9**.

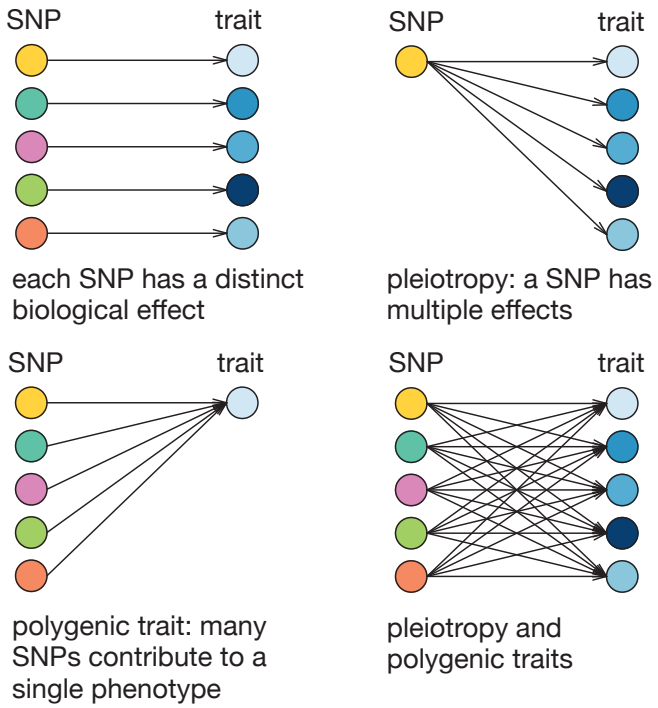


Figure 9. Pleiotropy and polygenicity. How variation in human genome (SNP; single-nucleotide polymorphism) can affect variation in the phenotype.

05 GOALS

5. Goals

The goal of all three studies in this thesis was to increase our understanding of the biology of human gestation and parturition. This work was done with a hope that minor increments in knowledge would eventually lead to effective preterm birth prevention methods: that children would not die and their parents would not suffer, that in the adulthood they would be whole and healthy, that society could spend more resources elsewhere and that researchers could concentrate on the next big problem.

Study I

In Study I our goal was to find genetic factors and processes modulating the pregnancy length. In particular: to identify SNPs that are statistically associated with gestational age and to explore which genes could be implicated by discovered SNPs.

Study II

In Study II we designed a new method that could provide reliable evidence of causality for the factors known to be associated with gestational age. Our goal was to test this method with maternal height data.

Study III

In Study III our goal was to find a biological explanation for the causal relationship between maternal height and the child's gestational age. In particular: to formulate a plausible falsifiable hypothesis and to design a quasi-experiment that could refute it.

06 DATA AND STUDY POPULATION

6. Data and Study Population

6.1. Genomic studies (I and II)

The dataset of Study I, and two additional datasets used in Study II are enriched in PTD cases and depleted from late preterm and early term pregnancies. Consequently, the phenotypic data (gestational age) follows a bimodal distribution (**Figure 10**) atypical to naturally occurring phenotype.

Figure 10: Bimodal distribution of gestational age at birth in three genomic cohorts is due to strategic sample selection.

The preterm birth oversampling was strategic. Since maximal statistical power in a case-control study is achieved with case-control ratio 50:50, a random sample from the population (with case-control ratio 5:95) was not optimal. All three genotype-phenotype datasets were created by different research groups/institutions with a single goal in mind - to detect genetic variants affecting the risk for preterm birth. So the choice to enrich in cases was logical.

The exclusion of "mid-range" gestational age values (255-279 days in DNBC[*]; 259-272 days in MoBa; 252-260 days in FIN) was also done for a reason. Having the inaccuracy of GA-dating methods in mind, all involved parties wanted to have the lowest rate of phenotypic misclassifications. As a general rule, with a constant sample size, the estimated effect size of a true genetic factor will be larger if cases and controls are defined from a continuous distribution using more extreme thresholds. At the same time, p-values for a true genetic factor will be lower. Since genotyping is a relatively costly data collection procedure, there is a need to optimize the sample selection strategy to gain maximal power with a given budget.

[*] *The DNBC, MoBa and FIN cohorts are described in detail in the same subchapter, below.*

6.1.1. MoBa data (Sudies I and II)

In Study I, all genotyped individuals are part of the Norwegian Mother and Child Cohort Study (MoBa)[78,79]. In brief, MoBa is a nationwide prospective pregnancy cohort, including more than 108,000 pregnancies during the period of 1999-2009. Women were recruited by postal invitation in connection with the routine ultrasound examination offered to all pregnant women in Norway at around 17 gestational weeks. Over a period of two decades, participants were asked to fill eleven questionnaires focused on health, lifestyle, diet, and anthropometrics (e.g., maternal height used in Study II). Pregnancy and birth records were added via personal number from the Medical Birth Registry of Norway[80] (e.g., gestational age used in Studies I and II, birth weight/length/head circumference used in Study II). Blood samples were collected from both parents during pregnancy and from mothers and children (umbilical cord) after birth. Biological material (DNA, RNA, whole blood, plasma and urine) was stored in a biobank.

Professor Bo Jacobsson's group did the first genotyping effort in the MoBa cohort in 2008, with a grant from the Research Council of Norway. It covered 2000 mothers and 1200 children, some of whom form the mother-child pairs (N pairs = 1017). The sample selection was done randomly in preterm (154-258 days) and term (273-286) strata of non-compromised singleton live-born pregnancies with spontaneous delivery. The case-control ratio was close to 50:50. Genotyping was done using Illumina's Human660W-quad_v1 DNA bead chips with more than 500,000 common genetic markers at Oslo University Hospital genotyping core facility.

6.1.2. DNBC data (Study II)

All genotyped individuals are participants enrolled by the Danish National Birth Cohort[81]. DNBC is a prospective nationwide cohort, in many aspects similar to MoBa. DNBC enrolled more than 100,000 pregnant women in the first trimester from 1996 to 2003.

The genotyped dataset contains approximately 1000 preterm and 1000 term singleton pregnancies. For each pregnancy, both mother and child are genotyped. The vast majority of children had parents and all four grandparents born in Denmark. There are no children born with recognized congenital or genetic abnormalities, and no mothers with declared medical conditions known to be associated with PTD. Also there are no pregnancies with major PTD-risk conditions: placenta previa, placental abruption, polyhydramnios, isoimmunization, placental insufficiency, preeclampsia/ eclampsia. All deliveries were spontaneous.

Blood samples (buffy coat) were collected for mothers and children. Genotyping was done using Illumina's Human660W-Quad_v1_A bead chips at Johns Hopkins University Center for Inherited Disease Research (Baltimore, Maryland, USA). For Study II, the DNBC data was retrieved from the public database dbGap[82].

6.1.3. FIN data (Study II)

The Finnish case-control cohort (FIN) was assembled specifically for genetic studies of preterm birth. Whole blood samples were collected from ~900 mother/child pairs from the Helsinki University Hospitals between 2004 and 2014. Gestational age at birth was evaluated using the UL method (crown-rump length, first ultrasound screening, 10-13 weeks). Collaborating nurses manually collected relevant phenotypic data. Only spontaneous, singleton births were included. Mothers and their children were of Finnish descent. The pairs were excluded due to PTD-related pregnancy problems, maternal medical problems, and fetal congenital problems. The final sample contained 783 mother-child pairs. Genotyping was done at Washington University, Vanderbilt University genotyping center with the grant support from the March of Dimes Foundation and the Fifth Third Foundation to Professor Louis J. Muglia. Several genotyping platforms were used: Affymetrix Genome-Wide Human SNP Array 6.0, Illumina HumanOmni2.5-8v1_A, Illumina HumanOmniExpress Exome-8v1-2_A, Illumina InfiniumOmniExpressExome-8v1-3_A. The number of SNPs on each chip type was 0.9, 2.2, 0.9, and 0.9 million, respectively. Imputation was used to ensure that all samples have genotypes from all SNPs. After imputation and quality control, all samples had genotypes for 9 million SNPs.

6.2. Epidemiological study (III)

6.2.1. Swedish Medical Birth Register data

All data in Study III comes from the Swedish Medical Birth Register. In Sweden, all births are compulsorily recorded in this register since 1973, regardless of mother's the birthplace and including even home deliveries. The register is managed by Sweden's National Board of Health and Welfare (Socialstyrelsen), which also stores and manages many other datasets linked by personal identity numbers. Researchers can apply to access the data but are never given personal identity numbers.

The register contains more than 4 million pregnancies and an extensive list of variables describing maternal anthropometrics, social and behavioral characteristics, medical conditions, pregnancy complications and characteristics, delivery complications and characteristics, child's weight at birth, gestational age, and many more.

Due to the missingness of some exclusion-filter indicators before the year 1990, only the pregnancies with births in year 1990 and onwards were included.

07 STUDY I: GWAS

7. Study I: GWAS

Literature-informed analysis of a genome-wide association study of gestational age in Norwegian women and children suggests involvement of inflammatory pathways

7.1. Background

7.1.1. GWAS principles

The genome-wide association study (GWAS) is a powerful tool to investigate the genetic architecture of complex traits. A typical GWAS one-by-one compares up to a million positions on a genome between those who have a trait of interest and the control individuals. If the control group at a particular position of the genome on average tends to have a different nucleotide than a case group, then researchers can suspect that this polymorphic site is associated with the trait, and that genes in a close proximity might be involved in a molecular aetiology of it.

The single-nucleotide polymorphisms (SNPs) that are used in genotyping are carefully pre-selected to represent a large fraction of inter-genome variability. The simplest form of analysis is done by comparing the frequencies of two alleles within each polymorphic position between the case and control groups of individuals. A statistical test is used to determine whether the difference in frequencies is improbable enough to be generated only by random sampling. The probability of such a difference to emerge only by chance, under the assumption that there is no real association, is represented by a p-value. The lower the probability - the smaller the p-value and the lower the confidence in the assumption of no association. While directly arguing against the null hypothesis[*], a low p-value indirectly argues for the alternative hypothesis.

The great value of GWAS is not at predicting individual's risk for certain diseases but identifying, with a high level of confidence, the genes that are influencing it. How could the knowledge of genetic factors be used for preterm birth prevention? After all, we cannot modify our genomes yet. The answer is simple and comforting. Even though genome editing is off the table, pharmacological interventions can affect gene transcription, can target gene transcripts (mRNA) and their products - proteins. The comforting part is that no gene functions in a solitude. Their transcripts and proteins act and interact in biochemical pathways, each consisting of numerous other players. Thus, a single implicated genetic variant can offer a large and varied flock of targets.

The more genetic risk factors we know, the better we can tell what role they are playing in biochemical pathways and the easier it is for us to find a safe way of

[*] *The null hypothesis is the default position, a claim that there is no association. In a classical significance testing approach (Ronald Fisher), we look for evidence against the null.*

stimulating and dimming the players of those pathways to restore the equilibrium specific to term-delivering women.

7.1.2. GWAS history

GWAS has already identified many SNP that are associated with the risk for lung cancer[83], celiac disease[84], Crohn's disease[85], Type 1[86] and Type 2[87] diabetes, Alzheimer's disease[88], Parkinson's disease[89], and thousands more[90]. Its popularity rose with an advent of genotyping microchips in 2005 (**Figure 11**).

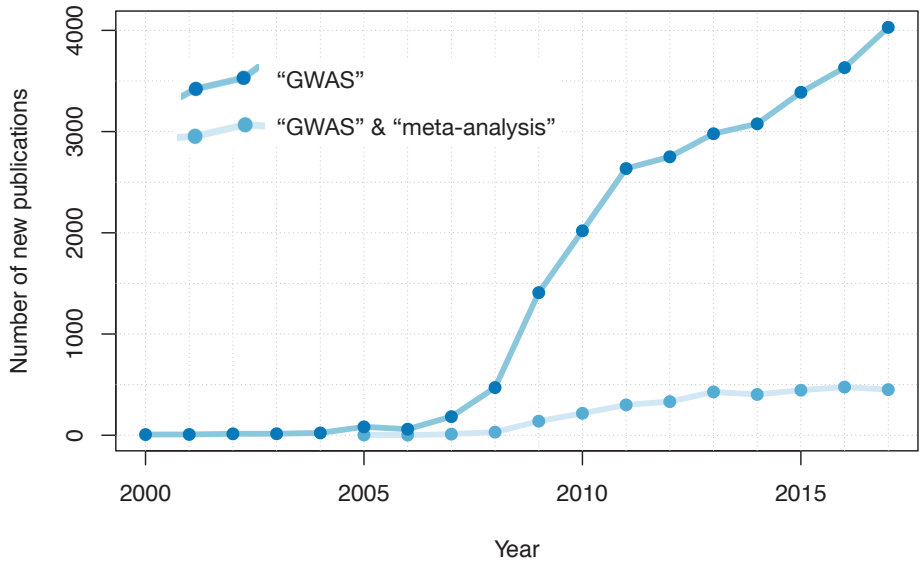


Figure 11. *Published scholarly work using GWAS. Data source - Scopus (accessed 2018 Sept).*

7.1.3. From SNPs to drugs

After a GWAS discovers a significant genetic variant (SNP) but before clinical pharmacology can take a shot at developing the drug, a lot remains to be done. The SNP might not be causal, but only correlate with some unobserved (not genotyped) truly causal mutation. This mutation must be shown to have even stronger association with the phenotype. Mutation must be mapped to a gene. Experimental evidence must be collected; indicating that specific allele of the mutation affects the expression of that gene. The direction of this effect must be determined. Association between gene expression levels and the phenotype must be confirmed. Gene implicates biochemical pathway. Often, a gene can be part of many pathways, thus more GWAS discoveries are needed to point out the pathway that is affecting the phenotype. Members of this pathway become the targets of clinical pharmacology research. To this day, there is only a small number of GWAS discoveries that went through all the way and reached the shelves in a form of drugs.

7.1.4. Statistical power

Statistical power calculations are often run before the experiment to determine how many study subjects would be needed to properly falsify the hypothesis. However, GWAS is a hypothesis-free experiment. In it, we do not know what exactly we are looking for. We do know that genetic effects are involved, but nothing more precise: neither SNP's effect size, nor its minor-allele frequency. **Figure 12** shows dependency of statistical power (the probability that a true causal SNP will be discovered) on the study size and two parameters describing the causal SNP. If someone would tell us that we are looking for a SNP with MAF=0.2 which decreases gestational age by 3 days per every copy of effect allele, then we would be certain that genotyping 2000 individuals should be sufficient to identify this SNP. But there is no one to reveal to us such hidden information. So instead, we run power calculations *after* GWAS, to show what types of SNPs *did have* a fair chance of being detected, but were not.

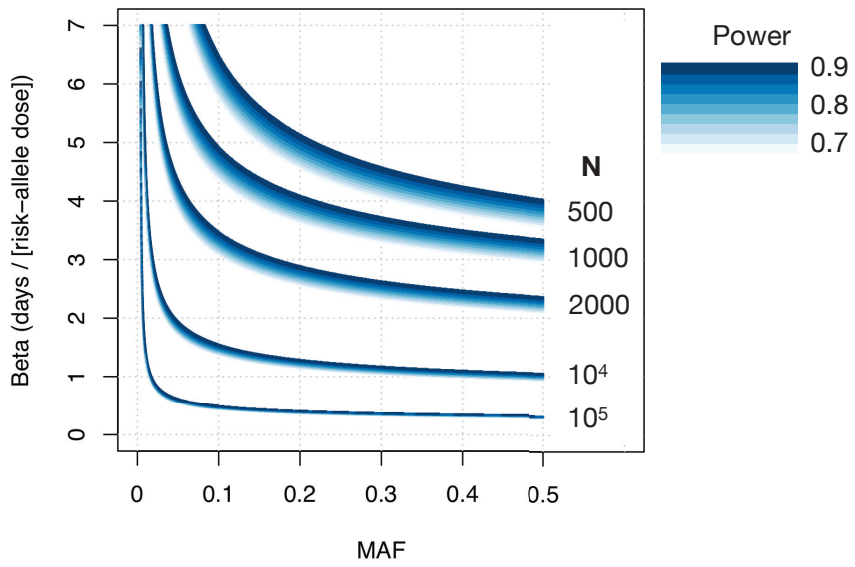


***Figure 12**. Power simulation for GWAS of gestational age. Statistical power depends on the study sample size (N), minor-allele frequency (MAF) of the SNP and its true effect size (Beta). The simulation assumed additive genetic mode of action, 1 million independent tests (i.e, genome-wide p-value threshold is 5e-8), normally distributed gestational age with mean=280 days and SD=9.7. Beta is interpreted as absolute difference in mean gestational age between genotypic groups AA vs AB, and AB vs BB. Power greater than 0.9 spans towards the upper right corner from every curve.*

7.1.5. The advent of large national datasets

The new and very powerful genome-editing tools would be useless without the knowledge about how our genomes function. Thus, there is a recent trend in national strategies aiming to collect as much phenotypic and genotypic information as possible. Today, governmental funding agencies of many countries are convinced, that massively genotyping or sequencing their population is a good idea in order to

tackle health issues (**Table 2**). Many GWAS studies will come soon as a result of well-funded efforts to genotype a significant fraction of populations. Similarly, an interest to participate in genome research is observed among private companies: 23andMe (USA), AstraZeneca (USA), BGI (China).

Table 2. Countries leading in genotyping a large fraction of population.

Country	Cohort name	Project size (individuals)	Fraction of population
Iceland	DeCODE	172 000*	51%
Finland	FinGen	500 000	9.1%
Norway	MoBa[79]	300 000	5.7%
Norway	HUNT[91]	250 000	4.8%
Estonia	EGCUT[92]	52 000	3.9%
UK	UK Biobank[93]	500 000	0.76%
USA	Million Veteran Program[94]	690 000†	0.2%
USA	"All Of Us"	1 000 000‡	0.3%

Project size reflects the existing or projected number of individuals with genome-wide genotyping data. Unofficial data from personal communications. The numbers are approximate and the list is not extensive.

7.1.6. Advantages of genomic studies

The methods investigating genotype-phenotype associations have an advantage over methods exploring biomarker-phenotype associations. The "biomarker" here is a placeholder for tissue- and time-specific entities or phenomena, e.g., proteins, gene transcripts (mRNA), DNA methylation patterns. Such non-GWAS methods are especially problematic in preterm birth studies.

Firstly, gene expression is widely varied across tissues. It is unclear which types of cells should be investigated when searching for biomarkers: leukocytes from blood, myocytes from myometrium, trophoblastic cells from placenta etc. This imposes an additional research hypothesis. A poor choice of study biomaterial will result in a whole budget wasted on non-significant results. But more importantly, the cell types that are most likely to hold a key to the secrets of human pregnancy are not accessible to analytical investigation due to potential risks to mother or the fetus. Biopsies of the human endometrium or fetal tissues would never even be considered by ethical review boards.

Secondly, unlike genomic studies, gene-expression studies investigate biomarkers that are changing in time. The levels of certain gene transcripts might follow the time passed since conception. But similarly, these levels might mark the time remaining till parturition. Without a possibility to collect bio-samples at a certain time-point before

* *Might currently be much closer to 100%.*

† *As of August 2018*

‡ *The first version of dataset is expected in 2019*

the delivery in all study subjects, any observed association between gene-expression levels and gestational age at birth will represent anything but the explanation for biology of preterm birth. Needless to say, such task of bio-sampling pregnancy exactly one week (for example) before delivery is not practically feasible.

Thirdly, unlike genomic studies, gene-expression studies do not imply causality in the discovered associations. There is a possibility, that a certain confounder, e.g. a regulating protein is triggering both the parturition pathway and also up-regulating an unrelated pathway that includes the gene being analyzed. All three shortcomings are also unavoidable in epigenetic studies, investigating the methylation (silencing and activation of certain regions) of the human genome.

The conclusion here is that GWAS studies do not fall victim to the aforementioned problems, because the genetic code is the same across all somatic tissues, is constant over time and, according to the central dogma of genetics, only causes things to happen and is not being caused.

7.1.7. Previous attempts at finding PTD genes

There were only two GWAS studies that investigated the timing of delivery before our Study I was published. None have identified SNPs that are robustly associated with PTD.

In 2013, three years before our Study I, the GWAS results from DNBC cohort were published[95]. The study investigated 22 autosomal chromosomes (without chromosome X) and included 849 preterm births and 949 term births as controls. Only fetal genomes were used. The authors did not find genome-wide significant hits. They also chose not to publish the list of top SNPs.

In 2015, one year before our Study I, the GWAS results from the Genomic and Proteomic Network (GPN) for Preterm Birth Research were published[96]. This study included subjects from the USA (only 68% of which were Caucasian) - both mothers and children. In the maternal group there were 935 cases and 946 controls, while the fetal group contained 916 cases and 935 controls. The spontaneous PTD was defined by gestational age at birth being less than 34 weeks; the control group was defined by birth between 39 and 42 weeks of gestation and with no history of PTD in any prior pregnancies. The researchers could not replicate their findings in a validation cohort.

7.2. Summary of the study

Since up to 30% of variation in human gestational age could be accounted for by heritability, we were motivated to identify some of the genetic factors using GWAS.

We screened 500,000 genotyped variants across all chromosomes without making any prior assumptions about their function. This is called a "hypothesis-free" approach, as opposed to a "hypothesis-driven" approach, in which we would only test SNPs in candidate genes. Both maternal and fetal genomes were tested. Separate analyses were run with deliveries initiated by labor and those initiated by prelabor rupture of membranes (PROM).

The GWAS did not reveal genome-wide significant associations. However, we observed a robust significant enrichment in phenotype-relevant candidate-gene sets. We interpreted this as an indication, that our GWAS ranked variants in a phenotype-relevant manner, and that increased sample size would confirm some of the "false negatives" to be "true positives". We also noticed, that a high fraction of top implicated genes were previously reported in a context relevant to preterm birth. Based on a literature search, many of the top implicated genes form interacting networks and often belong to the infection-inflammation pathways, which are lately emerging as an important etiology element of human parturition.

The value (application) of this study is twofold: 1) it provides a reasonably short list of polymorphisms that could be used as candidates in replicating studies with limited sample size: due to a low burden of multiple-tests small studies might still have enough statistical power to confirm our reported associations. 2) GWAS might benefit from our study design: inclusion of recessive and dominant genetic models was advantageous, because allelic interactions (dominance effects) implicated approximately 90% (more than expected by chance) of genes with biological relevance; similarly, 30% of genes (more than expected by chance) would have been overlooked if a minor-allele frequency filter (MAF > 0.1) were to be applied, and over 50% would have been lost if a GWAS sample size were to be "increased" by mixing PROM-delivering mothers (N = 336) and mothers with labor-initiated deliveries (N = 1407).

7.3. Novelty

Despite no robustly implicated genes in Study I, it is worth highlighting some novel aspects implemented in this study. We realised that the earlier study designs[95,96] could be improved in numerous ways, by either avoiding some obvious shortcomings or designing new features.

The first decision was to recognize that dichotomizing a continuous phenotype reduces its variance and, in most cases, should reduce statistical power of the study. This recognition was lacking in the previous studies. As a result, we chose to use continuous gestational age instead of its dichotomized form.

The second novel aspect was acknowledging the existence of different etiologies of parturition. If delivery was initiated by PROM, then the resulting gestational age becomes less informative of genetic effects that affect the "risk" of labor. Similarly, if delivery was initiated by labor, the resulting gestational age is less informative of genetic effects that affect the risk of PROM. Without a clear distinction between these two endophenotypes the statistical power will suffer. For that reason we ran analyses separately in pregnancies that started with PROM and labor.

The third novelty was eliminating prior assumptions about the genetic model. Typically, a GWAS study would assume an additive allelic effect, i.e., that the three possible genotypes will have an effect on the phenotype ordered in magnitude depending on the dosage of a risk allele: AA > AB > BB. While this is a good rule of thumb, there might be scenarios in which such assumption would result in reduced statistical power. In particular, dominance models where AA=AB>BB, or

AA>AB=BB. We repeated every GWAS two additional times corresponding to these two dominance models.

The fourth methodological improvement was the use of adaptive permutations to adequately control the Type I error rate inflated due to the use of dominance models and continuous phenotype with a skewed distribution.

7.4. Limitations

Upon reflection, Study I has limitations. We did not attempt to replicate our findings in an independent cohort. No imputation was done (statistically inferring non-genotyped SNPs). Some arbitrariness still remains, e.g., the number of top SNPs used in enrichment. We also speculated that the results might not replicate in the cohorts from other countries, because many reported genes are infection-related, and infection rates and types are likely specific to different geographical areas. The decision not to use the MAF filter might disproportionately increase the number of erroneously genotyped SNPs. Case-oversampling and unnatural phenotypic distribution complicates comparability of the effect size with the results from other studies. The use of linear mixed models (LMM) would have allowed us to include more individuals who now were excluded due to excessive cryptic relatedness. Similarly, with LMM, the phenotypes from other pregnancies of the same mother could have been utilized. Since gestational age by its nature is a "time-to-event" phenotype, it would be most appropriate to use Cox model (proportional hazards regression, or survival analysis); this would also allow having PROM-initiated deliveries and labor-initiated deliveries in the same analysis.

7.5. Methodological aspects

7.5.1. Genomic-data cleaning: mistakes in data

The most time-consuming stage of any genomic study is data cleaning, also called "quality control" (QC). The goal of this process is to minimize the error rate in the data received from the scanner.

Genotyping is done on a massive automated scale but in micro dimensions. All downstream analyses are based on an implicit assumption that all genomic positions of all individuals were evaluated (genotyped) correctly. Needless to say, a lot of things could go wrong in the genotyping facility: DNA samples could have too low concentrations, or be contaminated with reagents or other DNA; some types of magnetic beads (probing a certain SNP) could fail to attach to the array, the tag sequence could fail to attach to the bead, or DNA-bead attachment could fail. On a massive automated scale, these problems always happen.

There are various QC filters invented to detect and eliminate mistyped genotypes. Unfortunately, none* of them can tell exactly which genotype is wrong. The reader

* *Except QC filter using family data and Mendelian Errors.*

could visualize a table with 10 individuals (rows) and 10 SNPs (columns); a genotype would be a single cell in such a table of 100 cells. Instead, a typical QC filter highlights the SNP, which has unlikely genotypes at many individual samples; and similarly the sample, which has unlikely genotypes at many SNPs. In other words, the QC detects the SNPs and samples that are not worthy of trust. There are numerous filters that indicate such poorly performing SNPs and samples (**Table 3**).

Table 3. Genomic QC filters detect SNPs and samples that are not worthy of trust.

SNPs	samples	Filter name	Description
+	+	Missingness	A large fraction of not called genotypes indicates low DNA quality or concentration, or failed SNP chemistry.
+		Minor-allele frequency	A too low frequency can lead to an unreliable genotype assignment during genotype calling (clustering step)
	+	Heterozygote frequency	A too high or too low fraction of heterozygous genotypes indicates a DNA sample contamination
+		Hardy-Weinberg equilibrium	Genotype frequencies are discordant with frequencies that are theoretically expected from allele frequencies.
+	+	Mendelian errors	Detects impossible genotypic combinations between close family members.
	+	Sex chromosomes	Mismatches between declared and genetically inferred sex indicate sample-identity problems.
	+	Identity by descent (IBD)	Discordant relatedness in declared and genetically inferred pedigrees indicates sample-identity problems.

The first two columns indicate whether the filter is designed to detect problems in indivduals (samples) or the features (SNPs). The list is not extensive.

Troublingly, a lot of arbitrariness is involved in the QC. The same QC filters can give different results depending on how they are applied. In the worst-case scenario, the QC will result in an over-conservative elimination of trustworthy genotypes, thus reducing the economic efficiency and statistical power of the study. This could be illustrated in a simulated example (**Figure 13**), where 3 out of 20 genotyped samples had low DNA concentration (and thus - low genotype calling rate). Similarly, a low call rate was enforced on 2 out of 20 genotyped SNPs. In the simulated QC, a 10% missingness filter was applied, meaning that any SNP or any individual with a fraction of called (non-missing) genotypes lower than 90% will be considered as "bad" and not to be trusted in analyses.

By design, the perfect QC would mark 3 samples and 2 SNPs as unworthy of trust, thus deleting only 94 genotypes ($3\times20 + 2\times20 - 2\times3$). The best-performing method (7) in the simulated example deleted 130 genotypes, thus sacrificing 36 additional genotypes for no good reason (9% of total). The worst-performing method (1) deleted 50% genotypes without evidence ((292-94)/400), while two standard methods (2 and 3) deleted 25% and 43%, respectively.

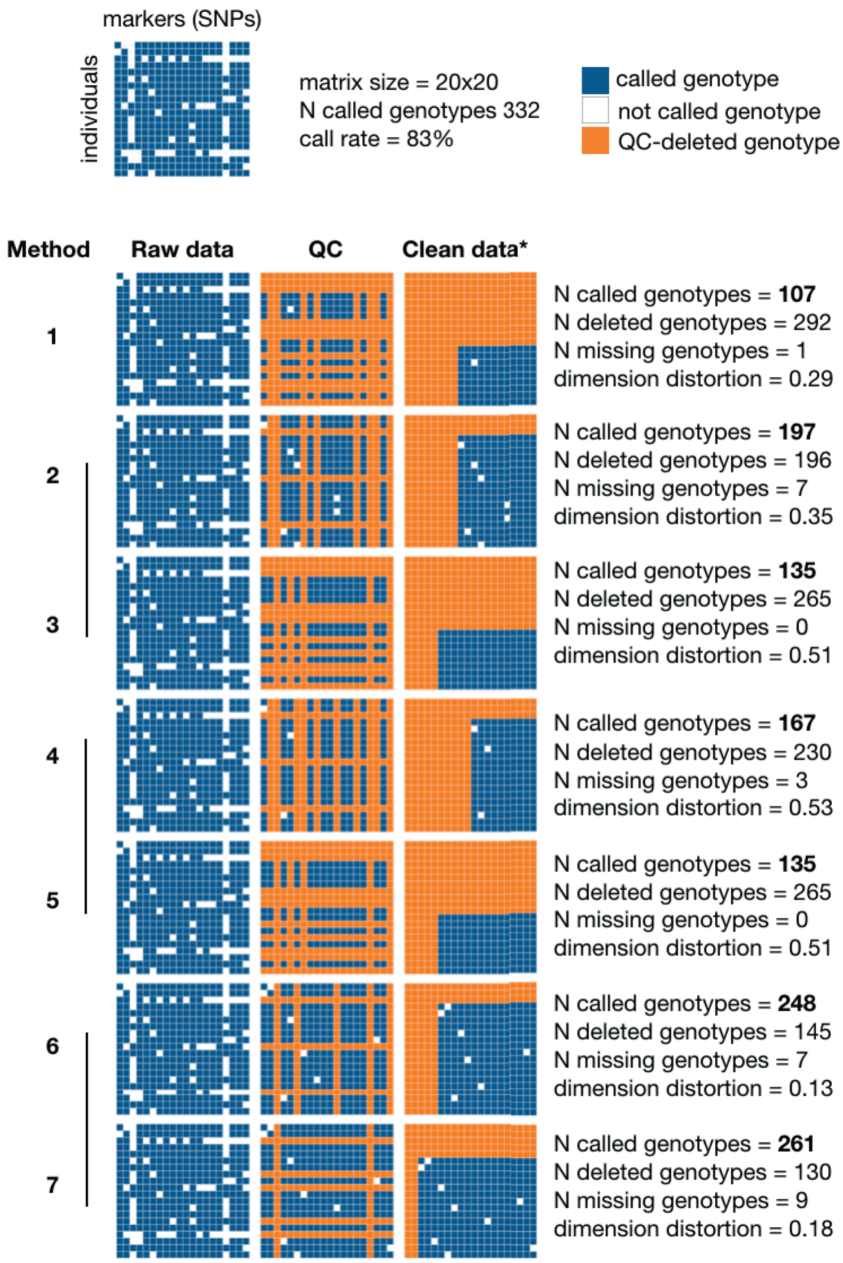


Figure 13. *Simulation "same QC - different results". The importance of the order in which SNP and sample filters are applied, also the effect of an iterative approach with a dynamic filter threshold (increasing stringency gradient). The final threshold for the missingness filter is 10%. Methods: (1) both filters applied simultaneously; (2) SNPs, then samples; (3) samples, then SNPs; (4) four cycles of method 2; (5) four cycles of method 3; (6) five cycles of method 2 with thresholds 50, 40, 30, 20, and 10%; (7) five cycles of method 3 with thresholds 50,40,30,20,10%. The dimension distortion parameter describes how much the cleaned dataset deviates from the raw dataset, with a value equal to "0" when the dimension ratio is kept similar to the original data. Asterisk (*) - for visual purposes SNPs and samples were reordered and clustered.*

This example shows the importance of using an iterative approach with a dynamic filter threshold. In this example, only the most basic QC filter was used, but the principle applies to any two filters, one pruning SNPs and the other pruning samples.

As a concluding remark, I would like to note that every step in a QC pipeline affects the outcome of the following steps. A careless ordering of the QC steps might lead to the poorly cleaned or overly pruned data. Since there are many types of filters, the final QC pipeline could quickly become very complex. A considerable amount of time in this project was spent on designing an optimal sequence of QC filters. To my knowledge, such thorough approach is rarely taken in the genomic data QC.

7.5.2. Genomic-data cleaning: bias in data

Besides not trusting the genotyping quality of certain samples and certain SNPs, there are more dangerous issues to consider. A different type of QC tries to avoid bias and inflated Type I error rate in the downstream analysis by not allowing some samples and some SNPs to enter the analysis stage.

Cryptic relatedness. Individuals randomly selected from a small population will be to some extent related among each other. However, that would violate the main assumption in association analysis - observations must be independent. It is therefore important to prune the dataset by eliminating one individual from each pair of related individuals. We treated each pair as "related" if genetic similarity was equivalent to or greater than that between cousins (PI_HAT>0.125). The challenge then was to determine which individual from a pair to remove. The choice should not be a random one, but rather be a function of genotyping data quality for that individual, the membership in a mother-child pair (more valuable), and phenotype missingness.

Population stratification. In GWA studies, whenever the phenotype has non-homogenous worldwide prevalence or distribution, the possibility of geographical confounding must be taken into account. In the simplest case, the study population is "homogenized" by excluding individuals who are "ethnic outliers" based on many polymorphisms in their genomes (principal components analysis, PCA). This procedure is sensitive to inclusion of related individuals (e.g., mothers and children). The standard bypass of this problem would be to run PCA on unrelated individuals who are "founders". However, such approach would be an oversight, as in our data it would mainly select mothers as analytical subjects for PCA, and then, if a mother is selected for exclusion, her child would automatically be also, since they share 50% of their genomes (and thus, at least 50% of "ethnicity"). The subtle problem with this approach lies in the fact that the fathers are not genotyped in our data. The child's genome represents both maternal and paternal ethnicities, while the mother's genome provides no clues about the father's ethnicity. The procedure would only eliminate confounding risk in GWAS of maternal genomes, but would not fully eliminate possibility of confounding in GWAS of children. My solution to this problem was to identify ethnic outliers in mothers and children separately.

7.5.3. Genome-wide association analysis

Permutation-based statistical significance testing is regarded as the gold standard[97]. Conventional asymptotic statistical tests are evaluated and validated by comparing their results to the ones generated using permutations. In Study I, the need for permutation-based association testing is illustrated by quantile-quantile (QQ) plot in **Figure 14**, where a clear inflation of Type I error is seen when a skewed phenotype and a recessive model are used. Since only the null hypothesis was designed to be valid for all the SNPs (due to random phenotype assignment), the transformed and ranked GWAS p-values were expected to fall on or very close to the black line. The "upward" deviation of observed p-values from their expected values is equivalent to Type I error, i.e., calling SNPs as being associated with the phenotype when, in fact, they are not. This is due to a violation of one assumption behind the linear regression. The inflation is mostly attributed to the skewed phenotype (setups B and D), but recessive model also adds to inflation when used together with the skewed phenotype (setup D).

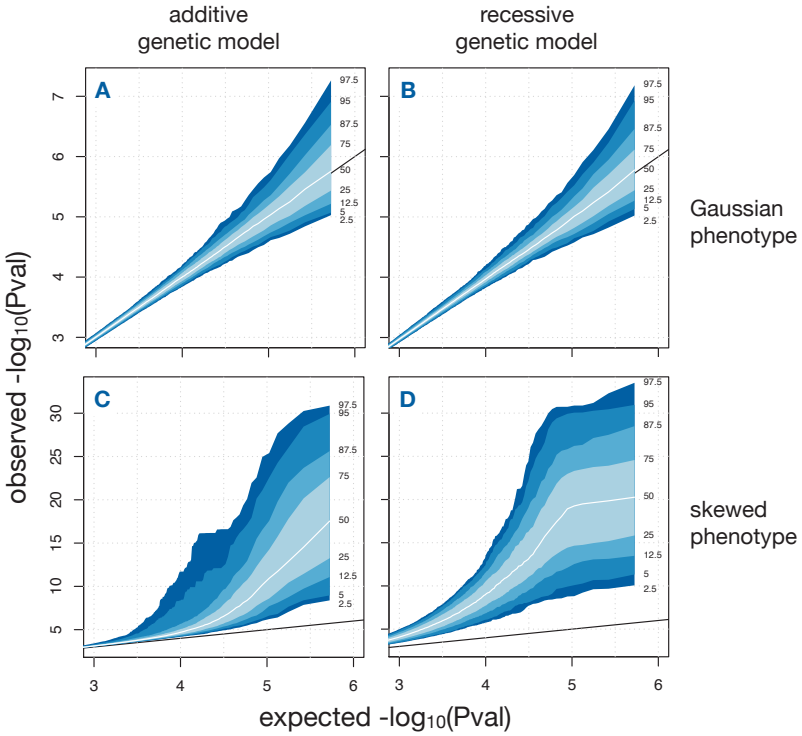
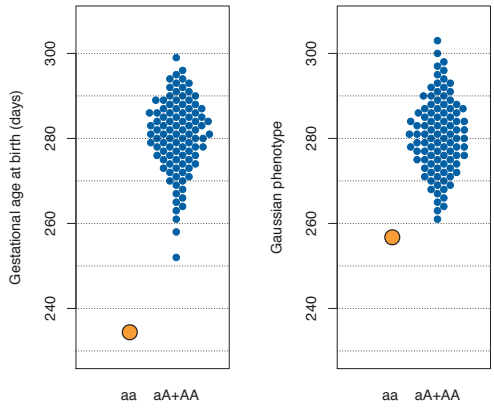


Figure 14. *Inflation of Type I error rate in GWAS due to skewed phenotype and recessive genetic model. We ran 500 GWA analyses for each setup (A-D). Each time in A and B the phenotype was randomly drawn from normal distribution, while in C and D it was randomly sampled from naturally skewed distribution of gestational age. Recessive model is defined with respect to minor allele. Genetic data is from Study I. In each GWAS, p-values were log-transformed and ranked. After 500 iterations, the distribution of -log$_{10}$(Pval) at each rank position (e.g., for the top SNP) was summarized by certain percentiles (small numbers). The chosen percentile colours give visual aid in reading the quantile-quantile plot (median line, white) and its empirical confidence intervals (50%, 75%, 90%, and 95%, from light blue to dark). Expected p-values were drawn from uniform distribution. The black line has an intercept 0 and beta 1.*

How could this undesirable phenomenon happen in the first place? And how can permutation-based regression solve it? Let's examine a simplified example in **Figure 15**, where the data consists of only 100 individuals. The SNP has a minor-allele frequency of 10%, thus the number of individuals in the minor-homozygote group is 1. The remaining 99 individuals belong to a single combined genotypic group. In both experiments, the lone individual happens to have the most extreme (the lowest) phenotypic value from the whole cohort. Thus, the p-value describing statistical significance behind genotype-phenotype association is the same as the answer to the school-level mathematics problem: "what is the probability to draw an orange from a basket with 99 apples and 1 orange?". In both experiments, the empirically correct level of surprise* that we should have, when observing such data with the null hypothesis in mind, is equivalent to 1/100. But even more surprisingly, this is not the p-value that the linear regression gives us.



phenotype	p-value by analysis method: linear regression	p-value by analysis method: permutation-based
Gestational age	0.0000007	0.010
Gaussian	0.008	0.010

Figure 15. *The importance of permutation-based association testing. Observations (individuals) are coloured based on their genotype, recessive model with respect to the minor allele is used. Both depicted scenarios are equally psychologically surprising under the null hypothesis of no association. Since they both argue equally strongly against the null, their p-values should be the same, as is indicated by empirically estimated p-values from permutation-based testing. In this example, linear regression grossly inflates statistical significance.*

The reason for this is that linear regression has certain assumptions, which are violated in the first experiment. Namely, when the phenotype is not distributed normally, in a setting where the predictor has only several levels and a low number of observations in one (or some) of them, the model residuals will be heteroscedastic and not normally distributed. One solution to this is enforcing the phenotype to have a normal distribution (undesirable). The other solution is getting rid of linear regression and using permutation-based regression instead.

* *The p-value is the probability to observe association as extreme as in experimental data, but under an assumption that association is not real and arose only due to chance. This probability summarizes our level of surprise in seeing unexpected patterns. The smaller the probability, the less we believe in the assumption of "pattern is due to random chance", and the stronger we believe in favour of an alternative explanation "pattern is due to causes".*

Similarly to the thinking described in a school-level mathematics problem, permutation-based regression shuffles the phenotypic values randomly and measures the strength of association. Then it repeats this null-hypothesis procedure a very large number of times. Finally, the strength of association witnessed in the real (non-shuffled) data is compared to the distribution of accumulated values from permuted data. The fraction of accumulated values that are equal to or exceeding the witnessed strength of association is the empirical p-value. It is the most intuitive, the least biased (in fact, not biased at all), and the most correct evaluation of statistical significance. However, it comes at a price of having to run hundreds of millions of tests per single SNP.

Multiple-test problem is notable in Study I, as we tested a half million genetic markers in eighteen GWA analyses each:

(3 genetic models) × (2 phenotypic subtypes + 1 mixed) × (2 types of genomes)

where subtypes are labor-initiated or PROM-initiated deliveries, and types of genomes are mothers and children. The exploratory nature of our study requires adequate correction for multiple testing. However, as many of the tests are not independent, a high correlation between tests (SNPs) and analyses should be taken into account before considering correction procedures. Since none of the GWAS revealed even remotely significant associations, we spared ourselves from estimating the total effective number of tests and used generated p-values only to rank the SNPs for the use in post-GWAS analyses.

7.5.4. Gene-set enrichment analysis

Objective phenotype-related candidate-gene sets. There is a very large number of gene-sets (lists of genes representing various biochemical pathways) ready to be used in enrichment analyses However, testing every available gene-set for enrichment would impose enormous burden of multiple-test correction and reduce statistical power. It is also impossible to avoid subjectivity when selecting only some of the available gene-sets to test against the GWAS top genes. Since we needed a small number of objectively relevant gene-sets covering various aspects of the phenotype, we developed a semantic PubMed abstract mining tool, which takes as an input a keyword and outputs a list of gene names that were found in the large body of abstracts containing that keyword. In this way we eliminated subjectivity, reduced the number of potential gene-sets and maximized their representativeness.

7.6. Impact and echoes

In 25 months of its online presence, Study I was cited four times by other scientific publications (a higher rate than the Impact Factor of *PLoS One* journal) and was viewed or downloaded by readers 2250 times.

As it became clear that Study I is statistically underpowered to detect GA-associated SNPs on its own, in 2017 we recycled the data in our new GWAS meta-analysis (not included in this thesis) with 40,000 genomes from the company *23andMe*'s research participants and two other birth cohorts from Denmark and Finland. This turned into an impactful publication in the *New England Journal of Medicine*[98], with 27 citations during its 12 months of online presence[*]. In here, we also re-used the literature-based candidate gene-set enrichment method that we developed in Study I. For the first time maternal genomic loci were found to be robustly associated with gestational length or preterm birth. The roles of these six loci showed significant biological relevance due to their known involvement in uterine development, maternal nutrition, and vascular control. Since then, we have replicated the findings in our yet another, currently ongoing GWAS meta-analysis with more than 100,000 maternal genomes.

As a side note, just recently, a successful GWAS on prolonged gestation (post-term birth) was published[99], also a large GWAS meta-analysis using fetal genotypes[100].

[*] *This publication was prominently featured in Bill Gates' speech during Presidential Symposium at American Society of Human Genetics meeting in Orlando, USA (October 18th, 2017).*

08 STUDY II: MENDELIAN RANDOMIZATION

Jonas Bačelis

8. Study II: Mendelian Randomization

Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis

8.1. Background

Among many environmental factors reported by observational studies as being associated with gestational age, one stands out as exceptional. It is maternal height*. Firstly, maternal height is a strong predictor of gestational age. Secondly, one hypothesis could elegantly explain this correlation: maternal height is only a proxy for uterine size. Due to fetal growth and limited uterine space, the fetal movements gradually decrease in frequency from the peak at 32 weeks[101]. Sometime at early term, fetal size starts to discomfort the uterine wall. Possibly mediated by inflammation[102], uterine distention then initiates the delivery. As a consequence, mothers with smaller uteri (and shorter stature) should, on average, deliver earlier. Hence, the causal link between uterine size and gestational age. Furthermore, there should be no causal link between uterine size and phenotypes like birth weight or birth length (adjusted for gestational age), since it is possible for the "gestational clock" to stop ticking, but not possible for the fetus to stop growing.

The aforementioned "causality via uterine stretch" hypothesis, if proven true, could be of huge value. Not because we want to "recommend" the pregnant women to be taller, but because the causal relationships have a potency to generate new hypotheses and probe the causal pathway deeper, until we find a component that we can modulate, thus reducing the risk of preterm birth.

The goal of any Mendelian randomization study is to imply causality between two variables (exposure and the outcome). It is important, because very often correlations are observed only due to confounding effects, i.e., there is no direct causal relationship between two variables. For example, the reported maternal height vs gestational age correlation could be confounded by maternal age or socio-economic status (nutrition). Only if we would be sure that the relationship is causal, we could start deconstructing and explaining the mechanisms leading from the risk factor to parturition.

As a background to Study II, it is important to mention that human height is a highly heritable trait. Approximately 80% of phenotypic variation can be explained by heritability. A very large GWAS meta-analysis[103] (with a total number of individuals exceeding 250,000) reported 697 independent height-affecting SNPs that explain a large fraction (20%) of variance in height.

* *From the perspective of the fetus, the mother and her traits are equivalent to environmental factors.*

It also might be of value to mention a general observation that gestation length in mammalian species (including humans) scales proportionally to body mass, suggesting that this trait is constrained by body size[16].

8.2. Summary of the study

Study II is an example of how genotyping data collected for GWA study can be used for a completely different purpose - to gain insights about environmental factors. The environmental factor was maternal height at pregnancy and our goal was to test whether it has a causal effect on birth outcomes (gestational age, birth weight, birth length).

We constructed a genetic score based on 697 SNPs known to be associated with adult height. In all three cohorts (MoBa, DNBC, FIN), this score explained the expected 20% of the variance in maternal height. We also developed a novel Mendelian randomization method, which utilizes the non-transmitted maternal alleles of these SNPs as a genetic instrument for maternal effect. This was done to avoid confounding due to genetic sharing between the mother and the infant. The crucial stage involved inferring parental transmission of the haplotypes, also known as genome phasing.

Gestational age was significantly associated with the non-transmitted haplotype score, while birth weight and birth length were not. These results demonstrate that the observed association between maternal height and gestational length is very likely causal. One possible way to explain such causal connection could be that mothers of shorter stature have smaller intrauterine space that imposes fetal-growth restrictions earlier than larger uteri of tall mothers, thus the child is likely to be born earlier.

8.3. Novelty

The major challenge in Study II was imposed by a fact that both mother and fetus can affect gestational duration. In a setting where the phenotype is a human pregnancy outcome, the only ethical (i.e., non-experimental) way to explore causal relationships is by using the Mendelian randomization method. Its strength lays in the randomness of allelic transmission. But in our setting, it also becomes a weakness: maternal and fetal genomes are correlated. This violates the principal assumption in Mendelian randomization - that genetic instrument (height-associated maternal SNPs) affects the outcome (gestational age) only via the suspected causal phenotype (maternal height). However, it is possible that instrument-outcome association arises only due to fetal effects created by the fetal genome.

The novelty of our study came from a shifted perspective in how we mentally model the pregnancy outcome. We either believe that gestational age is coupled with the maternal genome, or the fetal genome. Instead, our perspective was to imagine maternal and fetal genomes as one pseudo-triploid genome, one third of which is only owned by the mother. With this thinking, the non-transmitted haplotype Mendelian randomization method was born. Genetic instrument created from non-transmitted maternal alleles will only represent the direct causal pathway: if we observe an

association between such genetic instrument and gestational age - we have seen the evidence of causality between maternal height and the child's gestational age.

8.4. Limitations

There was a statistically significant correlation between height genetic scores derived from the transmitted and non-transmitted maternal haplotypes, which means that assortative mating must be in action. This would not be surprising, as the positive correlation between the statures of spouses is common in western populations[104]. However, we took precautions to avoid bias arising from the correlation between transmitted vs non-transmitted alleles in mother, and maternal vs paternal alleles in the child.

Heterogeneity must also be mentioned: in some analyses, the results of the three cohorts did not agree perfectly (S7 Table, Study II). In fact, the main conclusion seems to be driven solely by DNBC cohort. This might be due to the differing cohort sizes and inclusion criteria, but also due to differing environmental backgrounds* and differing reliability of reported height.

Biological pleiotropy could not be ruled out, where genetic instrument influences the birth outcomes through other mechanisms. For example, the SNPs associated with adult height can also influence fetal growth rate, and this might bias the UL-dated gestational age. Even though we took precautions, a complete exclusion of biological pleiotropy is generally not possible when a large number of genetic variants are used to construct the genetic instrument[105].

The UL-method used for GA-dating assumes that all fetuses have the same growth rate as the reference cohort used to generate the reference growth curves. Thus, gestational age of children with tall mothers could be overestimated. Ironically, the less accurate GA-dating method LMP would not suffer from such bias.

* *In a very healthy population, the phenotypic variance will be explained less by environmental factors and more by genetic influences, in contrast to a less healthy population with a large variation in the levels of environmental exposures.*

8.5. Methodological aspects

8.5.1. Phenotypic data cleaning

In Study II, the quality control of phenotypic data was done by the author. Maternal height data in the MoBa cohort was self-reported. Mothers received questionnaires by post and filled them in private. Later, the hand-filled questionnaires were read using optical character recognition system. We noticed a fraction of incredible maternal height and weight values. After closer inspection it was clear that these impossible values form distinct clusters (**Figure 16**) that are unlikely to be random mistakes.

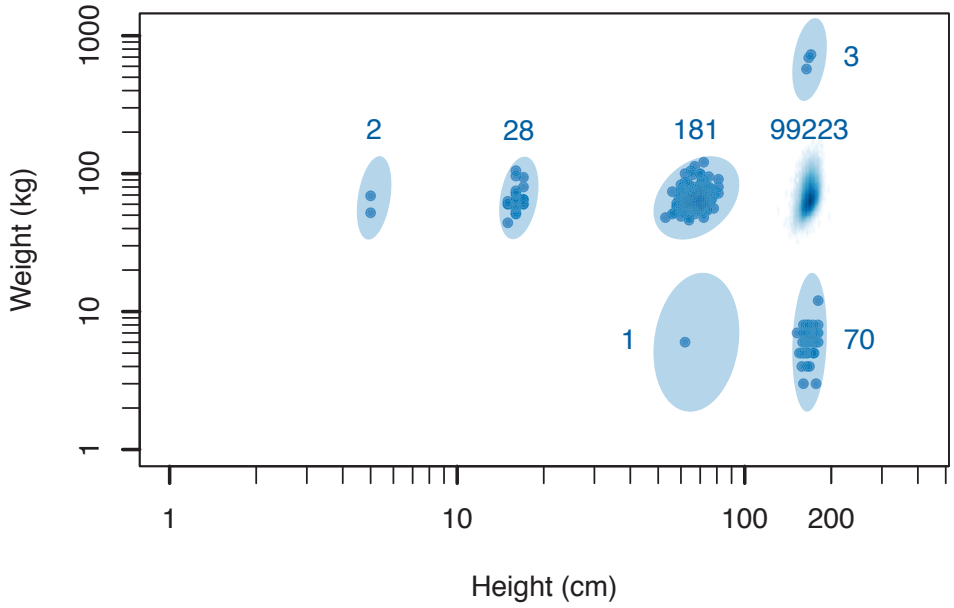


Figure 16. *Systematic errors introduced by optical character recognition algorithm while reading hand-written numbers. Blue points indicate errors (impossible values); blue cloud on the center-right shows the density of correct observations; blue-shaded areas are expected positions of errors if they were to be systematic, e.g., due to certain types of digit loss. The number of observations is shown next to each cluster. Data from a self-reported maternal Questionnaire-1 in MoBa.*

Instead, we realized that due to irregularity of hand-written numbers, in certain cases optical character recognition algorithm tends to loose "1" as the first or the last digit in the three-digit height values. For example, an entered height "173" can become "73" or "17". Similarly, when maternal weight was hand-written as a three-digit number with a decimal point (e.g., "62.5" kg), the algorithm interpreted it as "625" kg. A less common error was entering height in feet units instead of centimeters. Without this discovery, we would have had to eliminate these samples from analyses.

8.5.2. The non-transmitted haplotype

The key component in our analyses was to determine which maternal alleles of height-associated SNPs were not transferred to the fetus. This would not be possible if data would contain only mothers or only children. Ideally, family triad (mother, father, child) would be used, but none of the three cohorts had paternal genetic data.

In the case of family dyad (mother and child), identifying a non-transmitted allele is very straightforward only for some SNPs: if either or both of the dyad members are homozygotes, the allele transmission can be unambiguously determined from their genotypes. However, when both the mother and her child are heterozygotes at a particular SNP, then inference of allele transmission is impossible without additional data and algorithmic solution **(Figure 17).**

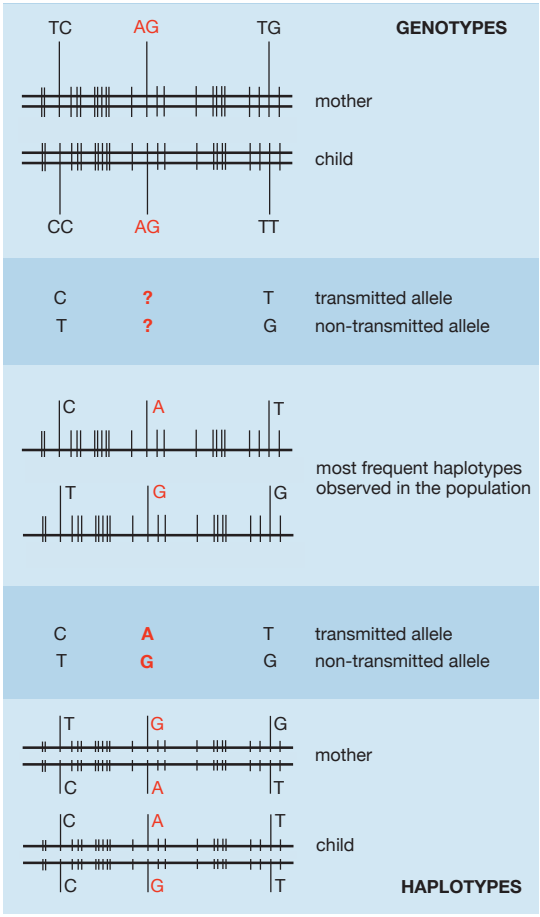


Figure 17. *Haplotype inference (phasing) in family-duo genotyping data. Stage 1: phasing using only Mendelian logic. Stage 2: phasing using additional information about haplotypes in the population (determined using sequencing methods). Horizontal lines represent chromosomes, vertical lines represent SNPs. Alleles highlighted in red cannot be phased using only Mendelian logic.*

For each of these twice-heterozygous SNPs, we constructed a long-range local haplotype and compared haplotype sharing to determine allelic transmission. In other words, the transmission was revealed by neighborhing SNPs that were not twice-heterozygous.

Unexpectedly, we also noticed that phasing software *Shapeit2*[106] (used for imputation procedures) unintentionally arranges its output in a way revealing maternal non-transmitted alleles. We used this accidental feature in the sensitivity analysis (not published in the paper), which agreed well with the in-house built method.

8.6. Impact and echoes

During the 37 months of its online presence, Study II was cited 30 times by other scientific publications from high-IF (Impact Factor) journals: *Nature Reviews Genetics* (IF=40), *Nature* (IF=40), *Science* (37), *Nature Genetics* (28), *New England Journal of Medicine* (79), *PLoS Medicine* (11), *PLoS Genetics* (6), *Human Molecular Genetics* (6), *PNAS* (10), *Epidemiology* (5). We achieved twice as high citation rate than the mean of *PLoS Medicine* journal. Study II was viewed or downloaded by readers 14,500 times and received good resonance in genetic conferences.

Importantly, one of the journals citing Study II was the *Nature Reviews Genetics*. In this review[107] of genomic causal inference tools, our method was highlighted in it's own new category in the section "Extensions of Mendelian randomization".

One *Nature* study from an independent research group was directly inspired by our novel non-transmitted haplotype score Mendelian randomization method[108]. Three other studies followed suit[109-111], the second one recommending it as a sensitivity analysis.

To our knowledge, nothing similar to our non-transmitted haplotype MR method has been applied before. The findings of Study II were replicated in a new larger study (currently a manuscript), where we demonstrate two orders of magnitude higher statistical significance than reported originally. A direct consequence of the findings in Study II was the Study III.

09 STUDY III: UTERINE DISTENTION

9. Study III: Uterine Distension

Uterine distention as a factor in birth timing: retrospective nationwide cohort study in Sweden.

9.1. Background

Study III was inspired by findings in Study II where we found evidence that maternal height causally affects a child's gestational age at birth. In Study II we already had a hypothesis on how this could be happening biologically. It involved several assumptions* and the central premise of uterine distention imposed by fetal growth. However, this hypothesis had to be challenged further[112].

The history of epidemiology is rich in "natural experiments", reaching the times of John Snow in the mid nineteenth century[113]. The idea of how to construct a falsification challenge to our hypothesis came from realization, that there are natural experiments modeling uterine distention in humans.

The most obvious natural experiment modeling two levels of uterine distention was twin and singleton pregnancies. At any time in pregnancy, twins should, on average, occupy more space than singletons. When near term, tension forces experienced by uterus carrying twins should be larger than in singleton pregnancies. If parturition is triggered by hypothetical stress signals coming from overstretched uterus (or the fetus experiencing restricted growth space), then twins should, on average, be born earlier than singletons. That is the first challenge to the hypothesis, and it passes the test: the mean GA in twin pregnancy is ~37 weeks[114], in triplets and quadruplets it is ~34 and ~31 weeks, respectively[115]. Together with the mean GA in singleton pregnancy (~40 weeks), these numbers demonstrate a textbook pattern of negative association between the human litter size and gestational age at birth[116]. We observed this pattern in our data (**Figure 18**). In fact, the association between the litter size and pregnancy duration is also well documented in other animal species[117,118].

Besides the twin vs singleton comparison, there was one additional component that eventually led us to the method used in Study III. Our next step in exploring uterine stretch hypothesis was to investigate the strength of association between maternal height and gestational age in various strata of both variables (**Figure 19**). We speculated that this association should be the strongest when the mother is relatively short and the child is born relatively early. Evidence suggested that this is exactly the case.

These ideas have led to the method used in the Study III, where we pushed the challenge to the hypothesis even further.

---

* *One assumption is that maternal height is correlated with uterine size and with tension forces experienced due to stretch imposed by uterine load.*

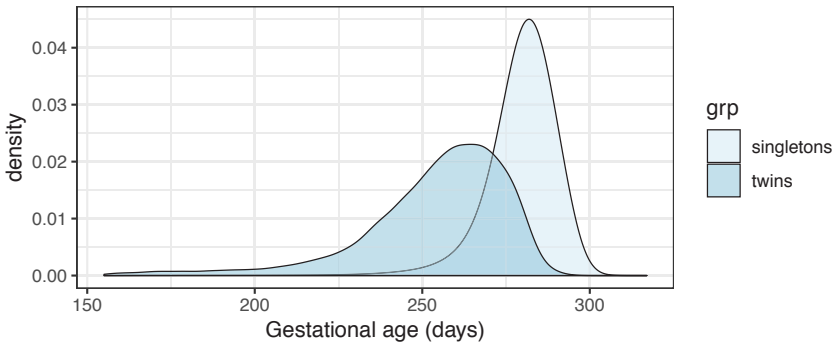


Figure 18. *Distribution of gestational age in twin and in singleton pregnancies. Data from the Swedish Medical Birth Register (1990-2013). Mean gestational age (GA) in twins and singletons is 254 and 280 days, respectively. Only spontaneous deliveries with live-born children, GA evaluated using ultrasound method.*

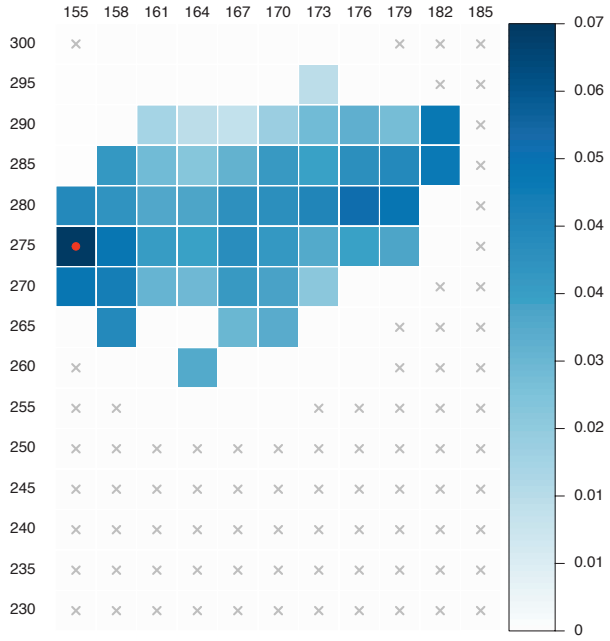


Figure 19. *Effect of maternal height on gestational age at various windows of each variable. Each cell represents a sub-population within a 10-cm-window in maternal height (columns) and 20-day-window in gestational age (rows), the center point of each window is shown on the axes. Colour scale indicates the effect size (days/cm). The cells that are not significant after multiple-test correction (15×11 tests) are assigned beta = 0 (white). The number of individuals in each cell determines the power (empirical probability) to detect effect size larger or equal to the effect size in the reference cell (marked with red dot): cells with <80% power are marked with grey crosses. Data from Swedish Medical Birth Register (1991-2012).*

9.2. Summary of the study

We used anthropometric data from the Swedish Medical Birth Registry to non-invasively test the hypothesis that gestation-imposed uterine stretch can trigger births in normal human pregnancies.

The quasi-experiment tested a hypothesis-derived prediction about a specific visual interaction pattern that should be observed when child's gestational age at birth is plotted against maternal height during pregnancy in two strata of uterine load. This pattern can be quantitatively evaluated using statistical significance of the interaction term between uterine load and maternal height in linear regression. Two natural models of large and small uterine load were found: twin vs singleton and large vs small for gestational age fetuses (LGA and SGA, respectively). At any time of pregnancy, twins and LGA fetuses should impose a larger uterine load than singletons and SGA fetuses. In both models, the observed (data-derived) interaction pattern was qualitatively identical to the predicted one.

9.3. Novelty

To our knowledge, none of the epidemiological studies investigating association between maternal height and child's gestational age at birth have gone further in exploring the causal mechanism. The novelty of our work is the insight that "uterine stretch" hypothesis predicts a specific visual interaction pattern that can be easily tested with the data accessible to epidemiologists in almost any country. If falsification challenge[112] to this prediction fails, it could be taken as evidence arguing for the hypothesis.

Another novel aspect in Study III is a regard to the competing background risks that can cause birth thus setting the gestational age[*].

9.4. Limitations

The strength of association between maternal height and the child's gestational age is dependent on the GA evaluation method, e.g., short and tall mothers have mean GA difference of 4.3 days when UL-based dating was used, but only 2.8 days when LMP-based dating was used[119]. This difference might be due to the fact that maternal height and fetal growth rates are correlated (due to genetic pleiotropy). Since UL method uses normal fetal growth rate in its assumptions, gestational age in tall mothers can be overestimated, and in short mothers - underestimated.

Fetal growth rate was estimated from birth weight and gestational age using reference longitudinal growth curves derived from healthy pregnancies with term deliveries. It is possible that such growth curves would be biased when assigning the birth weight Z-score (equivalent to growth rate) to preterm-born children.

These complex correlations between three main variables in Study III might have unintended effects on the interaction model.

[*] *see "Bathtub model" in the section Methodological aspects, "Simulation"*

Lastly, it is important to emphasize another hypothesis that could explain the observed interaction pattern. Instead of gestation-imposed uterine distention, it describes gestation-imposed metabolic constraints[120].

9.5. Methodological aspects

9.5.1. Interaction model

A particularly challenging task was to find the best way to detect and visualize the interaction pattern. Eventually, the simplest method* was chosen out of half a dozen less traditional ones. Some of them are presented in **Figure 20**.

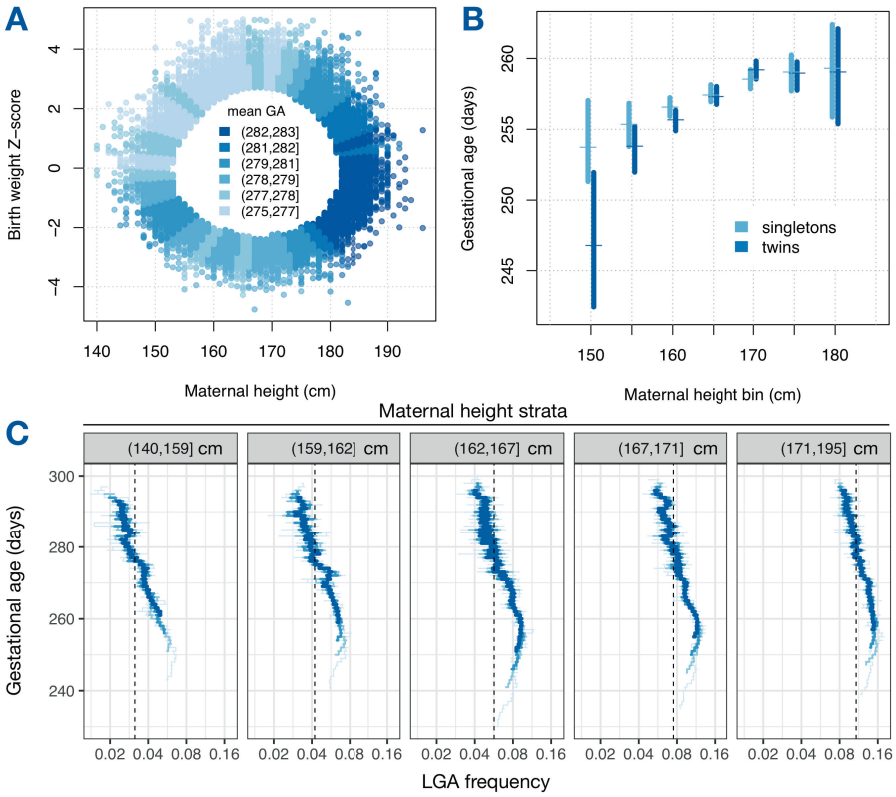


Figure 20. *Three alternative approaches explored for data analysis and visualization.* ***Panel A:*** *only extreme observations are analysed, where "extreme" is defined by the distance from the center (median) in a bivariate plot. There are 50 radial groups and 500 observations in each. Observations are coloured based on mean gestational age of the group. Birth weight Z-score is adjusted for gestational age using Marsal et al. method [121], and then adjusted for maternal height.* ***Panel B:*** *Singleton pregnancies are matched to twin pregnancies based on gestational age at birth. After that, seven bins of maternal height are formed, each 5 cm in size. Mean gestational age and 90% confidence intervals are estimated using bootstrap method for all 14 resulting groups of pregnancies. [continued]*

* *Study III, Figure 2.*

***Panel C**: Five boxes represent strata of increasing maternal height. X-axis is LGA frequency, where LGA is defined as birth weight Z-score >1.5 (SD); Z-score is adjusted for gestational age using Marsal et al. method. Vertical dotted lines show the LGA frequency in each stratum. Y-axis is gestational age at birth. Blue lines are smoothed moving averages, where colour encodes window size (from light to dark): 1000, 2000, 3000, and 5000 observations. Data: Study III.*

9.5.2. Simulation

The use of simulation models is not unusual in preterm birth research[122-124]. Although not included in the Study III publication, the simulation was the key element that convinced us to publish the findings. It allowed us to use the classical hypothesis-testing framework - a cornerstone of the scientific method.

Simulation 1 was the background risk alone. In pregnancy, the background risk symbolizes a natural gestational clock, for millions of years fine-tuned by evolution. It represents all unobserved genetic factors shaping the duration of normal human pregnancy, regardless of maternal height, uterine size and fetal growth rate.

Simulation 2 was informally named "the bathtub model". In it, metaphorically speaking, the time at which water spills out of the bathtub (gestational age at birth) is determined only by the volume of the bathtub (uterine size) and flow rate of water (fetal growth rate). No other factors were in place. Variation in spill time is deterministically caused by variation in volumes and flows.

Simulation 3 was a biological modification of the bathtub model. In it, an additional component was introduced - a background risk. The background risk competes with the bathtub model, and the one, which happens to trigger the birth first, determines gestational age.

As seen in **Figure 21**, only simulation 3 (Panel C) showed an interaction pattern qualitatively identical to the one observed in real data (Study III, Fig 2). Importantly, this was the model that we believed to be of closest resemblance to nature.

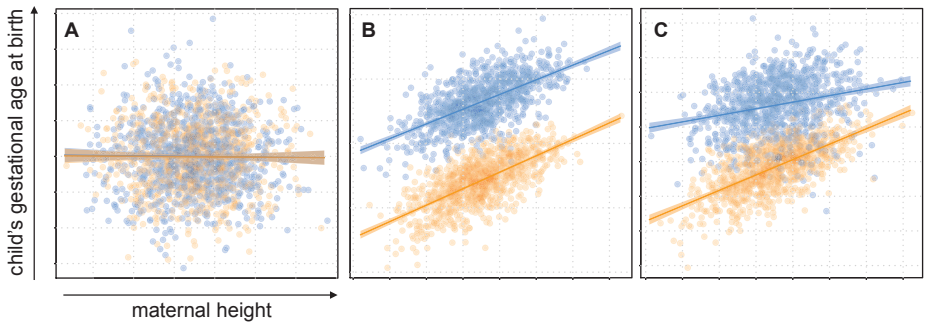


***Figure 21.** Simulation results. **Panel A:** birth is triggered only by a baseline risk (unknown environmental and genetic factors); **Panel B:** birth is triggered only by interaction (fetal size reaching the uterine size capacity); **Panel C:** birth is triggered by competing baseline and interaction risks. Pregnancies with slow fetal growth are depicted in blue circles and modelled by blue regression line with 95% confidence intervals; pregnancies with fast fetal growth are depicted in orange circles and modelled by orange regression line with 95% confidence intervals. Simulation used 10000 observations in each scenario.*

9.6. Impact and echoes

Study III was published in *BMJ Open* journal in October 2018, a month before the defence of this thesis. No data on citations, views or downloads is available at the time of writing.

Working on the ideas behind Study III inspired our extended research group to use the insight about competing risk events that determine pregnancy length in other projects. We hope that this insight will be useful to other researchers.

Study III indirectly implicated inflammation-related candidate genes for a currently ongoing genotype-phenotype association study.

10 ETHICAL APPROVALS

10. Ethical Approvals

Individuals used in Studies I-III are anonymous. The personal identification numbers that are used to combine various data sources are known only by the data managers, e.g., the Norwegian Institute of Public Health (for MoBa data) and Swedish National Board of Health and Welfare (for the Swedish Medical Birth Register data).

In **MoBa** cohort, written consent was received from all study participants[*] during the years of enrolment. The consent covers the use of biological samples, questionnaire data and medical records for scientific research purposes. Study participants have the right to opt-out from the study at any time, in which case their data would be deleted and biological samples destroyed. The informed-consent form for data collection was approved by the Norwegian Data Inspectorate (ref No. 01/4325) and by the Regional Committee for Medical Research Ethics (ref No. S-97045 and S-95113). Our study was approved by the Regional Committee for Medical Research Ethics in South-Eastern Norway (S-06075) and the Norwegian Data Inspectorate (05/016784). The approvals also apply to the Figures 2, 10, and 16 of this thesis.

In the **DNBC** cohort, the genomic and phenotypic data collection for medical research purposes of preterm birth was approved by the Research Ethics Committee of the Capital Region (Copenhagen) under the ID number H-A-2007-0017. The approval also applies to Figure 10 of this thesis. All adult participants gave written informed consent.

In **FIN** cohort, the Ethics Committee of Oulu University Hospital and Ethics Committee of Helsinki University Central Hospital approved the data collection. Permission to use genetic data for preterm birth studies was given by Cincinnati Children's Hospital Institutional Review Board (Nr: 2011-2047). The approvals also apply to Figure 10 of this thesis. All adult participants gave written informed consent.

As is the case in most genomic studies, if researchers would discover a carrier of highly detrimental mutation, they would not be ethically (or legally) allowed to contact and inform this study subject. Such decision and action could only be made by the data manager and only with a special decision of regional ethical review board.

The **Swedish Medical Birth Register** is a national population-wide dataset, thus the informed consent is not required. The collected data is considered a public property but released to researchers only with a permission of regional ethical review board and after data manager approves the research plan. Study III was approved by ethical review board of Western Health Care Region in Sweden (Nr: 968-14). The same approval also applies to Figures 1-5, and 18-20 of this thesis. Additional approval (Nr: 576-13) applies to Figures 6 and 7.

[*] *For children, consent was given by their mothers.*

11 FUTURE CHALLENGES

Jonas Bačelis

11. Future Challenges

It seems that currently we find ourselves in an episode of scientific progress, which Thomas Kuhn[125] referred to as "normal science" (juxtaposing it to the revolutionary episode - "paradigm shift"). Today, the scientists work within the central paradigm by doing "puzzle-solving", i.e., the progress is defined less by ingenious ideas but more by mundane tasks of applying known methods to new datasets. Thus, the steady progress relies on large investments in genotyping, sequencing, biobanking, infrastructure, analytical tools and pipelines.

The field of biomedicine is fertile with small and large enigmas, all ready to be deciphered from the four-letter alphabet code to a meaningful language of pathways and mechanisms. Every biological trait has thousands of causal genetic variants, and technological advancements of the current Renaissance of Genetic Era are likely to find most of them. From the current perspective it seems that genetic epidemiologists will have enough work for at least a couple more decades, performing an important but rather dull task - doing more of the same, only with larger datasets and more reliable methods.

But there are real challenges in the future, and in this chapter I would like to focus on them.

11.1. Sharing and harmonization

In the context of genomics, statistical power is our ability to correctly detect causal genetic factors. The power of statistical investigation monotonically increases with the number of study subjects used in it. If we want to find most of the risk-increasing mutations, we must be able to detect even those that impose a very small effect on the phenotype. Moreover, we should also be able to detect mutations with large effect but very rare in the population. More of correctly detected factors mean a clearer picture of biology behind the disease. And the more we know about its biology - the more molecular targets we can aim at with pharmacological missiles.

As the number of genotyped or sequenced human genomes will continue to grow, we will be faced with a problem of who owns the data and how easy it is to access it. Imagine that 1000 research groups each have genotyped 1000 individuals and discovered the most obvious genetic culprits (the "low-hanging fruits"). They published 1000 articles, each about one disease and each unable to account for any significant fraction of the total genetic component. However, the same amount of data has a tremendous power for big discoveries, but only if these groups would collaborate or decide to make their data open. For example, if 1000 genomes gives you a 80% power to detect a common genetic variant with minor-allele frequency (MAF) of 50% and effect size odds ratio (OR) >1.7, then, without any decrease in statistical power, 1000×1000 genomes could enable you to detect a much wider range of genetic variants: with MAF reaching as low as 5% and effect size as small as OR=1.05 (or also MAF=1% and OR=1.1). This is a stunning improvement.

For such cooperation and data sharing to flourish there must be a legal support and political will. Ethical framework will be needed that clarifies whether my genome is my property, or whether it is a collective data. Currently, researchers and their institutes do not have good reasons to share data, since access to the data defines their advantage over other researchers without it. Correct incentives by grant foundations could fix this.

The next question is how to harmonize the phenotype data. Each genotyping project would have to include much wider variety of phenotypic questions to make their data useful to other research groups.

11.2. Unintended consequences

The most reliable way to eliminate an illness from a population is to wait long enough until the natural selection drives the causal alleles to extinction. Naturally, the waiting time is too long for us to benefit, as evolutionary changes could take longer than a lifespan of a civilization. Currently, our best method of coping with illness is, in essence, curing only the *symptoms* of what is often a genetic problem; meanwhile the true genetic cause is never eliminated and propagates through generations. The reality is even grimmer: since the dawn of medicine, we have constantly tinkered with our survival probabilities, gradually becoming invisible to the natural selection. At the same time, random mutations are introduced at a steady pace of 50 *de novos* per person[126], thus the number of health-impairing mutations accumulates in the gene pool, and their frequencies, absent the selective pressure, drifts randomly.

Such interference with self-regulating evolutionary process has never happened in a history of Life before. And someday we might be facing the consequences. Our strive for health and happiness might come back to haunt us in unexpected ways: at future extreme environments, in unpredictable events and catastrophes, or population bottlenecks (e.g., colonization of Mars). With reduced medical supplies, power shortages, no medical service, we might find ourselves extremely vulnerable to diseases of genetic origin that we once though we had tamed. Very far in the future, our descendants riddled with autism, arthritis, asthma and atherosclerosis might look back at us as irresponsible innovators. Maybe, in a similar way as we look back at our ancestors, who thought that fossil fuels have no side effects.

Is there anything that could be done? The answers balance on an edge of modern ethics. If some form of artificial allele-frequency adjustment was to be introduced into our species, we should be very cautious about mistakes carelessly done in the beginning of the 20th century and look for means that do not trample the human rights and dignity. The knowledge of genetic risk factors can be used to introduce selective incentives for risk-allele carriers, e.g., to choose adoption instead of biological parenting in exchange for generous subsidies. We can only hope that such incentives will be wise and compassionate, because without humane artificial selection we will be stuck with the second option - genome-editing technologies. These carry great risks of unintended biological consequences and are prone to becoming the next source of inequality (i.e., "best genes for best money").

11.3. Publishing for people

The phenomenon of academic piracy serves as a barometer for injustice in the field of scholarly publishing. The current copyright laws are especially detrimental for education and research and have led to civil disobedience acts, such as "Sci-Hub", a searchable online database of stolen scholarly journal articles created by a "pirate in hiding" - Alexandra Elbakyan[127]. Earlier, a similar altruistic effort to make scientific research open and free to public have claimed a life of a brilliant programmer and activist Aaron Swartz[128].

According to the 2011 data, only 17% of articles were available as open-access, most of them (12%) immediately but some (5%) within 6-12 months of publication[129]. More than a third of all journals still publish papers behind a strict paywall (subscription-only model). Less than half have adopted a hybrid model, where papers are immediately made free to read for a fee if the authors wish, but most studies are still kept behind a paywall. Injustice comes from a fact that in a great majority of cases, the public tax money pays for the scholarly work: experiments, analyses, writing, reviewing (public service), editing, and publishing. At the same time, to access the material, university libraries are required to meet expensive subscription fees, while those outside the paywall are charged $20 - $50 to read a single article. Needless to say, this is unfair. Anyone should be able to access published results without the need to pay for it. Such injustice disproportionately affects low-income countries, by obstructing education and research and propagating inequality.

During the last weeks of writing this thesis, eleven European funding organizations (including the Research Council of Norway and the Swedish Research Council for Sustainable Development) announced an open-access initiative that requires grantees to make resulting papers free to download, read, translate or otherwise reuse the work under Creative Commons copyright license CC BY* immediately on publication.

The plan does not oppose the article-processing charges but wants to cap them. It also wants funders and universities to pay, not the authors. This initiative was inspired by open-access policy of the *Bill & Melinda Gates Foundation*, which also demands immediate open-access publishing. A few other funding agencies also use similar policies: the Wellcome Trust (UK) and the National Institutes of Health (USA).

In a near future, publicly funded researcher will not be published in *Nature*, *Science*, *Cell*, *The Lancet* and other hybrid journals, unless they change the reading and publishing subscriptions into a single fee. The big challenge will be to implement this plan without a drop in quality of peer-review work, without transferred costs disincentivizing the less-funded or independent scientists, and without pushing high-quality and high-prestige journals to extinction.

All three studies listed in this thesis were published in open-access journals.

* *https://creativecommons.org/licenses/*

12 OTHER CO-AUTHORED PUBLICATIONS

12. Other Co-authored Publications

Liu X, [...], Bacelis J, [...], Jacobsson B, [...], Feenstra B. **Variants in the fetal genome near pro-inflammatory cytokine genes on 2q13 are associated with gestational duration.** [bioRxiv: doi.org/10.1101/423897; 2018 Sept 23. Submitted to Nat Genet, ref.Nr.: NG-A49980].

Warrington NM, Beaumont RN, [...], Bacelis J, [...], Jacobsson B, [...], Freathy RM. **Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors.** [Manuscript. Submitted to Nat Genet, ref.Nr.: NG-A49297].

Murray SR, [...], Bacelis J, [...], Jacobsson B. **Geographical differences in preterm delivery rates in Sweden: a population-based cohort study.** Acta Obstet Gynecol Scand, 2018. 00:1–11. doi:10.1111/aogs.13455.

Zhang G, [...], Bacelis J, [...], Jacobsson B, Muglia LJ. **Genetic studies of gestational duration and preterm birth**. Best Pract Res Clin Obstet Gynaecol, 2018. [in press].

Papadopoulou E, [...], Bacelis J, [...], Jacobsson B, Sengpiel V. **Maternal caffeine intake during pregnancy and childhood growth and overweight: results from a large Norwegian prospective observational cohort study.** BMJ Open, 2018. 8:e018895.

Bohman A, [...], Bacelis J, Bende M, Torinsson NÅ. **A family-based genome-wide association study of chronic rhinosinusitis with nasal polyps implicates several genes in the disease pathogenesis.** PLoS One, 2017. 12:e0185244.

Zhang G, Feenstra B, Bacelis J, [...], Jacobsson B, Muglia LJ. **Genetic Associations with Gestational Duration and Spontaneous Preterm Birth.** N Engl J Med, 2017. 377:1156-67.

Juodakis J, Bacelis J, [...], Jacobsson B. **Time-Variant Genetic Effects as a Cause for Preterm Birth: Insights from a Population of Maternal Cousins in Sweden.** G3, 2017. 7:1349-56.

Middeldorp CM, [...], Bacelis J, [...], Jacobsson B, [...], Boomsma DI. **A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts.** J Am Acad Child Adolesc Psychiatry, 2016. 55:896-905 e6.

Bustamante M, [...], Bacelis J, [...], Jacobsson B, Sunyer J. **A genome-wide association meta-analysis of diarrhoeal disease in young children identifies FUT2 locus and provides plausible biological pathways.** Hum Mol Genet, 2016. 25:4127-42.

Okbay A, [...], Bacelis J, [...], Jacobsson B, [...], Benjamin DJ. **Genome-wide association study identifies 74 loci associated with educational attainment.** Nature, 2016. 533:539-42.

Paternoster L, [...], Bacelis J, [...], Jacobsson B, [...], Weidinger S. **Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis.** Nat Genet, 2015. 47:1449-56.

Sengpiel V, Bacelis J, [...], Jacobsson B. **Folic acid supplementation, dietary folate intake during pregnancy and risk for spontaneous preterm delivery: a prospective observational cohort study.** BMC Pregnancy Childbirth, 2014. 14:375.

Östensson M, Montén C, Bacelis J, Torinsson-Naluai Å. **A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease.** PLoS One, 2013. 8:e70174.

Sengpiel V, Elind E, Bacelis J, [...], Jacobsson B, Brantsaeter AL. **Maternal caffeine intake during pregnancy is associated with birth weight but not with gestational length: results from a large prospective observational cohort study.** BMC medicine, 2013. 11:42.

13 REFERENCES

13. References

1. Reich D. Who We Are and How We Got Here: Ancient DNA and the new science of the human past. Oxford University Press; 2018.
2. Church G. Please reanimate. Sci Am 2013;309:12.
3. Zhang W, Zhao G, Luo Z, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. Science 2017;355.
4. Champer J, Buchman A, Akbari OS. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nat Rev Genet 2016;17:146-59.
5. Wang Y, Shen Y, Gu Y, Zhu S, Yin Y. Genome Writing: Current Progress and Related Applications. Genomics Proteomics Bioinformatics 2018;16:10-6.
6. Norman JE, Marlow N, Messow CM, et al. Vaginal progesterone prophylaxis for preterm birth (the OPPTIMUM study): a multicentre, randomised, double-blind trial. Lancet 2016;387:2106-16.
7. Berghella V, Saccone G. Fetal fibronectin testing for prevention of preterm birth in singleton pregnancies with threatened preterm labor: a systematic review and metaanalysis of randomized controlled trials. Am J Obstet Gynecol 2016;215:431-8.
8. Jukic AM, Baird DD, Weinberg CR, McConnaughey DR, Wilcox AJ. Length of human pregnancy and contributors to its natural variation. Hum Reprod 2013;28:2848-55.
9. Mahendru AA, Wilhelm-Benartzi CS, Wilkinson IB, McEniery CM, Johnson S, Lees C. Gestational length assignment based on last menstrual period, first trimester crown-rump length, ovulation, and implantation timing. Arch Gynecol Obstet 2016;294:867-76.
10. Delpachitra P, Palmer K, Onwude J, et al. Ultrasound Reference Chart Based on IVF Dates to Estimate Gestational Age at 6-9 weeks' Gestation. ISRN Obstet Gynecol 2012;2012:938583.
11. Rysavy MA, Li L, Bell EF, et al. Between-hospital variation in treatment and outcomes in extremely preterm infants. N Engl J Med 2015;372:1801-11.
12. Quinn JA, Munoz FM, Gonik B, et al. Preterm birth: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine 2016;34:6047-56.
13. Spong CY. Defining "term" pregnancy: recommendations from the Defining "Term" Pregnancy Workgroup. JAMA 2013;309:2445-6.
14. Steer P. The epidemiology of preterm labour. BJOG : an international journal of obstetrics and gynaecology 2005;112 Suppl 1:1-3.
15. Goldenberg RL, Gravett MG, Iams J, et al. The preterm birth syndrome: issues to consider in creating a classification system. Am J Obstet Gynecol 2012;206:113-8.
16. Phillips JB, Abbot P, Rokas A. Is preterm birth a human-specific syndrome? Evol Med Public Health 2015;2015:136-48.
17. Patel RR, Steer P, Doyle P, Little MP, Elliott P. Does gestation vary by ethnic group? A London-based study of over 122,000 pregnancies with spontaneous onset of labour. Int J Epidemiol 2004;33:107-13.
18. Blencowe H, Cousens S, Oestergaard MZ, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 2012;379:2162-72.
19. Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 2015;385:430-40.
20. Blencowe H, Lee AC, Cousens S, et al. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. Pediatr Res 2013;74 Suppl 1:17-34.
21. Rossling H, Rosling Rönnlund A, Rosling O. Factfulness: Ten Reasons We're Wrong About the World and Why Things Are Better Than You Think. Flatiron Books; 2018.
22. Moster D, Lie RT, Markestad T. Long-term medical and social consequences of preterm birth. N Engl J Med 2008;359:262-73.
23. Dimes Mo, PMNCH, Children St, WHO. Born Too Soon: The Global Action Report on Preterm Birth. Geneva: World Health Organization; 2012.
24. Escobar GJ, McCormick MC, Zupancic JA, et al. Unstudied infants: outcomes of moderately premature infants in the neonatal intensive care unit. Arch Dis Child Fetal Neonatal Ed 2006;91:F238-44.
25. Escobar GJ, Joffe S, Gardner MN, Armstrong MA, Folck BF, Carpenter DM. Rehospitalization in the first two weeks after discharge from the neonatal intensive care unit. Pediatrics 1999;104:e2.
26. Joffe S, Escobar GJ, Black SB, Armstrong MA, Lieu TA. Rehospitalization for respiratory syncytial virus among premature infants. Pediatrics 1999;104:894-9.
27. Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. Lancet 2012;379:445-52.
28. Marlow N, Wolke D, Bracewell MA, Samara M, Group EPS. Neurologic and developmental disability at six years of age after extremely preterm birth. N Engl J Med 2005;352:9-19.
29. Serenius F, Ewald U, Farooqi A, et al. Neurodevelopmental Outcomes Among Extremely Preterm Infants 6.5 Years After Active Perinatal Care in Sweden. JAMA Pediatr 2016;170:954-63.

30. Farooqi A, Hagglof B, Sedin G, Gothefors L, Serenius F. Chronic conditions, functional limitations, and special health care needs in 10- to 12-year-old children born at 23 to 25 weeks' gestation in the 1990s: a Swedish national prospective follow-up study. Pediatrics 2006;118:e1466-77.
31. Gabrielson J, Hard AL, Ek U, Svensson E, Carlsson G, Hellstrom A. Large variability in performance IQ associated with postnatal morbidity, and reduced verbal IQ among school-aged children born preterm. Acta Paediatr 2002;91:1371-8.
32. Isaacs EB, Edmonds CJ, Chong WK, Lucas A, Morley R, Gadian DG. Brain morphometry and IQ measurements in preterm children. Brain 2004;127:2595-607.
33. WHO recommendations on antenatal care for a positive pregnancy experience: ; 2016.
34. Ferrero DM, Larson J, Jacobsson B, et al. Cross-Country Individual Participant Analysis of 4.1 Million Singleton Births in 5 Countries with Very High Human Development Index Confirms Known Associations but Provides No Biologic Explanation for 2/3 of All Preterm Births. PLoS One 2016;11:e0162506.
35. Iams JD, Goldenberg RL, Meis PJ, et al. The length of the cervix and the risk of spontaneous premature delivery. National Institute of Child Health and Human Development Maternal Fetal Medicine Unit Network. N Engl J Med 1996;334:567-72.
36. Abbott DS, Hezelgrave NL, Seed PT, et al. Quantitative fetal fibronectin to predict preterm birth in asymptomatic women at high risk. Obstet Gynecol 2015;125:1168-76.
37. Esplin MS, Elovitz MA, Iams JD, et al. Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. JAMA 2017;317:1047-56.
38. Jung EJ, Byun JM, Kim YN, et al. Antenatal magnesium sulfate for both tocolysis and fetal neuroprotection in premature rupture of the membranes before 32 weeks' gestation. J Matern Fetal Neonatal Med 2018;31:1431-41.
39. Iams JD, Romero R, Culhane JF, Goldenberg RL. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. Lancet 2008;371:164-75.
40. Simhan HN, Caritis SN. Drug therapy: Prevention of preterm delivery. The New England journal of medicine 2007;357:477-87.
41. Kenyon SL, Taylor DJ, Tarnow-Mordi W. Broad-spectrum antibiotics for preterm, prelabour rupture of fetal membranes: the ORACLE I randomised trial. ORACLE Collaborative Group. Lancet 2001;357:979-88.
42. Kenyon S, Pike K, Jones DR, et al. Childhood outcomes after prescription of antibiotics to pregnant women with spontaneous preterm labour: 7-year follow-up of the ORACLE II trial. Lancet 2008;372:1319-27.
43. Hassan SS, Romero R, Vidyadhari D, et al. Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo-controlled trial. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 2011;38:18-31.
44. Partridge EA, Davey MG, Hornick MA, et al. An extra-uterine system to physiologically support the extreme premature lamb. Nat Commun 2017;8:15112.
45. Partridge EA, Davey MG, Hornick MA, Flake AW. An EXTrauterine environment for neonatal development: EXTENDING fetal physiology beyond the womb. Semin Fetal Neonatal Med 2017;22:404-9.
46. Portmann A. Biologische Fragmente zu einer Lehre vom Menschen, 3rd exp. ed. Basel: Schwabe; 1969.
47. Rosenberg K, Trevathan W. Bipedalism and human birth: The obstetrical dilemma revisited. Evolutionary Anthropology: Issues, News, and Reviews 1995;4:161-8.
48. Muglia LJ, Katz M. The enigma of spontaneous preterm birth. The New England journal of medicine 2010;362:529-35.
49. Murray SR, Juodakis J, Bacelis J, et al. Geographical differences in preterm delivery rates in Sweden: a population-based cohort study. Acta obstetricia et gynecologica Scandinavica 2018.
50. Vladutiu CJ, Marshall SW, Poole C, Casteel C, Menard MK, Weiss HB. Adverse pregnancy outcomes following motor vehicle crashes. Am J Prev Med 2013;45:629-36.
51. Saurel-Cubizolles MJ, Zeitlin J, Lelong N, et al. Employment, working conditions, and preterm birth: results from the Europop case-control survey. J Epidemiol Community Health 2004;58:395-401.
52. Launer LJ, Villar J, Kestler E, de Onis M. The effect of maternal work on fetal growth and duration of pregnancy: a prospective study. Br J Obstet Gynaecol 1990;97:62-70.
53. Pompeii LA, Savitz DA, Evenson KR, Rogers B, McMahon M. Physical exertion at work and the risk of preterm delivery and small-for-gestational-age birth. Obstet Gynecol 2005;106:1279-88.
54. Han Z, Mulla S, Beyene J, Liao G, McDonald SD, Knowledge Synthesis G. Maternal underweight and the risk of preterm birth and low birth weight: a systematic review and meta-analyses. Int J Epidemiol 2011;40:65-101.
55. Jeffcoat MK, Geurs NC, Reddy MS, Cliver SP, Goldenberg RL, Hauth JC. Periodontal infection and preterm birth: results of a prospective study. J Am Dent Assoc 2001;132:875-80.
56. Meis PJ, Goldenberg RL, Mercer B, et al. The preterm prediction study: significance of vaginal infections. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol 1995;173:1231-5.
57. Copper RL, Goldenberg RL, Das A, et al. The preterm prediction study: maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol 1996;175:1286-92.
58. Andres RL, Day MC. Perinatal complications associated with maternal tobacco use. Semin Neonatol 2000;5:231-41.
59. Thompson JM, Irgens LM, Rasmussen S, Daltveit AK. Secular trends in socio-economic status and the implications for preterm birth. Paediatr Perinat Epidemiol 2006;20:182-7.
60. Smith LK, Draper ES, Manktelow BN, Dorling JS, Field DJ. Socioeconomic inequalities in very preterm birth rates. Arch Dis Child Fetal Neonatal Ed 2007;92:F11-4.

61. Jackson RA, Gibson KA, Wu YW, Croughan MS. Perinatal outcomes in singletons following in vitro fertilization: a meta-analysis. Obstet Gynecol 2004;103:551-63.
62. Astolfi P, De Pasquale A, Zonta L. Late childbearing and its impact on adverse pregnancy outcome: stillbirth, preterm delivery and low birth weight. Rev Epidemiol Sante Publique 2005;53 Spec No 2:2S97-105.
63. Conde-Agudelo A, Rosas-Bermudez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. JAMA 2006;295:1809-23.
64. Scholl TO, Johnson WG. Folic acid: influence on the outcome of pregnancy. Am J Clin Nutr 2000;71:1295S-303S.
65. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;42:441-7.
66. Clarke TK, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry 2017;22:1376-84.
67. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016;533:539-42.
68. Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun 2017;8:599.
69. Domschke K, Maron E. Genetic factors in anxiety disorders. Mod Trends Pharmacopsychiatry 2013;29:24-46.
70. Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B, Melbye M. Maternal contributions to preterm delivery. American journal of epidemiology 2009;170:1358-64.
71. Clausson B, Lichtenstein P, Cnattingius S. Genetic influence on birthweight and gestational length determined by studies in offspring of twins. BJOG : an international journal of obstetrics and gynaecology 2000;107:375-81.
72. Wilcox AJ, Skjaerven R, Lie RT. Familial patterns of preterm delivery: maternal and fetal contributions. American journal of epidemiology 2008;167:474-9.
73. Wu W, Witherspoon DJ, Fraser A, et al. The heritability of gestational age in a two-million member cohort: implications for spontaneous preterm birth. Hum Genet 2015;134:803-8.
74. Chaudhari BP, Plunkett J, Ratajczak CK, Shen TT, DeFranco EA, Muglia LJ. The genetics of birth timing: insights into a fundamental component of human development. Clin Genet 2008;74:493-501.
75. McElroy JJ, Gutman CE, Shaffer CM, et al. Maternal coding variants in complement receptor 1 and spontaneous idiopathic preterm birth. Hum Genet 2013;132:935-42.
76. Plunkett J, Feitosa MF, Trusgnich M, et al. Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. Hum Hered 2009;68:209-19.
77. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell 2011;147:32-43.
78. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). International journal of epidemiology 2006;35:1146-50.
79. Magnus P, Birke C, Vejrup K, et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol 2016;45:382-8.
80. Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. Acta obstetricia et gynecologica Scandinavica 2000;79:435-9.
81. Olsen J, Melbye M, Olsen SF, et al. The Danish National Birth Cohort--its background, structure and aim. Scand J Public Health 2001;29:300-7.
82. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007;39:1181-6.
83. Lan Q, Hsiung CA, Matsuo K, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nature genetics 2012;44:1330-5.
84. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. Nature genetics 2010;42:295-302.
85. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119-24.
86. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature genetics 2009;41:703-7.
87. Cho YS, Chen CH, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nature genetics 2012;44:67-72.
88. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nature genetics 2011;43:429-+.
89. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nature genetics 2010;42:781-U75.
90. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research 2017;45:D896-D901.
91. Krokstad S, Langhammer A, Hveem K, et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol 2013;42:968-77.
92. Leitsalu L, Haller T, Esko T, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol 2015;44:1137-47.
93. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.
94. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016;70:214-23.
95. Wu W, Clark E, Manuck T, Esplin M, Varner M, Jorde L. A Genome-Wide Association Study of spontaneous preterm birth in a European population [version 1; referees: 2 approved with reservations]. F1000Research 2013;2.

96. Zhang H, Baldwin DA, Bukowski RK, et al. A genome-wide association study of early spontaneous preterm delivery. Genet Epidemiol 2015;39:217-26.
97. Berry KJ, Mielke PW, J.E. J. Permutation Statistical Methods: An Integrated Approach. Springer; 2016.
98. Zhang G, Feenstra B, Bacelis J, et al. Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. N Engl J Med 2017;377:1156-67.
99. Schierding W, Antony J, Karhunen V, et al. GWAS on prolonged gestation (post-term birth): analysis of successive Finnish birth cohorts. J Med Genet 2018;55:55-63.
100. Liu X, Helenius D, Skotte L, et. a. Variants in the fetal genome near pro-inflammatory cytokine genes on 2q13 are associated with gestational duration. [biorxiv preprint] 2018.
101. Kuwata T, Matsubara S, Ohkusa T, et al. Establishing a reference value for the frequency of fetal movements using modified 'count to 10' method. J Obstet Gynaecol Res 2008;34:318-23.
102. Adams Waldorf KM, Singh N, Mohan AR, et al. Uterine overdistention induces preterm labor mediated by inflammation: observations in pregnant women and nonhuman primates. Am J Obstet Gynecol 2015;213:830 e1-e19.
103. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173-86.
104. Stulp G, Simons MJ, Grasman S, Pollet TV. Assortative mating for human height: A meta-analysis. Am J Hum Biol 2017;29.
105. Evans DM, Brion MJ, Paternoster L, et al. Mining the human phenome using allelic scores that index biological intermediates. PLoS Genet 2013;9:e1003919.
106. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 2013;10:5-6.
107. Pingault JB, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nat Rev Genet 2018.
108. Horikoshi M, Beaumont RN, Day FR, et al. Genome-wide associations for birth weight and correlations with adult disease. Nature 2016;538:248-52.
109. Richmond RC, Timpson NJ, Felix JF, et al. Using Genetic Variation to Explore the Causal Effect of Maternal Pregnancy Adiposity on Future Offspring Adiposity: A Mendelian Randomisation Study. PLoS Med 2017;14:e1002221.
110. Lawlor D, Richmond R, Warrington N, et al. Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: Sources of bias and methods for assessing them. Wellcome Open Res 2017;2:11.
111. Beaumont RN, Warrington NM, Cavadino A, et al. Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. Hum Mol Genet 2018;27:742-56.
112. Popper K. The Logic of Scientific Discovery. 1934.
113. Smith GD. Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain. Int J Epidemiol 2002;31:920-32.
114. Groutz A, Yovel I, Amit A, Yaron Y, Azem F, Lessing JB. Pregnancy outcome after multifetal pregnancy reduction to twins compared with spontaneously conceived twins. Hum Reprod 1996;11:1334-6.
115. Berkowitz RL, Lynch L, Stone J, Alvarez M. The current status of multifetal pregnancy reduction. Am J Obstet Gynecol 1996;174:1265-72.
116. Louise CK, Jenny EM. Obstetrics by Ten Teachers. 20 ed. CRC Press; 2017.
117. Okkens AC, Teunissen JM, Van Osch W, Van Den Brom WE, Dieleman SJ, Kooistra HS. Influence of litter size and breed on the duration of gestation in dogs. J Reprod Fertil Suppl 2001;57:193-7.
118. Dewar AD. Litter size and the duration of pregnancy in mice. Q J Exp Physiol Cogn Med Sci 1968;53:155-61.
119. Myklestad K, Vatten LJ, Magnussen EB, Salvesen KA, Romundstad PR. Do parental heights influence pregnancy length?: A population-based prospective study, HUNT 2. BMC Pregnancy Childbirth 2013;13:33.
120. Dunsworth HM, Warrener AG, Deacon T, Ellison PT, Pontzer H. Metabolic hypothesis for human altriciality. Proceedings of the National Academy of Sciences of the United States of America 2012;109:15212-6.
121. Marsal K, Persson PH, Larsen T, Lilja H, Selbing A, Sultan B. Intrauterine growth curves based on ultrasonically estimated foetal weights. Acta Paediatr 1996;85:843-8.
122. Juodakis J, Bacelis J, Zhang G, Muglia LJ, Jacobsson B. Time-Variant Genetic Effects as a Cause for Preterm Birth: Insights from a Population of Maternal Cousins in Sweden. G3 (Bethesda) 2017;7:1349-56.
123. Snowden JM, Basso O. Causal inference in studies of preterm babies: a simulation study. BJOG : an international journal of obstetrics and gynaecology 2018;125:686-92.
124. Suarez EA, Landi SN, Conover MM, Jonsson Funk M. Bias from restricting to live births when estimating effects of prescription drug use on pregnancy complications: A simulation. Pharmacoepidemiol Drug Saf 2018;27:307-14.
125. Kuhn TS. The Structure of Scientific Revolutions. University of Chicago Press; 1962.
126. 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;488:471-5.
127. Rosenwald MS. This student put 50 million stolen research articles online. And they're free. The Washington Post 2016;Sect. 2016 March 30.
128. Gustin S. Aaron Swartz, Tech Prodigy and Internet Activist, Is Dead at 26. Time 2013.
129. Laakso M, Bjork BC. Anatomy of open access publishing: a study of longitudinal development and internal structure. BMC medicine 2012;10:124.