



Biological insights into multiple birth: genetic findings from UK Biobank

Hamdi Mbarek¹ · Margot P. van de Weijer¹ · Mathijs D. van der Zee¹ · Hill F. Ip¹ · Jeffrey J. Beck² · Abdel Abdellaoui^{1,3} · Erik A. Ehli² · Gareth E. Davies² · Bart M. L. Baselmans¹ · Michel G. Nivard¹ · Meike Bartels¹ · Eco J. de Geus¹ · Dorret I. Boomsma¹

Received: 21 June 2018 / Revised: 15 November 2018 / Accepted: 5 January 2019 / Published online: 13 February 2019
© European Society of Human Genetics 2019

Abstract

The tendency to conceive spontaneous dizygotic (DZ) twins is a complex trait with important contributions from both environmental factors and genetic disposition. In earlier work, we identified the first two genes as maternal susceptibility loci for DZ twinning. The aim of this study was to identify genetic variants influencing multiple births and to genetically correlate the findings across a broad range of traits. We performed a genome-wide association study (GWAS) in 8962 participants with Caucasian ancestry from UK Biobank who reported being part of a multiple birth, and 409,591 singleton controls. We replicated the association between *FSHB*, *SMAD3* and twinning in the gene-based (but not SNP-based) test, which had been established in previous genome-wide association analyses in mothers with dizygotic twin offspring. Additionally, we report a novel genetic variant associated with multiple birth, *rs428022* at 15q23 ($p = 2.84 \times 10^{-8}$) close to two genes: *PIAS1* and *SKORI*. Finally, we identified meaningful genetic correlations between being part of a multiple birth and other phenotypes (anthropometric traits, health-related traits, and fertility-related measures). The outcomes of this study provide important new insights into the genetic aetiology of multiple births and fertility, and open up novel directions for fertility and reproduction research.

Introduction

Uncovering the mechanisms underlying ovarian function and follicle growth is essential for our understanding of human reproduction and female (in)fertility. Central to our

understanding of these traits is spontaneous multiple birth: the conception and development of two or more independent zygotes in one pregnancy, which might indicate increased fertility. In addition, multiple birth is associated with increased risks for both the mother and her offspring, such as increased risk of preterm birth [1] and increased maternal morbidity [2]. By improving our knowledge on the genetic basis and physiological mechanisms underlying this trait, we can make important advances in the outcomes for mother and offspring and reveal novel possibilities for fertility treatment.

Among multiple births, twinning is the most common outcome, with the prevalence of twins varying over time and geographic location. In Western Europe, the twinning rate is about 15–18 per 1000 maternities, and this number can increase to up to 40 per 1000 maternities in Africa [3]. The incidence of monozygotic (MZ) births is relatively stable over the world (3–4 per 1000 births) [4], but variations in, e.g., isolated populations worldwide are seen. The incidence of twin births has substantially increased since the 1970s for two reasons. First, the advent of assisted reproductive technologies (ARTs), such as in-vitro fertilisation (IVF), where multiple embryos are often transferred to increase the

These authors contributed equally: Hamdi Mbarek, Margot P. van de Weijer

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-019-0355-z>) contains supplementary material, which is available to authorised users.

✉ Hamdi Mbarek
h.mbarek@vu.nl

¹ Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, The Netherlands

² Avera Institute for Human Genetics, Avera McKennan Hospital and University Health Center, Sioux Falls, SD, USA

³ Department of Psychiatry, Academic Medical Center, Amsterdam, The Netherlands

likelihood of pregnancy, have resulted in higher rates of multiple gestation [5]. IVF also is associated with increased MZ twinning, though the mechanisms are poorly understood. Second, the age at which women get children has increased, a factor that is associated with higher rates of twinning [6].

While evidence indicating that MZ twinning is influenced by genetic factors is complex, it is long established that dizygotic (DZ) twinning is a heritable trait. Earlier studies on the inheritance of DZ twinning examined the twinning rate in families and found that relatives of mothers of twins report higher twinning rates compared with the general population [7, 8]. Although having DZ twins was at some point considered to be most consistent with an monogenic model [9], it is now established that DZ twinning is a complex polygenic trait. The search for twinning genes began with the first genome-wide association (GWA) study [10, 11]. GWA analyses were performed in 1980 mothers of spontaneous DZ twins and 12,953 controls, which led to the identification of two genetic variants increasing the chance of spontaneous DZ twinning. One of these variants (rs11031006) is near *FSHB*, a gene involved in the release of follicle-stimulating hormone (FSH) that controls ovarian folliculogenesis and ovulation. The other variant (rs17293443) is located in *SMAD3*, which regulates the response of the ovaries to FSH.

While the identification of the first genetic variants influencing spontaneous DZ twinning are an important development in the field of human reproduction, there is still much unknown fact about the mechanisms and genetic pathways underlying this trait. Most importantly, in previous GWA analyses we see a number of other loci that reach suggestive levels of significance that are near likely multiple birth candidate genes (e.g. near/in *INHB* and *SMAD4*) [10]. Yet, to establish whether these suggestive loci are in association with multiple births, larger sample sizes are needed. UK Biobank [12] (UKB), a cohort-based health resource, contains multiple birth and genotype data for an extensive number of participants. Employing these data, genetic variants associated with being part of a multiple birth can be used as a proxy of the genetic variants for giving birth to multiple birth offspring. The aim of this study was to perform a GWA study (GWAS) in “spontaneous” (see Materials and methods) DZ twins to identify genetic variants influencing multiple births and to genetically correlate our findings across a broad range of fertility-related traits.

Materials and methods

Discovery cohort—UKB

For this study, we analysed data from UKB release 2. The UKB cohort contains data for 488,377 participants from

across the United Kingdom aged between 40 and 69 years, collected between 2006 and 2010. The database was established to power investigations of the genetic and non-genetic determinants of human disease. Participants filled out questionnaires on many socioenvironmental, health, and lifestyle variables, and, additionally, provided blood, saliva, and urine samples. Detailed information on data collection and protocol are publicly available on the UKB website (<http://www.ukbiobank.ac.uk/>). Data access permission was granted under UKB application 25472 (PI Bartels).

Genetic data in UKB

An extensive description of the genetic data in UKB is described in Bycroft et al. [13]. In brief, participants were genotyped on two similar genotyping arrays (95% overlap of markers): the Applied Biosystems UK BiLEVE Array by Affymetrix (807,411 markers) and the Applied Biosystems UK Biobank Axiom Array (825,927 markers). The genotype data were subjected to a standardised quality control (QC) pipeline that was designed to address challenges specific to this dataset described in the paper by Winkler et al. [14]. For marker QC, the thresholds were set such that only strongly deviating markers fail the tests, allowing researchers to apply their own QC procedures on remaining markers. In sample QC, only duplicates and laboratory mishandled samples were removed. Other dubious samples were kept in the dataset, but information on these samples was made available to researchers. Single nucleotide polymorphisms (SNPs) were imputed from both the 1000 Genomes Phase 3 and the haplotype reference consortium (HRC) reference panel, but when SNPs were present in both panels, HRC imputation was the preferred option.

Subject selection

In order to avoid bias due to population stratification, we limited our GWA analysis to Caucasian participants (total $N = 8\,962$). We divided the sample into three separate groups where, within each group, none of the participants had a closer genetic relationship than fourth cousin with one another (see Bycroft et al. [13], for more details on the kinship coefficients). We first selected all participants that self-report to be “White British” and have similar genetic ancestry based on principal component analysis (PCA) (see Bycroft et al. [13], for more details on the PCA) and took the maximal set of unrelated participants ($N = 7036$ cases, $N = 325,773$ controls). Next, from the remaining set of Caucasian UK participants, we again selected a second maximal group of unrelated subjects ($N = 1364$ cases, $N = 56,507$ controls). Finally, we selected all Caucasian participants that did not self-report to be “White British” and did

not show genetic similarity to UK participants ($N = 562$ cases, $N = 27,311$ controls).

Multiple birth in UKB

To assess whether participants were part of a multiple birth, they were presented with the question: “Are you a twin, triplet or other multiple birth?” (UKB questionnaire field ID 1777). The options were: “Yes”, “No”, “Do not know”, and “Prefer not to answer”. In our analyses, we focused on 8962 participants with Caucasian ancestry that reported being part of a multiple birth, and 409,591 controls.

Identity-by-state (IBS) information (UKB questionnaire field ID 22013) is available for a subgroup of UKB participants that identify as being genetically related (UKB questionnaire field ID 22011), and can be used to assess the genetic relationship. A kinship coefficient, reflecting the possibility that two alleles sampled at random from two individuals are identical-by-descent, associated with each pair of participants is also available (UKB questionnaire field ID 22012). We identified and removed MZ twins from the same pair by plotting their kinship coefficient against the proportion of IBS-0 (proportion of no IBS sharing). If only one individual of a pair participated in UKB, it was not possible to identify him/her as an MZ or DZ twin based on kinship- and IBS information. Moreover, to remove twins potentially conceived as a result of clomifene, IVF or other ART we removed twins born after 1967, the year clomifene was introduced in the UK. In total, we excluded 358 cases based on zygosity and 174 cases based on year of birth.

Statistical analyses

Genome-wide association analyses

We performed GWA analyses in PLINK [15] using logistic regression under an additive genetic model with adjustment for age, sex, genotyping chip, and 40 principal components reflecting genetic ancestry, supplied by UKB. The results from these analyses were followed-up by post-GWA QC procedures, where we excluded structural variants, indels, monomorphic SNPs, SNPs with minor allele frequency (MAF) < 0.005 , and SNPs with missing or invalid data. Next, we aligned all SNPs to a reference file (<http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/>) and removed SNPs with allele mismatches and SNPs where the absolute difference between the reported effect allele frequency (EAF) and reference EAF was larger than 0.2. Finally, to meta-analyse the summary statistics from the three groups, we performed an N-weighted GWA meta-analysis (GWAMA), correcting for relatedness between the two UK samples based on the linkage disequilibrium (LD)

score cross trait intercept [16] ($N = 8962$ cases, $N = 409,591$ controls, $N_{\text{SNPs}} = 8,532,721$).

For functional annotation of our GWAS results, we used FUMA [17] (FUncTional Mapping and Annotation) to define genomic risk loci. SNPs that reached a significance threshold of (5×10^{-8}) were considered genome-wide significant. If two or more SNPs were genome-wide significant and independent from each other at $r^2 < 0.6$ or $r^2 < 0.1$ they were considered independent significant SNPs, or lead SNPs, respectively. Independent significant SNPs, which were closer than 250 kb were merged together in one genomic risk locus. SNPs in LD with these independent significant SNPs were considered candidate SNPs and these determined the borders of the genomic risk loci. We used LocusZoom to plot the genomic risk loci [18].

Gene-based test

In addition to single SNP analyses, we performed a gene-based GWA analysis (GWGAS), which combines SNP p -values within a gene into a gene test-statistic to increase power when the effects of individual markers are too weak to detect. We used the MAGMA (Multi-marker Analysis of GenoMic Annotation) function implemented in FUMA to perform a gene-based test [19]. We used the SNP-based p -values as input and annotated them to 18,187 known protein-coding genes. The Bonferroni-corrected significance threshold was defined at $\alpha = 0.05/18,187 = 2.75 \times 10^{-6}$.

Gene mapping

To map the associated variants to genes, we made use of the three mapping strategies in FUMA: (1) positional mapping, where we mapped SNPs to genes that are a maximum of 10 kb distance from the genomic locus, (2) expression Quantitative Trait (eQTL) mapping, where we mapped SNPs to genes whose RNA expression level they influence, and (3) chromatin interaction mapping, where we mapped SNPs to genes when there is a three-dimensional (3D) DNA–DNA interaction between a SNP region and another gene region. These interactions are the result of the packaging of genomes in the 3D nucleus, so that genomic regions interact in the same, or even distinct, chromosomes. If the SNP region interacts with a region that contains multiple genes, the SNP was mapped to all genes.

Genetic correlations

To quantify the shared genetic contribution between multiple birth and several other traits, we performed explorative genetic correlation analyses in LD Hub [20]. We included publicly available data from 687 traits, based on multiple

published GWASs and the available traits in UKB. LD Hub calculates genetic correlations between user-defined summary statistics of a trait of interest and predefined categories of other traits using LD score regression [21]. This method distinguishes between bias and inflation from a true polygenic signal by examining the relationship between linkage disequilibrium and test statistics. For the health-related and anthropometric traits ($N = 70$), the two categories most likely related to fertility and reproduction, we calculated the false discovery rate (FDR)-adjusted p -values as a means of assessing significance (threshold = 0.05). The genetic correlations were visualised using the ggplot2 [22] package in R [23].

Results

Genome-wide association analyses

We carried out a GWAS for being part of a multiple birth in the UKB discovery samples including a total of 8962 cases and 409,591 controls (see Fig. 1 and Supplementary Figure 1). We identified one region on chromosome 15 containing a genome-wide significant SNP, rs428022 (hg19 chr15:g.68249135A>G, $p < 5 \times 10^{-8}$). The region contains 33 candidate SNPs, with one independent significant lead SNP (see Table 1 and Supplementary Table 1). The strongest signal rs428022 ($p = 2.84 \times 10^{-8}$, odds ratio (OR) = 1.04) is an intergenic SNP, flanked by *PIAS1* [OMIM:

603566] and *SKOR1* [OMIM: 611273] (see Fig. 2). The gene-based test identified another genome-wide significant gene, *FSHB/ARL14EP* ($p = 1.17 \times 10^{-7}$) [OMIM: 612295] (see Supplementary Figure 2 and 3).

In a GWAS of spontaneous DZ twinning, we previously identified *FSHB* (1.54×10^{-9}) and *SMAD3* (1.57×10^{-8}) as maternal susceptibility loci for DZ twinning [10]. These two loci were replicated in this study with a significance threshold after Bonferroni correction ($p < 0.05/2$) and in the same direction (see Supplementary Table 2).

Gene mapping

The positional, eQTL, and chromatin interaction gene-mapping results of the new signal in 15q23 (rs428022) can be found in supplementary table 3, 4 and 5, respectively. Using positional mapping, we identified nine genes that are a maximum 10 kb up- or downstream of the genomic locus. Four genes were found through eQTL mapping. Finally, we found the same nine genes in chromatin interaction mapping as in positional mapping. Two genes were significant across all three mapping methods: *PIAS1* and *CALML4*.

Genetic correlation analyses

We calculated genetic correlations between being part of a multiple birth and all available traits in LD Hub. Supplementary Table 6 shows the complete results for all traits. Table 2 shows the 70 associations with a FDR threshold

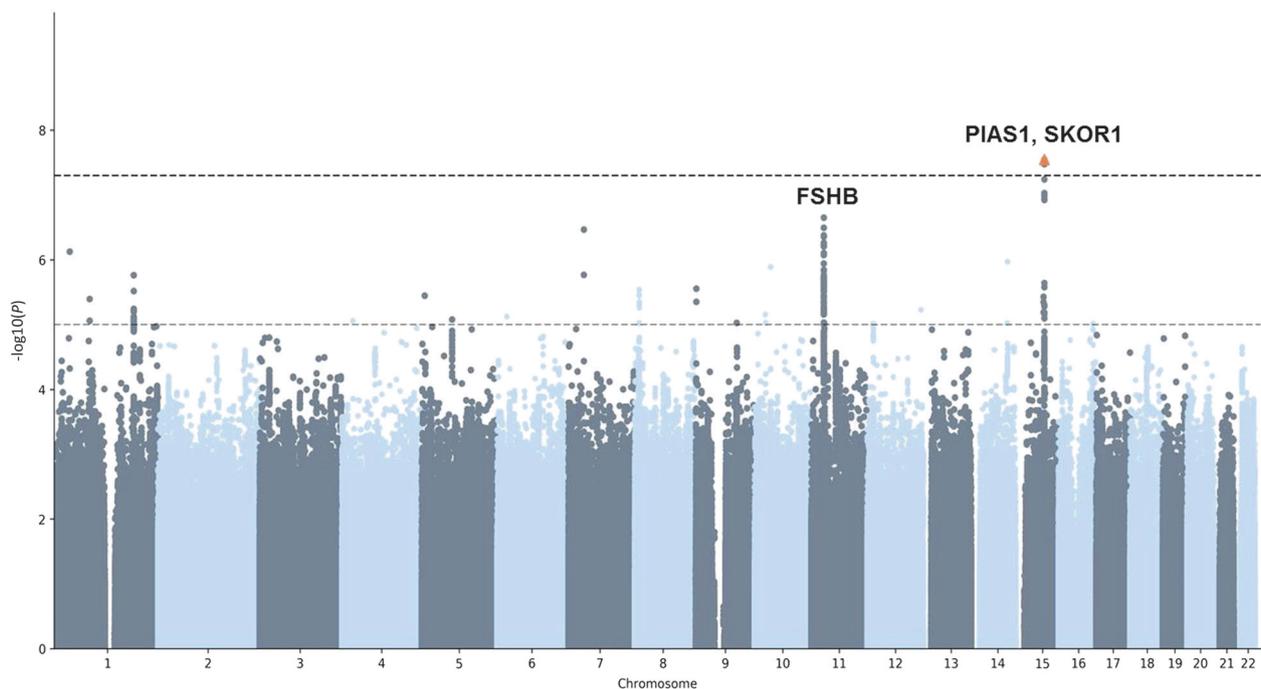


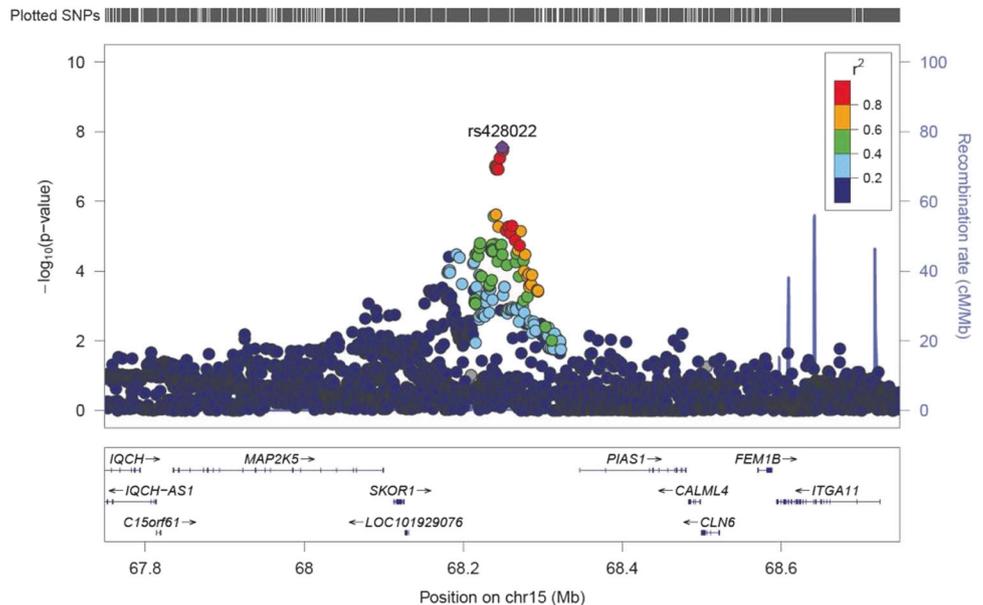
Fig. 1 Manhattan plot genome-wide association study (GWAS) multiple birth on 8962 cases and 409,591 controls

Table 1 Genome-wide significant SNPs in the GWAS multiple birth ($n = 8962$ cases versus 409,591 controls)

SNP	Locus	Position ^a	Gene	Annotation	Risk allele	RAF	OR (95% CI)	<i>P</i>
rs428022	15q23	68249135	PIAS1-SKOR1	Intergenic	A	0.34	1.04 (1.03–1.06)	2.84×10^{-8}
rs434545	15q23	68249752	PIAS1-SKOR1	Intergenic	A	0.34	1.04 (1.03–1.06)	3.35×10^{-8}

RAF risk allele frequency, OR odds ratio, 95% CI 95% confidence interval, *P* *p*-value, GWAS genome-wide association study

^aSNP position according to NCBI Human Genome Build 37

Fig. 2 Regional association plot for the top SNP rs428022

equal to or lower than 0.05. Of these associations, 32 can be classified as anthropometric traits (Fig. 3). We found positive genetic associations with measures of body mass, both with whole-body mass measures such as body mass index (BMI) ($r_g = 0.20$, $FDR = 0.017$), and specific body parts such as leg fat mass (left: $r_g = 0.20$, $FDR = 0.017$; right: $r_g = 0.20$, $FDR = 0.017$). We identified negative associations between multiple birth and anthropometric traits for impedance measures, again both whole-body (impedance of whole-body $r_g = -0.22$, $FDR = 0.017$) and body part specific (e.g., impedance of left arm $r_g = -0.22$, $FDR = 0.017$). We also found genetic associations with 38 other health-related traits (Fig. 4). These include a variety of traits, among which cardiovascular measures (e.g., acute myocardial infarction; $r_g = 0.24$, $p = 0.007$), fertility-related measures (e.g., age at menarche; $r_g = -0.22$, $FDR = 0.017$), and glucose-related traits (e.g., diabetes $r_g = 0.27$, $FDR = 0.017$).

Discussion

In this study, we replicated the association between *FSHB*, *SMAD3*, and twinning, which has been established in

previous GWA analyses in mothers with DZ twin offspring. In addition, we report a novel genetic variant associated with multiple births, *rs428022* at 15q23.

It is important to note that the significant association observed for the *FSHB* gene reconfirms the important role it plays in both male and female fertility. Recently, Rull and colleagues [24] proposed that *FSHB* -211 G>T variant (association *p*-value in this study = 2.02×10^{-5}) represents a key genetic modulator of circulating gonadotropin, leading to various possible downstream effects on reproductive physiology. The novel genome-wide hit on chromosome 15 identified in this study, *rs428022*, is an intergenic SNP flanked by *PIAS1* and *SKOR1*, and was additionally mapped to *CALML4* in all three mapping strategies. *PIAS1* (protein inhibitor of activated STAT 1) acts as a regulator of the androgen receptor (AR), dysregulation of which might lead to prostate cancer [25, 26]. In line with this role, it has been shown that *PIAS1* is upregulated upon androgenic stimulation [26] and in prostate cancer tumours [27]. The AR plays an important role in male fertility as testosterone exerts its action through this receptor, and variants in the AR gene may cause male infertility [28, 29]. In addition, it has been shown that protein inhibitors of activated signal transducers and activators of transcription (PIAS) proteins

Table 2 Genetic correlations FDR threshold < 0.05

Trait 1	Trait 2	PubMed ID	Ethnicity	r_g	se	z	p	FDR
Multiple birth	Arm fat mass (right)	0	European	0.194	0.0747	2.5964	0.0094	0.017
Multiple birth	Arm fat-free mass (left)	0	European	0.2293	0.0829	2.7668	0.0057	0.017
Multiple birth	Arm fat-free mass (right)	0	European	0.2352	0.0836	2.8142	0.0049	0.017
Multiple birth	Arm predicted mass (left)	0	European	0.234	0.0825	2.8365	0.0046	0.017
Multiple birth	Arm predicted mass (right)	0	European	0.2234	0.0814	2.7452	0.006	0.017
Multiple birth	Body mass index (BMI)	0	European	0.2009	0.0731	2.7465	0.006	0.017
Multiple birth	Hip circumference	0	European	0.2272	0.0819	2.7733	0.0055	0.017
Multiple birth	Impedance of arm (left)	0	European	-0.218	0.0721	-3.025	0.0025	0.017
Multiple birth	Impedance of arm (right)	0	European	-0.1887	0.0706	-2.6732	0.0075	0.017
Multiple birth	Impedance of leg (left)	0	European	-0.2065	0.0777	-2.6586	0.0078	0.017
Multiple birth	Impedance of leg (right)	0	European	-0.2154	0.0778	-2.7699	0.0056	0.017
Multiple birth	Impedance of whole body	0	European	-0.2189	0.0745	-2.9395	0.0033	0.017
Multiple birth	Leg fat mass (left)	0	European	0.1985	0.0734	2.703	0.0069	0.017
Multiple birth	Leg fat mass (right)	0	European	0.196	0.0738	2.6572	0.0079	0.017
Multiple birth	Leg fat-free mass (left)	0	European	0.2497	0.0863	2.8917	0.0038	0.017
Multiple birth	Leg fat-free mass (right)	0	European	0.2528	0.0869	2.9101	0.0036	0.017
Multiple birth	Leg predicted mass (left)	0	European	0.2506	0.0864	2.9003	0.0037	0.017
Multiple birth	Leg predicted mass (right)	0	European	0.2529	0.087	2.9065	0.0037	0.017
Multiple birth	Trunk fat-free mass	0	European	0.2308	0.0882	2.6153	0.0089	0.017
Multiple birth	Trunk predicted mass	0	European	0.2333	0.089	2.6216	0.0088	0.017
Multiple birth	Weight	0	European	0.2312	0.0818	2.8271	0.0047	0.017
Multiple birth	Whole-body fat-free mass	0	European	0.2411	0.0869	2.7752	0.0055	0.017
Multiple birth	Whole-body water mass	0	European	0.237	0.0858	2.7638	0.0057	0.017
Multiple birth	Parents age at death	27015805	European	-0.6775	0.2261	-2.9965	0.0027	0.017
Multiple birth	Diabetes diagnosed by doctor	0	European	0.2669	0.0999	2.6727	0.0075	0.017
Multiple birth	Diagnoses—main ICD10: I21 acute myocardial infarction	0	European	0.6467	0.2384	2.7125	0.0067	0.017
Multiple birth	Diagnoses—main ICD10: K44 diaphragmatic hernia	0	European	0.7996	0.2947	2.7132	0.0067	0.017
Multiple birth	Fathers age at death	0	European	-0.4794	0.1582	-3.0315	0.0024	0.017
Multiple birth	Illnesses of father: diabetes	0	European	0.5267	0.1925	2.7366	0.0062	0.017
Multiple birth	Illnesses of siblings: diabetes	0	European	0.4341	0.1651	2.6294	0.0086	0.017
Multiple birth	Maximum heart rate during fitness test	0	European	-0.6201	0.2362	-2.6253	0.0087	0.017
Multiple birth	Overall health rating	0	European	0.2332	0.0882	2.6437	0.0082	0.017
Multiple birth	Angina (diagnosed by doctor)	0	European	0.39	0.1493	2.6123	0.009	0.017
Multiple birth	Wheeze or whistling in the chest in last year	0	European	0.3069	0.1006	3.0516	0.0023	0.017
Multiple birth	Bilateral oophorectomy (both ovaries removed)	0	European	0.476	0.1841	2.5858	0.0097	0.017
Multiple birth	Emphysema/chronic bronchitis (diagnosed by doctor)	0	European	0.6061	0.2106	2.8778	0.004	0.017
Multiple birth	Age at menarche	25231870	European	-0.2195	0.0816	-2.6886	0.0072	0.017
Multiple birth	Basal metabolic rate	0	European	0.2432	0.086	2.8278	0.0047	0.017
Multiple birth	Vitamin and mineral supplements: vitamin A	0	European	1.08	0.3814	2.8315	0.0046	0.017
Multiple birth	Vitamin and mineral supplements: vitamin E	0	European	0.6599	0.2507	2.6324	0.0085	0.017
Multiple birth	Arm fat mass (left)	0	European	0.1866	0.0742	2.5137	0.0119	0.02
Multiple birth	Number of operations_ self-reported	0	European	0.3209	0.1276	2.5146	0.0119	0.02
Multiple birth	Obesity class 1 (BMI \geq 30 kg/m ²)	23563607	European	0.287	0.1162	2.4701	0.0135	0.021
Multiple birth	Whole-body fat mass	0	European	0.1789	0.0725	2.4669	0.0136	0.021
Multiple birth	Type 2 diabetes self-reported	0	European	0.5198	0.2105	2.4696	0.0135	0.021
Multiple birth	Comparative body size at age 10	0	European	0.1973	0.0809	2.4369	0.0148	0.023
Multiple birth	Ever had hysterectomy (womb removed)	0	European	0.5011	0.208	2.4089	0.016	0.024
Multiple birth	Heart attack diagnosed by doctor	0	European	0.4727	0.1965	2.4054	0.0162	0.024
Multiple birth	Sitting height	0	European	0.183	0.0773	2.3658	0.018	0.025
Multiple birth	Breastfed as a baby	0	European	-0.3459	0.1454	-2.3791	0.0174	0.025
Multiple birth	Diagnoses—main ICD10: O75 other complications of labour and delivery: not elsewhere classified	0	European	0.8404	0.3618	2.3225	0.0202	0.027
Multiple birth	Vitamin and mineral supplements: vitamin B	0	European	0.5249	0.226	2.3226	0.0202	0.027
Multiple birth	Obesity class 2 (BMI \geq 35 kg/m ²)	23563607	European	0.2937	0.131	2.2413	0.025	0.032
Multiple birth	Trunk fat mass	0	European	0.1631	0.0723	2.2576	0.024	0.032
Multiple birth	Difficulty not smoking for 1 day	0	European	0.4818	0.2165	2.2248	0.0261	0.033

Table 2 (continued)

Trait 1	Trait 2	PubMed ID	Ethnicity	r_g	se	z	p	FDR
Multiple birth	Ever had bowel cancer screening	0	European	-0.4202	0.1939	-2.1676	0.0302	0.038
Multiple birth	Diagnoses—main ICD10: I25 chronic ischaemic heart disease	0	European	0.3898	0.1818	2.1437	0.0321	0.039
Multiple birth	Rheumatoid arthritis	24390342	European	0.2495	0.1176	2.1211	0.0339	0.041
Multiple birth	Waist circumference	0	European	0.1553	0.0737	2.1068	0.0351	0.042
Multiple birth	Cancer code_ self-reported: prostate cancer	0	European	0.438	0.2103	2.0827	0.0373	0.043
Multiple birth	Illnesses of father: lung cancer	0	European	0.3582	0.1721	2.0813	0.0374	0.043
Multiple birth	Squamous cell carcinoma self-reported	0	European	0.9864	0.4793	2.0579	0.0396	0.044
Multiple birth	Hiatus hernia self-reported	0	European	0.4857	0.2349	2.0675	0.0387	0.044
Multiple birth	Cigarettes smoked per day	20418890	European	0.5047	0.2484	2.032	0.0422	0.045
Multiple birth	Illnesses of mother: high blood pressure	0	European	0.2479	0.1219	2.0339	0.042	0.045
Multiple birth	Age completed full time education	0	European	-0.1745	0.0867	-2.0135	0.0441	0.047
Multiple birth	Overweight (BMI \geq 25 kg/m ²)	23563607	European	0.2111	0.1067	1.9791	0.0478	0.049
Multiple birth	Lung cancer	27488534	European	0.3226	0.1629	1.9802	0.0477	0.049
Multiple birth	Diagnoses—main ICD10: S66 injury of muscle and tendon at wrist and hand level	0	European	0.8488	0.4306	1.9712	0.0487	0.049
Multiple birth	Pulse rate	0	European	-0.216	0.1101	-1.962	0.0498	0.05

r_g genetic correlation, se standard error, z z-score, p p-value, FDR false discovery rate

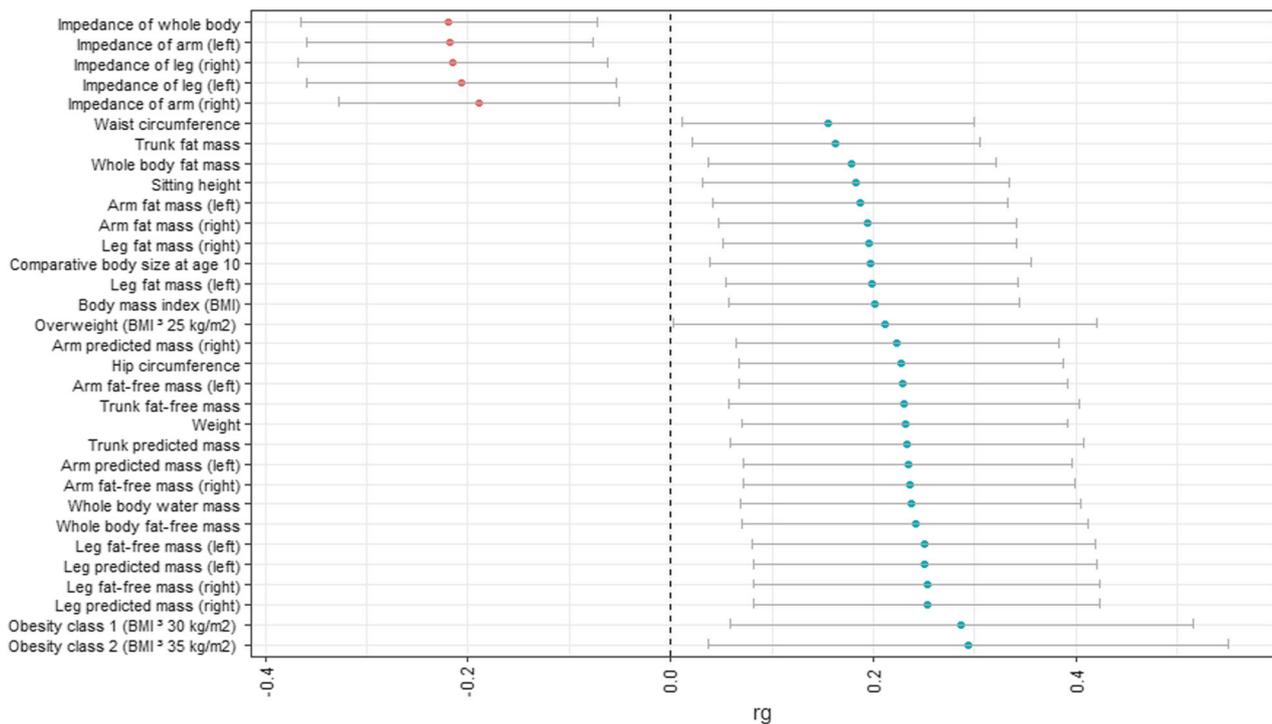


Fig. 3 Genetic correlations (r_g + 95% confidence interval (CI)) with anthropometric traits

interact with the transforming growth factor (TGF)-beta pathway and regulate *SMAD*-mediated transcriptional activity [30, 31].

The other close gene to rs428022, *SKOR1*, also known as *Fussel-15* (functional *SMAD* suppressing element on chromosome 15) interacts with *Smad1*, *Smad2*, and *Smad3* molecules and has been identified as molecular regulator of

bone morphogenetic protein (BMP) signalling [32]. The BMP family of proteins regulates many aspects of reproductive system development and biology. Animal studies showed that variants in two *BMP* genes (*GDF9* and *BMP15*) in sheep were associated with increased ovulation rate [33], and this was recapitulated in the marmoset [34], which has a high rate of twinning. The reduced activity of

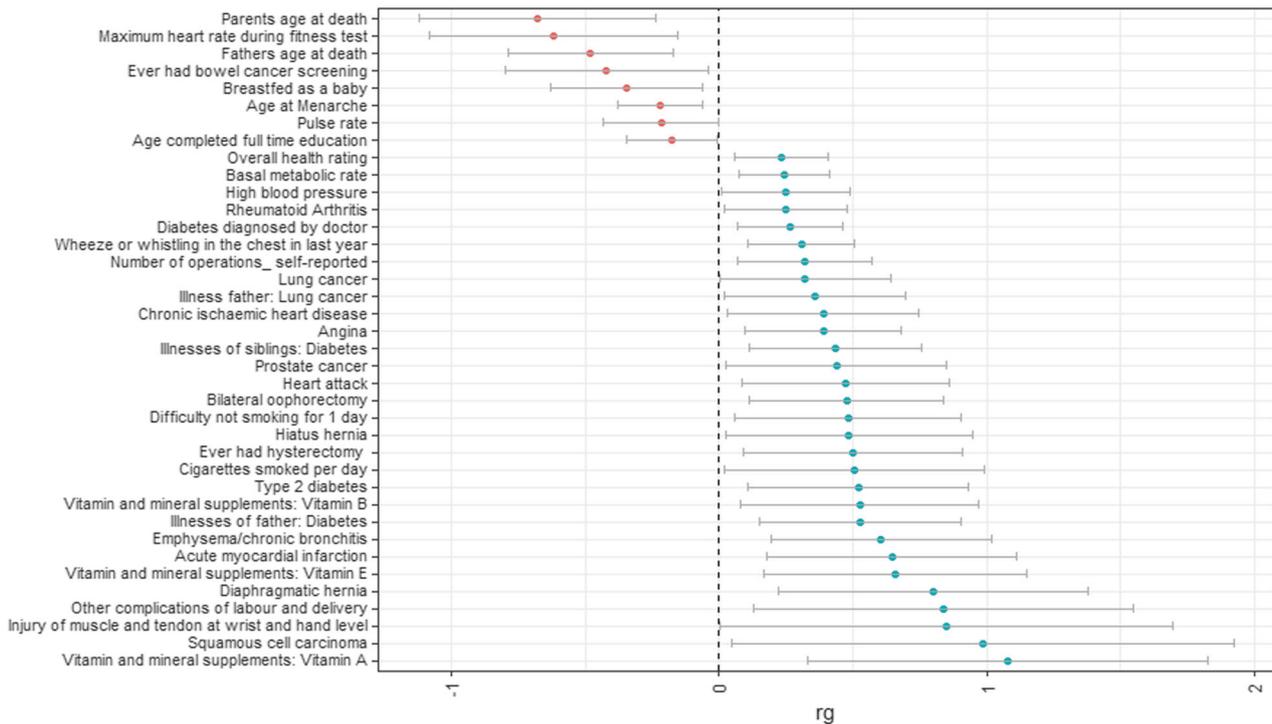


Fig. 4 Genetic correlations (r_g + 95% confidence interval (CI)) with other health-related traits

the BMP signaling system in the ovary leads to decreases in granulosa cell mitosis and its inhibiting action on FSH sensitivity. This in turn leads to selection of more follicles, increased ovulation rate, and multiple births.

One other gene was implicated in all three mapping strategies: *CALML4* (calmodulin-like 4). This gene is a protein-coding gene coding for calmodulin-like protein 4. However, very little information is present in the literature concerning the role of this gene. Although the role of our identified SNP in relation to *PIAS1* and *SKOR1* is in need of further investigation, a better understanding of the role of this gene and its interaction with *FSHB* and *SMAD3* in twinning and fertility could lead to new insights in basic and clinical reproductive physiology research.

We also identified possible genetic correlations between being part of a multiple birth and other phenotypes. The genetic correlations between multiple birth and several anthropometric traits such as BMI, weight, and hip circumference are in line with the previous epidemiological studies that linked these traits to a higher relative risk of having twins [3]. We replicated findings from previous investigations into the genetics of twinning with several fertility-related traits, such as a negative genetic association with age at menarche [10]. The negative association with age at death relates to theories on lifespan and fertility stating that higher longevity is at a cost of lesser reproductive success and vice versa [35], for which we now also found genetic evidence. The genetic associations we found with cardiovascular traits are interesting in light of a recent

paper by Byars and colleagues, where they examined whether coronary artery disease (CAD)-linked selection signals are linked to traits influencing reproduction [36, 37]. They found that CAD loci are enriched for effects on female lifetime reproductive success. The positive association we found also suggests these antagonistic pleiotropic effects as a higher genetic risk for twinning is associated with a higher genetic risk for myocardial infarction. Yet, it should be kept in mind that these analyses were explorative and that these correlations do not survive a Bonferroni-corrected threshold for multiple testing.

While this study provides important new insights into the genetic aetiology of multiple births, further work is required to establish the functional/biological pathway through which the identified genetic variant influences multiple births. At the moment, *PIAS1* and *SKOR1* seem likely candidate genes through which the SNP influences multiple births. The findings may be somewhat limited because of the constraints of the UKB database. Although we did what we could to exclude MZ twins (based upon identical by descent (IBD) = 2 for complete pairs) we could not exclude all single MZ twins. In addition, Yengo et al. [38] recently pointed out the limitation of potential intercept inflation of bivariate LD score regression in large samples. Therefore, due to large sample size, sample overlap might have been overestimated in the current study.

To summarise, in this study we replicated previous findings for multiple birth, and identified new potential genes influencing multiple birth (*PIAS1/SKOR1*). In

addition, we examined the genetic overlap between being part of a multiple birth and several other traits and identify many possible genetic associations with diverse health and anthropometric traits. While this study provides new insights into the genetic aetiology of multiple birth, further work is required to establish the functional pathway through which the identified genetic variant influences multiple birth.

Acknowledgements This research has been conducted using the UK Biobank Resource under application number 25472 (PI Bartels). HM and MDvdZ were supported through the VU-Avera Collaborative Agreement between the Vrije Universiteit Amsterdam and the Avera McKennan d/b/a Avera Institute for Human Genetics. MPvdW was supported by a University Research Fellow (URF) of the Vrije Universiteit, Amsterdam to DIB. MGN was supported by a Royal Netherlands Academy of Science Professor Award to DIB (PAH/6635), ZonMw grant: “Genetics as a research tool: A natural experiment to elucidate the causal effects of social mobility on health.” (pnr: 531003014) and ZonMw project: “Can sex- and gender-specific gene expression and epigenetics explain sex-differences in disease prevalence and etiology?” (pnr: 849200011). HFI was supported by the “Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies” (ACTION) project. ACTION receives funding from the European Union Seventh Framework Program (FP7/2007–2013) under grant agreement no 602768. Computational facilities on Cartesius were supplied by NWO via the grant: “Population scale genetic analysis” 2018/EW/00408559.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Blondel B, Kogan MD, Alexander GR, Dattani N, Kramer MS, Macfarlane A, et al. The impact of the increasing number of multiple births on the rates of preterm birth and low birthweight: an international study. *Am J Public Health*. 2002;92:1323–30.
- The ESHRE Capri Workshop Group. Multiple gestation pregnancy. *Hum Reprod*. 2000;15:1856–64.
- Hoekstra C, Zhao ZZ, Lambalk CB, et al. Dizygotic twinning. *Hum Reprod Update*. 2008;14:37–47.
- Smits J, Monden C. Twinning across the developing world. *PLoS ONE*. 2011;6:e25239.
- Sunderam S, Kissin DM, Crawford SB, et al. Assisted reproductive technology surveillance — United States, 2015. *MMWR Surveill Summ*. 2018;67:1–28.
- Beemsterboer SN, Homburg R, Gorter NA, Schats R, PGA Hompes, Lambalk CB. The paradox of declining fertility but

- increasing twinning rates with advancing maternal age. *Hum Reprod*. 2006;21:1531–32.
- Bulmer MG. *The biology of twinning in man*. Oxford: Clarendon Press; 1970.
- Parisi P, Gatti M, Prinzi G, Caperna G. Familial incidence of twinning. *Nature*. 1983;304:626–8.
- Meulemans WJ, Lewis CM, Boomsma DI, et al. Genetic modelling of dizygotic twinning in pedigrees of spontaneous dizygotic twins. *Am J Med Genet*. 1996;61:258–63.
- Mbarek H, Steinberg S, Nyholt DR, et al. Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility. *Am J Hum Genet*. 2016;98:898–908.
- Mbarek H, Dolan CV, Boomsma DI. Two SNPs associated with spontaneous dizygotic twinning: effect sizes and how we communicate them. *Twin Res Hum Genet*. 2016;19:418–21.
- 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.
- Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–9.
- Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9:1192–212.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
- Baselmans BML, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP, et al. Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet*. 2019. <https://doi.org/10.1038/s41588-018-0320-8>
- Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826.
- Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–37.
- de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11:e1004219.
- Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33:272–9.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5.
- Wickham H. *Ggplot2: elegant graphics for data analysis*. Springer, 2009.
- R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- Rull K, Grigorova M, Ehrenberg A, et al. FSHB -211 G>T is a major genetic modulator of reproductive physiology and health in childbearing age women. *Hum Reprod*. 2018;33:954–66.
- Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev*. 2004;25:276–308.
- Puhr M, Hoefler J, Eigentler A, et al. PIAS1 is a determinant of poor survival and acts as a positive feedback regulator of AR signaling through enhanced AR stabilization in prostate cancer. *Oncogene*. 2016;35:2322–32.
- Hoefler J, Schäfer G, Klocker H, et al. PIAS1 is increased in human prostate cancer and enhances proliferation through inhibition of p21. *Am J Pathol*. 2012;180:2097–107.

28. Dohle GR, Smit M, Weber RFA. Androgens and male fertility. *World J Urol.* 2003;21:341–5.
29. Ferlin A, Vinanzi C, Garolla A, et al. Male infertility and androgen receptor gene mutations: clinical features and identification of seven novel mutations. *Clin Endocrinol (Oxf).* 2006;65: 606–10.
30. Liang M, Melchior F, Feng X-H, Lin X. Regulation of Smad4 sumoylation and transforming growth factor- β signaling by protein inhibitor of activated STAT1. *J Biol Chem.* 2004;279:22857–65.
31. Long J, Matsuura I, He D, Wang G, Shuai K, Liu F. Repression of Smad transcriptional activity by PIASy, an inhibitor of activated STAT. *Proc Natl Acad Sci USA.* 2003;100: 9791–6.
32. Arndt S, Poser I, Moser M, Bosserhoff A-K. Fused-15, a novel Ski/Sno homolog protein, antagonizes BMP signaling. *Mol Cell Neurosci.* 2007;34:603–11.
33. Fabre S, Pierre A, Mulsant P, et al. Regulation of ovulation rate in mammals: contribution of sheep genetic models. *Reprod Biol Endocrinol.* 2006;4:20.
34. Harris RA, Tardif SD, Vinar T, et al. Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc Natl Acad Sci USA.* 2014;111:1467–72.
35. Westendorp RGJ, Kirkwood TBL. Human longevity at the cost of reproductive success. *Nature.* 1998;396:743–6.
36. Byars SG, Huang QQ, Gray L-A, et al. Genetic loci associated with coronary artery disease harbor evidence of selection and antagonistic pleiotropy. *PLoS Genet.* 2017;13:e1006328.
37. Corbett S, Courtiol A, Lummaa V, Moorad J, Stearns S. The transition to modernity and chronic disease: mismatch and natural selection. *Nat Rev Genet* 2018;1:419–30.
38. Yengo L, Yang J, Visscher PM. Expectation of the intercept from bivariate LD score regression in the presence of population stratification. *bioRxiv* 2018:310565. <https://doi.org/10.1101/310565>