nature COMMUNICATIONS

ARTICLE

Received 23 Aug 2013 | Accepted 11 Apr 2014 | Published 23 May 2014

DOI: 10.1038/ncomms4857

Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men

Zhibin Hu[1,2,*], Zheng Li[3,*], Jun Yu[1,4,*], Chao Tong[5,*], Yuan Lin[1,2,*], Xuejiang Guo[1,4,*], Feng Lu[2], Jing Dong[2], Yankai Xia[1,6], Yang Wen[1,2], Hao Wu[1,4], Honggang Li[7], Yong Zhu[3], Ping Ping[3], Xiangfeng Chen[3], Juncheng Dai[2], Yue Jiang[1,2], Shandong Pan[2], Peng Xu[8], Kailing Luo[9], Qiang Du[10], Bing Yao[11], Ming Liang[12], Yaoting Gui[13], Ning Weng[8], Hui Lu[3], Zhuqing Wang[3], Fengbin Zhang[14], Xiaobin Zhu[3], Xiaoyu Yang[1,15], Zhou Zhang[16], Han Zhao[17], Chenliang Xiong[7], Hongxia Ma[2], Guangfu Jin[2], Feng Chen[2], Jianfeng Xu[18], Xinru Wang[1,6], Zuomin Zhou[1,4], Zi-Jiang Chen[17], Jiayin Liu[1,15], Hongbing Shen[1,2] & Jiahao Sha[1,4]

Male factor infertility affects one-sixth of couples worldwide, and non-obstructive azoospermia (NOA) is one of the most severe forms. Our previous genome-wide association study (GWAS) identified three susceptibility loci for NOA in Han Chinese men. Here we test promising associations in an extended three-stage validation using 3,608 NOA cases and 5,909 controls to identify additional risk loci. We find strong evidence of three NOA susceptibility loci ($P<5.0\times10^{-8}$) at 6p21.32 (rs7194, $P=3.76\times10^{-19}$), 10q25.3 (rs7099208, $P=6.41\times10^{-14}$) and 6p12.2 (rs13206743, $P=3.69\times10^{-8}$), as well as one locus approaching genome-wide significance at 1q42.13 (rs3000811, $P=7.26\times10^{-8}$). In addition, we investigate the phenotypic effect of the related gene (*gek*, orthologous to *CDC42BPA*) at 1q42.13 on male fertility using a *Drosophila* model. These results advance our understanding of the genetic susceptibility to NOA and provide insights into its pathogenic mechanism.

[1] State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 210029, China. [2] Department of Epidemiology and Biostatistics and Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [3] Shanghai Human Sperm Bank, Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China. [4] Department of Histology and Embryology, Nanjing Medical University, Nanjing 210029, China. [5] Institute of Preventive Veterinary Medicine, Zhejiang Provincial Key Laboratory of Preventive Veterinary Medicine, Zhejiang University, Hangzhou 310058, China. [6] Department of Toxicology and Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [7] Family Planning Research Institute, Center of Reproductive Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430032, China. [8] Jinghua Hospital, Shenyang Dongfang Medical Group, Shenyang 110004, China. [9] Reproductive Medical Center, The Second Hospital of Nanning, Nanning 530022, China. [10] Department of Reproduction, Shengjing Hospital, China Medical University, Shenyang 110004, China. [11] Department of Andrology, Nanjing Jinling Hospital, Nanjing 210029, China. [12] The Second Affiliated Hospital, Shandong University of Traditional Chinese Medicine, Shandong 250033, China. [13] Guangdong Key Laboratory of Male Reproductive Medicine and Genetics, Peking University Shenzhen Hospital, FuTian District, Shenzhen 518035, China. [14] The Women's Hospital, School of Medicine, Zhejiang University, Hangzhou 310006, China. [15] Center of Clinical Reproductive Medicine, First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, China. [16] Maternal and Child Health Hospital of Shanxi Province, Xian 300013, China. [17] Center for Reproductive Medicine, Shandong Provincial Hospital, Shandong University, Jinan 250021, China. [18] Department of Urology, Huashan Hospital, Fudan University, Shanghai 200052, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.H. (email: zhibin_hu@njmu.edu.cn) or to Z.-J.C. (email: chenzijiang@hotmail.com) or to J.L. (email: jyliu_nj@126.com) or to J.S.(email: shajh@njmu.edu.cn).

Infertility is one of the most frequently diagnosed diseases in reproductive health and male-related problems that account for approximately half of all infertility cases[1,2]. A significant proportion of male infertility is accompanied by idiopathic azoospermia, most often presenting as non-obstructive azoospermia (NOA), which occurs in ~1% of all adult men[3]. A few studies have reported that genetic factors including chromosome number defects, Y-chromosome microdeletions and autosomal mutations or polymorphisms in multiple biological pathways are involved in the development of NOA[4–6].

We recently conducted a multistage genome-wide association study (GWAS) of NOA in Han Chinese based on genotyping 587,347 single nucleotide polymorphisms (SNPs) in 981 NOA cases and 1,657 controls with a two-stage validation using 1,946 NOA cases and 4,077 controls[7]. We identified three well-replicated susceptibility loci (rs12097821 at 1p13.3 for *PRMT6*, rs2477686 at 1p36.32 for *PEX10* and rs10842262 at 12p12.1 for *SOX5*)[7].

Here we evaluate promising associations in an extended three-stage validation using 3,608 NOA cases and 5,909 controls, and we focus on the SNPs that have *P* values ranging from 10^{-5} to 10^{-7} in the GWA scan (Supplementary Fig. 1 and Supplementary Data 1). In addition, we knocked down the genes that map to susceptibility loci that also have *Drosophila* homologues to investigate the phenotypic effect on male fertility. We find strong evidence of three NOA susceptibility loci ($P<5.0\times10^{-8}$) at 6p21.32 (rs7194 in *HLA-DRA*, $P=3.76\times10^{-19}$), 10q25.3 (rs7099208 close to *ABLIM1*, $P=6.41\times10^{-14}$) and 6p12.2 (rs13206743 between *MIR133B* and *IL17A*, $P=3.69\times10^{-8}$), as well as one locus approaching genome-wide significance (rs3000811 at 1q42.13, upstream of *CDC42BPA*, $P=7.26\times10^{-8}$). In addition, we show the phenotypic effect of the related gene (*gek*, orthologous to *CDC42BPA*) on male fertility by using *Drosophila* as an animal model. These results advance our understanding of the susceptibility to NOA in Chinese men.

Results

Susceptibility loci of NOA in validation studies. Seventy-seven SNPs met the selection criteria for the first-stage validation (Methods, Supplementary Fig. 1 and Supplementary Data 1). Additive models of logistic regression analyses were used to estimate the *P* values of association analyses. The 77 SNPs identified in the GWA scan and the three validations are shown in Table 1, Supplementary Fig. 1 and Supplementary Data 1. Among them, 9 SNPs (rs12023502, rs3000811, rs13206743, rs7194, rs7099208, rs4903393, rs1990264, rs4791224 and rs6055276) were consistently replicated in 1,144 NOA cases and 2,373 male controls with the same direction of significant associations as those in the GWA scan (Table 1 and Supplementary Data 1). For the second-stage validation, additional 1,662 NOA cases and 2,535 male controls were genotyped to verify the significant associations of the 9 loci; 6 of the 9 loci (rs3000811, rs13206743, rs7194, rs7099208, rs1990264 and rs4791224) were consistently associated with NOA risk (Table 1 and Supplementary Data 1). We then included an independent GWAS of NOA in Han Chinese[8] as the third-stage validation with 802 NOA cases and 1,001 male controls to confirm our associations. Four SNPs (rs3000811, rs7194, rs7099208 and rs13206743) were associated with NOA risk with the same direction as those in our GWA scan and the first two validations (Table 1, Supplementary Fig. 1 and Supplementary Data 1).

Combined analysis of the susceptibility loci. In the combined analysis, meta-analysis was used to combine the results of GWAS scan and three validations. Three SNPs reached genome-wide significance ($P<5.0\times10^{-8}$) for NOA susceptibility in 4,589 cases and 7,566 controls; these SNPs are located at 6p12.2 (rs13206743, $P_{combined}=3.69\times10^{-8}$, odds ratio (OR) = 1.35), 6p21.32 (rs7194, $P_{combined}=3.76\times10^{-19}$, OR = 1.30) and 10q25.3 (rs7099208, $P_{combined}=6.41\times10^{-14}$, OR = 1.41). Another SNP (rs3000811 at 1q42.13) has a *P* value approaching the genome-wide significance ($P_{combined}=7.26\times10^{-8}$, OR = 1.19) (Table 1). After examining heterogeneity in ORs across the initial GWAS and validation stages, no significant heterogeneity was observed for the four SNPs.

Stratified analyses by study regions. As the subjects were collected from different regions that are geographically distant and most likely genetically differentiated, we performed stratified analyses by four study regions (Southeastern China: Nanjing and Shanghai; Central China: Wuhan; Northern China: Shenyang; and Southern China: Nanning). Similar association strengths were shown between regions in absence of significant heterogeneity for each locus (Supplementary Table 1).

Imputation analysis in the GWAS. An imputation analysis in the GWA scan of 981 cases and 1,657 controls identified associations of SNPs with NOA risk at $P\leq1.0\times10^{-5}$ (imputed $r^2>0.3$, quality threshold >0.9, minor allele frequency >0.05, located at 500 kb up- or downstream of the four lead SNPs). For 6p21.32, a series of SNPs reaching $P\leq1.0\times10^{-5}$ were in strong linkage disequilibrium (LD) with rs7194 (six SNPs: $r^2=0.997$–0.998, $P=1.51\times10^{-6}$–3.99×10^{-6}, Fig. 1a and Supplementary Table 2). However, for 10q25.3, 6p12.2 and 1q42.13, we did not identify any untyped SNPs in high LD with the lead SNP reaching $P\leq1.0\times10^{-5}$ (Fig. 1b–d and Supplementary Table 2).

Functional relevance of the *Drosophila CDC42BPA* orthologue. Rs3000811 is located upstream of *CDC42BPA* (CDC42-binding protein kinase alpha), which has an orthologous gene named *gek* (*CG4012*) in *Drosophila*. We then performed phenotypic analysis on its related gene in *Drosophila* using GAL4-driven UAS-short hairpin RNA system. We used *nos-GAL4*, *bam-GAL4* and *c729-GAL4* to drive *UAS-gek^{shRNA}* expression, and we found that *nos>gek RNAi* and *bam>gek RNAi* males were fertile and most *c729>gek RNAi* males were infertile at 28 °C. To confirm the results, the single male fertility tests (one RNAi F1 male and three wild-type virgins in a tube) were conducted; all *nos>gek RNAi* ($n=50$) and *bam>gek RNAi* ($n=50$) males were fertile, and 96.15% ($n=52$) of *c729>gek RNAi* males were found to be completely infertile (Fig. 2a). Dissection results showed that ~54.35% ($n=46$) of the *c729>gek RNAi* testes have each stage of germ cells (Fig. 2b–k), but only 2.17% have mature motile sperms (Fig. 2l,m). Nuclear dye staining also showed decreased elongated spermatid bundles in the *c729>gek RNAi* testes (Fig. 2n–p). Analysis of testicular *gek* mRNA expression showed significant suppression of *gek* in all three types of *gek RNAi* testes (Fig. 2q). For the other three SNPs (rs13206743, rs7194 and rs7099208) showing significant associations with NOA, we also analysed genes adjacent to these SNPs and did not find orthologous genes in *Drosophila*.

Discussion

During manuscript preparation, an independent GWAS of NOA in Han Chinese was conducted, involving 802 azoospermia cases (from Shandong province in Northern China based on Illumina OmniExpress BeadChip) and 1,863 controls (1,000 men from Shanghai in Southeastern China based on the Illumina OmniExpress BeadChip and 863 women from Shandong province based on the Affymetrix SNP Array 6.0) for discovery and a two-stage validation of 1,424 cases and 2,713 controls[8]. Interestingly, Zhao

Table 1 | Association of 4 SNPs to azospermia in the GWA scan and validation studies.

Chr. (cytoband)	SNP	Associated gene	Study	Cases*	Controls*	MAF†		OR_{add}	P_{add}‡	P§
						Cases	Controls	(95% CI)‡		
1q42.13	rs3000811 A/G‖	*CDC42BPA*	GWAS	47/430/498	81/533/1043	0.27	0.21	1.40(1.22–1.60)	1.52×10^{-6}	
			Validation I	56/398/690	88/778/1506	0.22	0.20	1.14(1.01–1.29)	3.37×10^{-2}	
			Validation II	92/584/977	115/850/1565	0.23	0.21	1.12(1.00–1.24)	4.25×10^{-2}	
			Validation III	27/276/443	45/273/616	0.22	0.19	1.18(1.00–1.39)	5.65×10^{-2}	
			Combined all¶	222/1688/2608	329/2434/4730	0.24	0.21	1.19(1.12–1.27)	7.26×10^{-8}	0.065
6p12.2	rs13206743 T/C‖	*MIR133B-L17A*	GWAS	58/45/858	1/139/1515	0.08	0.04	1.68(1.36–2.07)	1.11×10^{-6}	
			Validation I	32/99/1013	43/175/2155	0.07	0.05	1.23(1.03–1.47)	1.97×10^{-2}	
			Validation II	7/161/1493	4/203/2316	0.05	0.04	1.27(1.04–1.56)	2.15×10^{-2}	
			Validation III	0/64/738	1/61/939	0.04	0.03	1.29(0.90–1.84)	1.69×10^{-1}	
			Combined all¶	97/369/4102	49/578/6925	0.06	0.04	1.35(1.22–1.51)	3.69×10^{-8}	0.131
6p21.32	rs7194 T/C‖	*HLA-DRA*	GWAS	94/421/466	90/631/935	0.31	0.24	1.37(1.21–1.56)	1.43×10^{-6}	
			Validation I	123/462/559	160/915/1296	0.31	0.26	1.27(1.14–1.41)	2.13×10^{-5}	
			Validation II	157/674/828	172/935/1414	0.30	0.25	1.24(1.13–1.37)	1.23×10^{-5}	
			Validation III	88/344/370	57/395/549	0.32	0.25	1.41(1.22–1.64)	4.01×10^{-6}	
			Combined all¶	462/1901/2223	479/2876/4194	0.31	0.25	1.30(1.23–1.38)	3.76×10^{-19}	0.404
10q25.3	rs7099208 T/C‖	*ABLIM1*	GWAS	7/236/698	10/267/1380	0.13	0.09	1.65(1.37–2.00)	1.94×10^{-7}	
			Validation I	9/242/892	8/374/1991	0.11	0.08	1.46(1.23–1.73)	1.32×10^{-5}	
			Validation II	20/330/1312	10/439/2086	0.11	0.09	1.27(1.09–1.47)	1.57×10^{-3}	
			Validation III	7/135/650	7/129/855	0.09	0.07	1.33(1.05–1.69)	1.90×10^{-2}	
			Combined all¶	43/943/3552	35/1209/6312	0.11	0.08	1.41(1.29–1.54)	6.41×10^{-14}	0.178

chr., chromosome; CI, confidence interval; GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphisms.
*Variant homozygote/heterozygote/wild-type homozygote.
†MAF.
‡Adjusted for age and the first principal component. OR_{add} and P_{add} were estimated in additive model.
§*P*: heterogeneity test between groups.
‖Major/minor alleles.
¶ORs (95% CI) and *P*-values are derived from meta-analysis.

et al.[8] found four SNPs residing in a 144-kb genomic region at 6p22 around *HLA-DRA* (major histocompatibility complex, class II, DR alpha) to be associated with NOA risk at genome-wide significant level. These variants were suggested to mediate the response to testicular microenvironmental antigens and cause testicular azoospermia through autoimmune inflammatory responses[8]. The lead SNP rs7194 in the current study was close to and in high LD with the reported SNPs in the GWAS by Zhao *et al.*[8] (rs7194 is only 834 base pair apart from and in complete LD with rs7192). Supportive evidence was also reported in previous candidate gene-based studies in which *HLA-DRB1* and -*DQB1* alleles were found to be associated with NOA in Japanese patients[9,10]. A recently published study showed strong association of HLA-DPB1*04:01 with Japanese NOA[11].

To further assess the association of *HLA* region with NOA, we performed an imputation of classical *HLA* alleles *in silico* and analysed their associations with NOA in the samples of the initial GWAS stage. Nine alleles other than *HLA-A* and *HLA-DPB1* alleles were associated with NOA, with *P*-values <0.05. All of the *HLA* alleles were in weak LD with rs7194, suggesting that the association of rs7194 with NOA risk may be independent from those classical *HLA* alleles (Supplementary Table 3). However, these results need to be confirmed in large samples with experimental data. In expression quantitative trait locus (eQTL) analysis (Methods), rs7194 was significantly associated with the mRNA expression levels of a series of genes coding major histocompatibility complex (that is, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*, *HLA-C*, *HLA-DQA1* and *HLA-DQB1*) in lymphoblastoid cell lines or monocytes as a *cis*-acting element. In addition, rs7194 could influence the expression of *C21orf127* and *ERG* (v-ets, erythroblastosis virus E26 oncogene homologue) as a *trans*-factor in monocytes. However, these results are preliminary and warrant further investigations.

The SNP rs7099208 is within the last intron of *FAM160B1* (family with sequence similarity 160, member B1) without known function and is 40 kb upstream of *TRUB1* (TruB pseudouridine (psi) synthase family member 1) at 10q25.3. The *TRUB1* products could be the human transfer RNA psi synthases, and its mRNA is reported to be widely expressed in various human tissues (especially heart, skeletal muscle and liver)[12]; however, no data to date are available regarding its function in spermatogenesis. This lead SNP is also 210 kb upstream of *ABLIM1* (actin-binding LIM protein 1). ABLIM1 was originally found in human retina and was postulated to regulate actin-dependent signalling[13]. In the testis, ABLIM1 may function in ectoplasmic specialization (ES), a testis-specific adherens junction. ES is known to be important for the connection between adjacent Sertoli cells and the formation of blood–testis barrier (BTB), as well as for the connection between Sertoli cells and elongating spermatids[14]. Moreover, silencing *ABLIM1* in RPE1 cells leads to spontaneous ciliogenesis and significantly promotes cilia elongation[15]. Cilia and sperm flagella both have microtubule-based axoneme structure[16], so *ABLIM1* may also regulate flagellum development during spermatogenesis.

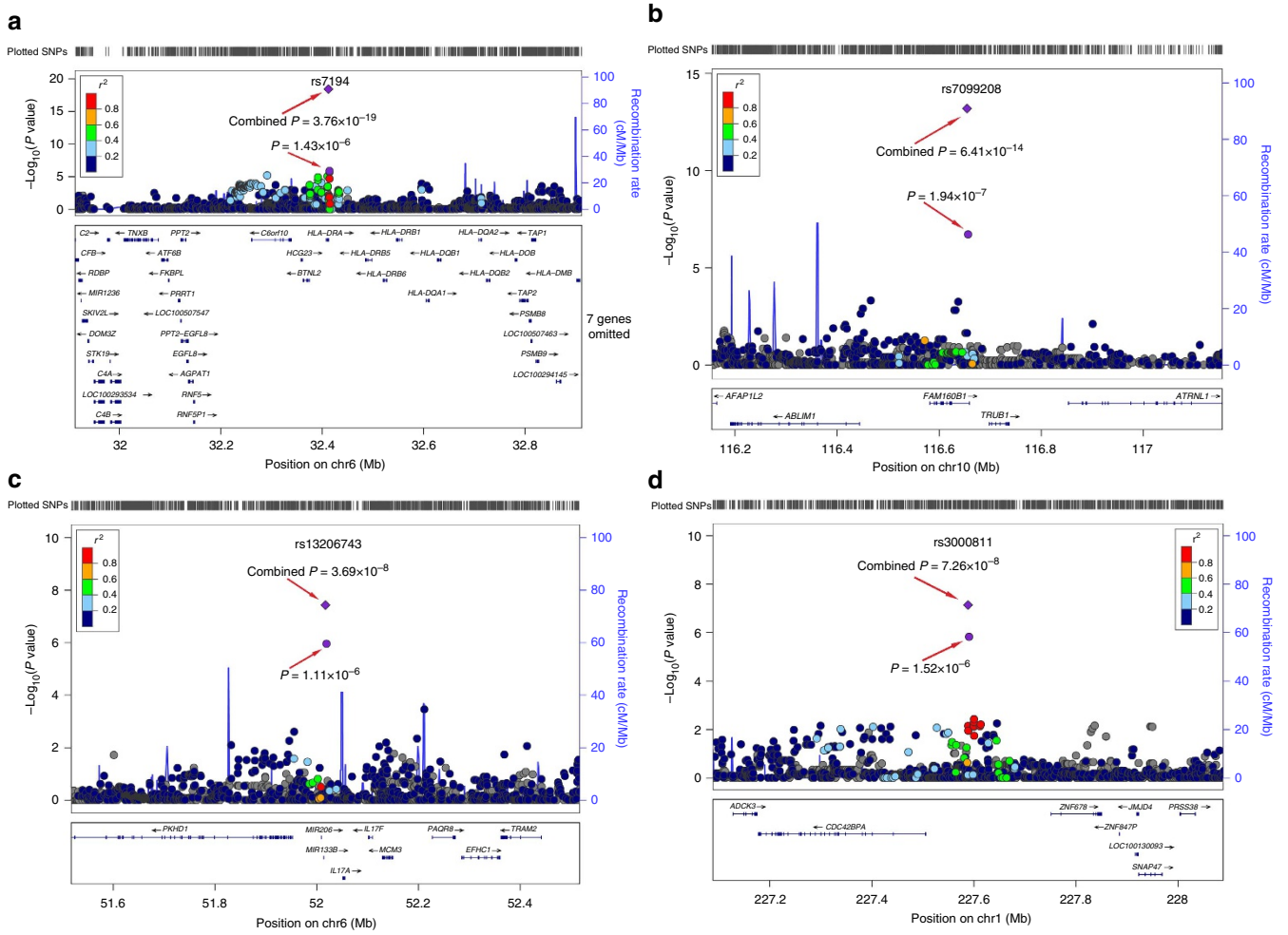

Figure 1 | Regional plot of the four identified lead SNPs. These SNPs are rs7194 at 6p21.32, rs7099208 at 10q25.3, rs13206743 at 6p12.2 and rs3000811 at 1q42.13. Additive models of logistic regression analyses were used to estimate the *P* values of association analyses in 981 NOA cases and 1,657 controls. The results ($-\log_{10} P$) are shown for SNPs in the region flanking 500 kb on either side of the marker SNPs. The marker SNPs are shown in purple, and the r^2 values of the rest of the SNPs are indicated by different colours. The genes within the region of interest are annotated, with arrows indicating transcription direction. For each plot, the recombination rates (right *y* axes) of the region are shown according to their chromosomal positions (*x* axis).

We also identified associations with NOA for rs13206743, located 3 and 7 kb downstream of *MIR133B* and *MIR206*, respectively. There is no correlative literature regarding the two microRNAs and spermatogenesis. The lead SNP is 34 kb upstream of interleukin *(IL)-17A* and 84 kb downstream of *IL-17F*. Studies have proven that testicular inflammatory disorders leading to impairment of spermatogenesis are the primary reasons for male infertility, and T helper cells play an essential role in this condition[17–19]. T helper 17 cells have been identified as a distinct lineage of CD4+ T cells and are characterized by the production of IL-17A, IL-17F and many other cytokines[20,21]. *IL-17* is expressed in normal human testis at base line levels, suggesting that *IL-17* may be involved in the maintenance of testicular immune privilege and spermatogenesis. However, an increased expression of *IL-17A* could be detected in azoospermic testis with chronic inflammation. Thus, overexpression of *IL-17A* could substantially damage the BTB and most likely destroy normal spermatogenesis and germ cells, which in turn could lead to azoospermia[22]. In addition, the SNP rs13206743 is ~65 kb upstream of *PKHD1* (polycystic kidney and hepatic disease 1), which may serve as the *PKHD1* promoter polymorphism. This gene is a disease-causing gene for autosomal recessive polycystic kidney disease[23].

Another lead SNP, rs3000811, lies in the 670 kb interval between *CDC42BPA* and *ZNF678* (zinc-finger protein 678). We identified *gek*, which is an orthologous gene of *CDC42BPA*, as playing a phenotypic effect on male fertility in *Drosophila*. Our results showed that inhibiting expression of *gek* could lead to male infertility. *C729-GAL4* is expressed in cyst cells, testis sheath and male accessory gland[24,25]. These findings suggest that the function of *gek* in soma is critical for male fertility. Although suppressing *gek* expression in germ cells with *nos-GAL4* and *bam-GAL4* does not affect male fertility, it is unclear whether *gek* is required in germ cells. Analysing null alleles of *gek* in the future is critical to solve the issues. According to these results, the function of *CDC42BPA* in human's NOA is worth deep exploration. The protein encoded by *CDC42BPA* is a member of the serine/threonine protein kinase family and has been shown to bind CDC42. It may function as a CDC42 downstream effector mediating CDC42-induced peripheral actin formation and promoting cytoskeletal reorganization[26]. CDC42 has been reported to regulate cell adhesion and junction function through regulating filopodia formation, protein trafficking and actin cytoskeleton[27].

In mammalian testes, extensive junction restructuring occurs in the seminiferous epithelium at the Sertoli–Sertoli and Sertoli–

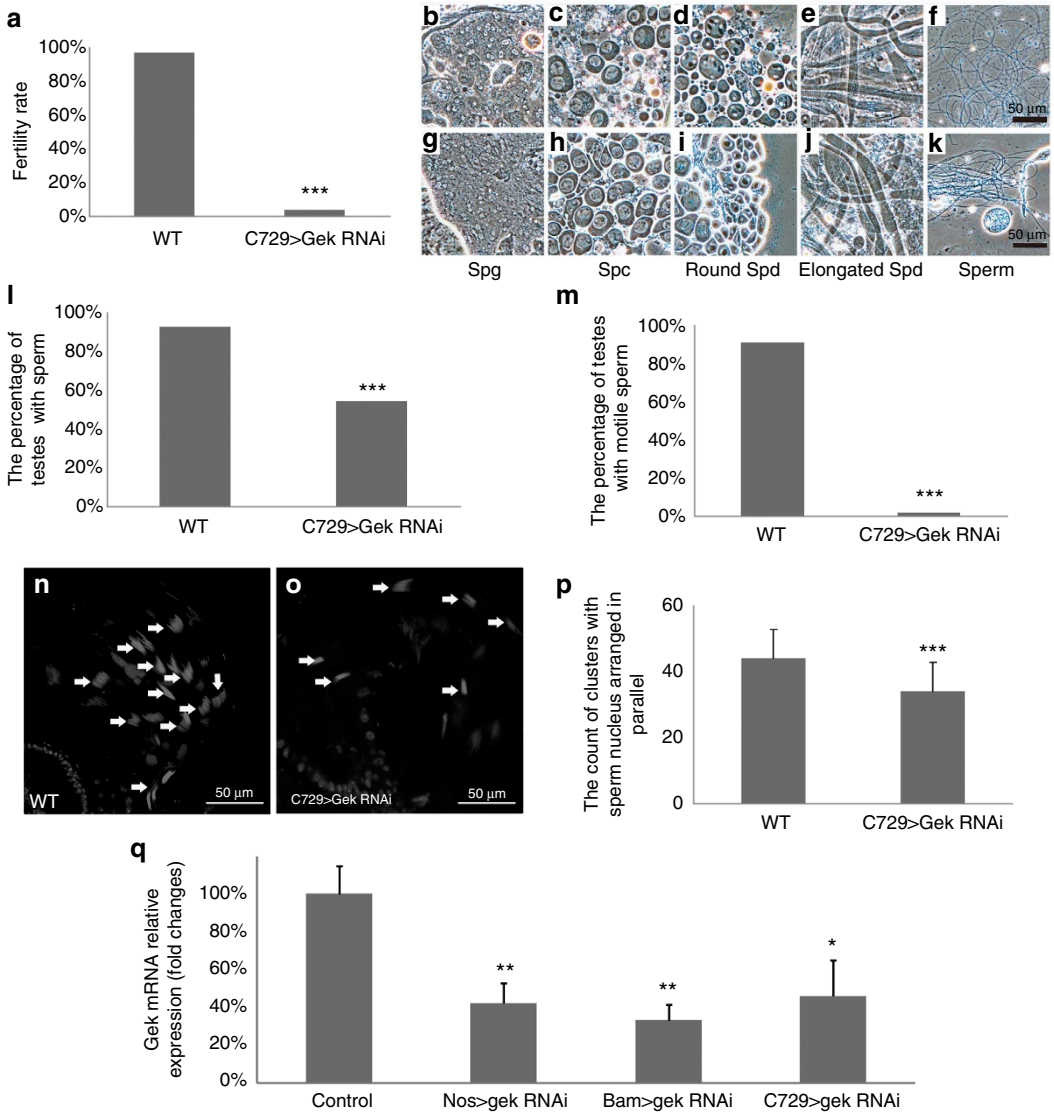

Figure 2 | *Gek* knockdown resulted in decreased male fertility. (**a**) A single male fertility test indicated that only 3.85% ($n=52$) of *c729>gek RNAi* males are fertile compared with 96.9% ($n=65$) of wild-type flies at 28 °C. Each stage of the germ cells could be found in the testes of wild-type (**b**–**f**) and *c729>gek RNAi* flies (**g**–**k**) (spg, spermatogonia; spc, spermatocyte; round spd, round spermatid; elongated spd, elongated spermatid). Only 54.35% ($n=46$) of *c729>gek RNAi* testes have morphologically mature sperm (**l**) and only 2.17% ($n=46$) of *c729>gek RNAi* testes have motile sperm (**m**). Hoechst staining (white) showed decreased elongated spermatid bundles (arrows) in the *c729>gek RNAi* testis (**n**) compared with that in wild-type testis (**o**). (**p**) Quantification of elongated spermatid bundles in the wild-type testes (43.97 ± 8.70, $n=62$) and the *c729>gek RNAi* testes (33.93 ± 8.78, $n=42$). χ^2 tests were used for the comparisons of ratios, and Student's *t*-test was used for the comparison of the quantification of elongated spermatid bundles. Significant differences are indicated by (***$P<0.001$). (**q**) Efficiency of *gek RNAi* knockdown evaluated by quantitative real-time PCR in *nos>gek RNAi*, *bam>gek RNAi* and *c729>gek RNAi* testes (Student's *t*-test: *$P<0.05$; **$P<0.01$). Error bars represent the s.d.

germ cell interface to facilitate the different cellular events of spermatogenesis, such as mitosis, meiosis, spermiogenesis and spermiation[14]. Recent studies have shown that CDC42 and components of the polarity protein complexes work in concert with laminin fragments, cytokines and testosterone to regulate the events of cell–cell interactions in the seminiferous epithelium via a local autocrine-based regulatory loop known as the apical ES–BTB–basement membrane axis. This functional axis coordinates various cellular events during different stages of the seminiferous epithelium cycle of spermatogenesis[28].

In conclusion, we identified four susceptibility loci (6p21.32, 10q25.3, 6p12.2 and 1q42.13) of NOA in the Chinese population, extending our previous findings and advancing the understanding of NOA susceptibility.

Methods

Study design. This study was approved by the institutional review boards of each participating institution, including the institutional review board from Nanjing Medical University, Renji Hospital, Center of Reproductive Medicine, Tongji Medical College of Huazhong University of Science and Technology, Jinghua Hospital, the second hospital of Nanning, Shengjing Hospital, Nanjing Jinling Hospital, the Second Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Peking University Shenzhen Hospital, Shandong Provincial Hospital and Huashan Hospital.

We performed a four-stage case-control analysis. The GWAS phase included 981 NOA cases recruited from the Nanjing Center of Reproductive Medicine and 1,657 male controls from Nanjing. For validations, we enlarged the sample size by testing 1,144 NOA cases and 2,373 healthy male controls (745 cases and 1,444 controls from Southeastern China: Nanjing and Shanghai; 106 cases and 527 controls from Central China: Wuhan; and 293 cases and 402 controls from Northern China: Shenyang) for the first-stage validation, and additional 1,662 NOA cases and 2,535 healthy controls (559 cases and 534 controls from

Southeastern China: Nanjing and Shanghai; 642 cases and 936 controls from Central China: Wuhan; 349 cases and 641 controls from Northern China: Shenyang; and 112 cases and 424 controls from Southern China: Nanning) for the second-stage validation. The 1,817 cases and 2,997 controls from the previous study were included in the first two validation studies. The 802 NOA cases and 1,001 male controls of the third-stage validation were from another independent GWAS of NOA in Han Chinese[8].

All infertile male subjects were genetically unrelated Han Chinese men and selected based on an andrological examination, including examination of medical history, physical examination, semen analysis, scrotal ultrasound, hormone analysis, karyotyping and Y-chromosome microdeletion screening. Those with a history of cryptorchidism, vascular trauma, orchitis, obstruction of the vas deferens, abnormalities in chromosome number or microdeletions of the azoospermia factor region on the Y chromosome were excluded from the study. Semen analysis for sperm concentration, motility and morphology was performed following World Health Organization (WHO) criteria (1999). Subjects with NOA had no detectable sperm in the ejaculate after evaluation of the centrifuged pellet. To differentiate from OA, only idiopathic azoospermic patients with small and soft testis, normal fructose and neutral alpha glucosidase in seminal plasma were included in the study. Those with a history of vasectomy were excluded. To ensure the reliability of the diagnosis, each individual was examined twice, and the absence of spermatozoa from both replicate samples was taken to indicate azoospermia. The GWAS control subjects were shared with the Nanjing Lung Cancer Study[29], and all of these males had fathered one or more healthy children. In addition, the controls for the validation stage had fathered one or more healthy child. A 5-ml sample of whole blood was obtained from each participant as a source of genomic DNA for further genotyping analysis and all participants would complete the informed consent in written before taking part in this research.

SNP selection and genotyping for validation. Because the control group was shared with the Nanjing Lung Cancer Study, the NOA cases were compared with two control groups separately or in combination with the discovery phase to avoid potential effects from lung cancer[7]. We selected SNPs meeting the following criteria for the first-stage validation: (i) SNPs had $1.0\times10^{-7}<P\leq1.0\times10^{-5}$ in the comparison between 981 NOA cases and 1,657 controls and had a consistent association at $P\leq1.0\times10^{-2}$ when the NOA cases were compared separately with the healthy male controls and the male lung cancer cases in additive models of logistic regression analyses; (ii) not located in the same chromosome region/gene of SNPs reported in our previous GWAS study; (iii) SNPs had clear genotyping clusters; and (iv) only the SNP with the lowest P value was selected when multiple SNPs were observed in a strong LD ($r^2\geq0.8$). A total of 77 SNPs met these criteria (Supplementary Data 1). Significantly associated SNPs ($P<0.05$) were further genotyped in validation samples.

Genotyping analyses were performed using the middle-throughput TaqMan OpenArray Genotyping Platform (Applied Biosystems, USA), the iPLEX Sequenom MassARRAY platform (Sequenom, CA, USA) and the TaqMan allelic discrimination Assay (Applied Biosystems). The primers and probes are available on request. The primers of nine SNPs using TaqMan allelic discrimination Assay to perform genotyping were provided in Supplementary Table 4. A series of methods were used to control the quality of genotyping: (i) case and control samples were mixed on each plate; (ii) genotyping was performed without knowing the case or control status; (iii) two water controls were used in each plate as blank controls; and (iv) five per cent of the samples were randomly selected for repeat genotyping.

Statistical analysis. The association analysis was performed using an additive model in a logistic regression analysis using the PLINK 1.07 program (http://pngu.mgh.harvard.edu/~purcell/plink/). Meta-analysis was performed in a combined analysis, and a random-effects model was used when heterogeneity between studies exists, that is, P value for heterogeneity test is <0.05; otherwise, a fixed-effect model was used. We used MACH 1.0 software (http://www.sph.umich.edu/csg/abecasis/MACH/index.html/) to impute ungenotyped SNPs using LD information from the hg18/1,000 Genomes Project database (CHB + JPT as a reference set; June 2010 release). The chromosome region was plotted using the LocusZoom 1.1 program (http://csg.sph.umich.edu/locuszoom/). Other analyses were performed with R software (version 2.11.1; The R Foundation for Statistical Computing; http://www.cran.r-project.org/). *HLA* alleles were predicted from dense SNP genotypes using the R software package, HIBAG (http://cran.r-project.org/web/packages/HIBAG/index.html).

eQTL analysis. We performed eQTLs analysis using the University of Chicago eQTL Browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/), which contains significant eQTLs identified in recent studies across multiple cell lines and populations.

Fly strains and genetics. The fly strains were obtained from the Bloominton Stock Center. Flies were cultured with standard culture medium. To enhance the knockdown efficiency, culture tubes were put into 28 °C incubators after egg laying. The male fertility tests were performed at room temperature. For single male fertility test, freshly hatched single males were crossed to multiple virgin Canton S flies and the presence of progeny (including larvae, pupae and adults) was scored after 10–15 days.

Testis analysis. Testis squashes were prepared in TB1 buffer (7 mM K_2HPO_4, 7 mM KH_2PO_4 [pH 6.7], 80 mM KCl, 16 mM NaCl, 5 mM $MgCl_2$, 1% PEG-6000) according to previous report[30]. Phase contrast light micrographs were taken, and mature and motile sperm were observed under a microscope. To count the elongated spermatid bundles, testes from freshly hatched males were dissected, fixed and stained with Hoechst 33342 dye. Confocal image stacks were taken, and elongated spermatid bundles were counted.

Gek mRNA expression in adult testes was analysed by quantitative real-time PCR with 18s mRNA as reference. The primers used were as follows, *gek* (forward), 5′-CGAAAGGGACTTCCGAGCAT-3′; *gek* (reverse), 5′-TCGGAGTTTGCAT CCACTTGT-3′; 18s RNA (forward), 5′-TGGTTTCCGGCAAGCTTCAA-3′; 18s RNA (reverse), 5′-ACTTCTTGAATCCGGTGGGC-3′.

References

1. De Kretser, D. M. & Baker, H. W. Infertility in men: recent advances and continuing controversies. *J. Clin. Endocrinol. Metab.* **84,** 3443–3450 (1999).
2. Dohle, G. R. *et al.* EAU guidelines on male infertility. *Eur. Urol.* **48,** 703–711 (2005).
3. Maduro, M. R. & Lamb, D. J. Understanding new genetics of male infertility. *J. Urol.* **168,** 2197–2205 (2002).
4. O'Flynn O'Brien, K. L., Varghese, A. C. & Agarwal, A. The genetic causes of male factor infertility: a review. *Fertil. Steril.* **93,** 1–12 (2010).
5. Dohle, G. R. *et al.* Genetic risk factors in infertile men with severe oligozoospermia and azoospermia. *Hum. Reprod.* **17,** 13–16 (2002).
6. Bashamboo, A. *et al.* Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *Am. J. Hum. Genet.* **87,** 505–512 (2010).
7. Hu, Z. B. *et al.* A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat. Genet.* **44,** 183–186 (2011).
8. Zhao, H. *et al.* A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *Am. J. Hum. Genet.* **90,** 900–906 (2012).
9. Matsuzaka, Y. *et al.* Susceptibility locus for non-obstructive azoospermia is localized within the HLA-DR/DQ subregion: primary role of DQB1*0604. *Tissue Antigens* **60,** 53–63 (2002).
10. Tsujimura, A. *et al.* Susceptibility gene for non-obstructive azoospermia located near HLA-DR and -DQ loci in the HLA class II region. *Hum. Genet.* **110,** 192–197 (2002).
11. Jinam, T. A. *et al.* HLA-DPB1*04:01 allele is associated with non-obstructive azoospermia in Japanese patients. *Hum. Genet.* **132,** 1405–1411 (2013).
12. Zucchini, C. *et al.* The human TruB family of pseudouridine synthase genes, including the Dyskeratosis Congenita 1 gene and the novel member TRUB1. *Int. J. Mol. Med.* **11,** 697–704 (2003).
13. Roof, D. J., Hayes, A., Adamian, M., Chishti, A. H. & Li, T. Molecular characterization of abLIM, a novel actin-binding and double zinc finger protein. *J. Cell Biol.* **138,** 575–588 (1997).
14. Wong, E. W., Mruk, D. D. & Cheng, C. Y. Biology and regulation of ectoplasmic specialization, an atypical adherens junction type, in the testis. *Biochim. Biophys. Acta* **1778,** 692–708 (2008).
15. Cao, J. *et al.* miR-129-3p controls cilia assembly by regulating CP110 and actin dynamics. *Nat. Cell Biol.* **14,** 697–706 (2012).
16. Zhou, J., Yang, F., Leu, N. A. & Wang, P. J. MNS1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet.* **8,** e1002516 (2012).
17. Weidner, W., Krause, W. & Ludwig, M. Relevance of male accessory gland infection for subsequent fertility with special focus on prostates. *Hum. Reprod. Update* **5,** 421–432 (1999).
18. Haidl, G., Allam, J. P. & Schuppe, H. C. Chronic epididymitis: impact on semen parameters and therapeutic options. *Andrologia* **40,** 92–96 (2008).
19. Schuppe, H. C. *et al.* Chronic orchitis: a neglected cause of male infertility? *Andrologia* **40,** 84–91 (2008).
20. Harrington, L. E. *et al.* Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type1 and 2 lineages. *Nat. Immunol.* **6,** 1123–1132 (2005).
21. Park, H. *et al.* A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin17. *Nat. Immunol.* **6,** 1133–1141 (2005).
22. Duan, Y. G. *et al.* Immunodeviation towards a Th17 immune response associated with testicular damage in azoospermic men. *Int. J. Androl.* **34,** e536–e545 (2011).
23. Onuchic, L. F. *et al.* PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. *Am. J. Hum. Genet.* **70,** 1305–1317 (2002).

24. Schulz, C. *et al.* A misexpression screen reveals effects of bag-of-marbles and TGF beta class signaling on the Drosophila male germ-line stem cell lineage. *Genetics* **167,** 707–723 (2004).
25. Hrdlicka, L. *et al.* Analysis of twenty-four Gal4 lines in *Drosophila* melanogaster. *Genesis* **34,** 51–57 (2002).
26. Wilkinson, S., Paterson, H. F. & Marshall, C. J. Cdc42-MRCK and Rho-ROCK signaling cooperate in myosin phosphorylation and cell invasion. *Nat. Cell Biol.* **7,** 255–261 (2005).
27. Ridley, A. J. *et al.* Cell migration: integrating signals from front to back. *Science* **302,** 1704–1709 (2003).
28. Wong, E. W. & Cheng, C. Y. Polarity proteins and cell-cell interactions in the testis. *Int. Rev. Cell Mol. Biol.* **278,** 309–353 (2009).
29. Hu, Z. B. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* **43,** 792–796 (2011).
30. Bate, M. & Martinez-Arias, A. *The Development of Drosophila melanogaster* (Cold Spring Harbor Press, 1993).

Acknowledgements

This work was funded by the National Key Basic Research Program Grant (2013CB945700, 2011CB944304, 2013CB911400, 2011CB944504, 2012CB966600 and 2010CB945200), the Chinese Natural Science Funds (81222006), the National Science Foundation for Distinguished Young Scholars of China (81225020) a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, and Key Project of Shanghai Municipal Science and Technology (10JC1409900)

Author contributions

Z.H. and J.S. directed the study, obtained financial support and were responsible for study design, interpretation of results and manuscript writing. Z.H. performed project management along with Z.L., X.G. and Y.L. and drafted the initial manuscript. Y.L., J. Dong and J. Dai were responsible for statistical analyses. Y.L., F.L., Y.W., Y.J. and S.P. were responsible for sample processing and managed the genotyping data. J.L., Z. Zhou, Y.X., B.Y., X.Y., H.M. and G.J. were responsible for subject recruitment and sample preparation for the Nanjing samples. Y.Z., P.P., X.C., H. Lu, Z.W. and X.Z. were responsible for subject recruitment and sample preparation for the Shanghai samples. Z.-J.C., H.Z. and J.X. were responsible for subject recruitment and sample preparation for the validation III study. J.Y., C.T., X.G. and H.W. were responsible for functional analysis in *Drosophila*. H.L. and C.X. were responsible for subject recruitment and sample preparation for the Wuhan samples. P.X., Q.D. and N.W. were responsible for subject recruitment and sample preparation for the Shenyang samples. K.L. was responsible for subject recruitment and sample preparation for the Nanning samples. J.L., Z.-J.C., M.L., Y.G., F.Z. and Z. Zhang contributed to clinical diagnosis and discussion. F.C. oversaw the statistical analyses process. H.S., J.L., Z.-J.C., X.W. oversaw the study design and contributed to interpretation of results. All authors approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Hu, Z. *et al.* Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. *Nat. Commun.* 5:3857 doi: 10.1038/ncomms4857 (2014).