ORIGINAL ARTICLE

Genome-wide association study identifies ERBB4 on 2q34 as a novel locus associated with sperm motility in Japanese men

Youichi Sato,1 Atsushi Tajima,2,3 Takehiro Sato,3 Shiari Nozawa,4 Miki Yoshiike,4 Issei Imoto,2 Aiko Yamauchi,1 Teruaki Iwamoto1,4,5

ABSTRACT

Background The decrease in sperm motility has a potent influence on fertilisation. Sperm motility, represented as the percentage of motile sperm in ejaculated sperms, is influenced by lifestyle habits or environmental factors and by inherited factors. However, genetic factors contributing to individual differences in sperm motility remain unclear. To identify genetic factors that influence human sperm motility, we performed a genome-wide association study (GWAS) of sperm motility.

Methods A two-stage GWAS was conducted using 811 Japanese men in a discovery stage, followed by a replication study using an additional 779 Japanese men.

Results In the two-staged GWAS, a single nucleotide polymorphism rs3791668 in the intron of gene erb-b2 receptor tyrosine kinase 4 (ERBB4) on chromosome 2q34 was identified as a novel locus for sperm motility, as evident from the discovery and replication results using meta-analysis (P=4.01 × 10−9, combined P=5.40×10−9).

Conclusions Together with the previous evidence that Sertoli cell-specific Erbb4-knockout mice display an impaired ability to produce motile sperm, this finding provides the first genetic evidence for further investigation of the genome-wide significant association at the ERBB4 locus in larger studies across diverse human populations.

INTRODUCTION

Approximately 10% couples display infertility issues, and half of these problems are related to men.1–2 Male factor infertility may arise from various medical conditions such as spermatozoal genetic failure, varicocele, obstructive azoospermia and congenital absence of vas deferens. Sperm motility—represented as the percentage of motile sperm in the ejaculated sperms—has a large influence on the fertilisation ability. Therefore, several studies are conducted to understand the factors that affect sperm motility.

Oxidative stress induced by alcohol consumption, cigarette smoking, obesity, diabetes, physical exercise, psychological stress, ageing, infection and environment factors (pollutants such as nitric oxide, lead and electromagnetic waves from cell phones) is one of the major factors responsible for the reduction in sperm motility.3–4 Genetic background has also been shown to be associated with sperm motility. The gr/gr subdeletion in the azoospermia factor c region of the Y chromosome was shown to be strongly associated with decreased sperm motility in men from Japanese population.5 Furthermore, polymorphisms in genes encoding cytochrome P450 family 19 subfamily A polypeptide 1,6 androgen receptor,7 follicle-stimulating hormone receptor,8 steroid 5α-reductase9 and oestrogen receptor10 were associated with sperm motility. These genes are related to the reproductive hormones and contribute to the testicular development and spermatogenesis; these genes have been proposed based on their functions. However, the genetic determinants for human sperm motility are poorly understood.

Genome-wide association study (GWAS) is an approach to find the genetic variations associated with disease or quantitative traits. To date, four GWASs associated with male infertility have been reported. These include the non-obstructive azoospermia or oligozoospermia in Caucasians or Chinese men12–15 and the family size or birth rate in Hutterite men in the USA.16 In the latter, 9 of the 41 single nucleotide polymorphisms (SNPs) were significantly correlated with the family size or birth rate and found to be associated with reduced sperm quantity and/or function in the subsequent validation study using 123 ethnically diverse men. However, there are no reports on GWAS of sperm motility. Here, we clarified the genetic determinants for human sperm quality by conducting a GWAS of sperm motility in 811 Japanese men, with a subsequent validation of the association in an additional 779 Japanese men.

METHODS

Subjects

We performed a two-staged genetic association study. The discovery stage included 816 men (20.7±1.7 years old, mean±SD) from the young Japanese population. These were recruited from university students in three study centres based in departments of urology at university hospitals in Japan (Kawasaki, Kanazawa and Nagasaki) as previously reported.17 The inclusion criteria were that the man was 18–24 years and that both he and his mother were born in Japan. The replication stage included 779 men (31.2±4.8 years old, mean±SD) of proven fertility recruited from the partners of pregnant women who attended obstetric clinics in Japan; 779 men (31.2±4.8 years old, mean±SD) of proven fertility recruited from the partners of pregnant women who attended obstetric clinics in Japan.
Statistical analysis

In discovery and replication stages, associations between each SNP and sperm motility were assessed using a multiple linear regression under an additive genetic model, with adjustments for age, BMI, ejaculatory abstinence period and time from masturbation to semen evaluation using PLINK or R V.3.1.2 software package (http://www.R-project.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.

The Manhattan and quantile–quantile plots were generated using qman package for the R software, while a regional plot was created by LocusZoom using the 1000 Genomes project Asian (ASN) data (November 2014).34 With the exception of annotations, linkage disequilibrium (LD) was calculated using PLINK software V.1.07 with genotype imputation data. Significant expression quantitative trait loci (eQTL) by SNP was searched on GTEx Portal database (http://www.gtexportal.org/). Since the raw value was closest to the normal distribution than some converted values, we decided to use the raw value for analysis in the present study. We set a suggestive threshold of P values <1×10⁻²⁰ in the discovery stage. The results were combined in a meta-analysis using the meta package for the R software. The extent of heterogeneity among studies was quantified by the I² statistic and statistically assessed by the Cochran’s Q test. No heterogeneity was observed in this study, as determined by the I² statistic <50% or P value >0.1; hence, a fixed-effect model using the inverse variance method was used. Genome-wide statistical significance was considered at P values <5×10⁻⁸.
RESULTS

We conducted a two-staged GWAS to identify genetic loci associated with human sperm motility. We enrolled 816 Japanese men from the university students for the discovery stage and 779 Japanese men from the partners of pregnant women for the replication stage of GWAS. After quality control of samples using initially genotyped 298,930 SNP data in the discovery stage, 811 Japanese men were selected. We performed imputation analysis, which provided typed and imputed genotypes for 390,1256 SNPs that passed quality control. Finally, 811 samples and 3,901,256 SNPs were included for the discovery stage. The characteristics of subjects are presented in table 1.

We performed GWAS between a total of 3,901,256 SNPs and sperm motility in 811 men in the discovery stage. Manhattan and quantile-quantile plots of GWAS are presented in figures 1 and 3 respectively. The genomic inflation factor ($\lambda$) was reported to be 1.0, indicating the likelihood of the inflation of the false-positive association. The top 50 GWAS candidate SNPs for sperm motility were presented in online supplementary table S1. We failed to find any SNPs to reach a genome-wide significance level ($P<5.0 \times 10^{-8}$) in the discovery stage. When setting a suggestive significance threshold of $P$ values $<1.0 \times 10^{-6}$, we identified that two SNPs, rs3791686 and rs1836719 on 2q34, were suggestively associated with sperm motility ($\beta=-4.25$, discovery $P=4.47 \times 10^{-7}$; $\beta=-4.22$, discovery $P=5.29 \times 10^{-7}$, respectively) (online supplementary table S1). These two SNPs are in strong LD ($r^2=0.99$); thus, we selected only the most significant SNP (rs3791686) for the subsequent replication genotyping.

In the replication study involving 779 proven fertile men, SNP rs3791686 on 2q34 showed a significant association with sperm motility ($\beta=-3.51$, replication $P=3.88 \times 10^{-3}$) (table 2). When we combined the discovery and replication results using meta-analysis, rs3791686 surpassed the threshold for genome-wide significance ($\beta=-4.01$, combined $P=5.40 \times 10^{-3}$) with no evidence of heterogeneity between the two studies. The variance in sperm motility explained by rs3791686 was 2.0%.

Figure 2 shows a regional association plot for the genomic region 400 kb upstream and downstream of the lead SNP rs3791686 in the discovery stage. Within the region, 24 genotyped and 289 imputed SNPs, including rs3791686, were associated with sperm motility, with discovery $P$ values $<0.05$ from the association analysis in discovery stage (online supplementary table S2). The sperm motility-associated genomic interval indexed by rs3791686 on 2q34 overlapped with a single known gene, erb-b2 receptor tyrosine kinase 4 (ERBB4), while the lead SNP rs3791686 was located in the intron of ERBB4. Of a total of 313 SNPs with discovery $P$ values $<0.05$ within the associated interval, none resulted in amino acid substitution or protein truncation or affected the splicing of ERBB4; one synonymous SNP (rs3748962) and seven SNPs in the 3′-untranslated region of ERBB4 were observed (online supplementary table S2). To obtain putative functional annotations of rs3791686 and other 13 SNPs in high LD ($r^2>0.80$ in East Asians from the 1000 Genomes Project) with rs3791686 (online supplementary table S3) within the associated interval, we used the following three databases: GTEx Portal,35 HaploReg36 and RegulomeDB.37 We assessed if the 14 SNPs, including rs3791686 on 2q34, were involved in eQTLs using the GTEx Portal database and found that no significant eQTLs were associated with all the SNPs examined. HaploReg and RegulomeDB databases search revealed that the 14 SNPs examined within the associated interval may be regarded as candidate regulatory SNPs (online supplementary table S3). In the HaploReg database, the lead SNP rs3791686 itself was associated with enhancer histone marks and DNase I hypersensitive region in embryonic stem-derived cells and resided in regulatory motifs of four transcription factors—Maf, Nkx2, Nkx3 and TATA-binding protein (online supplementary table S3). Of the 13 SNPs in high LD with the lead SNP rs3791686, five were associated with enhancer histone marks and/or DNase I hypersensitive regions in various types of cells and tissues, while 12 SNPs had the potential to alter nucleotide sequences of several regulatory motifs. The RegulomeDB database provided the experimental evidence that three SNPs (rs13003941, rs1836720 and rs1836719) were associated with enhancer histone marks and/or DNase I hypersensitive regions in various types of cells and tissues, while 12 SNPs had the potential to alter nucleotide sequences of several regulatory motifs. The RegulomeDB database provided the experimental evidence that three SNPs (rs13003941, rs1836720 and rs1836719) were located in DNase I hypersensitive and/or TF-binding regions in various cells. The iGSEA4GWAS analysis identified 421 significant pathway (FDR $<0.05$) (online supplementary table S4). Numerous pathways were identified in this analysis; this finding suggests that sperm motility ability is likely to affect by a complicated process involving interaction between multiple genes and pathways.

DISCUSSION

In the first two-staged GWAS of sperm motility in Japanese men, we identified a novel sperm motility-associated locus at

**Table 2** Genome-wide significant SNP identified in GWAS for sperm motility

| SNP (effect/reference) Chr | Position (hg 19) | Gene locus | Stage | Genotype | EAF | Beta (SE) | P value | Var (%)* |
|---------------------------|------------------|------------|-------|----------|-----|----------|---------|---------|
| rs3791686 (C/T)           | 2                | 212221870 | ERBB4 intron | Discovery | 0.24 | $-4.25$ (0.84) | $4.47 \times 10^{-7}$ |         |
|                           |                  |           |       | Replication | 0.23 | $-3.51$ (1.21) | $3.88 \times 10^{-3}$ |         |
|                           |                  |           |       | Meta-analysis |     | $-4.01$ (0.69) | $5.40 \times 10^{-3}$ | 2.0      |

Data are shown as the estimated linear regression statistic beta, SE and P value using an additive genetic model with adjustments for age, BMI, ejaculation abstinence and time from masturbation to test.

* Var (%): percentage of phenotypic variance explained by SNP.

BMI, body mass index; Chr, chromosome; EAF, effect allele frequency; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.
**ERBB4** on chromosome 2q34. The most strongly associated SNP was typed by imputation analysis. In this study, the subjects of discovery stage were genotyped using the Illumina HumanCore V1.0 DNA Analysis Kit with a total of 298,930 SNPs. Subsequently, to enhance the coverage, untyped SNPs were imputed. Sometimes, imputation methods may be less accurate for typing of SNPs. To confirm the accuracy of this imputation method, randomly selected samples were directly genotyped for the GWAS-lead SNP rs3791686. The result of imputation analysis was validated by the genotyping.

SNP rs3791686 lies in the intron of **ERBB4** gene, which is a member of the receptor tyrosine kinase family and epidermal growth factor receptor subfamily. **ERBB4** is expressed in several tissues, including kidney, breast, cerebrum, heart, bone, ovary and testis. On activation by its ligands, ERBB4 forms a dimer on the cell surface. Following cleavage of the ERBB4 ectodomain by a disintegrin and metalloprotease domain 17 (ADAM17) and γ-secretase, the intracellular domain of ERBB4 is translocated into the nucleus. Inside the nucleus, ERBB4 is involved in the regulation of cell proliferation and differentiation.39–42 ERBB4 is thought to be both necessary and sufficient to trigger an antiproliferative response in human breast cancer cells.43 Kim et al44 reported that the SNP rs13393577 in **ERBB4** is associated with breast cancer risk in Koreans by GWAS. In addition, previous GWASs in the National Human Genome Research Institute (NHGRI) GWAS Catalog demonstrate that SNPs in **ERBB4** are genome-wide significantly associated with polycystic ovary syndrome (lead SNP rs1351592)45 and BMI (lead SNP rs7599312).46 The lead SNPs at rs3791686 and show no pairwise LD (r² <0.01 in East Asians) with rs3791686. This indicates a novel association for sperm motility at **ERBB4** on 2q34, which is independent of other human diseases and traits.

The expression of ERBB4 is evident in male reproductive tissues, including testis. In the testicular tissue, ERBB4 is expressed in both somatic cells (Sertoli cells and Leydig cells) and germ cells.47 It is notable that Sertoli cell-specific Erbb4-knockout mice exhibit a developmental defect in the organisation of the testicular seminiferous tubules, which reduces male fertility. Aberration in the testicular cell adhesion machinery caused by Erbb4 deficiency leads to a compromised capacity of the testes to produce motile sperms.47 Thus, ERBB4 signalling in the Sertoli cells may influence the sperm motility, suggestive of the promising functional role of **ERBB4** in sperm motility. The lead SNP rs3791686 identified in this GWAS is an intronic SNP of **ERBB4** and displays the potential to act as a functional regulatory SNP based on the multiple functional annotations. As the functional annotation analyses reveal an association between other SNPs in high LD with rs3791686 and potential regulatory domains and motifs, the sperm motility locus at **ERBB4** may have a role in the regulation of ERBB4 expression via a cis-regulatory mechanism. Sandholm et al,48 reported that a cis-eQTL for **ERBB4** in tubulointerstitial-enriched kidney biopsies maps to intronic **ERBB4** SNPs, rs17418640 and rs17418814. Both of these SNPs are proxies for rs7588550, representing a suggestive association with diabetic nephropathy; however, these eQTL SNPs are not in LD (r²<0.01 in East Asians) with the sperm motility-lead SNP rs3791686. Further studies are warranted to assess the potential contribution of the sperm motility-associated locus indexed by rs3791686 to the regulation of EBRR4 expression. These studies will also help explore the possible involvement of this locus in...
the expression regulation on a genome-wide scale via transregulatory mechanisms.

Liu et al.⁹ have reported that five SNPs (rs215702, rs6476866, rs10129354, rs2477686 and rs10841496) were significantly correlated with sperm progressive motility. However, present study did not detect the variants associated with sperm motility including the region 400 kb upstream and downstream of these five SNPs. Previously, we also have reported that four SNPs as being significantly associated with risk factors for non-obstructive azoospermia (NOA) by Chinese GWAS⁵ were not associated with NOA in Japanese population. The reason for these may be that there are small genetic differences between Han Chinese and Japanese population by a principal component analysis using genotypic data of the HapMap CHB and JPT (online supplementary figure S1). Additionally, we found a strong association between Y-haplogroup and sperm motility in the same Japanese populations. However, none of the SNPs on Y chromosome display a significant association (P<0.05) with sperm motility in this study. The Illumina Human Core V.1.0 DNA analysis kit includes 1943 Y-chromosome markers. However, of these, only 177 markers could be examined in the discovery stage. Because this kit does not include Japanese Y-haplogroup specific markers, we did not find a significant association between Y-chromosome variants and sperm motility in this study.

Several limitations of this study should be noted. In this study, men of proven fertility were used, instead of randomly selected subjects as the replication samples. These were the only samples available for the current replication analysis. Using samples selected on the basis of fertility may cause bias. In fact, abstinence periods were significantly different between two cohorts (table 1). In general, longer abstinence period is correlated with lower sperm motility. As the previous study described, abstinence period was negatively correlated with sperm motility in both cohorts. To reduce the influence of the abstinence period on sperm motility, we included this as a covariate for a multiple linear regression analysis. Therefore, we think that the effect of abstinence period on the power to detect sperm motility-associated SNPs is minimized in this study. Additionally, all the participants in the current two-staged GWAS were Japanese men. Independent validation studies are required to test the observed association between $ERBB4$ SNPs and sperm motility using other general populations and ethnicities. The transethnic association analyses at the $ERBB4$ locus will also enable us to narrow the association signal to smaller sets of SNPs, when leveraging differences in LD structures across diverse populations. The limited statistical power of this two-staged GWAS prevented the detection of other true positive associations at a genome-wide significance level because the sample size was not large. We believe that other genetic loci may account for the interindividual variation in sperm motility, and therefore, larger scale GWAS analyses may be expected to identify novel associations between genetic variants and sperm motility.

It is one of the limitations that sperm motility may show sometimes intraindividual variation between samples from the same individual. When phenotypic repeatability is low, setting the upper boundary of heritability of a trait may decrease sensitivity to detect genetic variant/variants associated with a trait. As aforementioned, sperm motility depends on the abstinence period; in general, abstinence period and sperm motility shows a negative relationship. In our samples, although there is a difference in the strength of association, the abstinence period was indicated to be negatively correlated with sperm motility in both cohorts,⁶ which is not contradictory. In this study, we set a significance threshold of $P$ values $<1 \times 10^{-8}$ in discovery stage and performed the replication analysis of the selected SNP. The strength of the SNP-trait association between cohorts was slightly different, but there was no significant heterogeneity. As well as intraindividual variation of sperm motility between individual samples, the measurement of sperm motility may have variability by operators (individual technicians). To reduce the between-centre variability, technicians from each centre were initially trained by one technician from St. Marianna University in Kawasaki. In addition, to statistically reduce the influence of differences in semen assessment between the centres, we added each centre as a covariate and further conducted an association analysis between sperm motility and rs3791686. We found that rs3791686 was associated with sperm motility in the discovery stage ($β=−4.35$, $P=1.62 \times 10^{-7}$) and in the replication stage ($β=−3.16$, $P=0.012$). When we combined two results using meta-analysis, rs3791686 was genome-wide significantly associated with sperm motility ($β=−3.99$, $P=6.60 \times 10^{-5}$). This finding was very similar to the result (table 2) from the association analysis without adjustment for semen analysis centre. Although the measurements of the semen analysis may not necessarily be representatives of individual sperm motility, together with the previous finding of Sertoli cell-specific $Erbb4$ knockout mice, we are confident that the results of our GWAS are valid.

In conclusion, this first two-staged GWAS for sperm motility identifies a novel sperm motility-associated locus at $ERBB4$ on 2q34. The genetic evidence suggests that $ERBB4$ is a promising candidate for future association studies in diverse populations with larger sample sizes. Further studies such as fine-scale genetic mapping are needed to uncover a functional variant at this locus as well as the underlying molecular mechanism.

Acknowledgements We thank all the volunteers who participated in this study. We are grateful to the late Professor Yukata Nakahori and Professors Etietseh Koh, Jiro Kanaya, Mikio Namiki, Kiyoumi Matsumiya, Akira Tsujimura, Kyoshi Komatsu, Naoki Itoh and Yuichi Eguchi for collecting blood samples from the participants. We also thank Professor Toyomasa Katagiri for his assistance with the AB GeneKpM PCR System 9700.

Contributors YS and AT conceived, designed the experiments, performed the experiments and wrote the paper. TS performed the imputation analysis. SN, MY and TI prepared and collected samples. IT contributed material and analysis tools. YS, AT, TS, II, AY and TI reviewed and revised the manuscript.

Funding This study was supported in part by the Ministry of Health and Welfare of Japan (1013201) (to TI), Grant-in-Aids for Scientific Research (C) (26462461) (to YS), (23510242) (to AT) and Grant-in-Aids for Scientific Research (B) (17H03311) (to YS), (15H04320) (to AT) from the Japan Society for the Promotion of Science, the European Union (BM4-CT96-0314) (to TI), the Takeda Science Foundation (to AT) and The Suzuki Urinary Foundation (to YS).

Competing interests None declared.

Patient consent Obtained.

Ethics approval This study was approved by the ethics committees of the University of Tokushima and St. Marianna Medical University. All participants provided written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1 Skakkebaek NE, Giwercman A, de Kretser D. Pathogenesis and management of male infertility. Lancet 1994;343:1473–9.
Complex traits

McLachlan RJ, de Kreter DM. Male infertility: the case for continued research. Med J Aust 2001;174:116–7.

Agarwal A, Virk G, Ong C, du Plessis SS. Effect of oxidative stress on male reproduction. World J Mens Health 2014;32:1–17.

Opower CS, Henkel RR. An update on oxidative damage to spermatozoa and oocytes. Biomed Res Int 2016;2016:1–11.

Sato Y, Iwamoto T, Shinko T, Nozawa S, Yoshikawa M, Koh E, Kanaya J, Namiki M, Matsumiya K, Tsujimura A, Komatsu K, Ishii N, Eguchi J, Imoto Y, Iamauchi Y. A replication study of a candidate locus for follicle-stimulating hormone levels and association analysis for semen quality traits in Japanese men. J Hum Genet 2016;61:911–5.

World Health Organization. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. 4th edn. Cambridge: Cambridge University Press, 1999.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Shincuk A: a tool for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.

Brown SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotyping clusters. Am J Hum Genet 2007;81:1084–97.

Browning BL, Browning SR. Imputation of millions with references of missing samples. Am J Hum Genet 2016;98:116–26.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68–74.

Sudmant PH, Rausch T, Gardiner EA, Handsaker RE, Abazovic A, Huddleston J, Yang Y, Kuhn G, Fritz MH, Kong X, Mahler A, Stutz AM, Shi X, Cusack BR, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MR, Walker M, McGeary S, Kashin S, Garrison E, Auton A, Lam HYK, Xu M, Alkan C, Antaki D, Bae T, Cervera E, Chines P, Chong Z, Clarke L, Dale D, Ding L, Emery S, Fan X, Guha M, Kahveci F, Kidd KM, Yong T, Lanejier EW, McCarthy S, Flicek P, Gibbs RA, March R, Mason CE, Melaneau A, Muzny DM, Nelson BJ, Noon A, Parrish N, Pendleton N, Pratadano A, Raeder B, Schadt EE, Romanovitch M, Schlaat H, Sebra R, Shabalin AA, Untergrasser A, Walker JA, Wang M, Yu F, Zhang C, Jiang Z, Zheng-Bradley X, Zhou W, Zichiner T, Sebat J, Baxter MA, McCarroll SA, Mills ER, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korbel JO; 1000 Genomes Project Consortium. An integrated map of structural variation in 2,504 human genomes. Nature 2015;526:75–81.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gledt TP, Boehnke M, Abecasis GR, Willer CJ, LocZumob: regional visualization of genomic association scan results. Bioinformatics 2010;26:2336–7.

GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2015;47:508–513.

Ware LD, Kellis M. HaplReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–40.

Boyle AP, Hong EL, Haritunian V, Cheng Y, Schaub MA, Kasowski M, Karcewski LJ, Park J, Hitzi BC, Weng S, Cheng JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790–7.

Chen G, Cui S, Chang S, Zhang L, Wang G, Wang J, Jiang X, iPSGEWAS, TCGA: a server for identification of pathways/genome sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic Acids Res 2010;38:W90–5.

Mante BM, Jeschke M, Graus-Porta D, Taverna D, Hofer P, Groner B, Yarden Y, Hynes NE. Neu differentiation factors/hegulin modulates growth and differentiation of HC11 mammary epithelial cells. Endocrinology 1995;134:14–19.

Yang Y, Spitzer E, Meyer D, Sachs M, Niemann C, Hartmann G, Weidner KM, Birchmeier C, Birchmeier W. Sequential requirement of hepatocyte growth factor and neurigulin in the morphogenesis and differentiation of the mammary gland. J Cell Biol 1995;129:215–26.

Carpenter G, Liao H. Tracking of receptor tyrosine kinases to the nucleus. Exp Cell Res 2009;315:115–20.

Roskoski R. The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacol Res 2014;79:34–74.

Sato Y et al. J Med Genet: first published as 10.1136/jmedgenet-2017-104991 on 16 February 2018. Downloaded from http://jmg.bmj.com/ on 20 July 2018 by guest. Protected by copyright.
selection inferred from genetic associations with polycystic ovary syndrome. Nat Commun 2015;6:8464.