Human Molecular Genetics, 2020, Vol. 29, No. 23 3845–3858

doi: 10.1093/hmg/ddaa255
Advance Access Publication Date: 8 December 2020
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Autozygosity mapping and time-to-spontaneous delivery in Norwegian parent-offspring trios

Pol Sole-Navais[1,*,†], Jonas Bacelis[1], Øyvind Helgeland[2,3], Dominika Modzelewska[1], Marc Vaudel[2,4], Christopher Flatley[3], Ole Andreassen[5,6,7], Pål R. Njølstad[2,4], Louis J. Muglia[8,9], Stefan Johansson[2,10,‡], Ge Zhang[8,9] and Bo Jacobsson[1,3,11]

[1]Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg 41685, Sweden, [2]Center for Diabetes Research, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway, [3]Division of Health Data and Digitalization, Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Oslo 0213, Norway, [4]Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen 5021, Norway, [5]NORMENT, University of Oslo, Oslo 0450, Norway, [6]Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0450, Norway, [7]Department of Psychiatry, University of California San Diego, San Diego, CA 92093, USA, [8]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA, [9]Division of Human Genetics, The Center for Prevention of Preterm Birth, Perinatal Institute, March of Dimes Prematurity Research Center Ohio Collaborative, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45267, USA, [10]Center for Medical Genetics, Haukeland University Hospital, Bergen 5021, Norway and [11]Department of Obstetrics and Gynecology, Sahlgrenska University Hospital, Gothenburg 41685, Sweden

*To whom correspondence should be addressed at: Tel: (+46)31-3436769; Fax: (+46)031-786 3573; Email: pol.sole.navais@gu.se

Abstract

Parental genetic relatedness may lead to adverse health and fitness outcomes in the offspring. However, the degree to which it affects human delivery timing is unknown. We use genotype data from ≃25 000 parent-offspring trios from the Norwegian Mother, Father and Child Cohort Study to optimize runs of homozygosity (ROH) calling by maximizing the correlation between parental genetic relatedness and offspring ROHs. We then estimate the effect of maternal, paternal and fetal autozygosity and that of autozygosity mapping (common segments and gene burden test) on the timing of spontaneous onset of delivery. The correlation between offspring ROH using a variety of parameters and parental genetic relatedness ranged between −0.2 and 0.6, revealing the importance of the minimum number of genetic variants included in an ROH and the use of genetic distance. The optimized compared to predefined parameters showed a ≃45% higher correlation between parental genetic relatedness and offspring ROH. We found no evidence of an effect of maternal, paternal nor fetal overall autozygosity on spontaneous delivery timing. Yet, through autozygosity mapping, we identified three maternal loci *TBC1D1*, *SIGLECs* and *EDN1* gene regions reducing the median time-to-spontaneous onset of delivery by ≃2–5% (P-value $< 2.3 \times 10^{-6}$).

[†]Pol Sole-Navais, http://orcid.org/0000-0002-3326-266X
[‡]Stefan Johansson, http://orcid.org/0000-0002-2298-7008
Received: July 3, 2020. **Revised:** November 21, 2020. **Accepted:** November 24, 2020

We also found suggestive evidence of a fetal locus at 3q22.2, near the *RYK* gene region (P-value $= 2.0 \times 10^{-6}$). Autozygosity mapping may provide new insights on the genetic determinants of delivery timing beyond traditional genome-wide association studies, but particular and rigorous attention should be given to ROH calling parameter selection.

Introduction

Offspring of genetically related parents may receive two copies of the same allele co-inherited from a common ancestor. Because alleles are likely to be inherited in long segments, the length of the genome covered by homozygous by descent alleles (autozygosity) rises. Yet, in each generation, recombination breaks autozygous segments into smaller ones, which will be subject to selection for a longer period of time. Consequently, long autozygous segments, a product of recent parental relatedness, are enriched in low frequency and rare deleterious homozygous variants (1,2). The effects of parental genetic relatedness on general health and fitness (3) have been observed both in plants (4,5) and animals (6,7), including humans (e.g. reproductive success, overall health, height, lung function and fluid intelligence) (8–10).

The mapping of autozygous segments has provided insights into recessive effects in particular regions of the genome (11–14) by capturing the effects of low frequency and rare non-genotyped damaging variants that lie within these segments (15), even in traits not affected by overall autozygosity (12,14,16).

Autozygous segments are generally detected by identifying segments of consecutive homozygous variants (runs-of-homozygosity, runs of homozygosity (ROH)) in dense genotype data (17–19). However, ROH may contain both homozygous alleles co-inherited from a common ancestor and non-autozygous alleles. The identification of truly autozygous segments requires the specification of various parameters (e.g. minimum ROH length, number of SNPs included), which despite being optimized to detect autozygosity (17,20), remain arbitrary. These methodologic differences make between-study comparisons difficult and impede proper inference drawing.

Being the leading cause of death among children under 5 years of age (21), preterm delivery has a heavy burden on global health. The physiological control of human delivery timing is poorly understood; no adequate animal models exist, thus providing limited information about delivery timing mechanisms. Shaped by the maternal and fetal genomes, which are distinct, but otherwise related genomes, delivery timing is a trait with relatively high heritability (~20–25%) (22,23). Yet, genome-wide association studies (GWAS) have had only limited success in discovering its genetic determinants, with some exceptions (23–25). Balancing selection, among other evolutionary forces, partly shapes the genetic architecture of preterm delivery (26,27). By increasing heterozygosity, balancing selection may shelter recessive deleterious variants (28) the effect of which could impact a specific phenotype, for example, in case of inbreeding. As suggested by epidemiological studies, preterm delivery may be one such phenotype (29).

Here, we optimize ROH calling parameters by using the relationship between parental genetic relatedness and offspring ROH in parent-offspring triads and estimate the effect of maternal, paternal and fetal autozygosity on the timing of delivery. We also adopted autozygosity mapping as an alternative to typical GWAS to identify segments of the genome potentially harboring low frequency and rare genetic variants with recessive effects on the spontaneous delivery timing.

Results

Parental genetic relatedness and offspring ROH

Before ROH calling in all triad members, we estimated parental genetic relatedness and selected the parameters (LD pruning, physical versus genetic distance, minimum number of genetic variants included in a segment and the allowance of heterozygous calls within the segment) that maximized the correlation coefficient between parental genetic relatedness and offspring ROH, using 108 different combinations of parameters. We observed a limited amount of parental genetic relatedness, as expected in an outbred population, with a median of 5 cM shared between parents (Q1, Q3: 2.2, 12.4 total cM) (Supplementary Material, Table S1).

The correlation coefficient between parental genetic relatedness and offspring ROH varied considerably depending on the sub-cohort and the parameters used to call ROHs in the offspring, ranging from −0.23 to 0.60 (Supplementary Material, Fig. S1). To understand the effect of the different parameters for each ROH call on the correlation coefficient, we ran a linear model using information from all cohorts. The minimum number of genetic variants to call an ROH (including a third-degree polynomial) was the parameter with the strongest effect on the correlation coefficient, followed by the use of genetic versus physical distance and LD pruning (Fig. 1A). Among the parameters maximizing the correlation coefficient, the use of genetic, as opposed to physical distance, was the only consistent parameter in all sub-cohorts (Fig. 1B). Details on the optimal parameters used for each sub-cohort can be found in Supplementary Material, Table S2. The correlation coefficient between the optimized parameters and parental genetic relatedness was R≃ 0.55 (min. $R = 0.5$; max. $R = 0.6$), depending on the sub-cohort (Fig. 2).

Estimated autozygosity in parent-offspring trios

Once the optimal parameters were identified for each sub-cohort (Supplementary Material, Table S2), we called ROHs in all family members using these optimized parameters. We sought to describe the distribution of F_{ROH} and the other measures of autozygosity between family members (Supplementary Material, Table S3). We identified 15 mothers, 13 fathers and 9 fetuses with extreme inbreeding ($F_{ROH} > 8\%$, proposed by Yengo *et al.* (30)). ROHs were identified in 63.6% of the mothers ($n = 23\,676$), 63.0% of the fathers ($n = 24\,805$) and 57.2% of the fetuses ($n = 23\,948$). F_{ROH}, the number of autozygous segments and the average autozygous segment length (Fig. 3), was lower in offspring than in both parents (Wilcoxon test, all P-values $< 10^{-16}$), but we observed no differences between parents (for these analyses we did not include maternal chromosome X). As already described by Nalls *et al.* (31), a decline in measures of autozygosity after each generation is expected with increasing globalization and urbanization. Similarly, F_{HOM} was higher in both parents than in the offspring [median (Q1, Q3): 0.0036 (−0.0009, 0.0079)], but was slightly higher in mothers [median (Q1, Q3): 0.0046 (0.0001, 0.0092)] than in fathers [median (Q1, Q3): 0.0039 (−0.0004, 0.0087)]. Median time to most recent common

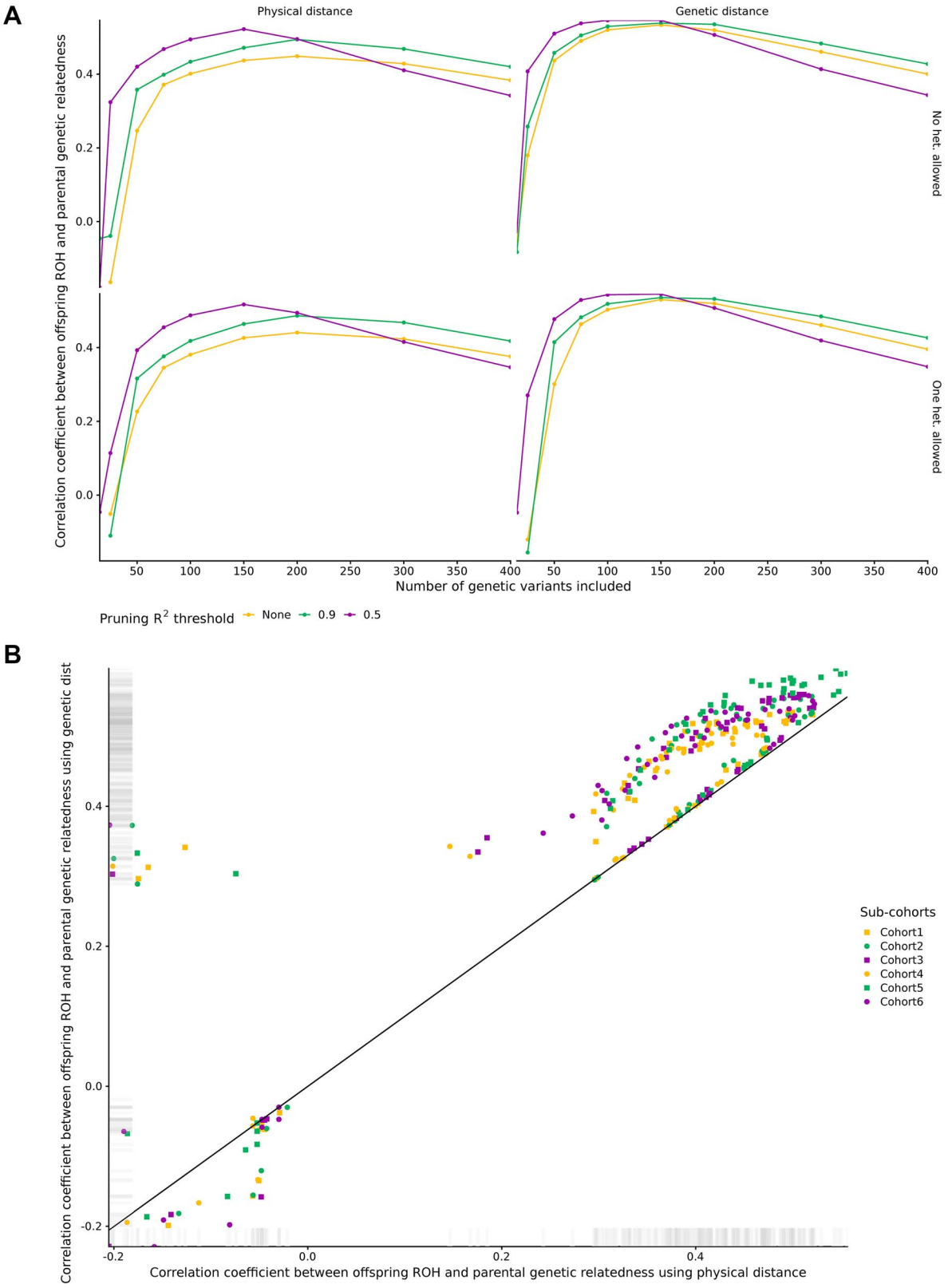

Figure 1. (**A**) Minimum number of homozygous genetic variants included in an ROH segment and correlation coefficient between parental genetic relatedness and offspring ROH. The correlation coefficient was averaged across all sub-cohorts for visualization purposes (n = 24 927). Upper and bottom rows show the correlation coefficient when no or one heterozygous call was allowed, respectively. Left and right columns show the correlation coefficient when using physical or genetic distance, respectively. (**B**) Correlation coefficient between parental genetic relatedness and offspring ROH using physical versus genetic distance. The correlation coefficient was averaged across all sub-cohorts for visualization purposes (*n* = 24 927). A total of 108 ROH calls were performed in offspring using different combinations of pruning, physical versus genetic distance, number of homozygous genetic variants and allowing one or no heterozygous calls within ROHs.

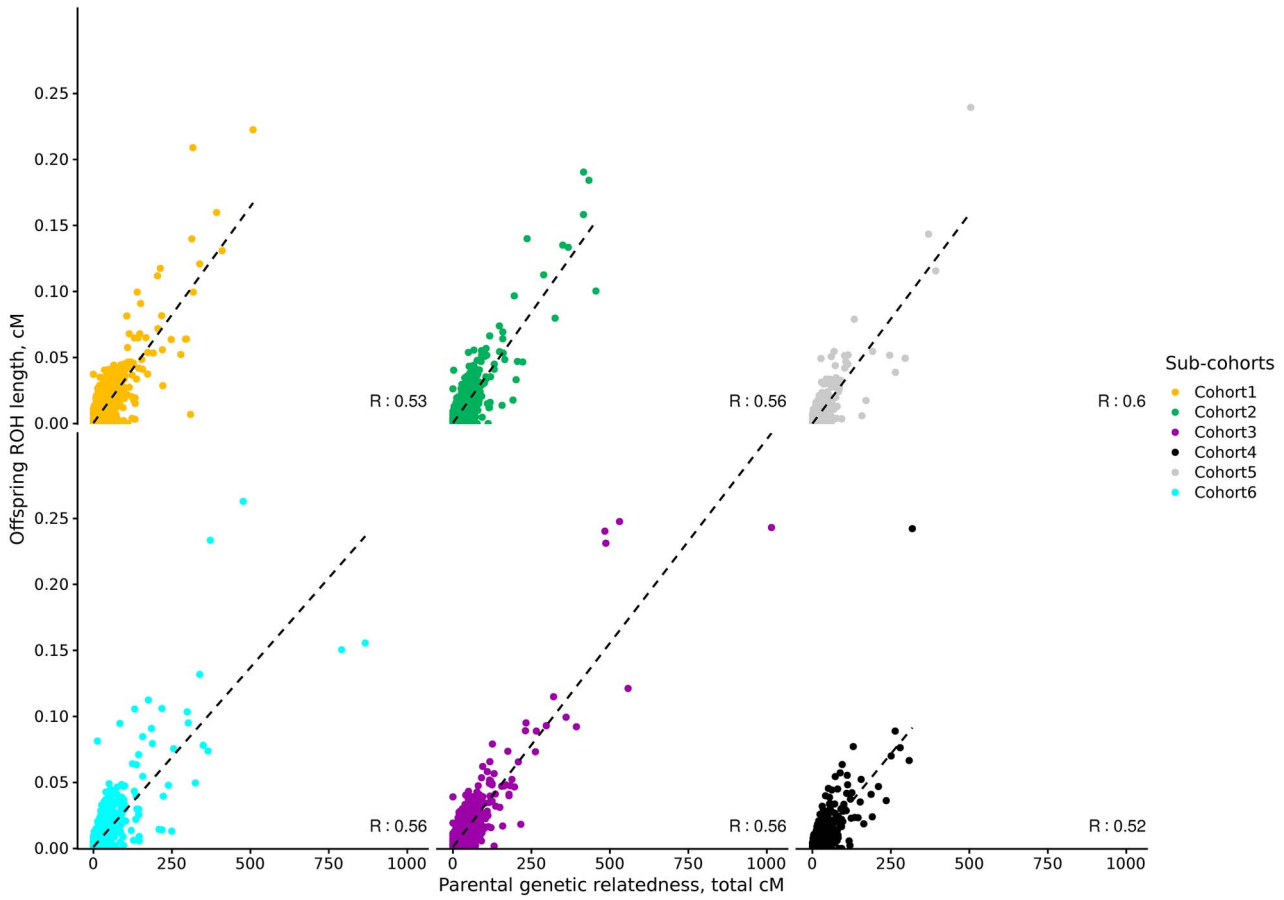

Figure 2. Offspring ROH and parental genetic relatedness using optimized ROH calling parameters. Total accumulated length of offspring ROH and parental genetic relatedness are in cM ($n = 24\,927$). For each sub-cohort, ROH calling parameters explaining the highest proportion of parental genetic relatedness were selected and are thus sub-cohort dependent.

ancestor was 16.1 generations in the mothers and fathers and 17.7 generations in the offspring.

Estimated autozygosity and delivery timing

To evaluate the effect of autozygosity on spontaneous delivery timing, we ran accelerated failure time (AFT) models in mothers, offspring and fathers, using different measures of autozygosity (Supplementary Material, Table S3). We observed no significant association between maternal (estimate: −0.06%; 95% CI: −0.17, 0.05%; *P*-value = 0.283, $n = 21\,815$, events = 19 690), paternal (estimate: 0.09%; 95% CI: −0.02, 0.19%; *P*-value = 0.110 n = 21 465, events = 19 338) nor fetal F_{ROH} (estimate: −0.10%; 95% CI: −0.27, 0.06%; *P*-value = 0.216, $n = 23\,503$, events = 20 789) and spontaneous delivery risk. Supporting the results from F_{ROH}, we observed no significant effect of F_{HOM}, or the other measures of autozygosity on spontaneous delivery risk.

Autozygosity mapping

We next investigated the frequency and effects of autozygous segments on spontaneous onset of delivery (Supplementary Material, Fig. S2). For each family member group, we split overlapping segments from all sub-cohorts into unique segments. Most autozygous segments had a very low frequency, with ∼98% having a frequency below 1% (Supplementary Material, Fig. S2). All segments with a frequency > 1% were identified in the *HLA* region (6p21), followed by segments in the *LCT* gene region (2q21.3), a region with recent positive selection (32), with some segments reaching a frequency of 0.8%.

For each of the identified autozygous segments, we ran an AFT model on spontaneous onset of delivery risk (Supplementary Material, Fig. S3) adjusting for sub-cohort, F_{ROH}, parity and the first 10 principal components. We obtained estimates for 60 391 (59 113 excluding chromosome X) maternal, 58 742 paternal and 44 741 fetal segments. We grouped segments into three different categories: high confidence (*P*-value below a Bonferroni corrected threshold; 0.05/number of segments), low confidence (*P*-value below a threshold using the effective number of segments) and no confidence segments (*P*-value above a threshold using the effective number of segments). The effective number of segments was calculated as the sum of eigenvalues explaining 99.5% of the variance of ROH (effective number of segments: 13 803, 13 328 and 1982 for maternal, paternal and fetal samples). We identified two independent maternal loci (10 segments in total, one surviving a strict Bonferroni correction and the other nine surviving after correcting for the effective number of segments; *P*-value $< 8.3 \times 10^{-7}$ and 3.6×10^{-6}, respectively). The high confidence locus is located in the *TBC1D1* gene region (4p14). This segment, of 0.24 cM, had a frequency of ∼0.3% and was associated with a 2% lower median time-to-spontaneous onset of delivery (*P*-value = 7.4×10^{-7}; Figs 4 and

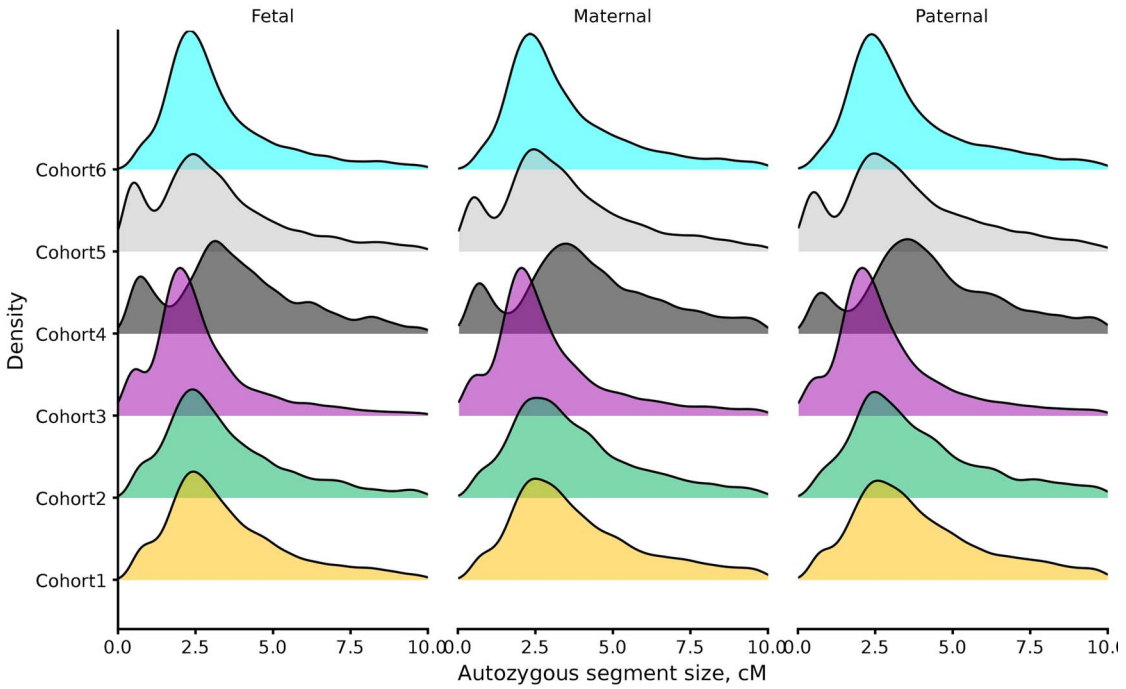

Figure 3. Distribution of the average segment length for each sub-cohort in maternal, paternal and fetal samples. Only subjects with detected autozygous segments are shown.

5, Table 1 and Supplementary Material, Fig. S4). *TBC1D1* encodes the TBC1 Domain Family Member 1 protein, which has a role in regulating cell growth and differentiation. *TBC1D1* is highly expressed in female reproductive organs (cervix, uterus and vagina) (33), but genetic variants affecting *TBC1D1* expression in these organs are yet to be identified. Nonetheless, eQTLs within the identified segments affecting *TBC1D1* expression have been identified in ovaries as well as in non-reproductive organs.

We repeated the analysis in an independent sample of ~2500 mothers, using predefined parameters (see Materials and methods), but we were unable to detect autozygous segments overlapping any of the high or low confidence segments.

We didn't identify any high confidence segments in fathers or fetuses (Supplementary Material, Fig. S5). Yet, using the effective number of autozygous segments (13 328 and 11 982, respectively, for fathers and fetuses), we obtained suggestive evidence for an additional independent locus in fathers and two independent loci in fetuses. The paternal low confidence autozygous segment was observed in a non-coding region (12q21), had a frequency of 0.03% and an effect on time-to-spontaneous onset of delivery of −2%. The fetal loci were located in chromosome 3 (3q22), near *RYK* and *AMOTL2* gene regions. The two segments had a frequency of 0.01%, and an estimated effect on the time-to-spontaneous onset of delivery of −2.7%. While the center of each segment is >1 cM apart from each other, the two fetal loci have the same effect size, standard error and frequency, suggesting that these are not independent. *RYK* encodes the Receptor Like Tyrosine Kinase, an atypical member of the family of growth factor receptor protein tyrosine kinases involved in stimulating Wnt signaling pathways. *RYK* might be involved in Robinow syndrome, as suggested by animal experiments using $Ryk^{-/-}$ homozygous mice, exhibiting a distinctive craniofacial appearance, shortened limbs and postnatal mortality (34). In addition, *RYK* has a low tolerance of loss-of-functions as exposed by the low gnomAD observed/expected constraint score (35). This means that this gene is in strong selection for loss-of-function variants, which could partly explain the results we have obtained. While we excluded subjects with congenital malformations and only included live births reported by the Norwegian Medical Birth Registry, we can't rule out the existence of errors in reporting that could drive the observed effect.

We did not expect an effect of paternal autozygous segments and therefore viewed autozygosity mapping in fathers as a negative control to detect bias due to uncontrolled confounding (e.g. population stratification). We discovered a paternal locus only after correcting for the effective number of segments. While this provides a substantial degree of assurance regarding the robustness of the maternal high confidence segment identified, the same cannot be argued for low confidence segments in neither mothers, fathers nor fetuses. This was confirmed by Q-Q plots of *P*-values (Supplementary Material, Fig. S3). As seen in rare variant analysis, this might be due to the overall low frequency of autozygous segments that are a more recent event and may tend to be geographically clustered (36).

Survival analysis of imputed data under a recessive model

Before running AFT models using single genetic variants, we attempted to identify homozygous subjects (mothers, fathers and fetuses) for imputed genetic variants within the high and low confidence segments with protein consequences. We inspected 10 maternal, one paternal and two fetal segments. None of the segments identified falls within the coding region of the gene overlapping the top maternal segment (*TBC1D1*). In maternal segments, we identified none but one mother homozygous for two missense variants (rs187076049 and rs10001580) within the *PGM2* gene region. This subject had a

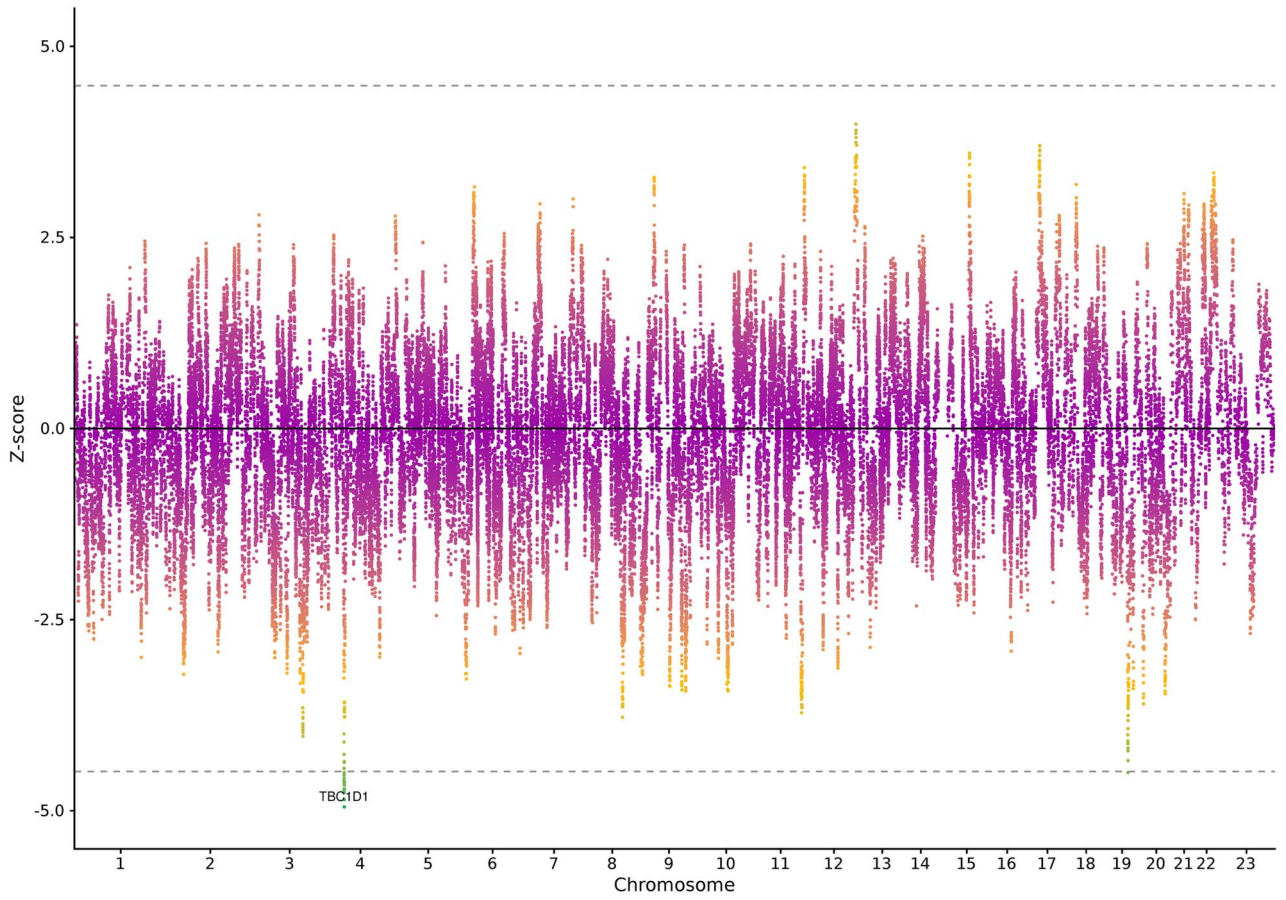

Figure 4. Associations between maternal autozygous segments and time-to-spontaneous delivery. Z-scores of maternal autozygous segments obtained from AFT models on time-to-spontaneous onset of delivery. The Bonferroni threshold for significance ($n = 23\,323$, n autozygous segments = 60 391, effective n of segments = 13 803) is indicated by the dotted line.

Table 1. Independent high and low confidence autozygous segments associated with time-to-spontaneous onset of delivery

Chr.	Start	End	Nearest gene	N	Freq.	Effect size (%)[a]	P-value	Mim
Maternal								
4	37 884 116	37 900 371	TBC1D1	23 323	0.003	−2.1	7.4×10^{-7}	–
4	37 586 331	37 590 112	c4orf19	23 323	0.002	−2.1	2.3×10^{-6}	–
Paternal								
12	76 086 784	76 099 541	KRR1	22 901	0.003	−1.9	3.2×10^{-6}	–
Fetal								
3	133 989 068	134 013 144	RYK	23 332	0.001	−2.7	2.0×10^{-6}	–
3	134 013 144	134 033 585	AMOTL2	23 332	0.001	−2.7	2.0×10^{-6}	–

High and low confidence autozygous segments (*P*-values below Bonferroni correction for effective number of segments) are shown. For each segment, the closest protein-coding gene is also shown, and for this gene, the Mim phenotype number for a phenotype with recessive inheritance. The total number of segments (effective number of segments) was 59 193 (13 803), 58 742 (13 328) and 44 741 (11 982) for maternal, paternal and fetal samples, respectively.
[a]The effect size reflects the effect of the autozygous segment on time-to-spontaneous onset of delivery, representing the percent difference in time-to-spontaneous onset of delivery between subjects affected by an autozygous segment in the region versus those not affected.

gestational duration below the median, and its removal slightly attenuated the effect size and increased the standard error, resulting in an increase in the *P*-value (from 1.9×10^{-6} including the subject to 4.1×10^{-6} after its removal). The rs10001580 genetic variant is associated with oestradiol levels (37) and is an eQTL for *PGM2* in cultured fibroblasts (33) and blood (38), and for *TBC1D1* in blood (38). None of the paternal or fetal segments fell within the coding region of any gene.

In this study, we assumed that the mechanism of action of genetic variants within autozygous segments is similar to that of mendelian disorders, we ruled out inspecting genetic variants outside the coding region of the gene. Nonetheless, with the intention to identify genetic variants underlying the effects observed in autozygosity mapping, we performed survival analysis under a recessive model using imputed genetic data for all variants within the high and low confidence segments (Supplementary Material, Table S5). Results included 580 maternal, 65 paternal and 164 offspring genetic variants. We identified an intronic maternal genetic variant in the *PGM2* gene associated with time-to-spontaneous onset of delivery after genome-wide

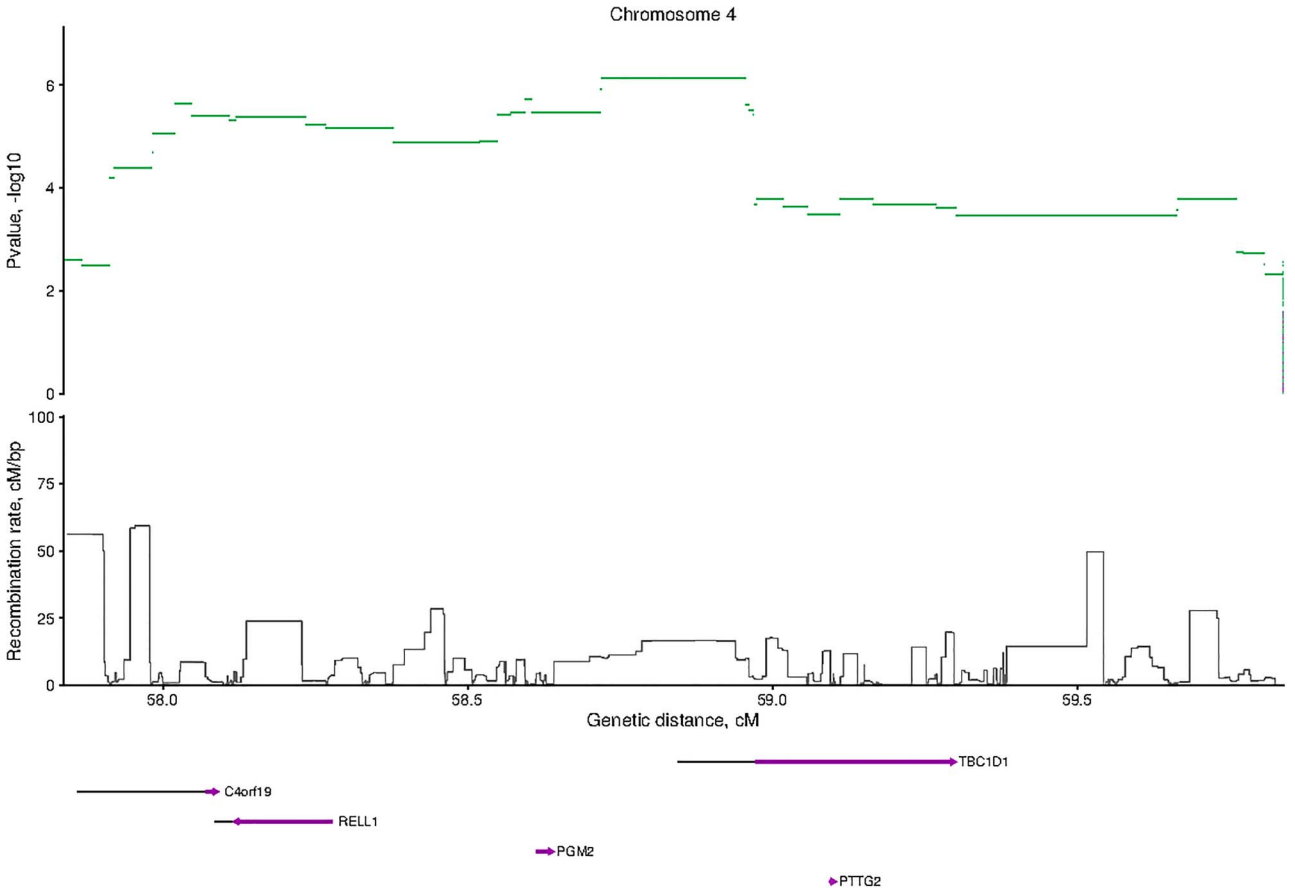

Figure 5. Associations between autozygous segments at the maternal 4p14 locus with time-to-spontaneous onset of delivery. *P*-values were obtained from AFT models on time-to-spontaneous onset of delivery ($n = 23\,323$). Gene names for the 10 longest genes in the region, recombination rate, transcript orientation (arrow) and coding region (purple) are also depicted.

Bonferroni correction (rs76770307, *P*-value = 1.2×10^{-9}), but with a very low homozygous count ($n = 1$). While the two genetic variants reported above (rs187076049 and rs10001580) are in high linkage disequilibrium with rs76770307 (D' = 1 for both genetic variants), none of them were associated with time-to-spontaneous onset of delivery using a recessive model (*P*-value = 0.431 and 0.828). Overall, 14 maternal genetic variants passed a relaxed Bonferroni correction (*P*-value < 0.05/580), with only two genetic variants with a homozygous count > 5 (rs10029748, located 5' UTR of *TBC1D1* and rs10008243, an intronic variant in *TBC1D1*). These genetic variants had a homozygous count of 0 in the replication dataset.

We identified no paternal genetic variants associated with time-to-spontaneous onset of delivery, even using a relaxed Bonferroni correction (0.05/65).

Along the same lines, no fetal genetic variants had a genome-wide significant *P*-value, and only one genetic variant (rs77926300) passed after applying a relaxed Bonferroni correction (0.05/164). However, this genetic variant had a homozygous count of 1.

Long segment gene burden analysis

Gene-level burden analysis of long autozygous segments provided evidence for additional genes (Supplementary Figure S6). We analysed a total of 18 675 maternal, 17 931 paternal and 17 621 fetal protein-coding genes. After controlling for multiple comparisons, we identified 41 different genes in mothers (Fig. 6) in the 19q13.41 locus (*P*-values < 2.0×10^{-6}), most probably as part of the same autozygous segment, and *EDN2* at the 1p34.2 locus (*P*-value = 2.3×10^{-6}). The 19q13.41 locus contains numerous *SIGLECs* and zinc finger protein genes. *SIGLECs* encode Sialic acid-binding immunoglobulin-type lectins, which are cell-surface proteins that bind sialic acid, and are mostly expressed in immune cells. Some of the 14 *SIGLEC* genes that exist are also expressed in villous and extravillous cytotrophoblasts, decidual cells, as early as eight gestational weeks, and in maternal uterine glands (39,40), suggesting a possible role of *SIGLECs* in mediating the immune tolerance at the feto-maternal interface. The *EDN2* gene (1p34.2 locus) is associated with almost 5% lower time-to-spontaneous onset of delivery (*P*-value= 2.3×10^{-6}), EDN2 encodes endothelin 2, a strong vasoconstrictor involved in follicular rupture and ovulation; it causes the contraction of the smooth muscle layer surrounding each follicle (41,42).

TBC1D1 gene, the top associated segment in autozygosity mapping, was nominally significant associated with time-to-spontaneous onset of delivery at the nominal level (*P*-value = 0.003) in the gene burden analysis.

Given the nature of this analysis targeting low-frequency segments, we were unable to detect an effect of these genes in the replication dataset.

We identified no paternal genes, and only one fetal gene after controlling for multiple comparisons (Supplementary Material, Fig. S7). The fetal gene identified (*P*-value = 6.0×10^{-8}), *DAOA* (13q33.2 region), had an estimated effect of −15% on

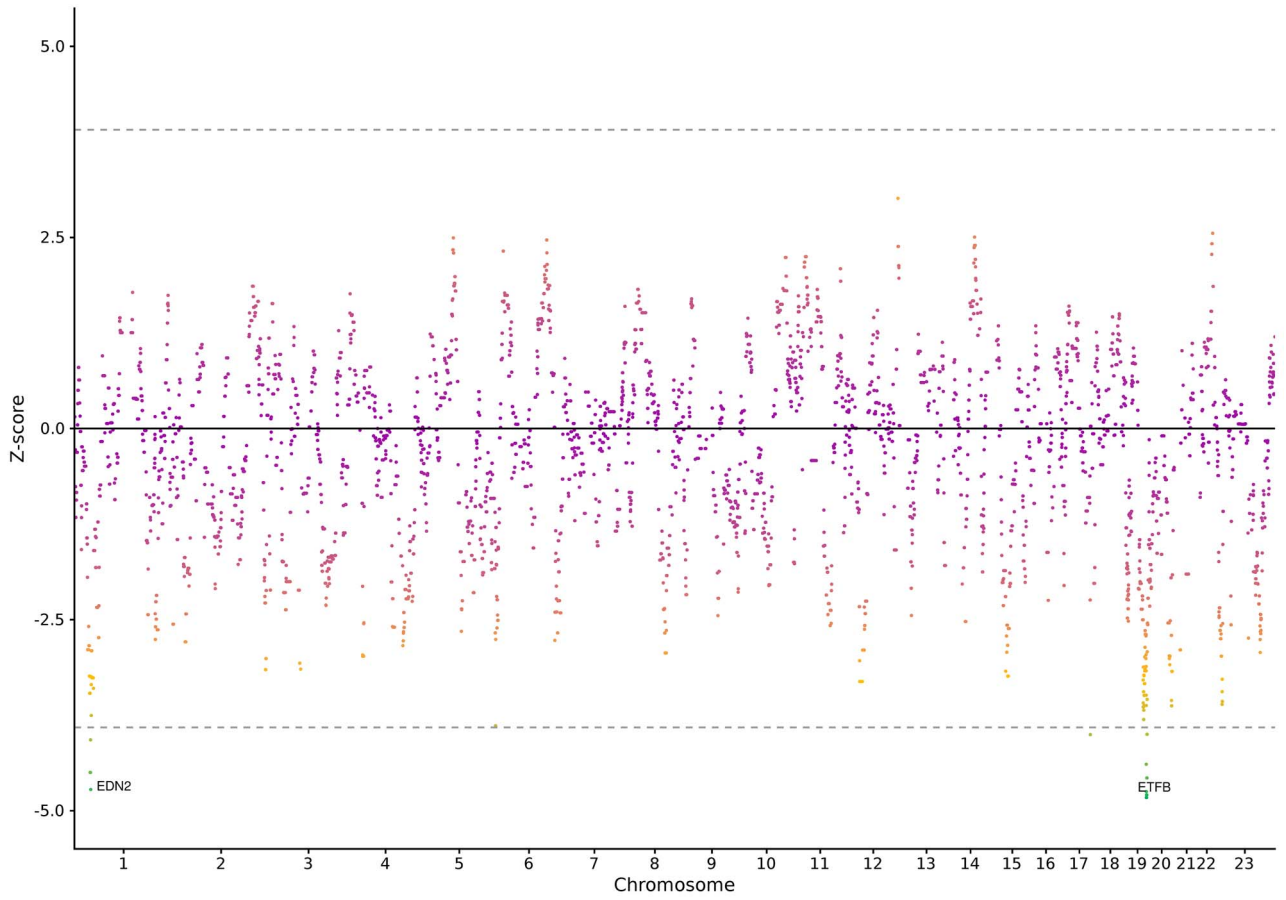

Figure 6. Maternal long segment gene burden test and time-to-spontaneous delivery. Z-scores of maternal genes obtained from AFT models on time-to-spontaneous onset of delivery are shown. Long autozygous segments were collapsed into protein-coding genes as a binary variable. The dotted line indicates the Bonferroni threshold for significance correcting for the effective number of genes ($n = 23\,323$, n genes = 18 675, effective n of genes = 1084).

median time-to-spontaneous delivery. Given the extreme P-value obtained for this gene (next lowest P-value = 2.7×10^{-4}) and considering it is a single hit in the region, we argue that this gene was identified due to uncorrected population stratification.

Comparison against predefined parameters

For completeness, we compared our approach with previously suggested (9) and widely used (10,30,43,44) parameters for ROH calling. Compared to the optimized parameters, predefined parameters led to an average 45% reduction in the correlation estimates between parental genetic relatedness and offspring ROH (Table 2 and Supplementary Material, Fig. S8).

We then assessed, for each individual, the overlap between segments called using optimized parameters and those using Joshi's parameters; the optimized parameters served as reference. We observed no overlapping segments in 29% of subjects and an overlap > 90% in 58% of subjects (Fig. 7).

Table 2. Correlation coefficients between offspring ROH and parental genetic relatedness using the optimized and predefined parameters

	Optimized parameters	Predefined parameters
Sub-cohort1	0.53	0.50
Sub-cohort2	0.56	0.38
Sub-cohort3	0.56	0.53
Sub-cohort4	0.52	0.36
Sub-cohort5	0.60	0.39
Sub-cohort6	0.56	0.53

Spearman rank-order correlation coefficients between offspring ROH and parental genetic relatedness.

Discussion

Low frequency and rare genetic variants may play an essential role in human delivery timing, but individual variants are difficult to identify due to a lack of information on gestational duration is already available large sequencing datasets (e.g. UK Biobank). Here, we systematically called ROH segments using over 20 000 parent-offspring trios and assessed the relationship between maternal, paternal and fetal autozygosity and spontaneous delivery risk. Calling ROHs using cohort-specific, rather than predefined, parameters increased the correlation between offspring ROH segments and parental genetic relatedness by 45%. While we observed no evidence of an effect of overall autozygosity on delivery timing, we identified three maternal autozygous segments in 4p14, 19q13.41 and 1p34.2 loci associated with time-to-spontaneous onset of delivery. We found no evidence supporting an effect of imputed genetic variants within these segments under a recessive mode.

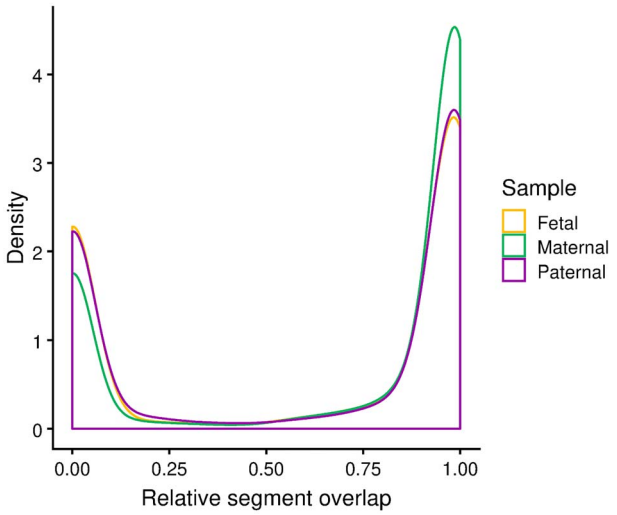

Figure 7. Autozygous segment overlap between optimized and predefined ROH parameters in mothers, fathers and fetuses. For each individual, we calculated the proportion of overlap between autozygous segments called using the optimized and predefined parameters. The optimized parameters served as the reference. Results for all cohorts merged are shown.

ROH calling parameter selection based on the relationship between offspring ROH length and parental genetic relatedness provided us with non-arbitrary F_{ROH} estimates. As a consequence, ROH calling parameters varied for each sub-cohort, leading to differences in the average length and the total number of autozygous segments between sub-cohorts. While the optimized parameters were far from optimal (correlation coefficient between offspring ROH and parental genetic relatedness, R ~0.55), our results indicate that the use of predefined parameters would have performed worse (R ~0.4). We observed this in samples genotyped with different arrays, but with a similar population structure; we expect differences to be larger in samples with a more diverse population background. We sought to remove arbitrariness in the selection of parameters, yet it is unavoidable to use subjective parameters, for example, when estimating parental genetic relatedness with GERMLINE. The use of an alternative minimum length for IBD segment identification would have probably led to different optimized parameters. We selected a minimum length of 2 cM, for which GERMLINE has a detection power and accuracy of ~70% (45). Another limitation was assuming that the optimized parameters we identified in the offspring would also be applicable to parents. While this assumption may not hold, a formal test would require parental and grand-parental genetic data, which is not available.

We observed no maternal, fetal nor paternal effect of autozygosity on time-to-spontaneous onset of delivery, indicating the effect is particularly modest. Despite a consensus for the use of F_{ROH} when estimating autozygosity, we used four additional alternative measures: all confirming a null effect. Given the small amount of autozygosity detected in outbred populations, large sample sizes are required for the detection of effects on any trait. Consistent with our findings, two recent studies in 200 000 and 1 M individuals found no evidence of an effect of autozygosity on their individual own birth weight (8,10), a measure strongly correlated with the duration of gestation. We did not classify ROHs according to their length, which would have helped to understand recent parental relatedness, background relatedness, or founder events in our sample (46). However, using our approach, we were targeting segments arising from recent parental genetic relatedness. Moreover, using 2 cM as the minimum threshold for IBD detection indicates that the common ancestor is 5–6 generations from parents (47).

IBD mapping offers better power than GWAS (48), which could also be true for autozygosity mapping. While we identified one region of high confidence in the TBC1D1 gene region, we find no associations from imputed genetic variants within this region using a recessive model. The underlying causal genetic variant, if any, might be rare and in low LD with other variants, suggesting that it is likely to affect protein sequence rather than the expression. Autozygosity mapping is similar to GWAS in the sense that it does not provide evidence of causality: non-controlled population structure can also result in significant differences in ROH rates in relation to time-to-spontaneous onset of delivery. As expected, and supporting a robust population stratification control, we did not observe a paternal effect of autozygosity, nor of autozygosity mapping using a stringent Bonferroni correction. However, we admit that we identified one paternal autozygous segment associated with time-to-spontaneous onset of delivery when controlling for the number of effective segments. As such, we only considered the results obtained controlling for the total number of autozygous segments.

Genomic proximity does not always guarantee correct gene identification (49). By assuming that autozygous segments act through similar mechanisms to monogenic traits, the mapping of autozygous segments to causal genes may appear to be less difficult than in GWAS. However, we expressly targeted long segments, which span multiple genes, making correct gene identification difficult. Hence, while the autozygous segment in *TBC1D1* gene region was the top associated segment, segments in the same region overlapping with other genes (*PGM2*) were also highly significant. Maternal autozygous segments in *TBC1D1* gene region were associated with 2.5% lower time-to-spontaneous onset of delivery. This gene, as well as *PGM2*, is highly expressed in the cervix, uterus and vagina, adding certain reassurance of a true positive. *TBC1D1* gene expression in the endometrium of cattle is increased during pregnancy (50), and *TBC1D1* is involved in muscle-contraction glucose uptake (51,52). While in this study, we targeted autozygous segments resulting from recent parental genetic relatedness, we decided to target long ROH and perform a gene burden test. This provided additional insights, involving two loci, which are also highly expressed in either placenta, uterus or ovaries: *SIGLECs* and *EDN2*. The effect size of these genes was twice as big as the effect size observed for the segments identified through autozygosity mapping. *SIGLEC-6* plays a role in placental excess proliferation and invasion (53), and its expression is increased in the basal plate and chorionic villi of preterm preeclamptic mothers, but not in those delivering at term (39). *EDN2* is a strong vasoconstrictor, and causes the contraction of the smooth muscle layer surrounding each follicle (41,42). *EDN2* mediates the ovulation by inducing the contraction of follicles for oocyte expulsion (36,37). The expression of MicroRNAs that target *EDN2* (miR-210) was decreased in placental villous tissues from preterm deliveries compared to term deliveries (54). Whether this gene can induce the contraction of other smooth muscles, such as endometrium, remains unknown. All these genes appear to be pointing toward either placental complications, (gestational) diabetes, or medical conditions linked to high blood pressure. However, in this study we explicitly excluded subjects with placental complications, diabetes or gestational diabetes, hypertension and pre-eclampsia, potentially ruling

out a mediation of the observed effects through any of these conditions.

In the absence of large sequencing data with high-quality information on pregnancy phenotypes, moving beyond traditional GWAS may prove to be useful to identify loci associated with delivery timing. However, special attention must be put on ROH calling parameters, particularly in the absence of parent-offspring data. Our observations suggest that there is no or a very small effect of autozygosity on the spontaneous delivery timing. Autozygosity mapping and gene burden tests highlighted three candidate maternal loci associated with time-to-spontaneous onset of delivery. We hope that future functional follow-up studies based on the observations presented here will yield novel insights and a better characterization of the mechanisms behind human delivery timing.

Materials and Methods

Study population

In this study, we used genotype data drawn from the Norwegian Mother, Father and Child Cohort Study (MoBa) (55,56). This family based cohort enrolled more than 114 000 children, 95 000 mothers and 75 000 fathers from 50 Norwegian Hospitals between 1999 and 2008. The MoBa Genetics infrastructure is a collaborative research effort consisting mainly of three major research projects, with a total of ∼25 000 parent-offspring trios genotyped. Four different genotyping arrays were used in each project: 11 000 parent-offspring trios were genotyped with Illumina HumanCoreExome at the Genomics Core Facility (Norwegian University of Science and Technology, Trondheim, Norway), 9000 parent-offspring trios were genotyped with Illumina Infinium Global Screening Array MD at the Erasmus Medical Center (Erasmus University, Rotterdam, The Netherlands), 3000 parent-offspring trios were genotyped using Infinium Global Screening Array 24 at deCode Genetics (Reykjavik, Iceland) and 5000 parent-offspring trios were genotyped using Illumina Infinium OmniExpress at deCode Genetics (Reykjavik, Iceland). For the replication of single genetic variant results, we obtained additional genetic data from 3000 mothers from the same MoBa cohort, genotyped in two batches (Illumina Infinium OmniExpress 24, genotyped at deCode Genetics, Reykjavik, Iceland).

We excluded multiple pregnancies, women with gestational or type 2 diabetes, women with pre-eclampsia or hypertension, pregnancies lasting <154 (considered not viable) or $\geq$308 days, with a birth weight below 1500 gr, missing gestational duration, conceived by in-vitro fertilization or affected by any of the following: polyhydramnios, oligohydramnios or congenital malformations. Gestational duration was estimated by ultrasound scan at 19–20 gestational weeks.

Phenotype and covariates

Maternal health information prior to and during pregnancy, including gestational duration and parity, as well as complications of pregnancy and birth were recovered from the Medical Birth Registry of Norway.

Descriptive characteristics of the different sub-cohorts can be viewed in Supplementary Material, Table S6. We defined spontaneous onset of delivery as a delivery initiated by spontaneous contractions or rupture of membranes. Deliveries initiated by induction methods, including the use of prostaglandins, oxytocin, amniotomy or any other induction procedure, or planned cesarean section were censored. Maternal education was defined as $\leq$12 years, 13–16 years or $\geq$17 years, and household incomes as none, one parent or both of parents having an income $>$ 300 000 Norwegian crowns.

Genotyping and quality control

Genotyping and quality control have been previously described (57) and performed accordingly for all samples included in this study, regardless of the genotyping array platform used. Samples with a call rate <0.98 or excess heterozygosity >4 SD, and variants with call rates $<98\%$, 10% GenCall-score <0.3, cluster separation <0.4, Theta AA standard deviation >0.4 and HWE P-value $<10^{-6}$ were excluded. We included autosomal markers for fathers and offspring, as well as autosomal chromosome X markers for mothers. Genome coordinates were mapped to the Genome Reference Consortium Human Build 37 (hg19). We excluded samples with recent ancestry different to European, and those with genetic relatedness with another sample greater than second to a third cousin (KING (58) cut-off $>$ 0.0884). After ROH calling in mothers, fathers and offspring, we excluded subjects with suspected uniparental disomy defined as having at least one chromosome with $>70\%$ covered by ROH. Subjects with extreme inbreeding, defined as ROH segments covering $>8\%$ of the genome, were also excluded.

Autozygosity parameter selection

ROH calling with PLINK (59) relies on the input of several parameters such as the minimum ROH length, the number of heterozygotes allowed or the number of genetic variants included, amongst others. Despite the fact that these parameters have been optimized for estimating autozygosity (17,20), the selection remains arbitrary. To overcome this, we selected PLINK v.1.9 parameters *a posteriori*. We called ROH segments 108 times in the offspring by modifying the LD pruning threshold, using cM or bp, varying the number of heterozygotes allowed and the minimum homozygous count to call an ROH (see Supplementary Material, Methods for details), and selected the parameters that maximized the Spearman rank-order correlation coefficient between parental genetic relatedness and offspring ROH length. These parameters were later used to call ROHs in mothers, fathers and offspring (see Autozygosity calling in family trios). Parental genetic relatedness was estimated as the total length of shared identical-by-descent (IBD) segments in cM using GERMLINE v.1.5.3 (47) (minimum length of IBD shared segments: 2 cM; remaining parameters set by default). Prior to IBD identification, haplotype phase was estimated using Eagle v.2.4.1 (60) using trio data.

We estimated the correlation between parental genetic relatedness offspring ROH length by using Spearman rank-order correlation. This approach was used for each of the sub-cohorts and for the 108 different combinations of ROH calling parameters. The parameters leading to the highest (positive) R were used for ROH calling in all family members.

We attempted to replicate the findings by using genetic data in mothers and fathers. Given offspring genotype data were not available in these additional batches, we used predefined ROH calling parameters, based on the parameters obtained using our approach. These were the ROH calling parameters employed for the replication sub-cohorts, using genetic distance and with moderate pruning ($R^2>0.5$):

–homozyg-window-snp 125 –homozyg-snp 125 –homozyg-kb 0.0000001 –homozyg-gap 5000 –homozyg-window-missing

125 * 0.05 –homozyg-window-threshold 0.0005 –homozyg-window-het 0 –homozyg-density 5000.

For completeness, we compared our approach with widely used ROH calling parameters previously proposed by Joshi (9). The following parameters were employed in non-pruned data:

–homozyg –homozyg-window-snp 50 –homozyg-snp 50 –homozyg-kb 1500 –homozyg-gap 1000 –homozyg-density 50 –homozyg-window-missing 5 –homozyg-window-het 1.

As for the optimized parameters, we assessed the correlation between parental genetic relatedness and offspring ROH using Joshi's parameters. Finally, we evaluated the overlap between autozygous segments called using the optimized and Joshi's parameters. We matched autozygous segments for each individual with at least one autozygous segment detected using the optimized parameters, with autozygous segments detected using Joshi's parameters. For each individual, we estimated the proportion of overlapping segments divided for the total segment length using the optimized parameters.

Autozygosity calling in family trios

The sub-cohort-specific parameters maximizing the coefficient of determination between offspring ROH length and parental genetic relatedness were used to call ROHs in all family members. We calculated F_{ROH} for each subject as the sum of total ROH length divided by the total mappable distance (either in cM or bp) (15).

There is no single best measure of autozygosity but we chose F_{ROH} as our primary measure and investigated the effects of other measures: excess homozygosity (F_{HOM}), the total number of segments (NSEG) and the average length of ROH. F_{HOM} was calculated on a SNP-by-SNP basis using PLINK—het flag. Additionally, we estimated the time to the most recent common ancestor in generations as $d/2$ k, where d is the genetic distance (100 cM) and k the individual average ROH length (61).

Analysis

In this study, we decided to study delivery timing using survival analysis. We used AFT models for a number of reasons (see Supplementary Methods). Basically, the covariates accelerate or decelerate survival time (62), affecting the rate at which an individual proceeds to the event. Time-varying effects, which contribute to explain gestational duration heritability (63), are not assumed, as opposed to in Cox survival models.

We estimated the effect of F_{ROH}, F_{HOM}, the total number of segments (NSEG), the average segment size, autozygosity mapping and gene burden on spontaneous onset of delivery risk in mothers, fathers and offspring in separate models using R 'survival' package. The time scale was days until delivery, and the event, a spontaneous onset of delivery; pregnancies with deliveries initiated by induction or a planned cesarean section were censored (see phenotype and covariates section for a detailed definition of spontaneous onset of delivery). Throughout the manuscript, we report an estimate of the percent difference in time-to-spontaneous delivery (i.e. (exp(beta)−1) × 100) for any survival time quantile. We ran a total of two models for both F_{ROH}, F_{HOM}, NSEG, the average length of segments and time to a most common ancestor for each family member group: a crude model adjusting for sub-cohort and model 1 adjusting for sub-cohort, parity (nulliparous versus multiparous), 10 principal components, maternal educational attainment and household income. Thus, we used a Bonferroni corrected significance threshold by accounting for the number of exposures in each family member, and considered significant if P-value was below 0.01 (0.05/5 autozygosity measures). Autozygosity mapping and gene burden models were adjusted for sub-cohort, F_{ROH}, parity and the first 10 principal components.

Code for data manipulation and analysis was structured using Snakemake (64) and is available at https://github.com/PerinatalLab/ROH.

Autozygosity mapping

After identifying ROH segments in each family member of each sub-cohort, we split segments into unique intersecting segments into all sub-cohorts. Segments shared across sub-cohorts or unique to a single sub-cohort were kept (i.e. no other segment, from any cohort, had its start or end position included in any other segment). These segments were transformed into a matrix of binary calls, indicating whether an individual is autozygous for that particular segment. Only segments shared by at least 20 subjects (1 in 1250 or ∼0.08%) were kept.

Splitting segments into multiple non-overlapping segments means that additional correlation between close segments (some segments differed only in one subject) was introduced. The use of a Bonferroni P-value threshold correction would be overly conservative. To avoid this, we estimated the effective number of segments to identify segments with suggestive evidence. For each member of the family, we ran principal components analysis on the segment matrix from each chromosome. The number of first eigenvalues explaining 0.995 of the variance in each chromosome was summed to obtain an effective number of segments. We identified high confidence segments using a Bonferroni corrected threshold of 0.05/total number of autozygous segments. Subsequently, low confidence segments were defined as segments surviving a Bonferroni correction of 0.05/effective number of autozygous segments and were not high confidence segments; remaining segments were considered of no confidence.

We created clumps of high and low confidence segments not farther from 0.5 cM from each other's central position and chose the segment with the lowest P-value within each clump as the top independent segment.

While we did not expect an effect from the paternal autozygous segments, we still ran survival models in fathers and used the results as a negative control.

Survival analysis of imputed genetic data under a recessive model

Genotype data from all participants included in this study were pre-phased using Shapeit v2.790 (65) and then imputed at the Sanger Imputation Server using the Haplotype Reference Consortium v1.1 reference panel (66). Single genetic variant association analysis using hard called genotypes was performed using a recessive model. We fit separate models for mothers, fathers and offspring for all genetic variants with an INFO score > 0.4 that lay within the high or low confidence segments. We did not restrict minor allele frequency. All models were adjusted for sub-cohort, parity and the first 10 principal components. Genetic variants surviving a genome-wide Bonferroni correction threshold (P-value $< 5 \times 10^{-8}$) were considered to be associated with time-to-spontaneous delivery and were subsequently tested for replication. A replication stage was performed using the same methods as for the discovery stage.

Long segment gene burden analysis

We conducted a gene burden analysis of long segments in all family members using the same methods as for autozygosity

mapping. For this analysis, we classified segment length according to Pemberton *et al.* (67). Briefly, we modeled the segment length distribution as a mixture of three Gaussian distributions: short segments may reflect ancient haplotypes, intermediate segments may have arised from background relatedness and long segments may result from recent parental relatedness. We classified segment length using *Mclust* from the *mclust* package for each cohort. Given that our segment length was already large, we used segments classified as intermediate or long length for the burden test. Boundaries between different autozygous segment sizes for each sample and sub-cohort can be viewed in Supplementary Material, Table S7. For each protein-coding gene, we encoded gene burden as a binary variable indicating whether an individual had an autozygous segment partially or fully overlapping the gene transcription start and end (University of California Santa Cruz Table browser (68)). Whenever an autozygous segment overlapped several gene regions, each gene was used in the analysis.

Segment and genetic variants annotation

All segments and genetic variants were mapped to the nearest protein-coding gene. Gene transcription coordinates were downloaded from the University of California Santa Cruz Table browser (68). Functional data of genetic variants using a recessive model, the expected mode of action in variants within our high confidence segments, are scarce. Thus, we mapped the identified genes with genes known to affect phenotypes with a recessive inheritance from the Online Mendelian Inheritance in Man (OMIM) data (69). We annotated variants within high confidence segments with moderate or high impact (according to Ensembl VEP (70)) using gnomAD version 2.0.1 (71).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgments

The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research, NIH/NIEHS (contract no N01-ES-75558), NIH/NINDS (grant no. 1 UO1 NS 047537-01 and grant no. 2 UO1 NS 047537-06A1). This work is supported by a grant from the Burroughs Wellcome Fund (10172896). We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authority and KG Jebsen Stiftelsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen and the Western Norway health Authorities (Helse Vest). L.J.M. and G.Z. were supported by the March of Dimes Prematurity Research Center Ohio Collaborative. We are grateful to all the participating families in Norway who take part in this ongoing cohort study.

Conflict of Interest statement. None declared.

Ethics statement

This study was approved by the Regional Committee for Medical and Health Research Ethics from Norway (2015/2425).

References

1. Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A. and Li, J.Z. (2013) Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.*, **93**, 90–102.
2. Pemberton, T.J. and Szpiech, Z.A. (2018) Relationship between deleterious variation, genomic autozygosity, and disease risk: insights from the 1000 genomes project. *Am. J. Hum. Genet.*, **102**, 658–675.
3. Charlesworth, D. and Willis, J.H. (2009) The genetics of inbreeding depression. *Nat. Rev. Genet.*, **10**, 783–796.
4. Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J. *et al.* (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.*, **6**, 6258.
5. Darwin, C. (1876) *The effects of cross and self fertilisation in the vegetable kingdom*. John Murray, London.
6. Huisman, J., Kruuk, L.E.B., Ellis, P.A., Clutton-Brock, T. and Pemberton, J.M. (2016) Inbreeding depression across the lifespan in a wild mammal population. *Proc. Natl. Acad. Sci. USA*, **113**, 3585–3590.
7. Pemberton, J.M., Ellis, P.E., Pilkington, J.G. and Bérénos, C. (2017) Inbreeding depression by environment interactions in a free-living mammal population. *Heredity*, **118**, 64–77.
8. Johnson, E.C., Evans, L.M. and Keller, M.C. (2018) Relationships between estimated autozygosity and complex traits in the UK biobank. *PLoS Genet.*, **14**, e1007556.
9. Joshi, P.K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., Jackson, A.U., Schurmann, C., Smith, A.V., Zhang, W. *et al.* (2015) Directional dominance on stature and cognition in diverse human populations. *Nature*, **523**, 459–462.
10. Clark, D.W., Okada, Y., Moore, K.H.S., Mason, D., Pirastu, N., Gandin, I., Mattsson, H., Barnes, C.L.K., Lin, K., Zhao, J.H. *et al.* (2019) Associations of autozygosity with a broad range of human phenotypes. *Nat. Commun.*, **10**, 4957.
11. Weiss, L.A., Pan, L., Abney, M. and Ober, C. (2006) The sex-specific genetic architecture of quantitative traits in humans. *Nat. Genet.*, **38**, 218–222.
12. Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R. and Malhotra, A.K. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci. USA*, **104**, 19942–19947.
13. Powell, J.E., Henders, A.K., McRae, A.F., Kim, J., Hemani, G., Martin, N.G., Dermitzakis, E.T., Gibson, G., Montgomery, G.W. and Visscher, P.M. (2013) Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet.*, **9**, e1003502.
14. Mukherjee, S., Guha, S., Ikeda, M., Iwata, N., Malhotra, A.K., Pe'er, I., Darvasi, A. and Lencz, T. (2014) Excess of homozygosity in the major histocompatibility complex in schizophrenia. *Hum. Mol. Genet.*, **23**, 6088–6095.
15. Keller, M.C., Visscher, P.M. and Goddard, M.E. (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, **189**, 237–249.
16. Johnson, E.C., Bjelland, D.W., Howrigan, D.P., Abdellaoui, A., Breen, G., Borglum, A., Cichon, S., Degenhardt, F., Forstner, A.J., Frank, J. *et al.* (2016) No reliable association between

runs of homozygosity and schizophrenia in a well-powered replication study. *PLoS Genet*., **12**, e1006343.
17. Howrigan, D.P., Simonson, M.A. and Keller, M.C. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics*, **12**, 460.
18. Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M. and Wilson, J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One*, **5**, e13996.
19. Ceballos, F.C., Hazelhurst, S. and Ramsay, M. (2018) Assessing runs of homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, **19**, 106.
20. Gazal, S., Sahbatou, M., Perdry, H., Letort, S., Génin, E. and Leutenegger, A.-L. (2014) Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. *Hum. Hered*., **77**, 49–62.
21. Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J.E., Cousens, S., Mathers, C. and Black, R.E. (2015) Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. *Lancet*, **385**, 430–440.
22. Wu, W., Witherspoon, D.J., Fraser, A., Clark, E.A.S., Rogers, A., Stoddard, G.J., Manuck, T.A., Chen, K., Esplin, M.S., Smith, K.R. *et al.* (2015) The heritability of gestational age in a two-million member cohort: implications for spontaneous preterm birth. *Hum. Genet*., **134**, 803–808.
23. Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L.M., Juodakis, J., Miller, D.E., Litterman, N., Jiang, P.-P., Russell, L. *et al.* (2017) Genetic associations with gestational duration and spontaneous preterm birth. *N. Engl. J. Med*., **377**, 1156–1167.
24. Huusko, J.M., Karjalainen, M.K., Graham, B.E., Zhang, G., Farrow, E.G., Miller, N.A., Jacobsson, B., Eidem, H.R., Murray, J.C., Bedell, B. *et al.* (2018) Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLoS Genet*., **14**, e1007394.
25. Rappoport, N., Toung, J., Hadley, D., Wong, R.J., Fujioka, K., Reuter, J., Abbott, C.W., Oh, S., Hu, D., Eng, C. *et al.* (2018) A genome-wide association study identifies only two ancestry specific variants associated with spontaneous preterm birth. *Sci. Rep*., **8**, 226.
26. LaBella, A.L., Abraham, A., Pichkar, Y., Fong, S.L., Zhang, G., Muglia, L.J., Abbot, P., Rokas, A. and Capra, J.A. (2020) Accounting for diverse evolutionary forces reveals mosaic patterns of selection on human preterm birth loci. *Nat. Commun*., **11**, 3731.
27. Karn, M.N. and Penrose, L.S. (1951) Birth weight and gestation time in relation to maternal age, parity and infant survival. *Ann. Eugenics*, **16**, 147–164.
28. Llaurens, V., Whibley, A. and Joron, M. (2017) Genetic architecture and balancing selection: the life and death of differentiated variants. *Mol. Ecol*., **26**, 2430–2448.
29. Mumtaz, G., Nassar, A.H., Mahfoud, Z., El-Khamra, A., Al-Choueiri, N., Adra, A., Murray, J.C., Zalloua, P. and Yunis, K.A. (2010) Consanguinity: a risk factor for preterm birth at less than 33 weeks' gestation. *Am. J. Epidemiol*., **172**, 1424–1430.
30. Yengo, L., Wray, N.R. and Visscher, P.M. (2019) Extreme inbreeding in a European ancestry sample from the contemporary UK population. *Nat. Commun*., **10**, 3719.
31. Nalls, M.A., Simon-Sanchez, J., Gibbs, J.R., Paisan-Ruiz, C., Bras, J.T., Tanaka, T., Matarin, M., Scholz, S., Weitz, C., Harris, T.B. *et al.* (2009) Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. *PLoS Genet*., **5**, e1000415.
32. Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol*., **4**, e72.
33. Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M. *et al.* (2019) The GTEx consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903.
34. Halford, M.M., Armes, J., Buchert, M., Meskenaite, V., Grail, D., Hibbs, M.L., Wilks, A.F., Farlie, P.G., Newgreen, D.F., Hovens, C.M. *et al.* (2000) Ryk-deficient mice exhibit craniofacial defects associated with perturbed Eph receptor crosstalk. *Nat. Genet*., **25**, 414–418.
35. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
36. Persyn, E., Redon, R., Bellanger, L. and Dina, C. (2018) The impact of a fine-scale population stratification on rare variant association test results. *PLoS One*, **13**, e0207677.
37. McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J.E., Aguirre, M. and Rivas, M.A. (2019) Global biobank engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics*, **35**, 2495–2497.
38. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S. *et al.* (2018) Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367.
39. Rumer, K.K., Uyenishi, J., Hoffman, M.C., Fisher, B.M. and Winn, V.D. (2013) Siglec-6 expression is increased in placentas from pregnancies complicated by preterm preeclampsia. *Reprod. Sci. Thousand Oaks Calif*, **20**, 646–653.
40. Sammar, M., Siwetz, M., Meiri, H., Fleming, V., Altevogt, P. and Huppertz, B. (2017) Expression of CD24 and Siglec-10 in first trimester placenta: implications for immune tolerance at the fetal-maternal interface. *Histochem. Cell Biol*., **147**, 565–574.
41. Ko, C., Gieske, M.C., Al-Alem, L., Hahn, Y., Su, W., Gong, M.C., Iglarz, M. and Koo, Y. (2006) Endothelin-2 in ovarian follicle rupture. *Endocrinology*, **147**, 1770–1779.
42. Cacioppo, J.A., Lin, P.-C.P., Hannon, P.R., McDougle, D.R., Gal, A. and Ko, C. (2017) Granulosa cell endothelin-2 expression is fundamental for ovulatory follicle rupture. *Sci. Rep*., **7**, 817.
43. Szpiech, Z.A., Mak, A.C.Y., White, M.J., Hu, D., Eng, C., Burchard, E.G. and Hernandez, R.D. (2019) Ancestry-dependent enrichment of deleterious homozygotes in runs of homozygosity. *Am. J. Hum. Genet*., **105**, 747–762.
44. Ceballos, F.C., Hazelhurst, S. and Ramsay, M. (2019) Runs of homozygosity in sub-Saharan African populations provide insights into complex demographic histories. *Hum. Genet*., **138**, 1123–1142.
45. Naseri, A., Liu, X., Tang, K., Zhang, S. and Zhi, D. (2019) RaPID: ultra-fast, powerful, and accurate detection of segments identical by descent (IBD) in biobank-scale cohorts. *Genome Biol*., **20**, 143.
46. Rosenberg, N.A., Pemberton, T.J., Li, J.Z. and Belmont, J.W. (2013) Runs of homozygosity and parental relatedness. *Genet. Med. Off. J. Am. Coll. Med. Genet*., **15**, 753–754.
47. Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M. and Pe'er, I. (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res*., **19**, 318–326.

48. Browning, S.R. and Thompson, E.A. (2012) Detecting rare variant associations by identity-by-descent mapping in case-control studies. *Genetics*, **190**, 1521.
49. Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D. *et al.* (2016) Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*, **538**, 523–527.
50. Forde, N., Duffy, G.B., McGettigan, P.A., Browne, J.A., Mehta, J.P., Kelly, A.K., Mansouri-Attia, N., Sandra, O., Loftus, B.J., Crowe, M.A. *et al.* (2012) Evidence for an early endometrial response to pregnancy in cattle: both dependent upon and independent of interferon tau. *Physiol. Genomics*, **44**, 799–810.
51. Whitfield, J., Paglialunga, S., Smith, B.K., Miotto, P.M., Simnett, G., Robson, H.L., Jain, S.S., Herbst, E.A.F., Desjardins, E.M., Dyck, D.J. *et al.* (2017) Ablating the protein TBC1D1 impairs contraction-induced sarcolemmal glucose transporter 4 redistribution but not insulin-mediated responses in rats. *J. Biol. Chem.*, **292**, 16653–16664.
52. Kjøbsted, R., Roll, J.L.W., Jørgensen, N.O., Birk, J.B., Foretz, M., Viollet, B., Chadt, A., Al-Hasani, H. and Wojtaszewski, J.F.P. (2019) AMPK and TBC1D1 regulate muscle glucose uptake after, but not during, exercise and contraction. *Diabetes*, **68**, 1427–1440.
53. Rumer, K.K., Post, M.D., Larivee, R.S., Zink, M., Uyenishi, J., Kramer, A., Teoh, D., Bogart, K. and Winn, V.D. (2012) Siglec-6 is expressed in gestational trophoblastic disease and affects proliferation, apoptosis and invasion. *Endocr. Relat. Cancer*, **19**, 827–840.
54. Mayor-Lynn, K., Toloubeydokhti, T., Cruz, A.C. and Chegini, N. (2011) Expression profile of microRNAs and mRNAs in human placentas from pregnancies complicated by preeclampsia and preterm labor. *Reprod. Sci. Thousand Oaks Calif*, **18**, 46–56.
55. Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A.K., Handal, M., Haugen, M., Høiseth, G., Knudsen, G.P. *et al.* (2016) Cohort profile update: the Norwegian mother and child cohort study (MoBa). *Int. J. Epidemiol.*, **45**, 382–388.
56. Paltiel, L., Anita, H., Skjerden, T., Harbak, K., Bækken, S., Kristin, S.N., Knudsen, G.P. and Magnus, P. (2014) The biobank of the Norwegian mother and child cohort study—present status. *Nor. Epidemiol.*, **24**, 1–2.
57. Helgeland, Ø., Vaudel, M., Juliusson, P.B., Lingaas Holmen, O., Juodakis, J., Bacelis, J., Jacobsson, B., Lindekleiv, H., Hveem, K., Lie, R.T. *et al.* (2019) Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nat. Commun.*, **10**, 4448.
58. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. and Chen, W.-M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics*, **26**, 2867–2873.
59. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.
60. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R. *et al.* (2016) Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.*, **48**, 1443–1448.
61. Thompson, E.A. (2013) Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*, **194**, 301–326.
62. Kleinbaum, D.G. and Klein, M. (2012) Survival Analysis: A Self-Learning Text. In *Statistics for Biology and Health*, 3rd edn. Springer-Verlag, New York, NY.
63. Juodakis, J., Bacelis, J., Zhang, G., Muglia, L.J. and Jacobsson, B. (2017) Time-variant genetic effects as a cause for preterm birth: insights from a population of maternal cousins in Sweden. *G3 Bethesda Md*, **7**, 1349–1356.
64. Köster, J. and Rahmann, S. (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **28**, 2520–2522.
65. Delaneau, O., Zagury, J.-F. and Marchini, J. (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods*, **10**, 5–6.
66. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
67. Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A. and Li, J.Z. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.*, **91**, 275–292.
68. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC table browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.
69. McKusick, V.A. (2007) Mendelian inheritance in man and its online version, OMIM. *Am. J. Hum. Genet.*, **80**, 588–604.
70. Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P. *et al.* (2018) Ensembl variation resources. *Database J. Biol. Databases Curation*, **2018**, bay119.
71. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2019) Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210.