Genome-wide meta-analysis of common variant differences between men and women

Vesna Boraska[1,2,*], Ana Jerončić[3], Vincenza Colonna[1,5], Lorraine Southam[1], Dale R. Nyholt[6], Nigel William Rayner[1,7,8], John R.B. Perry[7,9,10], Daniela Toniolo[11], Eva Albrecht[12], Wei Ang[13], Stefania Bandinelli[14], Maja Barbalic[15], Inês Barroso[1,16], Jacques S. Beckmann[18,19], Reiner Biffar[20], Dorret Boomsma[22], Harry Campbell[23], Tanguy Corre[11], Jeanette Erdmann[24,25], Tõnu Esko[27,28,29], Krista Fischer[27], Nora Franceschini[30], Timothy M. Frayling[9], Giorgia Girotto[31], Juan R. Gonzalez[32,33], Tamara B. Harris[34], Andrew C. Heath[35], Iris M. Heid[36,37], Wolfgang Hoffmann[21], Albert Hofman[38,40], Momoko Horikoshi[7,8], Jing Hua Zhao[17], Anne U. Jackson[41,42], Jouke-Jan Hottenga[22], Antti Jula[43], Mika Kähönen[44,46], Kay-Tee Khaw[47], Lambertus A. Kiemeney[48], Norman Klopp[52], Zoltán Kutalik[18,54], Vasiliki Lagou[7,8], Lenore J. Launer[34], Terho Lehtimäki[45,46], Mathieu Lemire[55], Marja-Liisa Lokki[56], Christina Loley[26], Jian'an Luan[17], Massimo Mangino[10], Irene Mateo Leach[58], Sarah E. Medland[6], Evelin Mihailov[28], Grant W. Montgomery[6], Gerjan Navis[59], John Newnham[13], Markku S. Nieminen[61], Aarno Palotie[1,57,62,63], Kalliope Panoutsopoulou[1], Annette Peters[53], Nicola Pirastu[31], Ozren Polašek[4], Karola Rehnström[1,57], Samuli Ripatti[57], Graham R.S. Ritchie[1,64], Fernando Rivadeneira[38,39,40], Antonietta Robino[31], Nilesh J. Samani[65], So-Youn Shin[1], Juha Sinisalo[61], Johannes H. Smit[66], Nicole Soranzo[1,10], Lisette Stolk[39,40], Dorine W. Swinkels[49], Toshiko Tanaka[68], Alexander Teumer[69], Anke Tönjes[70,71], Michela Traglia[11], Jaakko Tuomilehto[72,73,74,75], Armand Valsesia[18,54,76], Wiek H. van Gilst[58], Joyce B.J. van Meurs[39,40], Albert Vernon Smith[77,78], Jorma Viikari[79,81], Jacqueline M. Vink[22], Gerard Waeber[82], Nicole M. Warrington[13], Elisabeth Widen[57], Gonneke Willemsen[22], Alan F. Wright[83], Brent W. Zanke[84], Lina Zgaga[23,85], Wellcome Trust Case Control Consortium (WTCCC), Michael Boehnke[41,42], Adamo Pio d'Adamo[31], Eco de Geus[22], Ellen W. Demerath[86], Martin den Heijer[50,51], Johan G. Eriksson[87,89,90,91,92], Luigi Ferrucci[93], Christian Gieger[12], Vilmundur Gudnason[77,78], Caroline Hayward[83], Christian Hengstenberg[94], Thomas J. Hudson[55,95], Marjo-Riitta Järvelin[96,97,98,99], Manolis Kogevinas[32,33,100,101], Ruth J.F. Loos[17], Nicholas G. Martin[6], Andres Metspalu[27,28,29], Craig E. Pennell[13], Brenda W. Penninx[60,67,102], Markus Perola[57,88], Olli Raitakari[80,103], Veikko Salomaa[87], Stefan Schreiber[104], Heribert Schunkert[24,25], Tim D. Spector[10], Michael Stumvoll[70,71], André G. Uitterlinden[38,39,40], Sheila Ulivi[105], Pim van der Harst[58], Peter Vollenweider[82], Henry Völzke[21], Nicholas J. Wareham[17], H.-Erich Wichmann[37,106,107], James F. Wilson[23], Igor Rudan[2,23], Yali Xue[1] and Eleftheria Zeggini[1]

---

*To whom correspondence should be addressed at: Wellcome Trust Sanger Institute, The Morgan Building, Wellcome Trust Genome Campus, CB10 1HH, Hinxton, Cambridge, UK. Tel: +44 1223 834 244; Email: vboraska@mefst.hr; vb2@sanger.ac.uk

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, [2]Department of Medical Biology, [3]Department for Research in Biomedicine and Health and [4]Department of Public Health, University of Split School of Medicine, Split, Croatia, [5]Institute of Genetics and Biophysics 'A. Buzzati-Traverso', National Research Council (CNR), Naples, Italy, [6]Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia, [7]Wellcome Trust Centre for Human Genetics and [8]Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Oxford, UK, [9]Genetics of Complex Traits, Peninsula Medical School, University of Exeter, UK, [10]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK, [11]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy, [12]Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, [13]School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia, [14]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy, [15]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA, [16]University of Cambridge Metabolic Research Labs and [17]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK, [18]Department of Medical Genetics, University of Lausanne, Switzerland, [19]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, Lausanne, Switzerland, [20]Department of Prosthetic Dentistry, Gerostomatology and Dental Materials and [21]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, [22]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands, [23]Centre for Population Health Sciences, University of Edinburgh, Scotland, UK, [24]Universität zu Lübeck, Medizinische Klinik II, Lübeck, Germany, [25]Deutsches Zentrum für Herz-Kreislauf-Forschung e. V. (DZHK) and [26]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany, [27]Estonian Genome Center and [28]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia, [29]Estonian Biocenter, Tartu, Estonia, [30]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA, [31]Institute for Maternal and Child Health - IRCCS 'Burlo Garofolo' – Trieste, University of Trieste, Italy, [32]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, [33]CIBER Epidemiologia y Salud Pública (CIBERESP), Barcelona, Spain, [34]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, [35]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA, [36]Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany, [37]Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, [38]Department of Epidemiology and [39]Department of Internal Medicine, Erasmus MC Rotterdam, The Netherlands, [40]Netherlands Consortium of Healthy Ageing, Rotterdam, The Netherlands, [41]Department of Biostatistics and [42]Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA, [43]National Institute for Health and Welfare, Department of Chronic Disease Prevention, Population Studies Unit, Turku, Finland, [44]Department of Clinical Physiology and [45]Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere, Tampere, Finland, [46]Tampere University Hospital, Tampere, Finland, [47]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK, [48]Department of Epidemiology, Biostatistics and HTA & Department of Urology, [49]Department of Laboratory Medicine, [50]Department of Endocrinology and [51]Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [52]Unit for Molecular Epidemiology and [53]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, [54]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [55]Ontario Institute for Cancer Research, Toronto, Canada, [56]Transplantation Laboratory, Haartman Institute and [57]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, [58]Department of Cardiology, [59]Department of Internal Medicine and [60]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [61]Division of Cardiology, Cardiovascular Laboratory, Helsinki University Central Hospital, Helsinki, Finland, [62]Department of Medical Genetics, University and University Central Hospital of Helsinki, Helsinki, Finland, [63]The Broad Institute of MIT and Harvard, Cambridge, USA, [64]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, [65]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK, [66]Department of Psychiatry, EMGO Institute, Center for Neurogenomic and Cognitive Research and [67]Department of Psychiatry, EMGO Institute, Institute for Neurosciences, VU University Medical Centre, Amsterdam, The Netherlands, [68]Clinical Research Branch, National Institute of Aging, National Institute on Aging, Baltimore MD

21250, USA, [69]Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany, [70]Department of Medicine and [71]IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany, [72]Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki 00271, Finland, [73]South Ostrobothnia Central Hospital, Seinäjoki 60220, Finland, [74]Red RECAVA Grupo RD06/0014/0015, Hospital Universitario La Paz, Madrid 28046, Spain, [75]Centre for Vascular Prevention, Danube-University Krems, Krems 3500, Austria, [76]Ludwig Institute for Cancer Research, Lausanne, Switzerland, [77]Icelandic Heart Association, Kopavogur, Iceland, [78]University of Iceland, Reykjavik, Iceland, [79]Department of Medicine and [80]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, [81]Turku University Hospital, Turku, Finland, [82]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland, [83]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, Scotland, UK, [84]Ottawa Hospital Research Institute, Ottawa, Canada, [85]Andrija Stampar School of Public Health, School of Medicine, University of Zagreb, Croatia, [86]Division of Epidemiology & Community Health, University of Minnesota, MN, USA, [87]Department of Chronic Disease Prevention, Chronic Disease Epidemiology and Prevention Unit and [88]Department of Chronic Disease Prevention, Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland, [89]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland, [90]University Central Hospital, Unit of General Practice, Helsinki, Finland, [91]Folkhalsan Research Centre, Helsinki, Finland, [92]Vasa Central Hospital, Vasa, Finland, [93]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA, [94]Klinik und Poliklinik für Innere Medizin II, Regensburg, Germany, [95]Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, Canada, [96]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK, [97]Institute of Health Sciences and [98]Biocenter Oulu, University of Oulu, Oulu, Finland, [99]National Institute for Health and Welfare, Oulu, Finland, [100]IMIM (Hospital del Mar Research Institute), Barcelona, Spain, [101]National School of Public Health, Athens, Greece, [102]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands, [103]The Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, [104]Institut für Klinische Molekularbiologie, Christian-Albrechts Universität, Kiel, Germany, [105]Institute for Maternal and Child Health - IRCCS 'Burlo Garofolo', Trieste, Italy, [106]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, [107]Klinikum Grosshadern, Munich, Germany

The male-to-female sex ratio at birth is constant across world populations with an average of 1.06 (106 male to 100 female live births) for populations of European descent. The sex ratio is considered to be affected by numerous biological and environmental factors and to have a heritable component. The aim of this study was to investigate the presence of common allele modest effects at autosomal and chromosome X variants that could explain the observed sex ratio at birth. We conducted a large-scale genome-wide association scan (GWAS) meta-analysis across 51 studies, comprising overall 114 863 individuals (61 094 women and 53 769 men) of European ancestry and 2 623 828 common (minor allele frequency >0.05) single-nucleotide polymorphisms (SNPs). Allele frequencies were compared between men and women for directly-typed and imputed variants within each study. Forward-time simulations for unlinked, neutral, autosomal, common loci were performed under the demographic model for European populations with a fixed sex ratio and a random mating scheme to assess the probability of detecting significant allele frequency differences. We do not detect any genome-wide significant ($P < 5 \times 10^{-8}$) common SNP differences between men and women in this well-powered meta-analysis. The simulated data provided results entirely consistent with these findings. This large-scale investigation across ∼115 000 individuals shows no detectable contribution from common genetic variants to the observed skew in the sex ratio. The absence of sex-specific differences is useful in guiding genetic association study design, for example when using mixed controls for sex-biased traits.

INTRODUCTION

The male-to-female sex ratio at birth is very constant across world populations, ranging between 1.02 and 1.08 (102–108 male to 100 female live births), with an average of 1.06 for populations of European descent (1–3). The sex ratio at birth is mainly determined by factors influencing the primary sex ratio, which is the sex ratio at conception, and

those influencing the survival of the embryo (4). Frequently reported primary sex ratio-determining factors include motility and survival time of X-bearing and Y-bearing sperm. A proportion of prenatal mortality can be attributable to immunological interaction between mother and embryo (4). Interestingly, more males are being born in spite of the fact that there is higher mortality of males than females during intrauterine life (4,5). In addition, the sex ratio is considered to be affected by numerous other biological (endogenous) and environmental (exogenous) factors, albeit their influence is generally thought to be of a small effect (1,6). These factors include gonadotropins and/or testosterone concentration at the time of conception, ovulation induction, parental age, parity, birth order, coital rates, infertility, parental illness, maternal malnutrition, smoking, exposure to certain chemicals, stress, war, socioeconomic status and many others (1,6–8). The variation in sex ratio was also observed in many animal and plant species (9). Studies of parasitoid wasps, particularly *Nasonia vitropennis*, identified several quantitative trait loci (QTL) associated with the sex ratio, pointing to a genetic contribution (9). In addition, many authors suggest that the human sex ratio also has a heritable component. Paternal effects have been proposed to play a role in the sex ratio, for example, men with more brothers tend to have more sons whereas men with more sisters tend to have more daughters (5,10,11). Based on population genetics modelling, Gellatly *et al*. (11) suggested that the sex ratio is determined by common inheritance of polymorphic autosomal genes that exert their effect through the male reproductive system. Another study of reproductive fitness in the Hutterite population suggested that genetic variants, both autosomal and X-linked, influence natural fertility in humans (12). Research of human births in two-child families observed that sexes of offspring do not follow a binomial model of inheritance where probability of having a boy equals probability of having a girl (13). This study also pointed to the lack of independence among sexes of children of the same parents (13). A couple of possible genetic mechanisms underlying this observation such as Y- and X-linked immunological incompatibilities between mother and embryo have been proposed (13,14).

In the present study, we test whether common variant genetic effects partly underlie the observed male-to-female sex ratio at birth. To address this, we investigate the presence of autosomal and chromosome X variant differences between men and women across 114 863 individuals through large-scale genome-wide association study (GWAS) meta-analysis. We also conduct a forward-time simulation study to assess the probability of observing significant allele frequency differences at autosomal markers between men and women. Our study has high power to detect loci with modest to small effect sizes.

RESULTS

GWAS meta-analysis results

Initial meta-analysis results pointed to an excess of associations compared with the null distribution (Fig. 1A). We examined all genome-wide significant SNPs with effective sample size >10 000 to check for false positives due to genotyping error or other artefacts. We investigated three main diagnostic
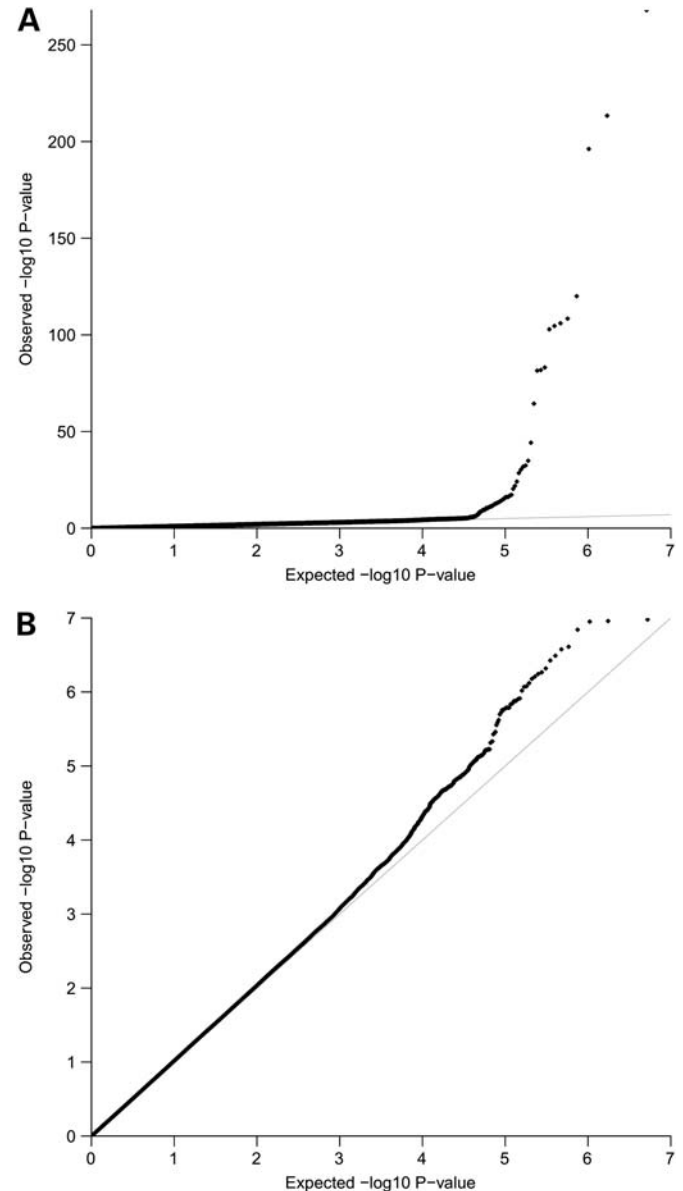
Figure 1. QQ plots for 2 623 828 directly genotyped and imputed SNPs: (**A**) for all examined SNPs; (**B**) after exclusion of poorly genotyped/called SNPs.

metrics: poor cluster plots in men or women (Supplementary Material, Fig. S1), sequence similarity on chromosome Y and exact Hardy–Weinberg equilibrium (HWE), $P < 1.0 \times 10^{-6}$ for men or women. Autosomal SNPs that lie in genomic regions that have sequence similarity on the Y chromosome may be incorrectly genotyped/called in men, but not in women, which may give rise to false-positive associations. This can be traced through several quality control (QC) checks in men: excess of heterozygosity, deviation from HWE and poor cluster plots in men and not in women. Some/all of these factors were observed for SNPs designated for exclusion from follow-up (all with highly significant association *P*-values). We excluded SNPs from the pseudoautosomal chromosome X boundary regions (within 55 kb on the short arm of chromosome X and 115 kb on the long arm that lie next to the non-pseudoautosomal regions) to guard

against genotyping error in men (for example, caused by the presence of truncated copies of genes and/or mappings to multiple places across the genome). After exclusion of all poorly genotyped/called SNPs, we detect a single genome-wide significant association at a non-pseudoautosomal chromosome X variant (rs12689384, $P = 2.66 \times 10^{-13}$), which is an intronic variant within *RBMX2*. As this SNP was imputed in all studies driving the association, we checked cluster plots of all directly typed variants 500 kb upstream and downstream of the associated SNP in all studies, excluded SNPs with poor clustering, re-imputed the region and re-run the meta-analysis. The significance of rs12689384 dropped by five orders of magnitude (allele G, OR = 1.18, 95% CI [1.11–1.25], $P = 3.27 \times 10^{-8}$) but remained nominally genome-wide significant. However, there are several factors that reduce the credibility of this finding. First, 12 studies in total contributed summary statistics for this variant (for a total of 33 259 individuals), out of which six WTCCC1 studies drive the association (Supplementary Material, Table S1). This SNP is imputed in all WTCCC1 studies. Secondly, as shown in the regional association plot (Supplementary Material, Fig. S2), this SNP lacks support for association from neighbouring variants. The next most statistically strong association in the region is modest ($P = 1.99 \times 10^{-4}$) and observed at an SNP (rs2294956) which is in perfect linkage disequilibrium with rs12689384 ($r^2 = 1$, D' = 1) based on HapMap CEU, on which imputation was based. This second SNP (rs2294956) is also imputed in the WTCCC1 studies, but in fact the meta-analysis includes data from over twice the sample size (31 studies, 67 162 individuals, Supplementary Material, Table S1). All studies contributing directly typed data for this variant (46 066 individuals with directly typed data) show no evidence for association, indicating that the signal observed at the imputed variant may be an artefact. We have therefore not considered this single associated SNP any further.

The distribution of association *P*-values after meta-analysis QC was consistent with the null (Fig. 1B). Our study has 80% power to detect OR of 1.13 (at $\alpha = 5 \times 10^{-8}$) for SNPs with minor allele frequency (MAF) >5%, assuming an additive model. We did not further examine SNPs with *P*-values above the genome-wide significance threshold since our study had sufficient power to detect associations of loci with small-to-modest effect sizes.

Simulation study results

Association analysis of 1 337 699 autosomal common and 135 988 autosomal low-frequency variants in the simulated case–control set matching the empirical study did not identify any differences in allele frequencies between men and women ($\alpha = 5 \times 10^{-8}$). Quantile–quantile (QQ) plots for simulated common and low-frequency variants are shown in Supplementary Material, Figure S3.

DISCUSSION

This large-scale investigation across 114 863 individuals identified no detectable contribution from common genetic variants to the observed skew in sex ratio at birth. This study combined the data from 51 cohorts and has excellent power to detect small-to-modest effect sizes at common loci. The sample sizes contributing to the analysis of chromosome X SNPs were lower due to limited overlap of directly typed SNPs across platforms (in the absence of imputed data across all studies). However, power remains high at over 80% to detect small-to-modest effect sizes. From the phenotypic aspect, sex is a well-characterized trait representing an additional strength of this study, which is unlikely to suffer from phenotype misclassification.

Our results, within the power constraints of our study, indicate that sex-specific selection against particular autosomal genetic variants is not a plausible explanation for the observed male-to-female sex ratio at birth and argue against the hypothesis that incompatible genotypes at common variants between the autosomes and sex chromosomes could lead to miscarriage, thus generating sex-specific genetic differences. We performed forward-time simulations of 1.3 million independent autosomal, common, neutral loci, conditioning on the male-to-female sex ratio at birth, in a cohort matching the original study sample. The lack of any significant allele frequency differences between men and women was in keeping with the findings of the GWAS meta-analysis for autosomal SNPs. We also tested the effects of low-frequency variants in the simulated data, and found no evidence for association with the observed sex ratio at birth. However, we cannot rule out the effects of rare, structural or chromosome Y variants since these were not analysed in our study.

Sex chromosome loci may be relevant for the sex ratio determination due to their expression in the reproductive system, their role in spermatogenesis, sperm morphogenesis and movement and male–female fertility in general (15,16). Therefore, we performed a comprehensive chromosome X analysis involving two main chromosome X regions: pseudoautosomal and non-pseudoautosomal. There are two pseudoautosomal regions (PAR1 and PAR2), which are homologous on X and Y chromosomes, and for which men and women carry two alleles per SNP, whereas for the non-pseudoautosomal region men carry only one allele per SNP. We investigated allele frequency differences between men and women in both chromosome X regions and we observed the association of one non-pseudoautosomal SNP (rs12689384) just below the genome-wide significance level. For various reasons expanded in the Results section, we believe that this variant may be an imputation artefact and have thus not taken it forward to further studies.

The investigated dataset consisted of more women (61 094) than men (53 769). Our meta-analysis incorporated summary statistics deriving from 51 collaborating studies and the vast majority of these studies (36 studies) are population based. The main difference in the sex ratio is driven by these population based studies and the reasons for having fewer men can be heterogeneous and study specific. Most likely the main reasons are the generally recognized lower male response to take part in epidemiological population-based studies (17) and/or sex differences in longevity where women have a higher expected lifespan (18). Fifteen of the 51 contributing studies are disease- rather than population-based and the sex ratio in these studies approximately corresponds to the disease sex ratio in the population. We were

driven by the rationale that the sex ratio at birth is constant throughout time and across all world populations, meaning that common variants are more likely to underlie the observed sex ratio at birth. Therefore, in the case of a higher male death rate, we would still have enough power to detect common variant differences due to a very large sample set. However, there are scenarios where this sampling difference between men and women might cause bias, for example there may be a genetic variant that is influencing both the sex ratio and longevity in men. In that case, higher male death rate would cause the removal of this specific genetic variant, thus masking the signal.

Our results have important implications for genetic association study design, for example regarding the selection of control sets for sex-biased traits such as prostate cancer in men or anorexia nervosa in women. The use of single-sex controls for sex-specific diseases generally decreases the sample size and power of a study. Our findings demonstrate that mixed sex controls can be used as an appropriate set in studies of sex-specific traits, when focusing on common loci. As one additional implication for genetic association study analyses, our study stresses the importance of careful pre- and post-analysis QC. QQ plots of our initial meta-analysis results showed high deviation from the null, yet, after QC we observe no inflation of signal. A robust and thorough QC pipeline is necessary to verify any positive association signals, especially in meta-analyses where many studies contribute data that were genotyped (and phenotyped) in many different settings.

We conclude that common genetic variants do not play a role in defining male-to-female sex ratio at birth. In this large-scale meta-analysis of ∼115 000 individuals, we found no allele frequency differences at common loci between men and women. Simulated data of autosomal neutral variants support these findings. Our results can be useful in informing GWAS study design, especially when using mixed controls for sex-biased traits.

MATERIALS AND METHODS

Study samples

We conducted genome-wide meta-analysis across 51 studies, comprising overall 114 863 individuals (61 094 women and 53 769 men) of European ancestry. The characteristics of samples from contributing studies are presented in Supplementary Material, Table S2.

Ethics statement

Each study obtained ethical approval from their respective research ethics committee and all participants gave signed informed consent in accordance with the Declaration of Helsinki.

Genotyping, imputation and QC

All samples were genotyped using commercially available Illumina (Illumina, Inc., San Diego, CA, USA) or Affymetrix (Affymetrix, Inc., Santa Clara, CA, USA) platforms. Imputation of missing genotypes was based on HapMap Phase II

genotypes for the European population (CEU). QC of directly typed and imputed variants was conducted separately in each study. Study-specific information on genotyping platforms, imputation methods and QC metrics is presented in Supplementary Material, Table S3. QC checks included tests for relatedness among samples within individual studies.

Genome-wide association analysis of autosomal variants

Case–control association analysis of autosomal SNPs was conducted under the additive model, for directly typed and imputed variants, within each study. Women were coded as cases and men as controls. Association analyses of imputed variants took genotype uncertainty into account, with the exception of the QIMR study which conducted analysis on best-guess genotypes. Where necessary, the first three genotype-based principal components were used as covariates. Studies with related individuals additionally adjusted analyses for family relatedness using linear mixed models. Study-specific association analysis software is presented in Supplementary Material, Table S3.

Chromosome X analysis

Each contributing study performed two separate chromosome X analyses, including pseudoautosomal and non-pseudoautosomal regions. Association analyses were performed, as per autosomes, under the additive model. Overall, 42 studies performed analysis of pseudoautosomal region, 11 of these imputed data using HapMap Phase II, all others used directly typed variants only. For non-pseudoautosomal region, 46 studies performed association analysis, 12 of these used HapMap Phase II imputed data whereas others used directly typed variants only. Study-specific chromosome X imputation/association analysis software is presented in Supplementary Material, Table S3.

GWAS meta-analysis

We performed fixed and random effects meta-analysis to synthesize summary statistics results across contributing studies to identify autosomal and chromosome X common SNP differences between men and women. For meta-analysis purposes, we used GWAMA [19]. Prior to meta-analysis, we excluded SNPs with MAF lower than 0.05 and SNPs with low imputation accuracy scores. Specifically, we used a cut-off of $rsq\_hat < 0.3$ for genotypes imputed with MACH [20], BEAGLE [21] and PLINK [22] software and a cut-off of proper info score <0.5 for IMPUTE [23] software. Overall, 2 623 828 directly genotyped and imputed SNPs passed QC criteria and were included in the meta-analysis. The genomic control (GC) inflation factor (lambda) was calculated and applied to correct the results for each study separately prior to the meta-analysis. The meta-analysis results were also corrected for overall lambda GC. The average GC inflation factor across studies was 1.005 for directly genotyped SNPs, 0.97 for imputed SNPs and 1.007 overall, suggesting little population stratification. To determine the effective number of individuals for each study, we calculated effective number of cases (N_eff_case) and multiplied it by 2. N_eff_case was derived

using the formula $N\_eff\_case = 2 \times N\_case \times N\_ctrl/(N\_case + N\_ctrl)$, where N_case and N_ctrl is the number of cases (women) and controls (men), respectively. We investigated evidence of heterogeneity using the I^2 statistic (24). Genome-wide significance was set to 5×10^{-8}. We created QQ plots to visualize meta-analysis association results. The power of our study was determined using QUANTO (25).

Simulation study

To exclude the possibility that our null results for autosomal variants are due to either sampling bias or data quality and to examine the probability of having false positives within the power constraints of our study, we sought a theoretical corroboration of our empirical results by conducting association analysis in an 'ideal' unbiased simulated dataset. Simulated genetic data were produced by means of forward-time simulation (26–28) under a model of a single population with two bottlenecks according to Schaffner *et al.* (29) with two exceptions: recent exponential growth of population size (instead of instantaneous changes) and final effective population size of 10^6 (instead of 10^5), as this has been shown to be the case for the European population (30). Demographic model parameters are given in Supplementary Material, Table S4. The generation time was assumed to be 25 years, and the mutation rate per site per generation was 1.5×10^{-8} (29). We applied a fixed sex ratio and a random mating scheme (i.e. parents are randomly selected irrespective of their genotype) validated by different genetic and demographic models (27). We set a probability of having a male offspring to 0.5122, which corresponds to a male-to-female ratio of 1.05. Simulations were run for 17 000 generations after which we randomly sampled women and men matching the original study for sample size (women = 61 094; men = 53 769).

We simulated unlinked, neutral, autosomal common variants with initial MAF of 0.02 in the founder population (27,31). The total number of simulated loci was 56 502 900, out of which 2.4% were common (MAF > 0.05). We performed allele-based chi-squared association tests on the 1 337 699 common loci. This figure matches the estimated number of independent SNPs in HapMap CEU samples of around 1 million (32). We additionally performed allele-based Fishers exact association tests on 135 988 low-frequency variants (MAF 0.01–0.05). Supplementary Material, Figure S4 shows the MAF spectrum for simulated data compared with the 1000 Genomes Project Pilot 3 CEU ($2n = 60$) data.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGMENTS

Conflict of Interest statement. None declared.

FUNDING

REFERENCES

1. Mackenzie, C.A., Lockridge, A. and Keith, M. (2005) Declining sex ratio in a first nation community. *Environ. Health Perspect.*, **113**, 1295–1298.

2. Kaba, A.J. (2008) Sex ratio at birth and racial differences: why do black women give birth to more females than non-black women? *Afr. J. Reprod. Health*, **12**, 139–150.
3. James, W.H. (1987) The human sex ratio. Part 1: a review of the literature. *Hum. Biol.*, **59**, 721–752.
4. Chahnazarian, A. (1988) Determinants of the sex ratio at birth: review of recent literature. *Soc. Biol.*, **35**, 214–235.
5. Curtsinger, J.W., Ito, R. and Hiraizumi, Y. (1983) A two-generation study of human sex-ratio variation. *Am. J. Hum. Genet.*, **35**, 951–961.
6. Dodds, L. and Armson, B.A. (1997) Is Canada's sex ratio in decline? *CMAJ*, **156**, 46–48.
7. James, W.H. (2004) Further evidence that mammalian sex ratios at birth are partially controlled by parental hormone levels around the time of conception. *Hum Reprod*, **19**, 1250–1256.
8. Fukuda, M., Fukuda, K., Shimizu, T., Andersen, C.Y. and Byskov, A.G. (2002) Parental periconceptional smoking and male: female ratio of newborn infants. *Lancet*, **359**, 1407–1408.
9. Pannebakker, B.A., Watt, R., Knott, S.A., West, S.A. and Shuker, D.M. (2011) The quantitative genetic basis of sex ratio variation in *Nasonia vitripennis*: a QTL study. *J. Evol. Biol.*, **24**, 12–22.
10. Trichopoulos, D. (1967) Evidence of genetic variation in the human sex ratio. *Hum. Biol.*, **39**, 170–175.
11. Gellatly, C. (2009) Trends in population sex ratios may be explained by changes in the frequencies of polymorphic alleles of a sex ratio gene. *Evol. Biol.*, **36**, 190–200.
12. Kosova, G., Abney, M. and Ober, C. Colloquium papers: heritability of reproductive fitness traits in a human population. *Proc. Natl Acad. Sci. USA*, **107**(Suppl 1), 1772–1778.
13. Carlton, M.A. and Stansfield, W.D. (2005) Making babies by the flip of a coin? *Am. Statist.*, **59**, 180–182.
14. Stansfield, W.D. and Carlton, M.A. (2007) Human sex ratios and sex distribution in sibships of size 2. *Hum. Biol.*, **79**, 255–260.
15. Burgoyne, P.S. (1998) The role of Y-encoded genes in mammalian spermatogenesis. *Sem. Cell Dev. Biol.*, **9**, 423–432.
16. Zhang, Y.E., Vibranovski, M.D., Landback, P., Marais, G.A. and Long, M. (2010) Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.*, **8**, pii: e1000494.
17. McCarty, C., Wilke, R., Giampietro, P., Wesbrook, S. and Caldwell, M. (2005) Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Person. Med.*, **2**, 49–79.
18. May, R.C. (2007) Gender, immunity and the regulation of longevity. *Bioessays*, **29**, 795–802.
19. Magi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
20. Li, Y., Ding, J. and Abecasis, G. (2006) Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.*, **79**, S2290.
21. Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.*, **81**, 1084–1097.
22. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
23. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
24. Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557–560.
25. Gauderman, W.J. (2003) Candidate gene association analysis for a quantitative trait, using parent-offspring trios. *Genet. Epidemiol.*, **25**, 327–338.
26. Peng, B., Amos, C.I. and Kimmel, M. (2007) Forward-time simulations of human populations with complex diseases. *PLoS Genet.*, **3**, e47.
27. Peng, B. and Kimmel, M. (2007) Simulations provide support for the common disease-common variant hypothesis. *Genetics*, **175**, 763–776.
28. Peng, B. and Kimmel, M. (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.
29. Schaffner, S.F., Foo, C., Gabriel, S., Reich, D., Daly, M.J. and Altshuler, D. (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.*, **15**, 1576–1583.
30. Coventry, A., Bull-Otterson, L.M., Liu, X., Clark, A.G., Maxwell, T.J., Crosby, J., Hixson, J.E., Rea, T.J., Muzny, D.M., Lewis, L.R. *et al.* (2010) Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.*, **1**, 131.
31. Reich, D.E. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet.*, **17**, 502–510.
32. Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.*, **32**, 381–385.