Identification of seven genes essential for male fertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in Drosophila

Jun Yu1,2,†, Hao Wu1,2,†, Yang Wen1,4,†, Yujuan Liu1,2,†, Tao Zhou1,2, Bixian Ni1,4, Yuan Lin1,4, Jing Dong1,4, Zuomin Zhou1,2, Zhibin Hu1,4,*, Xuejiang Guo1,2,*, Yuan Lin1,4, Jing Dong1,4, Zuomin Zhou1,2, Zhibin Hu1,4,*, Xuejiang Guo1,2,*, Jiahao Sha1,2,*, and Chao Tong3,*

1State Key Laboratory of Reproductive Medicine, 2Department of Histology and Embryology, Nanjing Medical University, Nanjing 210029, China, 3Life Sciences Institute and Innovation Center for Cell Biology, Zhejiang University, Hangzhou 310058, China, and 4Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China

Abstract

Non-obstructive azoospermia (NOA) is a complex and severe condition whose etiology remains largely unknown. In a genome-wide association study (GWAS) of NOA in Chinese men, few loci reached genome-wide significance, although this might be a result of genetic heterogeneity. Single nucleotide polymorphisms (SNPs) without genome-wide significance may also indicate genes that are essential for fertility, and multiple stage validation can lead to false-negative results. To perform large-scale functional screening of the genes surrounding these SNPs, we used in vivo RNA interference (RNAi) in Drosophila, which has a short maturation cycle and is suitable for high-throughput analysis. The analysis found that 7 (31.8%) of the 22 analyzed orthologous Drosophila genes were essential for male fertility. These genes corresponded to nine loci. Of these genes, leukocyte-antigen-related-like (Lar) is primarily required in germ cells to sustain spermatogenesis, whereas CG12404, doublesex-Mab-related 11E (dmrt11E), CG6769, estrogen-related receptor (ERR) and sulfateless (sfj) function in somatic cells. Interestingly, ERR and sfj are also required for testis morphogenesis. Our study thus demonstrates that SNPs without genome-wide significance in GWAS may also provide clues to disease-related genes and therefore warrant functional analysis.

Introduction

More than 10% of couples of childbearing age have fertility problems. Male factor infertility accounts for approximately half of these cases (1). The causes of male infertility are complex and diverse, but non-obstructive azoospermia (NOA), the most severe form of male infertility, occurs in 1% of all adult men (1). These patients have no sperm in their semen. In the past few years,
some genes associated with male infertility have been identified (2,3), yet we still know little about NOA.

Since the first genome-wide association study (GWAS) was successfully performed in 2005 (4), thousands of single nucleotide polymorphisms (SNPs) have been found to be associated with common diseases (5). GWAS help to connect SNPs with certain diseases and can be used to improve risk assessment and accelerate drug development. However, most GWAS have lacked systematically functional verification, limiting their role in guiding our understanding of human diseases. In addition, huge sample sizes and strict quality control are needed to uncover loci with significant genome-wide associations (P-value < 5 x 10^-8) (6,7). Large numbers of SNPs with weak associations are filtered out owing to a potential limiting factor in the statistical power to detect weak genetic loci (8). Furthermore, GWAS do not attempt to identify functional SNPs but rather ‘tag’ the approximate location of disease variants, and they can identify disease-associated alleles only when such alleles have strong linkage disequilibrium with tag-SNPs (tSNPs). Such associations will be weak for rare disease-causing variants (9). Rare variants tend to show greater geographic clustering than common ones and can display spatial structures in populations that are not well controlled in common variant analyses (10). Therefore, tSNPs with lower association scores may be excluded by multiple validation analysis, resulting in false-negative results.

In our previous work, we performed a three-stage analysis of a NOA GWAS in Han Chinese men (a total of 2927 individuals with NOA and 5734 control cases) (11). After the combined analysis, only three SNPs (rs12097821, rs2477686 and rs10842262) conferred a significant genome-wide risk of NOA (P-value < 5 x 10^-8) were identified. Fx10, a gene adjacent to the risk locus rs2477686, was reported to be essential for male fertility in Drosophila (12), indicating the effectiveness of GWAS at identifying genes required for male fertility. Additionally, our extended validation sample size of 2806 cases and 4334 controls revealed three additional NOA susceptibility loci with genome-wide significant associations as well as one locus approaching genome-wide significance (rs3000811, upstream of CDC42 binding protein kinase alpha; CDC42BPA). We also found that gek, an ortholog of CDC42BPA, is essential for male fertility in Drosophila (13).

Because the identified susceptibility loci far from fully explain the complexity of human NOA, we hypothesize that SNPs without genome-wide association (or without the support of multiple validation analysis) may also be important indicators of causal genes for NOA. To test this hypothesis, we developed a strategy using a Drosophila in vivo RNA interference (RNAi) knockdown system to systematically explore and identify functional spermatogenesis genes from tSNPs identified by GWAS.

Drosophila is an excellent model organism in which to screen genes essential for male fertility (14). Drosophila has a short life cycle, and genome-wide gene-knockdown resources are available, which make high-throughput functional screening possible. Adult Drosophila testes contain spermatagonia, spermatocytes, spermatids and sperm, the same populations of germ cells that exist in the human testis. At the apical tip of the Drosophila testis, germline stem cells (GSCs) divide to generate two cells, a new stem cell and a gonialblast that proliferates and differentiates into spermatocytes (15). The cyst cells provide an environment for germ cell proliferation, growth and differentiation, similar to the role of Sertoli cells in the human testis (16). Many fertility-related genes are conserved between Drosophila and mammals (17), and many mutants of homologous human and Drosophila genes have similar testicular phenotypes (14,18).

Using the Drosophila in vivo RNAi model, here we carried out a systematic analysis of genes surrounding the loci associated with NOA and identified seven genes (7/22, 31.8%) essential for male fertility.

**Results**

**A strategy for male fertility-essential genes by GWAS of NOA followed by RNAi-mediated functional screening in Drosophila**

In our previous NOA GWAS screen, 109 tSNPs were found to be associated with NOA (P-value < 10^-5). After several rounds of validation, we found strong evidence of genome-wide significance (P-value < 5 x 10^-8) for NOA susceptibility for six SNPs, as well as one locus that approached genome-wide significance. Although the other SNPs did not reach genome-wide significance (P-value < 5 x 10^-8) by multiple validation analysis, they may still indicate surrounding genes essential for fertility that confer a risk for NOA.

A recent study showed that the majority (~93%) of disease- and trait-associated SNPs identified by GWAS lie within non-coding sequences. Approximately 32% of GWAS SNPs are located in deoxyribonucleic acid (DNA) regulatory DNA that can regulate genes within 100 kb (19). To cover more regulated genes, we considered genes flanking the tSNPs within 200 kb. For the 103 tSNPs that did not reach genome-wide significance after multiple validations but had a P-value < 10^-5, genes were classified into two classes: Class I, genes with the tSNP located inside the gene body; Class II, adjacent genes located within the 200 kb flanking the tSNPs. In total, 142 candidate genes were found around the tSNPs in the human genome. For Drosophila ortholog annotation, only orthologs (homology type: one to one and one to many) with sequence identity >20% were considered for functional studies. The analysis found 38 orthologous Drosophila genes through the Ensembl database (http://www.ensembl.org/), corresponding to 37 human genes and 38 susceptible tSNPs. Of these genes, Nfl and sina are essential for male fertility according to Flybase (http://flybase.org/) and the MGI database (http://www.informatics.jax.org/). We have previously shown that gek is essential for fertility (13). Additionally, mice or Drosophila carrying mutations in six other genes (Nuk, Fak56D, Oseg5, CG9220, CG3925 and IA-2) are known to be fertile. UAS-RNAi lines for another seven orthologous Drosophila genes (corresponding to four human genes and four candidate tSNPs) are not available. These genes were excluded, and the remaining 22 genes were subjected to functional studies (Fig. 1 and Supplementary Material, Table S1).

The Drosophila UAS-Gal4 system-based RNAi silencing method has previously been used for fertility analysis (20). Because different Gal4 drivers have different expression levels and patterns, seven different Gal4 lines were used to drive UAS-RNAi expression in the fly testis (see Fig. 2B). Nos-Gal4 has germ cell-specific expression, primarily in early stage germ cells (21,22). Bag of marbles-Gal4 (bam-Gal4) expression is restricted to 2- to 16-cell spermatagonia (22,23). CG387-Gal4 is expressed in cyst cells (24). C729-Gal4 is expressed in cyst cells, the testis sheath and the male accessory gland (13). Esq-Gal4 is expressed in the somatic cells of the developing gonad (25), and upd-Gal4 is expressed in the hub cells (22). We also used the weak but ubiquitously expressed Gal4 driver tubulin-Gal4 (tub-Gal4) to knockdown gene expression (22). We crossed these Gal4 lines with the individual UAS-RNAi lines to allow expression of shRNAs in different testicular cell types and examined the fertility of the resulting adult male flies. To further characterize the RNAi phenotypes,
we also analyzed squashed testes under phase-contrast microscopy and stained for different markers to distinguish various cell types within the testes. In total, we screened 29 transgenic RNAi lines corresponding to 22 fly genes at two heat-stress temperatures (Supplementary Material, Table S2).

Fertility tests revealed seven genes essential for male fertility (Fig. 2A and Table 1): Lar, papi, ERR, CG12404, dmrt11, CG6769 and sfl (Fig. 2C and Supplementary Material, Table S2). Lar is required in the germ cells, whereas ERR, CG12404, dmrt11, CG6769 and sfl are essential in the somatic cells of the testis. For papi, only individuals expressing tub-Gal4-driven RNAi had reduced fertility (Supplementary Material, Table S2).

Lar is required for spermatogenesis in germ cells

Lar is a transmembrane receptor protein with tyrosine phosphatase activity that is important for neuron growth and development (26,27). Interestingly, Lar is also required for oogenesis and spermatogenesis in Drosophila gonads (28,29). Lar is expressed in the GSC membrane at the interface between the hub and the GSCs. Loss of Lar results in loss of GSCs in the testis stem cell niche (29). In this study, we found that Lar is also essential for later stages of spermatogenesis in flies. When we used bam-Gal4 to knockdown Lar expression in 2- to 16-cell spermatogonial, we found that almost all (93.3%) at 25°C, N = 60; 96.9% at 28°C, N = 96) resulting males were sterile (Fig. 2C). Phase-contrast visualization of squashed adult testes showed that bam> Lar RNAi testes (N = 26, T = 25°C) had defective elongating and elongated spermatids and no mature sperm (Fig. 3G and H), indicating that Lar plays a role in the late stage of spermatogenesis. In bam> Lar RNAi testes, some cysts close to the testis apex were vasa negative, indicating a loss of germ cells (Fig. 3A and D). Bam-positive spermatogonial regions were also reduced in size and closer to the testis tip than those of controls (Fig. 3B and E). In addition, more Eya-positive mature cyst cells accumulated in Lar RNAi testes than in the controls (see Fig. 3C, F and I), most likely due to enhanced somatic cell proliferation and differentiation in compensation for the lost germ cells.

Five genes are required in testicular somatic cells to sustain male fertility

Somatic cells are critical for germ cell differentiation, proliferation and maintenance (30). In our screen, we identified the somatic cell genes CG12404, dmrt11E, CG6769, sfl and ERR as required to maintain male fertility. CG12404 encodes a Yip1 domain-containing protein that is involved in vesicular transport from the Golgi complex to the endoplasmic reticulum (31). Two independent RNAi lines, CG12404 RNAi1 (#1463 from VDRC) and CG12404 RNAi2 (#14989 from Bloomington), were used to target CG12404 in our screen. The c729> CG12404 RNAi1 males were totally infertile (N = 49 at 25°C and N = 24 at 28°C), and 61.9% (N = 63, T = 25°C) and 72.6% (N = 62, T = 28°C) of c729> CG12404 RNAi2 males were infertile (Fig. 2C). Defects were observed from the round spermatid stage in c729>CG12404 RNAi1 testes (N = 24, T = 28°C). The spermatozoa interlaced with each other and lost their typical rapid movement (Fig. 4A and B). dmrt11E encodes a transcription factor-like protein containing a highly conserved DNA-binding DM domain that plays a critical role in sex determination (32). A total of 43.7% (N = 32, T = 28°C) of c729> dmrt11E RNAi males were infertile, and no mature sperm were observed in the squashed testes (N = 13, T = 28°C). Mild defects at the round spermatid stage were detected (Fig. 4A and C), indicating that dmrt11E is required for a late stage of spermatogenesis. CG6769 encodes a zinc finger protein that most likely binds to RNA. Its physiological function is unknown. 84.0% (N = 50,
T = 28°C) of c729>CG6769 RNAi males were sterile. However, these testes contained germ cells of all stages, with no obvious morphological defects (N = 24, T = 28°C) (Fig. 4A and D).

**ERR and sfl are required for testis morphogenesis**

ERR is a nuclear receptor that may bind to steroid hormones (33). However, its function is not well understood. In our screen, 56.9% (N = 58, T = 25°C) and 70.2% (N = 114, T = 28°C) of c729>ERR RNAi males were infertile. We also used c587-Gal4, a cyst cell-specific Gal4 driver, to knockdown ERR expression. 45.5% (N = 77, T = 25°C) and 74.6% (N = 59, T = 28°C) of c587>ERR RNAi male flies were sterile (Fig. 2C), suggesting that ERR expression in cyst cells is critical for male fertility. Interestingly, 88.2% (N = 85, T = 25°C) of the c729>ERR RNAi males and 76.7% (N = 60, T = 25°C) of the c587>ERR RNAi males had abnormal testis morphology.

**Table 1. Summary of essential fertility genes identified in the screen**

<table>
<thead>
<tr>
<th>Gene (fly)</th>
<th>CG No.</th>
<th>Chr (fly)</th>
<th>Homology type (fly versus human)</th>
<th>Identity (%)</th>
<th>Classification</th>
<th>Gene (human)</th>
<th>Chr (human)</th>
<th>Indicating SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>papi</td>
<td>CG7082</td>
<td>2L</td>
<td>One to one</td>
<td>27</td>
<td>Class I</td>
<td>TDRKH</td>
<td>1</td>
<td>rs11204885</td>
</tr>
<tr>
<td>CG12404</td>
<td></td>
<td>2L</td>
<td>One to many</td>
<td>47</td>
<td>Class II</td>
<td>YIPF5</td>
<td>5</td>
<td>rs312620</td>
</tr>
<tr>
<td>CG6769</td>
<td></td>
<td>X</td>
<td>One to one</td>
<td>45</td>
<td>Class II</td>
<td>ZNF622</td>
<td>5</td>
<td>rs163065</td>
</tr>
<tr>
<td>sfl</td>
<td>CG8339</td>
<td>3L</td>
<td>One to many</td>
<td>45</td>
<td>Class II</td>
<td>NDS1</td>
<td>5</td>
<td>rs253296</td>
</tr>
<tr>
<td>Lar</td>
<td>CG10443</td>
<td>2L</td>
<td>One to many</td>
<td>48</td>
<td>Class I</td>
<td>PTFRD</td>
<td>9</td>
<td>rs10978121</td>
</tr>
<tr>
<td>ERR</td>
<td>CG7404</td>
<td>3L</td>
<td>One to many</td>
<td>37</td>
<td>Class II</td>
<td>ESRRB</td>
<td>14</td>
<td>rs4903393</td>
</tr>
<tr>
<td>ERR</td>
<td>CG7404</td>
<td>3L</td>
<td>One to many</td>
<td>36</td>
<td>Class II</td>
<td>ESRRG</td>
<td>1</td>
<td>rs2221285</td>
</tr>
<tr>
<td>dmr11E</td>
<td>CG15749</td>
<td>X</td>
<td>One to one</td>
<td>26</td>
<td>Class II</td>
<td>DMRT2</td>
<td>9</td>
<td>rs10756006</td>
</tr>
</tbody>
</table>

Genes were classified into two classes: Class I, genes with the tSNP located inside the gene body; Class II, adjacent genes located within the 200 kb flanking the tSNPs.
changing from the typical long tubular shape to a small and oval-shaped structure. Upon dissection, all germ cell stages were observed not only in most (75% of c729>ERR RNAi and 92% of c587>ERR RNAi) of the morphologically normal but also in many (18.8% of c729>ERR RNAi and 85.7% of c587>ERR RNAi) of the morphologically abnormal testes (Supplementary Material, Fig. S1 and S3C).

To analyze whether the loss of ERR affects the fate and proliferation of germ cells or somatic cells, we stained the testes with different cellular markers. Germ cells, hub cells and mature cysts were observed in control testes (A–C) and bam>Lar RNAi (D–F) testes immunostained with anti-Vasa antibody (a germ cell marker, green in A and D) and anti-FasIII antibody (a hub cell marker, red in A and D). White arrows indicate a cyst of cells without Vasa staining. DE-cadherin marks hub cells and cyst cells (red in B–C and E–F), and Bam marks spermatogonia (green in B and E). In bam>Lar RNAi testes, the distance between the Bam-positive spermatogonia and hub cells is smaller than in the control (B and E, white lines). Eya, a marker of mature cyst cells (green in C and F) is increased in bam>Lar RNAi testes. (G and H) Phase-contrast view of control (G1–G5) and bam>Lar RNAi (H1–H4) testes. Different stages of spermatogenesis in normal testes are indicated, including spermatogonia (G1), spermatocytes (G2), round spermatids (G3), elongating and elongated spermatids (G4) and mature sperm (G5). The germ cells of bam>Lar RNAi testes were normal from spermatogonia to the round spermatid stage (H1–H4). Defects in elongating and elongated spermatids were observed, and no mature sperms were present. (I) Eya-positive cyst cell numbers were increased in bam>Lar RNAi testes. Temperature stress (25°C). Scale bars: 20 µm. (Student’s t-test: *P-value < 0.05). Error bars represent the SD.

(Fig. 5A1–E1 and Supplementary Material, Fig. S3A), changing from the typical long tubular shape to a small and oval-shaped structure. Upon dissection, all germ cell stages were observed not only in most (75% of c729>ERR RNAi and 92% of c587>ERR RNAi) of the morphologically normal but also in many (18.8% of c729>ERR RNAi and 85.7% of c587>ERR RNAi) of the morphologically abnormal testes (Supplementary Material, Fig. S1 and S3C). To analyze whether the loss of ERR affects the fate and proliferation of germ cells or somatic cells, we stained the testes with different cellular markers. Germ cells, hub cells and mature cysts...
cells were present in the ERR RNAi testes (Fig. 5A2–E2, A3–E3, A6–E6). However, the number of premeiotic germ cells (gonialblasts, spermatogonia and spermatocytes) was strikingly reduced in the c729>ERR RNAi testes with morphological defects (Fig. 5C2 and C4). In addition, more phospho-Histone H3 staining was detected in the ERR RNAi testes than in wild-type controls (Fig. 5A5–E5), indicating more cell proliferation in the RNAi testes. Interestingly, knockdown of sfl in the somatic cells during gonadal development by esg-Gal4 resulted in oblong-shaped testes similar to the c729>ERR RNAi or c587>ERR RNAi testes. sfl encodes a glucosamine N-sulfotransferase that is essential for development and organ morphogenesis (34, 35). When we knocked down sfl by tub-Gal4, esg-Gal4 (T = 28°C) or upd-Gal4, all the flies died during early stages. However, esg>sfl RNAi male flies raised at 25°C were sterile, but 97.3% (N = 110, T = 25°C) were sterile. 51.4% (N = 109, T = 25°C) of the esg>sfl RNAi testes lost the tubular shape and had morphological defects (Fig. 6A1–C1 and Supplementary Material, Fig. S3B). Germ cells at each spermatogenic stage were detected in the squashed testes of most esg>sfl RNAi males (N = 22 with morphologically defective testes and N = 50 with normal testes) (Supplementary Material, Fig. S2 and S3D). Whole-mount testes stained with DNA dye also showed clusters of elongated spermatid nuclei (Supplementary Material, Fig. S4), regardless of whether the testis morphology was defective. Similar to the c729>ERR RNAi and c587>ERR RNAi testes, increased phospho-Histone H3 staining was observed (Fig. 6A5–C5). The similarity between the ERR and sfl knockdown animals suggests that they may work in the same pathway to regulate testis morphogenesis.

**Discussion**

With the rapid progress made by GWAS, many risk SNPs have been identified in the past few years (5). However, the biological significance of these SNPs is largely unknown. Considering the highly heterogeneous human population structure and properties of tSNPs, many SNPs surrounding causal genes may have been missed due to low association values and multiple validations. Here, we have presented a systematic screening strategy using model organisms and human NOA GWAS data.

In this study, of the 22 genes screened, 7 (31.8%) were required for male fertility, indicating the usefulness of functional screening to control for false negatives in the SNP data. It has been estimated that ~1500 genes (~11% of the fly genome) are required for male fertility in Drosophila genome based on several large-scale screen data (14). A 31.8% hit rate is significantly greater than what one would expect due to chance (Fisher’s exact test P = 0.008), suggesting that the candidates selected based on our GWAS data are enriched with genes required for male fertility. In this study, we started with 142 human genes corresponding...
to 84 fly orthologs because many human genes did not have fly orthologs. Conservation analysis showed that some human genes may have multiple orthologs in Drosophila, which making functional testing impossible or difficult. Besides, the sequence similarity between some human genes and the Drosophila orthologs are very low. Since sequence similarity is an important parameter for functional conservation (36,37), only Drosophila orthologs with sequence identity over 20% were considered in this study, which is especially important for the orthologous relationship of ‘one to many’. Thus, of the 84 Drosophila orthologs, only 38 matched these criteria, and were subjected to the following functional analysis. The limited homology between the human genes and the Drosophila genes indeed make testing every single candidate impossible. But, comparing with other model systems suitable for fast and large-scale analysis such as yeast and worm, Drosophila is the closest to human.

The seven identified fertility genes in flies correspond to eight human genes and nine loci associated with NOA (Table 1), including two Class I SNPs (rs10978121 and rs11204885) located inside the genes PTPRD and TDRKH, the human orthologs of Lar and papi. In our Drosophila studies, we found that some fertility-related genes are surrounded by multiple NOA-associated SNPs. For example, fly ERR, a gene encoding an estrogen receptor-related receptor (ESRR) family protein, is essential for male fertility. ERR is an ortholog of human ESRR gamma (ESRRG) and ESRR beta (ESRRB), two genes that are surrounded by three NOA-associated SNPs. Multiple NOA-associated SNPs surrounding the same gene are evidence for its important function in male fertility.

We used seven different Gal4 drivers in our screen, each with distinct expression timing, expression pattern and strength. Although some have overlapping expression patterns, they did not produce similar phenotypes in some cases when crossed with individual RNAi lines. For example, both nos-Gal4 and bam-Gal4 are expressed in germ cells. However, nos> Lar RNAi males were fertile, while the majority of bam> Lar RNAi males were sterile. Nos-Gal4 is primarily expressed in the GSC and gonialblast and is substantially down-regulated at later stages (22), whereas bam-Gal4 expression is restricted to 2- to 16-cell spermatagonia (22,23). In addition to the expression pattern differences, the expression timing and levels of these Gal4 lines are also different, which might explain why the RNAi phenotypes are
distinct. In previous studies, Lar was reported to be required for GSC maintenance (29). The lack of fertility defects in the germ cells of nos-Gal4 Lar knockdowns are most likely due to low knockdown efficiency.

Loss of dmrt11E in somatic cells led to spermatogenesis defects at late stages. dmrt11E encodes a protein containing a DM domain that is highly conserved in the Doublesex and mab-3 related transcription (DMRT) protein family. Interestingly, DMRT factor 1 (DMRT1) and DMRT factor 2 (DMRT2), the mammalian orthologs of dmrt11E, map to a region of the genome associated with gonadal dysgenesis and XY sex reversal, which make them candidates for male fertility-required genes (38–40). In addition, loss-of-function mutation of DMRT1 has been identified as a risk factor and potential genetic cause of human spermatogenic failure (41). Loss of DMRT2 causes neonatal lethality, hindering its use in the study of fertility in adult animals (42). Our study provides direct evidence that DMRT family genes do indeed play critical roles in male fertility.

In our screen, knockdown sf or ERR expression in somatic cells resulted in morphological testis defects that phenocopy the drosophila Wnt oncogene analog 2 (Dwnt-2) mutant. The formation of a spiral-shaped testis during pupation requires the gonad and genital disc to grow toward and fuse with each other (43). The gonadal pigment cells and the precursors of smooth muscle cells migrate to form a bilayered sheath that covers the mature gonad and seminal vesicle (43). Dwnt-2 is a member of the Wnt family (44). Male flies with Dwnt-2 mutations are sterile, with small and abnormally oblong testes. Dwnt-2 is expressed in gonadal somatic cells and is required for male-specific pigment cell specification and muscle precursor migration (45). Interestingly, sf is required for the biogenesis of heparan sulfate proteoglycans, which are essential for Wg signaling (46). It will be interesting to determine whether it also regulates Dwnt-2 signal transduction. ERR encodes a member of the ESRR family, which is part of the nuclear hormone receptor superfamily. ERR is an orphan nuclear receptor whose ligand is not known (47). It has been reported that ERR plays a central role in carbohydrate metabolism and directs a critical metabolic transition during development (48). There is no evidence that ESRR family proteins regulate Wnt signaling or vice versa. Both ESRR and Wnt family proteins are involved in oncogenic pathways. It will be interesting to test whether there is any interplay between ESRR proteins and the Wnt signaling pathway.

Interestingly, several of the seven essential fertility genes identified in this study are highly conserved but have very little functional annotation. For example, there have been few studies of CG12404 and CG6769 in flies. YIF5S, the human ortholog of CG12404, is required for endoplasmic reticulum and Golgi structure maintenance in cultured cells (49). ZFNN22, the human ortholog of CG6769, has been reported to regulate apoptosis (50). However, the physiological function of both genes is unknown. Here, we found that both genes are required for male fertility in flies. Further investigation will be required to determine the underlying molecular mechanisms through which CG12404 and CG6769 regulate male fertility.

By exploring the RNA-Seq data set generated by the Genotype-Tissue Expression (GTEx) project (http://www.gtexportal.org/home/), we found that six (TDKH, YIF5S, ZNF622, NDST1, PTPRD and ESRRB) out of eight genes identified in our study were expressed in human testis samples. Although ESRRG and DMRT2 were not present in the GTEx data set, we found that ESRRG is listed as a testis-expressing gene in the UniGene database (http://www.ncbi.nlm.nih.gov/unigene/) and DMRT2 was reported to locate in a region that is critical for testis development and is expressed in the adult human testis (40). The presence of the genes in human testes suggested that these genes and its orthologs might be required for male fertility not only in flies but also in humans.
Indeed, it has been reported that some of the genes identified in this study have differential expression levels in the NOA patients and the healthy controls (51). The testicular specimens from both the NOA patients (n = 18) and the healthy controls (n = 4) express all of the genes identified here. Among them, TDRKH and ESRRB have lower expression levels while YIPF5 has higher expression level in NOA patients than those in the healthy controls (P < 0.05) (Supplementary Material, Table S3). Their results were strongly in line with our previous GWAS screening study. So, these expression data in NOA patients showed a possibility that these genes are involved in NOA pathogenesis. There are high genetic heterogeneities in human. With larger tissue sample size, expression analysis may be able to reveal more human fertility-related genes. The elevated YIPF5 level in the NOA patients are surprising because we identified this gene as a fertility-essential gene based on a RNAi knockdown assay. It would be interesting to test whether or not YIPF5 overexpression in flies could lead to male sterility.

The tissue-specific knockdown system we used in this study is very powerful because it not only bypassed the requirement of the gene in other organs but also allow us to dissect the gene’s function in specific cells and developmental stages in germine. One concern about this system is that the broadly expressed genes we identified might be essential in other organs than testis and might not be relevant to NOA, because NOA is asymptomatic in human. Published studies showed pleiotropic nature of many broadly expressed genes in tissues. In flies, the weak alleles of many broadly expressed essential genes lead only to male sterility (52), suggesting that the broadly expressed genes could lead to sterility phenotype without other symptoms. However, further validating experiments in human can help elucidate causal genes of NOA.

In summary, we show that Drosophila RNAi screening is a powerful tool for controlling for false negatives in multiple stage population data. This technique could be further modified by adding tissue-specific long-range regulation and expression quantitative trait loci data (53). The seven essential fertility genes we found warrant further study and may help reveal the mechanisms underlying NOA.

**Materials and Methods**

**Identification of Drosophila gene orthologs surrounding tSNPs associated with NOA in humans**

The tSNPs associated with human NOA that were considered in this study all had a P-value < 10^-8 in a previous GWAS of NOA performed in Han Chinese men (1000 individuals with NOA (cases) and 1703 male controls) genotyped for 906 703 SNPs with Affymetrix Genome-Wide Human SNP Array 6.0 chips (11). To obtain candidate fertility-related genes, we classified SNP-associated human genes into two classes: Class I, genes that intersected with the SNPs and Class II genes located within 200 kb upstream/downstream of the SNPs. We then selected the corresponding unique homologous Drosophila genes (homology type: orthologous one to many or one to one) of the above associated human genes for candidate targets to be verified. We only considered Drosophila genes with at least 20% sequence identity with their human orthologs. The genome backgrounds for human and Drosophila were GRCh37 (hg19) and BDGP R5 (dm3), respectively. Gene descriptions and genome localizations for genes and SNPs were batch-obtained from the UCSC genome browser (http://genome.ucsc.edu/). Orthologous relationships between human and Drosophila genes were annotated via BioMart (http://www.biomart.org/) based on Ensembl orthologs.

**Statistical methods**

Fisher exact tests were conducted in SAS 9.3 (SAS Institute Inc., Cary, NC) to compare the rate of genes involved in male sterility in our study and the genome-wide rate estimated by Hackstein’s study (14).

The Student’s t test was used for differences between groups. *P*-value < 0.05; **P*-value < 0.01; ***P*-value < 0.001. Error bars represent the SD.

**Fly stocks and screening crosses**

All UAS-RNAi transgenic fly lines were obtained from the VDRC, the Bloomington Drosophila Stock Center (BDSC) and the Tsinghua Fly Center (THFC). Gal4 stock lines ordered from BDSC and the Drosophila Genetic Resource Center (DGRC) were as follows: esg-Gal4 (DGRC, #109126), nos-Gal4 (BDSC, #4937), tub-Gal4 (BDSC, #5138), c279-Gal4 (BDSC, #6983), upd-Gal4 (BDSC, #26796). Bsm-Gal4 and c587-Gal4 were gifts from D.H. Chen and X. Huang, respectively. Canton S flies were used as wild-type flies. We generated two-round crosses and set two temperature stress conditions (25 and 28°C). Transgenic UAS-RNAi males were crossed to virgin females carrying seven different Gal4 drivers at room temperature. To improve the efficiency of RNAi silencing, we employed a higher temperature stress (28°C) after enough eggs had been laid and incubated the eggs at 28°C until adults hatched out. We performed single male fertility tests using a single F1 RNAi adult male fly enclosed for 3 days in a cross with three wild-type virgin females at room temperature.

**Dissection and phase-contrast view**

Testes were dissected in 1 x phosphate-buffered saline (PBS) and washed. Shredded testes were observed directly on slides by phase-contrast microscopy after gentle squashing with a cover slip.

**Immunofluorescence**

Antibodies used were as follows: rat anti-Vasa (Developmental Studies Hybridoma Bank; DSHB, 1:20), mouse anti-EYA (DSHB, 1:20), mouse anti-Fas III (DSHB, 1:50), rat anti-DE-cadherin (DSHB, 1:20) and rabbit anti-Bam C (a gift from DH Chen, 1:2000). Secondary antibodies conjugated to A488, Cy3 or A647 (Molecular Probes and Jackson Immunologicals) were diluted at 1:20. Mouse anti-Fas III, rat anti-DE-cadherin, and the secondary antibodies conjugated to A488, Cy3 or A647 were diluted at 1:20, respectively. Human Molecular Genetics, 2015, Vol. 24, No. 5

**Supplementary Material**

Supplementary Material is available at HMG online.
Authors’ Contributions

C.T., J.S., X.G. and Z.H. directed the study, obtained financial support and were responsible for the study design, interpretation of results and manuscript writing. C.T., J.S., X.G. and Z.H. performed project management along with J.Y., H.W. and Y.-J.L. Y.W., T.Z., B.N., Y.L., J.D. and Z.Z. were responsible for data annotation and analysis. J.Y., H.W. and J.-Y.L. were responsible for the Drosophila experiments. All authors approved the final version of the manuscript.

Acknowledgements

The authors wish to thank all study participants, research staff and students who assisted with this work. We are grateful to BDSC, THFC, VDRC and DGERC for providing fly lines and Dahua Chen (Institute of Zoology, CAS) and Xun Huang (Institute of Genetics and Developmental Biology, CAS) for sharing reagents and stocks.

Conflict of Interest statement. None declared.

Funding

This work was supported by the National Basic Research Program of China (2014CB943100, 2011CB966600 and 2015CB943003) and the Chinese Natural Science Fund (81222006 and 2015CB943003) and the Chinese Natural Science Fund (81222006).

References


