

**Supplemental Data**

**Identification of Common Genetic Variants**

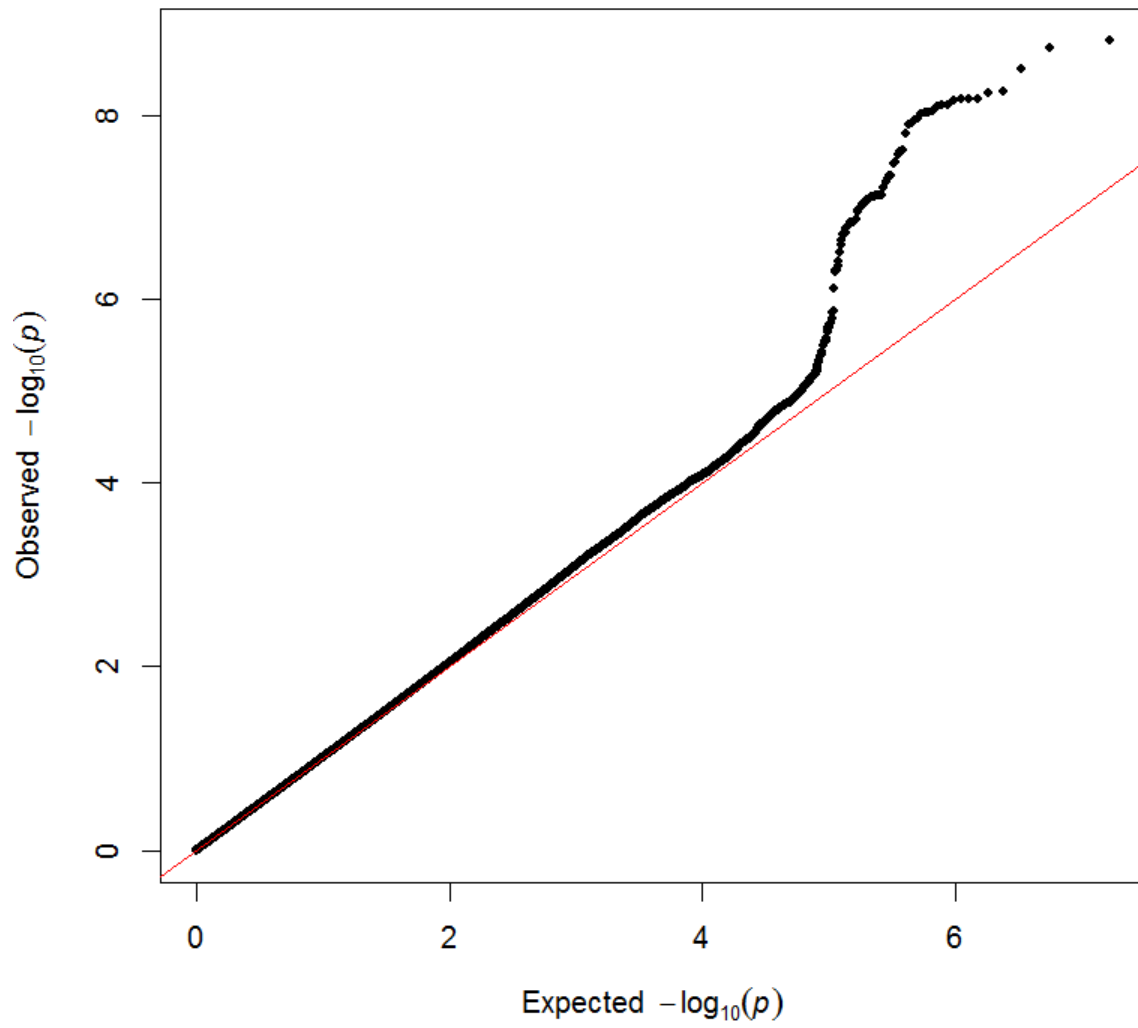
**Influencing Spontaneous Dizygotic Twinning**

**and Female Fertility**

**Hamdi Mbarek, Stacy Steinberg, Dale R. Nyholt, Scott D. Gordon, Michael B. Miller, Allan F. McRae, Jouke Jan Hottenga, Felix R. Day, Gonneke Willemsen, Eco J. de Geus, Gareth E. Davies, Hilary C. Martin, Brenda W. Penninx, Rick Jansen, Kerrie McAloney, Jacqueline M. Vink, Jaakko Kaprio, Robert Plomin, Tim D. Spector, Patrik K. Magnusson, Bruno Reversade, R. Alan Harris, Kjersti Aagaard, Ragnar P. Kristjansson, Isleifur Olafsson, Gudmundur Ingi Eyjolfsson, Olof Sigurdardottir, William G. Iacono, Cornelis B. Lambalk, Grant W. Montgomery, Matt McGue, Ken K. Ong, John R.B. Perry, Nicholas G. Martin, Hreinn Stefánsson, Kari Stefánsson, and Dorret I. Boomsma**

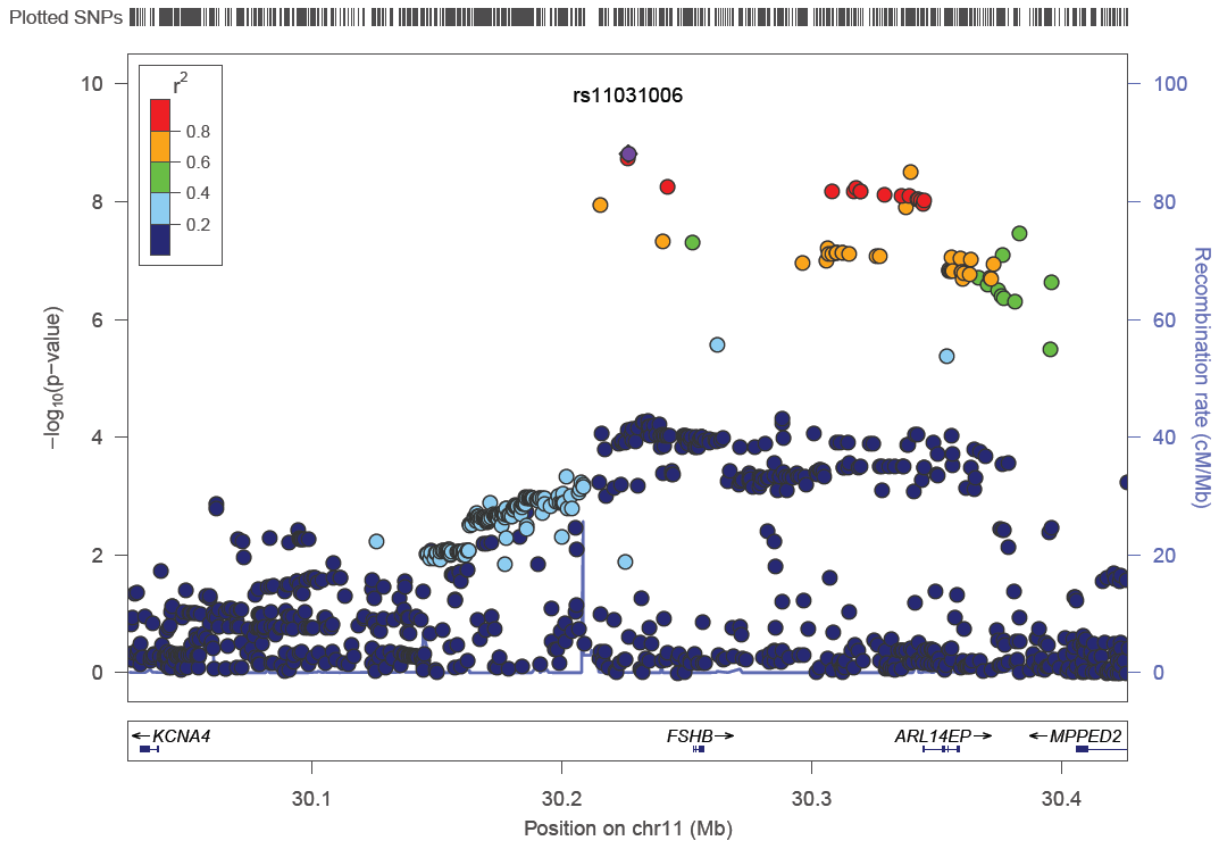
SUPPLEMENTAL DATA

**Figure S1.** Quantile-Quantile (Q-Q) plot of observed vs. expected  $P$  values of the association results in the meta-analysis (Lambda =1.01)

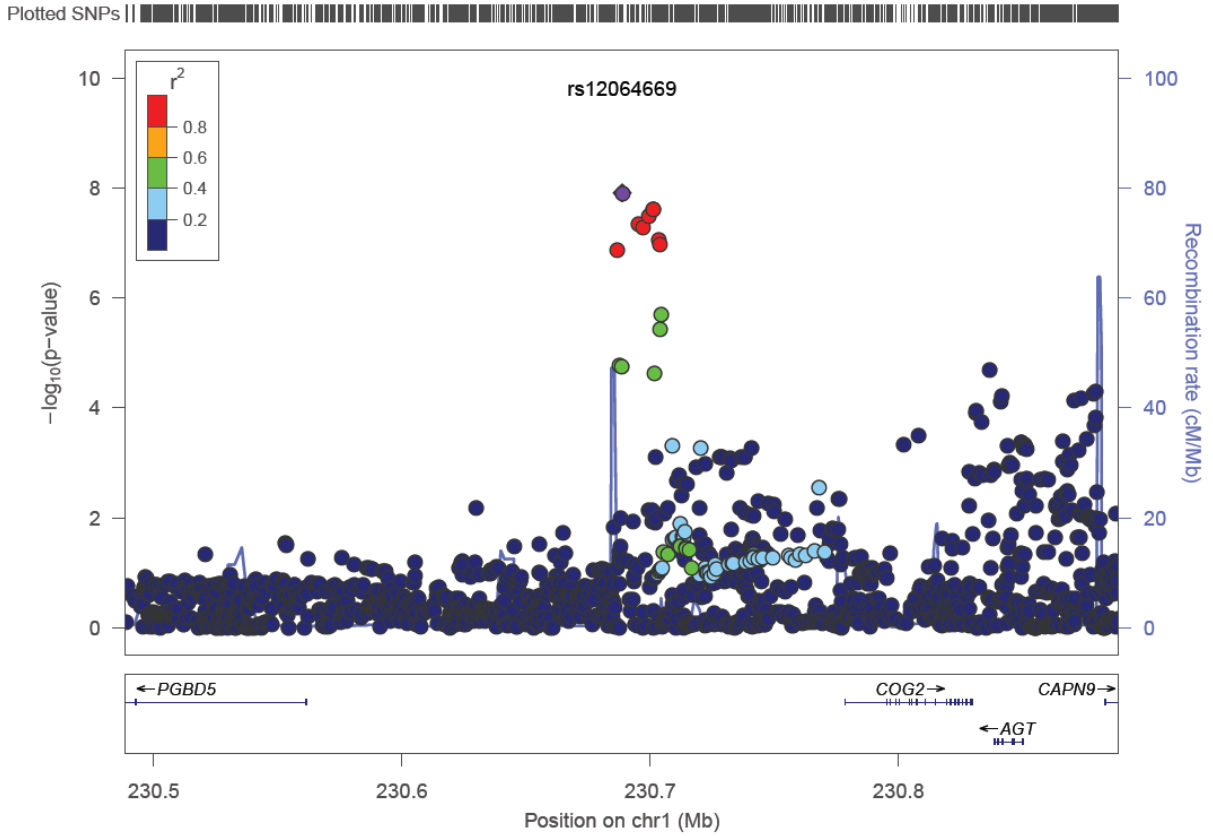


**Figure S2 A, B, C.** Regional association plots corresponding to the meta-analysis results for the top 3 loci. For each plot, the  $-\log_{10} P$  values (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). The loci are at *11p13* (b), *1q42.13* (c), *15q22.33* (d). The estimated recombination rates from the 1000 Genomes Project March 2012 release are shown as blue lines, and the genomic locations of genes within the regions of interest in the NCBI Build 37 human assembly are shown as arrows. SNP color represents LD with the most highly associated SNP at each locus. The figures were created with LocusZoom (<http://csg.sph.umich.edu/locuszoom/>). Mb, megabases.

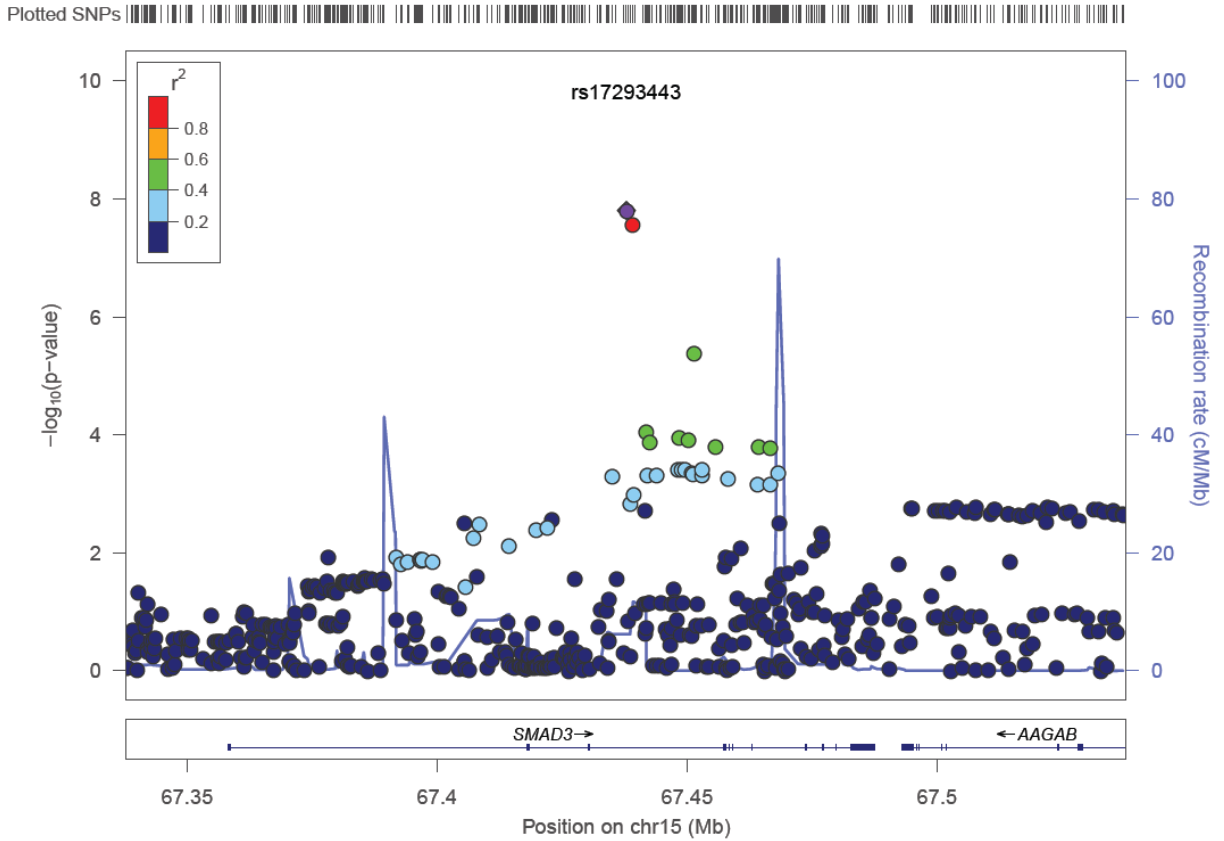
**Figure S2A.** Regional association plot of the region *11p13* (rs11031006)



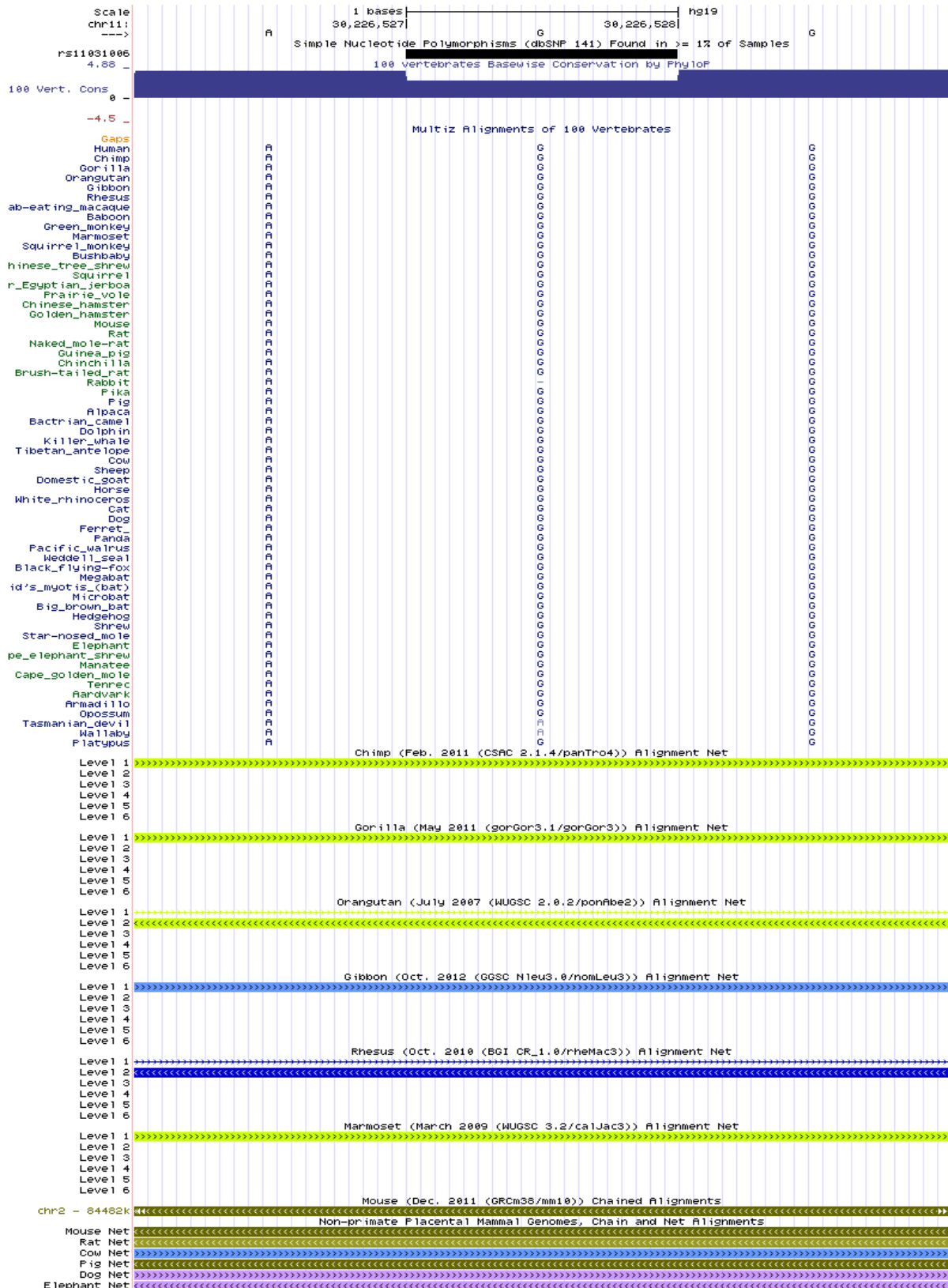
**Figure S2B.** Regional association plot of the region *1q42.13* (rs12064669)



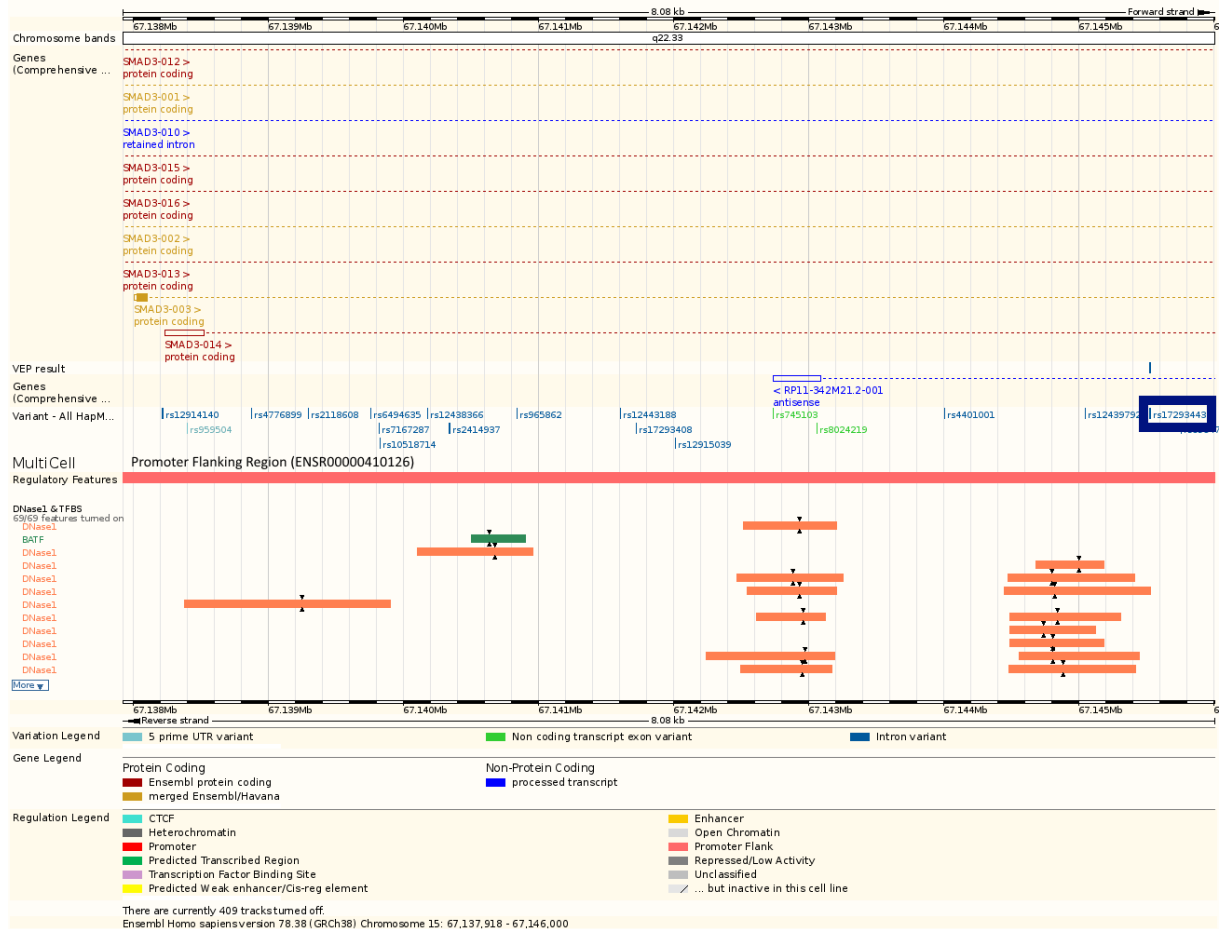
**Figure S2C.** Regional association plot of the region *15q22.33* (rs17293443)



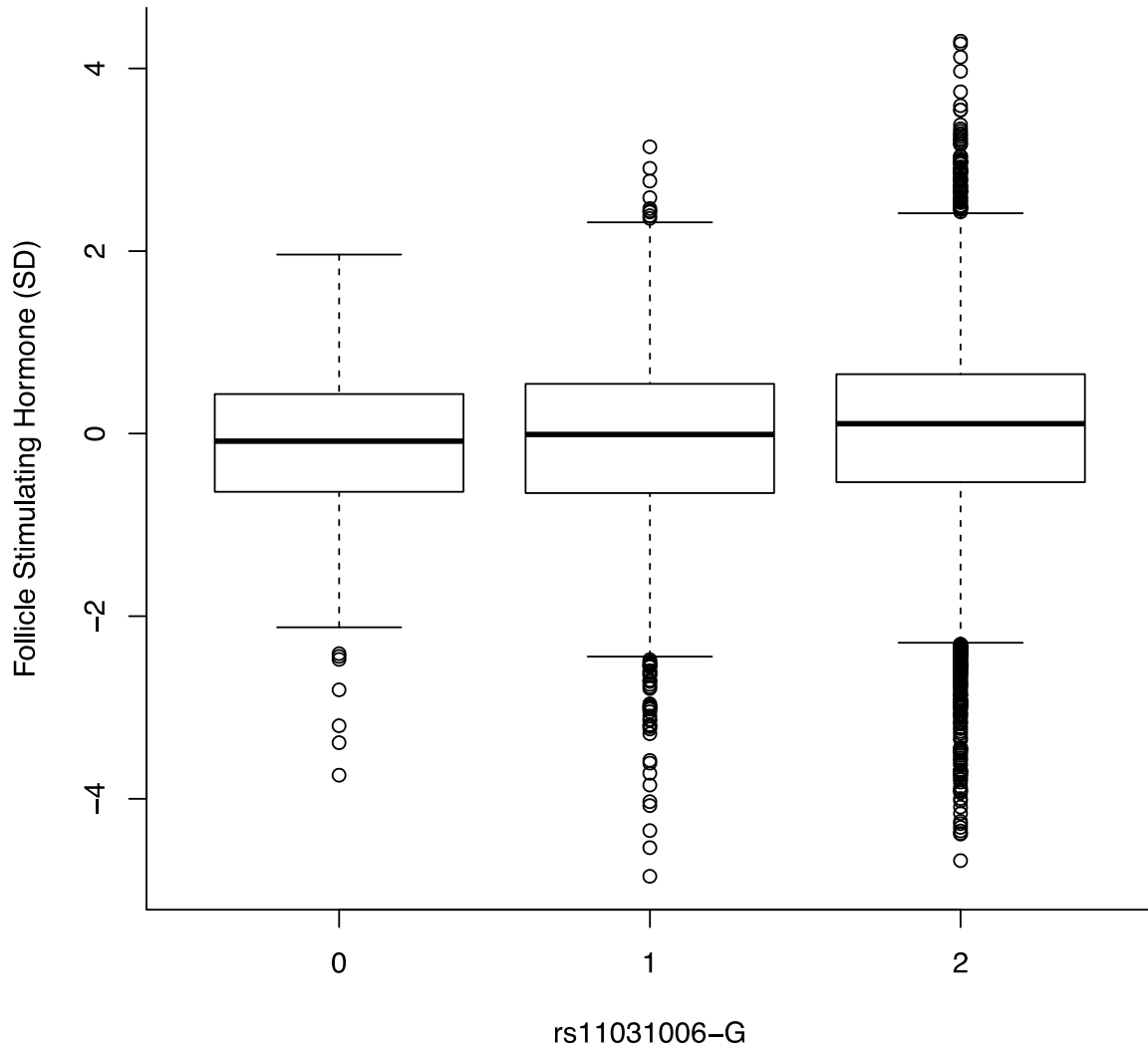
**Figure S3.** UCSC Browser view of rs11031006 sequence conservation. Multiz alignments and Alignment Nets to the NCBI GRCh37/UCSC hg19 human assembly show a high degree of conservation at the rs11031006 locus across all sequenced mammals.



**Figure S4.** Ensembl Browser view of rs17293443 regulatory features. Based on Ensembl regulatory annotations of GRCh38, rs17293443 lies within a Promoter Flanking Region (ENSR00000410126) and is contained in a DNase I hypersensitive site suggesting open chromatin<sup>2</sup>.

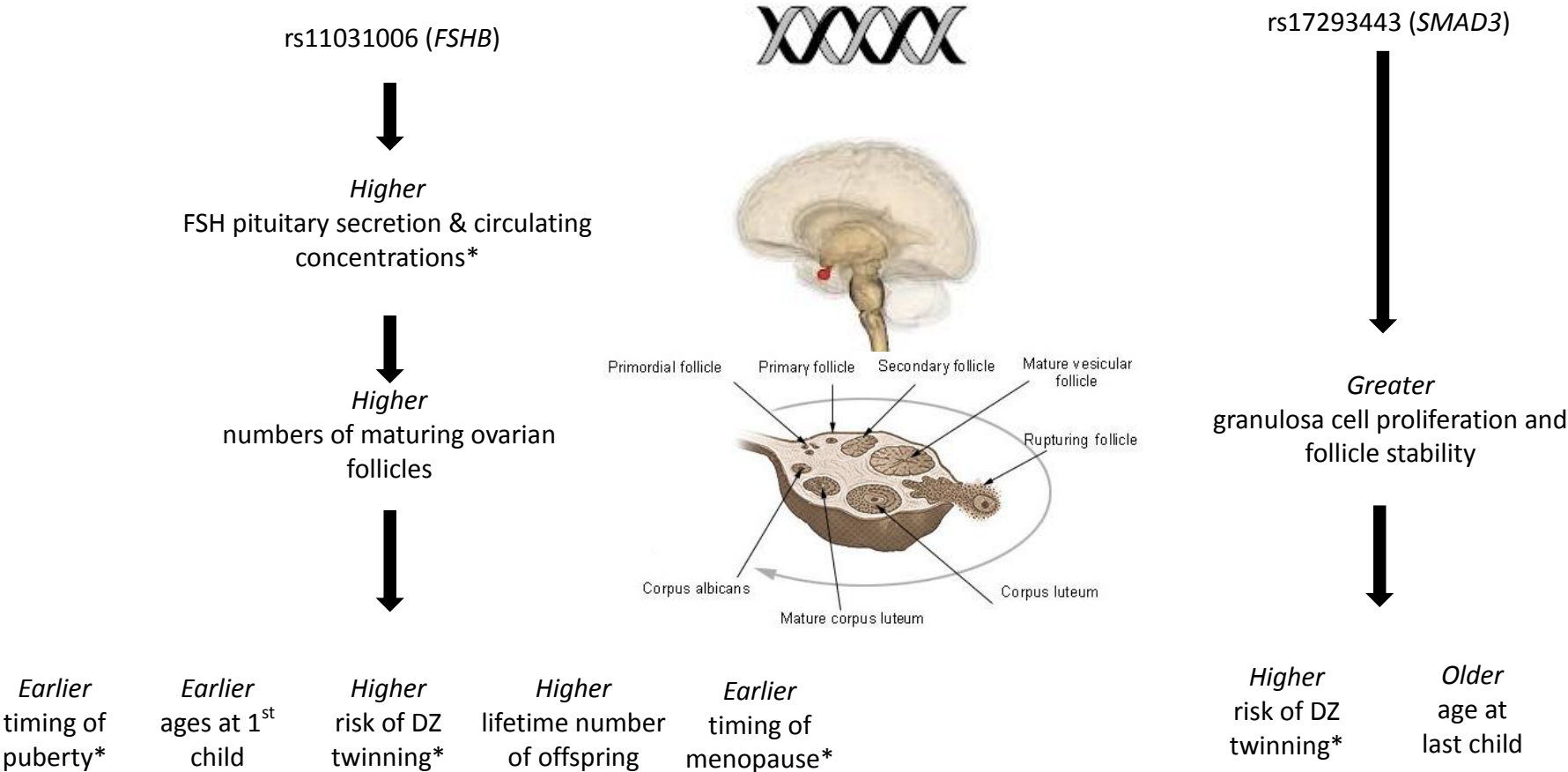


**Figure S5.** Boxplot of FSH level as a function of the number of rs11031006-G alleles based on data from 11,003 genotyped Icelanders. FSH level was significantly associated with rs11031006-G count ( $p=2.3\times 10^{-10}$ ). SD (standard deviation)



**Figure S6.** Schematic depiction of the effects of *FSHB* and *SMAD3* variants on DZ twinning and other reproductive outcomes.

\*indicates observed or reported genome-wide significant associations.





**Table S1.** Information on genotyping methods, quality control of SNPs, imputation, and statistical analysis for genome-wide association study cohort

Study	GENOTYPING						IMPUTATION			ASSOCIATION ANALYSIS	
	Platform	Genotype calling algorithm	Inclusion Criteria			SNPs which met QC criteria	Imputation software	Inclusion Criteria		SNPs in meta-analysis	Analyses software
			MAF	Call rate	P for HWE			MA F	Imputation quality		
<b>QIMR</b>	Illumina 317K; HumanCNV370; HumanCNV370-Quad; Human610-Quad; Human660W-Quad	GenomeStudio	≥1%	≥95%	≥10 <sup>-6</sup>	273,158	MACH  Minimac	≥1%  ≥0.3		8,396,183	PLINK
<b>NTR</b>	Affymetrix Perlegen5.0; Illumina660; Illumina Omni Express 1M; Affymetrix 6.0	Beadstudio,  Birdseed	≥1%	≥95%	≥10 <sup>-5</sup>	356,057-  1,135,535	MACH  Minimac	≥1%  ≥0.3		8,084,689	PLINK
<b>MCTFR</b>	Illumina Human660W-Quad	GenomeStudio	≥1%	≥99%	≥10 <sup>-7</sup>	527,829	Beagles  Minimac	≥1%  ≥0.3		7,903,594	PLINK

**Table S2.** Association results of the top SNPs having  $p < 5 \times 10^{-8}$  in the discovery, sorted by chromosome and position

CHR	SNP	BP <sup>a</sup>	Gene	Risk allele	RAF	OR	SE	P	P heterogeneity <sup>b</sup>
1	rs12064669	230688643	90kb 5' of COG2	C	0.10	1.43	0.06	$1.23 \times 10^{-8}$	0.81
1	rs11122535	230695062	83kb 5' of COG2	C	0.10	1.41	0.06	$4.51 \times 10^{-8}$	0.89
1	rs12063832	230699392	79kb 5' of COG2	G	0.09	1.41	0.06	$3.25 \times 10^{-8}$	0.81
1	rs12068056	230701078	77kb 5' of COG2	C	0.10	1.41	0.06	$2.38 \times 10^{-8}$	0.81
1	rs12060563	230701288	77kb 5' of COG2	A	0.10	1.41	0.06	$2.44 \times 10^{-8}$	0.82
11	rs11031002	30215261	37kb 5' of FSHB	T	0.87	1.40	0.06	$1.13 \times 10^{-8}$	0.16
11	rs11031005	30226356	26kb 5' of FSHB	T	0.85	1.41	0.06	$1.81 \times 10^{-9}$	0.18
11	rs11031006	30226528	26kb 5' of FSHB	G	0.85	1.41	0.06	$1.54 \times 10^{-9}$	0.19
11	rs11031010	30240178	12kb 5' of FSHB	C	0.85	1.35	0.05	$4.58 \times 10^{-8}$	0.13
11	rs74485684	30242287	10kb 5' of FSHB	T	0.83	1.36	0.05	$5.51 \times 10^{-9}$	0.15
11	rs10835638	30252352	209bp 5' of FSHB	G	0.85	1.35	0.06	$4.85 \times 10^{-8}$	0.12
11	rs75525300	30307989	37kb 5' of ARL14EP	A	0.83	1.36	0.05	$6.57 \times 10^{-9}$	0.15
11	rs12278989	30316822	28kb 5' of ARL14EP	A	0.83	1.35	0.05	$6.55 \times 10^{-9}$	0.15
11	rs11031040	30317733	27kb 5' of ARL14EP	T	0.83	1.36	0.05	$5.71 \times 10^{-9}$	0.15
11	rs12271187	30319259	25kb 5' of ARL14EP	G	0.83	1.35	0.05	$6.71 \times 10^{-9}$	0.15
11	rs12271300	30319433	25kb 5' of ARL14EP	G	0.83	1.35	0.05	$6.69 \times 10^{-9}$	0.15
11	rs10835649	30329175	15kb 5' of ARL14EP	A	0.83	1.35	0.05	$7.51 \times 10^{-9}$	0.15
11	rs4071563	30336002	8.6kb 5' of ARL14EP	G	0.83	1.35	0.05	$7.76 \times 10^{-9}$	0.15
11	rs11031047	30337523	7.1kb 5' of ARL14EP	C	0.83	1.35	0.05	$1.21 \times 10^{-8}$	0.18

11	rs7947350	30338842	5.8kb 5' of ARL14EP	A	0.83	1.35	0.05	$8.00 \times 10^{-9}$	0.15
11	rs7929660	30339461	5.2kb 5' of ARL14EP	A	0.83	1.37	0.05	$3.05 \times 10^{-9}$	0.16
11	rs12363824	30342150	2.5kb 5' of ARL14EP	G	0.84	1.37	0.05	$8.99 \times 10^{-9}$	0.15
11	rs6484479	30342495	2.2kb 5' of ARL14EP	G	0.83	1.35	0.05	$9.12 \times 10^{-9}$	0.14
11	rs3858429	30343757	890bp 5' of ARL14EP	C	0.83	1.35	0.05	$9.50 \times 10^{-9}$	0.14
11	rs4071558	30344591	56bp 5' of ARL14EP	C	0.83	1.35	0.05	$1.08 \times 10^{-8}$	0.13
11	rs4071559	30344725	ARL14EP	C	0.84	1.37	0.05	$9.31 \times 10^{-9}$	0.15
11	rs12294104	30382899	23kb 3' of MPPED2	C	0.84	1.37	0.06	$3.39 \times 10^{-8}$	0.46
15	rs17293443	67437863	SMAD3	C	0.24	1.27	0.04	$1.57 \times 10^{-8}$	0.52
15	rs72743459	67439023	SMAD3	T	0.23	1.27	0.04	$2.70 \times 10^{-8}$	0.41

<sup>a</sup>Position according to human reference NCBI37/hg19, <sup>b</sup>Cochran's *Q* test heterogeneity *P* value

**Table S3.** Twenty one SNPs with  $p < 5 \times 10^{-8}$  spanning 167.638 kb around *FSHB* adjusted for the top SNP rs11031006 (30226528) in the 5' of *FSHB*

SNP	Position	Meta-analysis		Meta-analysis adjusted for rs11031006		LD with rs11031006
		OR (SE)	P	OR (SE)	P	Pairwise $r^2$
rs11031005	30226356	1.41 (0.06)	$1.81 \times 10^{-9}$	-	-	1
rs7929660	30339461	1.37 (0.05)	$3.05 \times 10^{-9}$	1.07 (0.12)	0.56	0.62
rs7485684	30242287	1.36 (0.05)	$5.51 \times 10^{-9}$	1.07 (0.14)	0.63	0.81
rs11031040	30317733	1.36 (0.05)	$5.71 \times 10^{-9}$	1.07 (0.12)	0.58	0.63
rs12278989	30316822	1.35 (0.05)	$6.55 \times 10^{-9}$	1.06 (0.12)	0.62	0.63
rs75525300	30307989	1.36 (0.05)	$6.57 \times 10^{-9}$	1.07 (0.13)	0.57	0.63
rs12271300	30319433	1.35 (0.05)	$6.69 \times 10^{-9}$	1.07 (0.12)	0.61	0.63
rs12271187	30319259	1.35 (0.05)	$6.71 \times 10^{-9}$	1.06 (0.12)	0.12	0.63
rs10835649	30329175	1.35 (0.05)	$7.51 \times 10^{-9}$	1.06 (0.12)	0.63	0.63
rs4071563	30336002	1.35 (0.05)	$7.76 \times 10^{-9}$	1.06 (0.12)	0.64	0.63
rs7947350	30338842	1.35 (0.05)	$8.00 \times 10^{-9}$	1.06 (0.12)	0.66	0.63
rs12363824	30342150	1.37 (0.05)	$8.99 \times 10^{-9}$	1 (0.12)	0.98	0.63
rs6484479	30342495	1.35 (0.05)	$9.12 \times 10^{-9}$	1.05 (0.12)	0.70	0.63
rs4071559	30344725	1.37 (0.05)	$9.31 \times 10^{-9}$	1.01 (0.12)	0.96	0.63
rs3858429	30343757	1.35 (0.05)	$9.50 \times 10^{-9}$	1.05 (0.12)	0.71	0.63
rs4071558	30344591	1.35 (0.05)	$1.08 \times 10^{-8}$	1.04 (0.12)	0.77	0.63
rs11031002	30215261	1.40 (0.06)	$1.13 \times 10^{-8}$	1.12 (0.19)	0.54	0.79
rs11031047	30337523	1.35 (0.05)	$1.21 \times 10^{-8}$	1.04 (0.12)	0.76	0.61

---

rs12294104	30382899	1.37 (0.05)	$3.39 \times 10^{-8}$	1.10 (0.09)	0.27	0.37
rs11031010	30240178	1.35 (0.05)	$4.58 \times 10^{-8}$	1.04 (0.11)	0.71	0.64
rs10835638	30252352	1.35 (0.06)	$4.85 \times 10^{-8}$	1.03 (0.10)	0.79	0.62

---

**Table S4.** SNP with  $p < 5 \times 10^{-8}$  in *SMAD3* adjusted for the top SNP rs17293443 (67437863 bp)

SNP	Position	Meta-analysis		Meta-analysis adjusted for rs17293443		LD with rs17293443
		OR (SE)	P	OR (SE)	P	Pairwise $r^2$
rs72743459	67439023	1.27 (0.04)	$2.70 \times 10^{-8}$	1.04 (0.19)	0.83	0.86

**Table S5.** Four SNPs with  $p < 5 \times 10^{-8}$  spanning 12.645 kb at *Iq42.13* adjusted for the top SNP rs12064669 (230688643 bp)

SNP	Position	Meta-analysis		Meta-analysis adjusted for rs12064669		LD with rs12064669
		OR (SE)	P	OR (SE)	P	Pairwise $r^2$
rs12068056	230701078	1.41 (0.06)	$2.38 \times 10^{-8}$	1.09 (0.31)	0.79	0.97
rs12060563	230701288	1.41 (0.06)	$2.44 \times 10^{-8}$	1.10 (0.31)	0.77	0.97
rs12063832	230699392	1.41 (0.06)	$3.25 \times 10^{-8}$	1.17 (0.34)	0.65	0.97
rs11122535	230695062	1.41 (0.06)	$4.51 \times 10^{-8}$	-	-	1

**Table S6.** Gene-based association results of the 63 candidate genes from Harris *et al.* in our MODZT GWAS

Gene	Name	uncorrected P-value
<b>BMP family and related</b>		
<i>BMP1</i>	Bone morphogenetic protein 1	0.09
<i>BMP2</i>	Bone morphogenetic protein 2	0.92
<i>BMP3</i>	Bone morphogenetic protein 3	0.34
<i>BMP4</i>	Bone morphogenetic protein 4	0.86
<i>BMP5</i>	Bone morphogenetic protein 5	0.45
<i>BMP6</i>	Bone morphogenetic protein 6	0.48
<i>BMP7</i>	Bone morphogenetic protein 7	0.98
<i>BMP8A</i>	Bone morphogenetic protein 8A	0.09
<i>BMP8B</i>	Bone morphogenetic protein 8B	0.85
<i>BMP10</i>	Bone morphogenetic protein 10	0.98
<i>BMP15</i>	Bone morphogenetic protein 15	NA
<i>BMPRIA</i>	Bone morphogenetic protein receptor, type IA	0.9
<i>BMPR1B</i>	Bone morphogenetic protein receptor, type IB	0.01
<i>BMPR2</i>	Bone morphogenetic protein receptor, type II	0.04
<i>BMPER</i>	BMP binding endothelial regulator	0.99
<i>BMP2K</i>	BMP2 inducible kinase	0.97
<i>BAMBI</i>	BMP and activin membrane-bound inhibitor homolog	0.06
<i>KCP</i>	Kielin/chordin like protein (cystein rich BMP regulator 2)	0.54
<b>GDF family</b>		
<i>GDF1</i>	Growth differentiation factor 1	0.16
<i>GDF2</i>	Growth differentiation factor 2	1
<i>GDF3</i>	Growth differentiation factor 3	0.89
<i>GDF5</i>	Growth differentiation factor 5	0.75
<i>GDF6</i>	Growth differentiation factor 6	0.33
<i>GDF7</i>	Growth differentiation factor 7	0.39
<i>GDF9</i>	Growth differentiation factor 9	0.07
<i>GDF10</i>	Growth differentiation factor 10	1
<i>GDF11</i>	Growth differentiation factor 11	0.4

<i>GDF5</i>	Growth differentiation factor 15	0.63
<b>IGF family</b>		
<i>IGF1</i>	Insulin-like growth factor 1	9.36×10 <sup>-3</sup>
<i>IGF1R</i>	Insulin-like growth factor 1 receptor	0.89
<i>IGF2</i>	Insulin-like growth factor 2	0.96
<i>IGF2BP1</i>	Insulin-like growth factor 2 mRNA binding protein 1	0.07
<i>IGF2BP2</i>	Insulin-like growth factor 2 mRNA binding protein 2	0.87
<i>IGF2BP3</i>	Insulin-like growth factor 2 mRNA binding protein 3	0.65
<i>IGF2R</i>	Insulin-like growth factor 2 receptor	0.08
<i>IGFALS</i>	Insulin-like growth factor binding protein, acid labile subunit	0.64
<i>IGFBP1</i>	Insulin-like growth factor binding protein 1	0.97
<i>IGFBP2</i>	Insulin-like growth factor binding protein 2	1
<i>IGFBP3</i>	Insulin-like growth factor binding protein 3	0.98
<i>IGFBP4</i>	Insulin-like growth factor binding protein 4	0.24
<i>IGFBP5</i>	Insulin-like growth factor binding protein 5	1
<i>IGFBP6</i>	Insulin-like growth factor binding protein 6	0.89
<i>IGFBP7</i>	Insulin-like growth factor binding protein 7	0.73
<i>IGFBPL1</i>	Insulin-like growth factor binding protein-like 1	0.97
<i>IGFL1</i>	IGF-like family member 1	0.95
<i>IGFL2</i>	IGF-like family member 2	0.92
<i>IGFL3</i>	IGF-like family member 3	0.83
<i>IGFL4</i>	IGF-like family member 4	0.95
<b>Follistatin and related</b>		
FST	Follistatin	0.47
FSTL1	Follistatin-like 1	0.99
FSTL3	Follistatin-like 3 (secreted glycoprotein)	0.95
FSTL4	Follistatin-like 4	0.47
FSTL5	Follistatin-like 5	0.67
WFIKKN1	WAP, follistatin/kazal, immunoglobulin, kunitz	0.11
<b>Hormones synthesis and secretion</b>		
<i>ACVR2B</i>	Activin A receptor type IIB	1
<i>FSHB</i>	Follicle-stimulating hormone beta polypeptide	1.74×10 <sup>-11</sup>
<i>FSHB</i>	Follicle-stimulating hormone receptor	5.72×10 <sup>-3</sup>

---



<i>FSHR</i>	Inhibin, alpha	0.98
<i>INHA</i>	Inhibin, beta A	0.1
<i>INHBA</i>	Inhibin, beta B	1
<i>INHBB</i>	Luteinizing hormone/choriogonadotropin receptor	0.07
<i>LHCGR</i>	Luteinizing hormone beta polypeptide	0.18
<i>LHB</i>	Growth hormone secretagogue receptor	0.14
<i>GHSR</i>		

---

**Table S9.** Association of MODZT polygenic risk scores with DZ twinning in Iceland (595 mothers of DZ twins; 41,716 mothers without twins)

$p^*$	OR	Nagelkerke $R^2$ (%)	P-value
0.0001	1.12	0.13	0.006
0.0003	1.15	0.19	0.001
0.001	1.13	0.14	0.004
0.003	1.11	0.11	0.01
0.01	1.11	0.10	0.01
0.03	1.10	0.10	0.02
0.1	1.10	0.10	0.02
0.3	1.10	0.10	0.02
1	1.10	0.10	0.02

\*  $p$ , fraction of causal markers used for LDpred run

**Table S10.** Association of MODZT polygenic risk scores with fertility outcomes in 41,946 women (LDpred run using  $p=0.0003$ )

Phenotype	Beta*	P-value
Has children (yes/no)	0.055	0.01
Number of children	0.0072	0.03
Age at first child	-0.015	0.02
Age at last child	-0.0066	0.20
Average birth interval	-0.0098	0.46

\*Beta is from logistic regression (has children), Poisson log-linear regression (number of children), or linear regression (age at first child, age at last child, average birth interval). Prior to analysis, age at first child, age at last child and average birth interval were transformed to have a standard normal distribution. Polygenic risk scores were based on effect sizes derived from an LDpred run using  $p=0.0003$  and were standardized to have a mean of 0 and a standard deviation of 1.

## References

1. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* 12, 996-1006.
2. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res* 42, D749-755.

## Acknowledgments

Support for the Netherlands Twin Register was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organization for Health Research and Development (ZonMW) grants, 904-61-193,480-04-004, 400-05-717, Addiction-31160008, 911-09-032, Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007); Royal Netherlands Academy of Science Professor Award (PAH/6635) to DIB; European Research Council (ERC-230374 and ERC-284167); Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH R01 HD042157-01A1). Part of the genotyping was funded by the Genetic Association Information Network (GAIN) of the

Foundation for the National Institutes of Health and Grand Opportunity grants 1RC2 MH089951). We acknowledge support from VU Amsterdam and the Institute for Health and Care Research (EMGO+). The Berghofer Medical Research Institute (QIMR) study was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia (241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496610, 496739, 552485, 552498, 1050208, 1075175). Dale R. Nyholt was supported by the Australian Research Council (ARC) Future Fellowship (FT0991022), NHMRC Research Fellowship (APP0613674) Schemes and by the Visiting Professors Programme (VPP) of the Royal Netherlands Academy of Arts and Sciences (KNAW). Allan F. McRae was supported by an NRMRC Career Development Fellowship (APP1083656). Grant W. Montgomery was supported by NIH grant (HD042157, a collaborative study of the genetics of DZ twinning) and NHMRC Fellowship (GNT1078399). The Minnesota Center for Twin and Family Research (MCTFR) was supported in part by USPHS Grants from the National Institute on Alcohol Abuse and Alcoholism (AA09367 and AA11886), and the National Institute on Drug Abuse (DA05147, DA13240, and DA024417).

We would like to thank also 23andMe's consented research participants for contributing data on age at menarche for the FSHB gene locus and the Twinning Gwas Consortium (TGC). Co-authors from: Finland (Anu Loukola, Juho Wedenoja, Emmi Tikkanen, Beenish Qaiser), Sweden (Nancy Pedersen, Andrea Ganna), United kingdom King's College London (Department of Twin Research & Genetic Epidemiology: Pirro Hysi, Massimo Mangino), Institute of Psychiatry, Psychology & Neuroscience, Medical Research Council Social, Genetic and Developmental Psychiatry Centre (Eva Krapohl, Andrew McMillan).