

ARTICLE

Received 5 May 2016 | Accepted 4 Nov 2016 | Published 20 Dec 2016

DOI: 10.1038/ncomms13840

OPEN

# Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility

Kevin Litchfield<sup>1</sup>, Max Levy<sup>1</sup>, Darshna Dudakia<sup>1</sup>, Paula Proszek<sup>2</sup>, Claire Shipley<sup>2</sup>, Sander Basten<sup>3</sup>, Elizabeth Rapley<sup>1</sup>, D. Timothy Bishop<sup>4</sup>, Alison Reid<sup>5</sup>, Robert Huddart<sup>5</sup>, Peter Broderick<sup>1</sup>, David Gonzalez de Castro<sup>2,6</sup>, Simon O'Connor<sup>2</sup>, Rachel H. Giles<sup>3</sup>, Richard S. Houlston<sup>1,7</sup> & Clare Turnbull<sup>1,8,9</sup>

Testicular germ cell tumour (TGCT) is the most common cancer in young men. Here we sought to identify risk factors for TGCT by performing whole-exome sequencing on 328 TGCT cases from 153 families, 634 sporadic TGCT cases and 1,644 controls. We search for genes that are recurrently affected by rare variants (minor allele frequency <0.01) with potentially damaging effects and evidence of segregation in families. A total of 8.7% of TGCT families carry rare disruptive mutations in the cilia-microtubule genes (CMG) as compared with 0.5% of controls ( $P = 2.1 \times 10^{-8}$ ). The most significantly mutated CMG is *DNAAF1* with biallelic inactivation and loss of *DNAAF1* expression shown in tumours from carriers. *DNAAF1* mutation as a cause of TGCT is supported by a *dnaaf1*<sup>hu255h</sup>(+/-) zebrafish model, which has a 94% risk of TGCT. Our data implicate cilia-microtubule inactivation as a cause of TGCT and provide evidence for CMGs as cancer susceptibility genes.

<sup>1</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, London SM2 5NG, UK. <sup>2</sup>Centre for Molecular Pathology, The Royal Marsden NHS Foundation Trust, London SM2 5NG, UK. <sup>3</sup>Department of Nephrology and Hypertension, Regenerative Medicine Center Utrecht, University Medical Center Utrecht, Uppsalalaan 6, Utrecht 3584CT, The Netherlands. <sup>4</sup>Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, Leeds LS9 7TF, UK. <sup>5</sup>Academic Radiotherapy Unit, The Institute of Cancer Research, London SM2 5NG, UK. <sup>6</sup>Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast BT9 7AE, UK. <sup>7</sup>Division of Molecular Pathology, The Institute of Cancer Research, London SM2 5NG, UK. <sup>8</sup>William Harvey Research Institute, Queen Mary University, London EC1M 6BQ, UK. <sup>9</sup>Department of Clinical Genetics, Guy's and St Thomas' NHS Trust, London SE1 9RS, UK. Correspondence and requests for materials should be addressed to C.T. (email: clare.turnbull@icr.ac.uk).

**T**esticular germ cell tumour (TGCT) is the most common cancer in men aged 15–45 years, with over 18,000 new cases diagnosed annually in Europe<sup>1–3</sup>. Cure rates for TGCT are typically high because of the extreme sensitivity of malignant testicular germ cells to chemotherapy; however, an increased risk of metabolic syndrome, infertility and secondary cancer associated with survivorship is now recognized<sup>4–6</sup>. Furthermore, there are limited options for patients with platinum-resistant tumours, for whom long-term survival remains poor. It is therefore anticipated that an increased understanding of TGCT pathogenesis will generate new therapeutic targets.

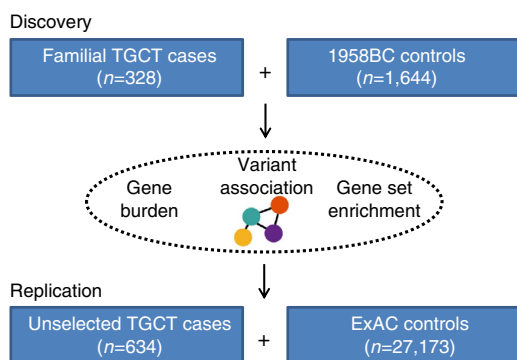
Family, twin and migrant studies support a strong inherited genetic basis to TGCT susceptibility, with brothers of cases having a four- to eightfold increased risk of TGCT<sup>7–10</sup>. While Mendelian susceptibility to TGCT has been inferred from the combination of the high familial risks and reports of multiplex TGCT families, no rare high-impact alleles have so far been identified<sup>11</sup>. In contrast to associations identified through genome-wide association studies, the identification of this class of susceptibility is especially important since such mutations are causal and thus provide direct insight into TGCT biology.

Here, we use whole-exome sequencing (WES) of germline DNA to identify novel high-impact TGCT risk variants, focusing our analysis on familial TGCT as such cases are enriched for genetic susceptibility. We identify rare disruptive mutations in cilia-microtubule function genes as determinants of susceptibility to TGCT. This is further supported by evidence of second somatic mutation in tumour tissue and functional data from zebrafish, which collectively suggest a model of cilia inactivation in promoting TGC tumorigenesis.

## Results

**Whole-exome sequencing of familial TGCT cases.** To identify rare germline variants involved in TGCT we performed WES of 328 TGCT cases from 153 independent families of European ancestry (Methods). For comparison we analysed WES data on 1,644 UK population controls from the 1958 Birth Cohort (1958BC) with no personal history of malignancy (Fig. 1). DNA from germline blood samples from cases and controls was sequenced using Illumina TruSeq exon capture or Nextera Rapid Capture in conjunction with Illumina Hi-Seq 2000 or 2500 technology (Methods). To avoid erroneous findings we performed alignment and variant calling of all samples simultaneously (Methods). Each captured base was sequenced to an average depth of  $49 \times$  across samples.

**Gene burden analysis of familial TGCT data.** We searched in familial cases for genes that were recurrently affected by rare



**Figure 1 | Study design.** Overview of patient samples and exome sequencing study design.

variants with presumptive damaging effects (nonsense, splice acceptor/donor and indel frameshift changes) that had a low burden of comparable variants in controls. We did not include potentially damaging missense variants because of the limited accuracy of *in silico* prediction tools in predicting pathogenicity of missense variants for human disease (American College of Medical Genetics<sup>12</sup>). We performed a collapsing T1 gene burden test imposing a maximal minor allele frequency (MAF) threshold of 1%, to select for rare high-impact variants. To ensure independent events, case counts were based on one individual per pedigree, which was randomly assigned as the proband. Significance was assessed by permutation. To prioritize genes for high-impact variants we filtered results to select only genes containing rare disruptive mutations which segregated with TGCT in at least two families (Table 1). There was no gene for which mutations were detected in more than three of the 153 families when the filters described for mutation type, frequency and segregation were applied. The top ranked gene exome-wide was *DNAAF1/LRRC50* (Dynein, Axonemal, Assembly Factor 1/Leucine-rich repeat containing 50), with the rare disruptive mutations p.Arg636Ter and p.Gly434ProfsTer4 (Fig. 2a) segregating with TGCT in PED-2331 and PED-2152 families (Fig. 3). *DNAAF1* forms a component of the microtubule outer dynein arm, stabilizing microtubule-based cilia<sup>13</sup>. A deleterious phenotype in humans for disruptive *DNAAF1* mutations has previously been established, with biallelic mutations causing recessive primary ciliary dyskinesia (PCD)<sup>14</sup>, which is characterized by impaired primary cilia function, chronic lung disease, male infertility and hearing impairment<sup>15,16</sup>. In addition to *DNAAF1*, mutations in the paralogue genes *LRRC6* (ref. 17) and *CNTRL* (centriolin; ref. 18) were identified in TGCT cases in three additional families. In total *DNAAF1* and its paralogues were mutated in nine cases, from five families with segregation of mutations detected in four families (Fig. 3).

**Gene set enrichment analysis of familial TGCT data.** To complement our T1 burden analysis of single genes, we conducted a Gene-Set-Enrichment-Analysis based on the GO Biological Process ontology in order to identify groups of genes based on specified biological processes/pathways associated with TGCT the signal for which would not have individually been detectable in exome-wide analysis. The top ranked set from analysis of 1,090 canonical gene sets was the gene set related to cilia-microtubule function ( $P = 2.1 \times 10^{-8}$ ,  $Q = 0.01$ , permutation test), containing genes *DNAAF1*, *DYNC2H1*, *DRC1*, *CEP290* and *MAP4* (Table 2 and Supplementary Table 1) which are each a cause of recessive ciliopathy (PCD—*DNAAF1/DRC1* (refs 14,19); asphyxiating thoracic dystrophy—*DYNC2H1* (ref. 20); Joubert syndrome—*CEP290* (refs 21,22); and Senior-Loken syndrome—*MAP4* via *TRAF3IP1* mutation<sup>23</sup>). Moreover, even excluding *DNAAF1*, the cilia-microtubule gene set remained the most significantly associated of the 1,090 gene sets ( $P = 9.1 \times 10^{-6}$ ,  $Q = 0.1$ , permutation test). The mutations we identified in the TGCT cases cluster predominantly in the same ciliopathic-associated protein domains as mutations causing the recessive diseases (Fig. 2b)<sup>24</sup>.

**Replication sequencing of unselected TGCT cases.** We next performed WES of 634 sporadic TGCT cases from the UK, comparing mutation frequencies with 27,173 ExAC non-Finnish European, cancer-free controls<sup>25</sup>. In gene-set enrichment analysis (GSEA), association was replicated in this data set for the cilia-microtubule gene set ( $P = 0.024$ , Fisher's Exact Test), with additional rare disruptive case variants identified in *DNAAF1*, *MAP4*, *DRC1*, *DYNC2H1* and *CEP290*. Remarkably as well as an

**Table 1 | Genes with rare MAF <1% disruptive mutations segregating in two or more familial TGCT pedigrees.**

Gene	TGCT Probands (n = 150)	Controls (n = 1,609)	OR	P-value	Total no. affected familial cases (n = 306)	Rank
<i>DNAAF1</i>	2	1	21.3	$9.9 \times 10^{-3}$	4	1
<i>ACSM1</i>	2	6	3.6	$6.4 \times 10^{-2}$	4	2
<i>TSNAXIP1</i>	3	11	2.9	$6.9 \times 10^{-2}$	6	3
<i>C1orf186</i>	2	8	2.7	$1.3 \times 10^{-1}$	4	4
<i>KIAA1586</i>	2	8	2.7	$1.3 \times 10^{-1}$	4	5
<i>ABCA10</i>	2	8	2.7	$1.6 \times 10^{-1}$	4	6
<i>PIK3C2G</i>	3	13	2.5	$2.4 \times 10^{-1}$	6	7
<i>C1orf168</i>	2	9	2.4	$2.5 \times 10^{-1}$	5	8
<i>RRP15</i>	3	18	1.8	$2.8 \times 10^{-1}$	6	9
<i>MUC4</i>	3	22	1.5	$2.9 \times 10^{-1}$	7	10
<i>OR6K2</i>	2	15	1.4	$3.0 \times 10^{-1}$	5	11
<i>KRTAP1-1</i>	2	15	1.4	$3.0 \times 10^{-1}$	4	12
<i>ABCC12</i>	3	23	1.4	$3.3 \times 10^{-1}$	7	13
<i>CALML4</i>	2	16	1.3	$3.4 \times 10^{-1}$	5	14

Abbreviations: MAF, minor allele frequency; OR, odds ratio; TGCT, testicular germ cell tumour. OR calculated based on proband vs control frequency. P-value from T1 gene burden test, with significance assessed by permutation test.

**Table 2 | Gene set enrichment analysis results in familial TGCT data set.**

Gene set	Number of genes	Enrichment Score	P-value	Q value	Rank
Cilia-microtubule function	8	2.05	$2.1 \times 10^{-8}$	0.01	1
Reactome loss of nlp from mitotic centrosomes	27	1.93	$3.1 \times 10^{-4}$	0.11	2
Chromosome segregation	16	1.84	$1.8 \times 10^{-3}$	0.37	3
Reactome recruitment of mitotic centrosome proteins and complexes	31	1.83	$6.7 \times 10^{-4}$	0.34	4
Reactome glutathione conjugation	10	1.82	$2.2 \times 10^{-3}$	0.34	5
Response to light stimulus	23	1.81	$2.1 \times 10^{-3}$	0.33	6
Respiratory gaseous exchange	6	1.80	$8.3 \times 10^{-4}$	0.31	7
Response to radiation	29	1.78	$2.8 \times 10^{-3}$	0.37	8
Kegg abc transporters	30	1.78	$2.1 \times 10^{-3}$	0.36	9
Response to ultraviolet	15	1.76	$6.1 \times 10^{-3}$	0.41	10
Photoreceptor cell maintenance	7	1.75	$3.8 \times 10^{-3}$	0.45	11
Chromosome organization and biogenesis	41	1.74	$1.7 \times 10^{-3}$	0.46	12
Regulation of endocytosis	6	1.72	$4.4 \times 10^{-3}$	0.55	13
Reactome mitotic g2 g2 m phases	35	1.72	$4.1 \times 10^{-3}$	0.54	14
Membrane fusion	11	1.71	$9.4 \times 10^{-3}$	0.52	15

Abbreviations: TGCT, testicular germ cell tumour. P-value significance assessed by permutation test.

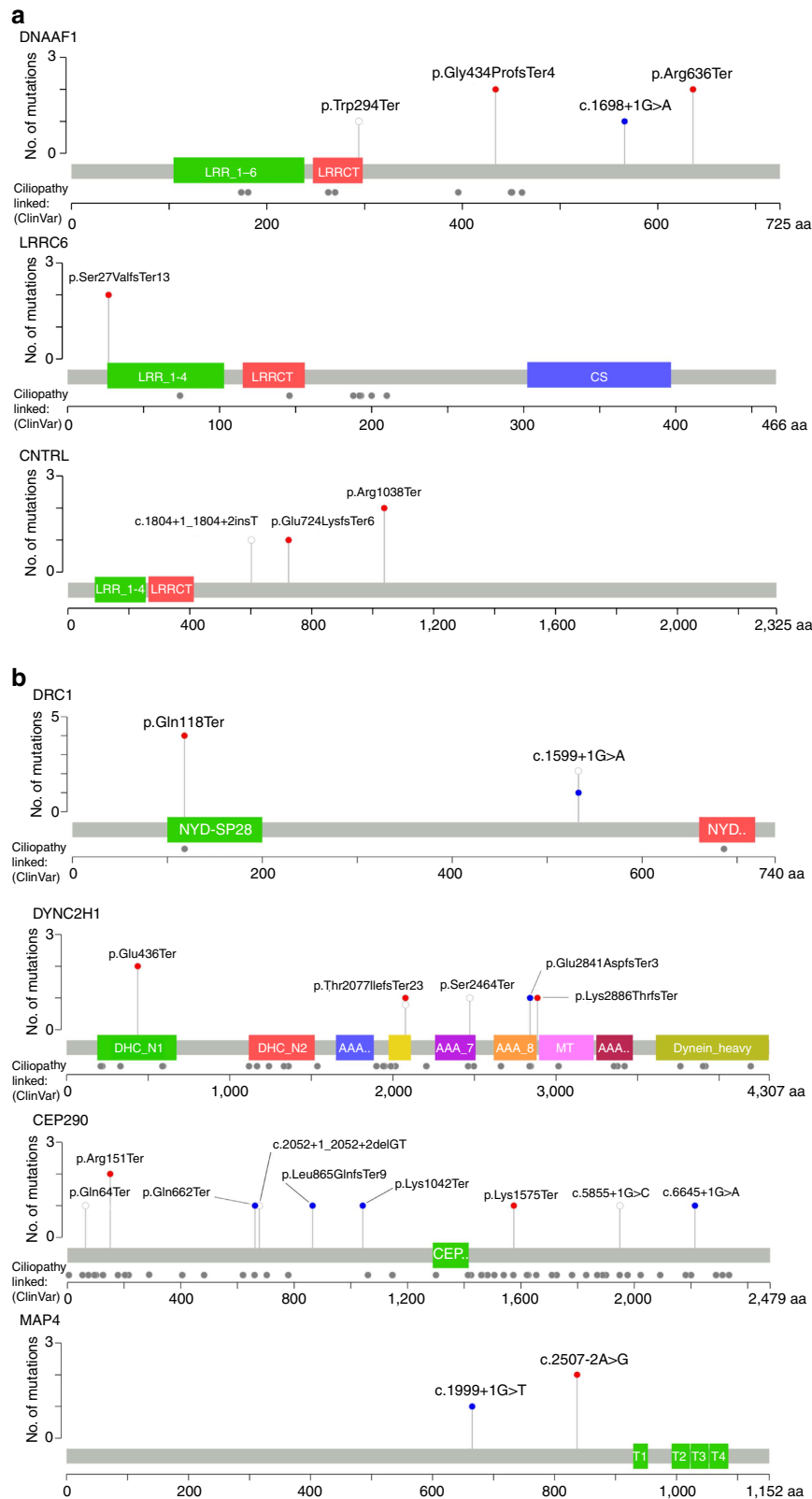
additional *DNAAF1* mutation, an additional *MAP4* mutation was observed, taking the total number of *MAP4* rare disruptive case mutations to three; whereas no rare disruptive mutations in *MAP4* were observed in 27,173 ExAC controls ( $P = 1.9 \times 10^{-5}$ , Fisher's Exact Test), the 1,644 UK controls or 4,300 European controls from the exome variant server project.

**Somatic alterations in mutation carriers.** Immunohistochemistry (IHC) staining for *DNAAF1* showed complete absence in 3/3 tumours available from *DNAAF1* mutation carriers, with presence of the protein demonstrable in normal surrounding tissue (Fig. 4). Evidence of inactivation of the second allele was also demonstrated on sequencing of two of the three tumours (Supplementary Fig. 1). Collectively these findings are compatible with *DNAAF1* having a tumour suppressor function.

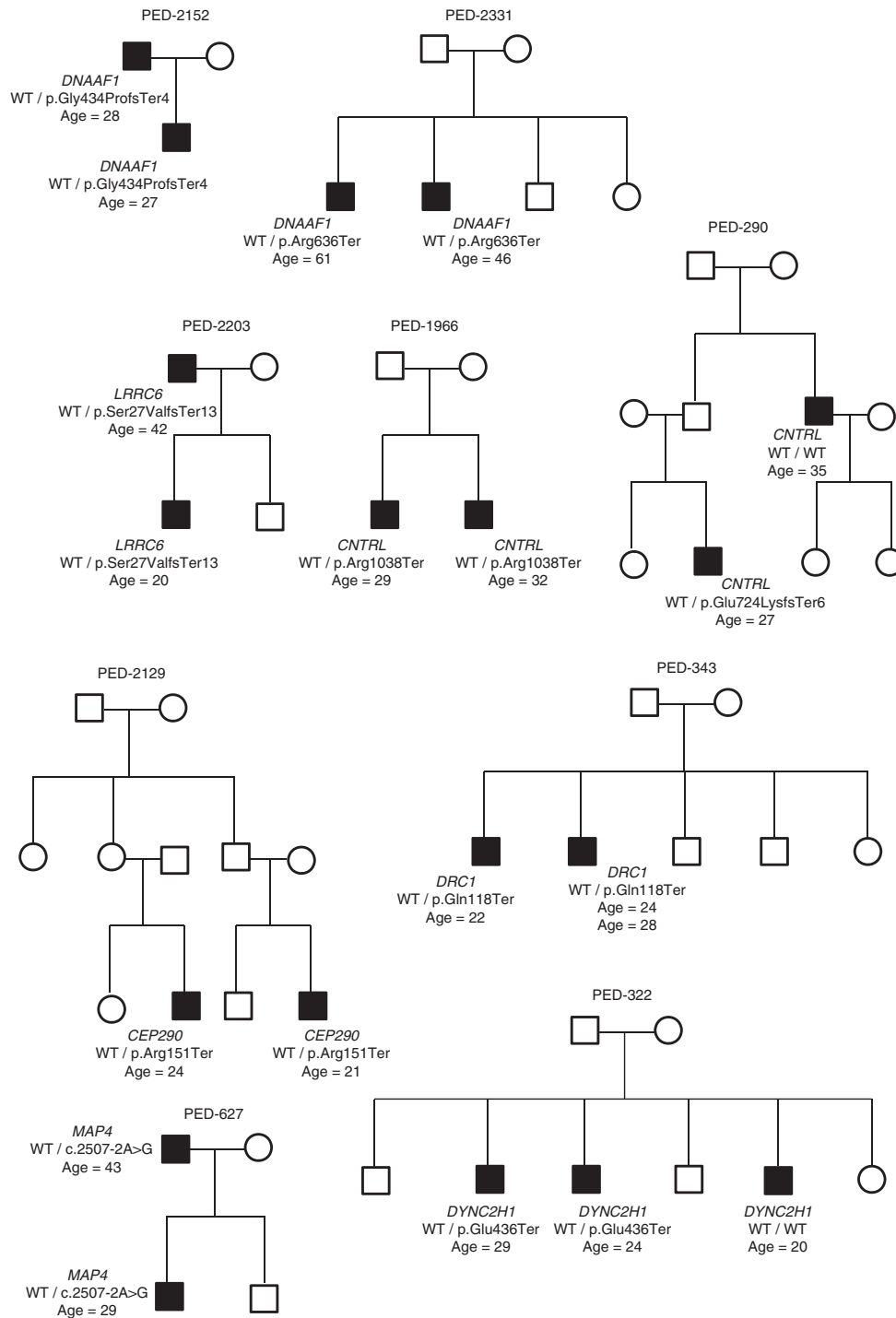
**Functional studies in zebrafish model.** We have previously implicated mutation of *DNAAF1* as a cause of TGCT in zebrafish ( $n = 30$ ) with loss of heterozygosity of *DNAAF1* demonstrated in the tumours<sup>26</sup>. To further explore the link between disruptive mutations in *DNAAF1*, ciliary function and TGCT, we conducted additional studies in *dnaaf1*<sup>hu255h</sup> mutant and wild-type zebrafish. We first examined the frequency and characteristics

of TGCTs in 136 heterozygotely mutated *dnaaf1*<sup>hu255h</sup> male zebrafish compared with 114 age-matched male wild-type fish: TGCT were observed in 94% (128/136) of *dnaaf1*<sup>hu255h</sup> (+/-) mutants as compared with 14% (16/114) of those with wild-type genotype ( $P = 3.5 \times 10^{-14}$ , Fisher's Exact Test) (Supplementary Fig. 2a). Tumours were characterized by severely reduced end-stage differentiated germ cells and an increase in early spermatogonial-like cells thus closely resembling human seminoma (Supplementary Fig. 2b). We have previously extensively demonstrated a morphological and histological analogy of zebrafish TGCTs to human TGCTs (Supplementary Fig. 2b), specifically for the subtype seminoma, based on established markers for early spermatogonia and spermatozoa. Nevertheless, despite the recognized similarities between the genera, we acknowledge there are discernible differences in gametogenesis between zebrafish and human germ cells, including distinct sex determination mechanisms of the germ cell lineage and markedly different architectural composition of the gonads in terms of somatic stem cell niche presence and a cystic expansion versus progressive tubular differentiation, respectively.

To explore the potential mechanism by which cilia function promotes TGCT formation, electron microscopy of



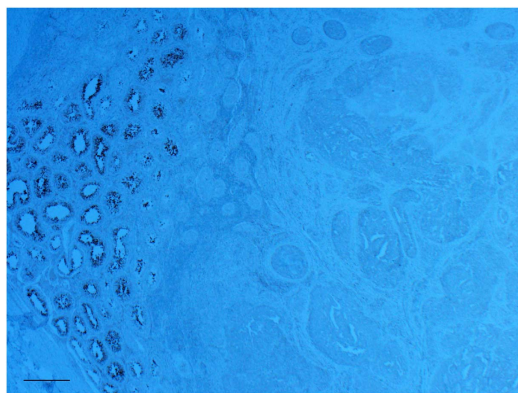
**Figure 2 | Disruptive germline mutations in cilia-microtubule pathway genes identified in TGCT cases. (a)** DnaaF1 and paralogue genes; **(b)** cilia-microtubule gene set. Red dots denote mutations identified in familial TGCT cases, blue dots denote mutations in unselected TGCT cases and white dots denote mutations in UK controls. Grey dots denote mutations catalogued by ClinVar<sup>24</sup> as a cause of recessive ciliopathy. Domain abbreviations: LRR, leucine-rich repeat; LRRCT, leucine-rich repeat C-terminal; CS, CHORD-containing proteins and SGT1; NYD-SP28, NYD-SP28 sperm tail; NYD., NYD-SP28\_assoc sperm tail C-terminal domain; DHC\_N1, dynein heavy chain, N-terminal region 1; DHC\_N2, dynein heavy chain; N-terminal region 1, AAA., hydrolytic ATP-binding site of dynein motor region D1; AAA\_7, P-loop containing dynein motor region D3; AAA\_8, P-loop containing dynein motor region D4; MT, microtubule-binding stalk of dynein motor; Dynein\_heavy pfama, dynein heavy chain and region D6 of dynein motor; CEP., coiled-coil region of centrosome protein; M1-4, Tau/MAP 1-4.



**Figure 3 | Segregating TGCT pedigrees of cilia-microtubule pathway gene carriers.** Circles, female; and squares, male. TGCT cases denoted by shaded symbols; ages refer to age at diagnosis of TGCT.

spermatogonia intercellular bridges in wild-type fish was performed. We noted that the intercellular bridges of early spermatogonia in wild-type zebrafish are flanked by ciliary structures (Supplementary Fig. 3), suggesting a role for cilia in spermatogonial stem cell differentiation, the failure of which is considered to be a fundamental precursor step in TGCT oncogenesis. This observation is also consistent with a model whereby impaired cilia function, for example through loss of *DNAAF1*, triggers structural instability preceding premature dissociation of early germ cells and concomitant differentiation disruption. Further functional studies are needed to confirm this hypothesis.

**Analysis of the cancer genome atlas data.** More than 40% of cancer susceptibility genes are found to be tumorigenic when mutated only in tumour DNA<sup>27</sup>, accordingly we sought to assess whether *DNAAF1* was also frequently lost somatically. Analysis of 150 human TGCTs publically available through the cancer genome atlas project (<http://cancergenome.nih.gov/>) showed significant focal somatic deletion at 16q23-16q24.3 ( $Q = 1.6 \times 10^{-4}$ , from GISTIC2 with significance assessed by permutation, corrected for multiple testing using the Benjamini-Hochberg method), encompassing *DNAAF1*, is a feature in 24.7% of tumours, which predominantly have seminoma histology.



**Figure 4 | IHC staining for DNAAF1 expression in available tumour tissue from mutation carriers.** IHC showing positive DNAAF1 expression in surrounding normal tissue (left) but loss of expression within the tumour (middle and right). Data are shown for tumour from PED-2152 (p.Gly434ProfsTer4). A comparable pattern was found in PED-2331 (p.Arg636Ter) and S-1645 (c.1698 + 1G > A). Scale bar, 5  $\mu$ m.

In addition *DNAAF1* is highly methylated in seminomas, as compared with non-seminomas ( $P = 2.3 \times 10^{-7}$ , Kruskal–Wallis test), and this is accompanied by downregulation of *DNAAF1* expression as compared with normal testis ( $P = 2.2 \times 10^{-16}$ , Kruskal–Wallis test). Intriguingly all of the five cases we identified with rare germline *DNAAF1* mutation had either seminoma ( $n = 4$ ) or mixed histology ( $n = 1$ ). Collectively these data are consistent with loss of *DNAAF1* having a more general impact on seminoma oncogenesis.

## Discussion

Here we report the largest WES study to date of familial and sporadic TGCT, identifying rare mutations in CMGs as determinants of disease susceptibility. A pertinent question relates to the overlapping phenotypic impact of CMG mutation, with the genetic and functional data presented here linking heterozygote mutation to TGCT risk, while homozygote mutations have been previously associated with a range of rare autosomal recessive ciliopathies. While we have identified descriptive case reports of TGCT coincident with PCD<sup>28,29</sup>, a robust statistical assessment of disease coincidence is not currently possible due to the rare nature of both conditions and the lack of systematic registry data. A model whereby heterozygote but not homozygote CMG mutation is primarily associated with TGCT risk is reconcilable, owing to the importance of timing in TGCT oncogenesis, with proliferative growth occurring in a specific post-pubertal time window. Under such a model, biallelic loss of function present at birth may trigger differing phenotypic consequences to heterozygote germline mutation followed by a second later somatic event. Finally infertility features as a common factor associated with both TGCT and PCD<sup>16,17</sup>, further supporting a potential phenotypic relationship, and while we cannot discount infertility as an intermediate phenotype promoting TGCT risk, the combined human and zebrafish rare deleterious mutational data strongly support genetic causality. In terms of functional mechanisms, we note that loss of cilia function is emerging more broadly as an important pathway in tumorigenesis in multiple cancer types<sup>30–32</sup>. The functional basis of inactivation of CMGs in oncogenesis remains to be established; however, ciliation and the cell cycle are mutually exclusive with both processes competing for the centrosome<sup>33</sup>. Hence cilia inactivation may bias towards cell cycle progression and proliferative growth.

The focus of our study was on disruptive, protein truncating mutations; with missense variants not included on account of insufficient tools to reliably predict pathogenicity<sup>12</sup>. This issue is exemplified by a previous study of *DNAAF1* (ref. 26), in which candidate missense variants in human TGCT cases were proposed as pathogenic; case–control analysis of these variants confirms that they are of equivalent frequency between cases and control series (Supplementary Table 2). In the broader context of understanding the genetic architecture of TGCT, while providing evidence for the role of rare variants as risk factors, our analysis is consistent with the previously proposed model of polygenic susceptibility, in which much of the heritable risk of TGCT is associated with common genetic variants<sup>34</sup>. Such a model is supported by the recent GWAS which have so far identified 25 risk loci which collectively account for 19% of the familial TGCT risk<sup>34–47</sup>.

In conclusion, we have provided evidence for the role of inherited mutations in CMGs as determinants of TGCT, identifying germline disruptive mutations in 9% of familial pedigrees, with additional evidence implicating disruptive mutations in *DNAAF1* in TGC tumorigenesis from IHC and sequencing studies of human tumours and a *dnaf1*<sup>hu255h</sup> (+/–) zebrafish model displaying a 94% frequency of TGCT. As well as revealing insights into disease pathways, our data provides a resource for contextualizing the impact of future candidate TGCT genes.

## Methods

**Ethics.** Written informed consent was obtained from all individuals with ethical review board approval (UK National Cancer Research Network Multi-Research Ethics Committee—MREC02/06/66, 06/MRE06/41) and the study was conducted in accordance with the declaration of Helsinki.

**Subjects and data sets.** TGCT cases were ascertained from (i) ‘The UK Genetics of Testicular Cancer Study’ and (ii) ‘Identification, epidemiological and molecular analyses of families with susceptibility to TGCT’ (recruitment via the UK Testicular Cancer Collaboration and International Testicular Cancer Linkage Consortium; Supplementary Notes 1 and 2) which were coordinated by The Institute of Cancer Research. All cases had self-reported European ancestry. Of the 328 familial cases 289 were of UK origin, and 39 were of non-UK European ancestry. All 634 sporadic cases were from the UK. The controls comprised 1,644 healthy individuals from the UK 1958 Birth Cohort<sup>48</sup>—974 from the ICR1000 data set (EGAD00001001021) and an additional 670 individuals (EGAS00001001667) also sequenced at Institute of Cancer Research under the same protocol.

**Whole-exome sequencing.** 1  $\mu$ g of DNA from each individual was fragmented using a Covaris E Series instrument (Covaris Inc. Woburn, MA, USA). Indexed paired-end libraries were prepared using Illumina TruSeq 62 Mb expanded exome enrichment kit (Illumina, San Diego, CA, USA). Forty-nine samples with insufficient input DNA were prepared using Illumina Nextera Rapid Capture 37 Mb exome enrichment kit. The  $2 \times 100$  bp sequencing was performed using Illumina HiSeq2000 or 2500 technology.

**Read mapping and variant analysis.** Paired-end fastq files were extracted using CASAVA software (v.1.8.1, Illumina) and aligned to build 37 (hg19) of the human reference genome using Stampy and BWA software<sup>49</sup>. Alignments were processed using the Genome Analysis Tool Kit (GATKv3) pipeline according to best practices<sup>50,51</sup>. Analysis was restricted for all samples to the capture regions defined in the Truseq 62 Mb bed file plus 100 bp padding. The Variant Effect Predictor (VEP) was used to provide annotations on the predicted impact of each variant. We additionally annotated with alignability of 100mers and distance from simple repeats defined by University of California, Santa Cruz (UCSC) browser tracks. Mean coverage of  $49 \times$  was achieved across targeted bases with 83% being covered at  $\geq 15 \times$ ; cases and controls had similar technical sequencing metrics. We excluded 57 subjects with low-quality data ( $< 50\%$  of bases covered with minimum  $15 \times$ ) or non-European ancestry. An additional 21 controls samples were excluded due to sex discrepancy and detection of cancer history during the course of the analysis.

We considered only canonical transcripts and for each variant, assuming the most deleterious predicted effect for each transcript according to VEP. For all analyses we imposed GATK internal calling thresholds excluding variants as per current best practice guidelines<sup>50,51</sup>—in the 99.5th truth tranche for single-nucleotide variants and  $> 99$ th tranche for indels. To minimize false positives we adopted an automated approach imposing:  $GQ \geq 30$ , for a heterozygous call an alternate depth  $\geq 3$  and  $\chi^2 < 10.83$  (that is,  $P > 0.0001$ ) for the observed versus

expected distribution of alternate/reference alleles (alt-ref-ratio), UCSC alignability (100 bp window size) = 1, not in simple repeat, Hardy–Weinberg equilibrium test ( $P > 1.0 \times 10^{-8}$ ) in cases and controls and an overall call rate  $\geq 75\%$  in both cases and controls. To eliminate false-positive associations caused by indels, genes featuring only this class of mutation type were filtered out. To ensure a high-quality variant set and reproducibility was verified for a subset of samples ( $n = 162$ ) which had also been genotyped using Illumina HumanCNV370-Duo bead arrays— $> 99\%$  concordance was observed between platforms. Following QC the final data set comprised 306 familial TGCT cases (from 150 independent pedigrees), 613 unselected TGCT cases and 1,609 controls. Variants previously known to cause recessive ciliopathies (as displayed in Fig. 2) were extracted from ClinVar<sup>24</sup>, filtering to select only ciliopathic phenotypes and to remove variants classified as benign.

**Statistical analyses.** To test whether rare mutations contribute to TGCT we performed a collapsing burden test imposing a maximal MAF threshold of 1%. To ensure independent events, case counts were based on one individual per pedigree randomly assigned as the proband. Significance levels were assessed using  $10^5$  permutations on case/control status. Study power to detect was calculated using disease allele frequency in controls as the baseline allele frequency, while the frequency in cases determined by a weighted average of the enrichment found in cases with one, two and three affected first-degree relatives. Allele counts were then sampled from frequencies between 0.00001 and 0.01 and relative risks between 1.75 and 10.0. A Fisher's test was then performed for each sampling of cases and controls. This process was performed 10,000 times for each frequency/relative risk combination and for each instance the frequency of tests that were significant at an exome-wide significance of  $8.0 \times 10^{-7}$  equated to study power. Statistical analyses were carried out using R3.0.2 and Stata (v12) (StataCorp, Lakeway Drive College Station, TX, USA) software.

**Pathway analysis.** To test for over-representation of damaging variants within genes mapping to a specific pathways or biological process we performed a GSEA. This comprised 1,090 canonical gene sets from KEGG, GO: biological processes and Reactome pathways, supplemented by an OMIM search-term-driven method including genes expressed in normal testicular tissue<sup>52</sup> and genes with evidence of somatic inactivation from TGCT. A pre-ranked GSEA was then performed for all gene sets, with ranking of genes based on their permuted  $P$  values for familial cases compared with controls. For replication only the significant ( $Q < 0.1$ ) cilia-microtubule pathway was evaluated.

**Immunohistochemistry.** Sections for each case were cut at 4  $\mu\text{m}$  and placed on Superfrost plus slides. All slides were baked in the oven at 72 °C for 40 min before staining. Slides were stained on the Ventana BenchMark ULTRA instruments using the Ventana OptiView DAB detection kit. Slides underwent a heat mediated antigen retrieval step using Ventana Cell Conditioning 1 reagent, heating slides to 100 °C for 64 min. Following this, slides were incubated with the DNAAF1 polyclonal antibody (Abnova, catalogue number PAB23762)) for 32 min, using a dilution 1:500.

**Tumour sequencing.** Rare disruptive variants in CMG found by WES were confirmed by Sanger sequencing, and examined for loss of heterozygosity in tumour samples relative to germline samples. Primers flanking variants were designed using Primer3 (ref. 53) (Supplementary Table 3). PCR amplicons of germline and tumour samples were bidirectionally sequenced using the BigDye Terminator Cycle sequencing kit and an Applied Biosystems 3730xl DNA Analyzer (Life Technologies). Sequence traces were inspected using Mutation Surveyor software (SoftGenetics, State College, PA, USA). Heterozygote variants confirmed in germline sequences were examined in corresponding tumour samples where tumour blocks were available.

**Zebrafish model.** All animal experiments were ethically approved by the Animal Care Committee of the Royal Dutch Academy of Science according to the Dutch legal ethical guidelines. The human tumour samples were approved by an institutional review board (MEC 02.981). Samples were used according to the 'Code for Proper Secondary Use of Human Tissue in the Netherlands,' developed by the Dutch Federation of Medical Scientific Societies<sup>26</sup>. Embryos were generated by natural pair-wise matings of heterozygous carriers and raised at 28.5 °C on a 14 h light/10 h dark cycle in a 100 mm<sup>2</sup> petri dish containing aquarium water. *dnaaf1<sup>hu255h</sup>* have a T/A mutation, changing a conserved leucine into a stop codon (L88X; ENSEMBL gene: ZGC::56169) and randomly selected control zebrafish of the same background were maintained according to standard protocols. Before tissue isolation, zebrafish were euthanized by overdose of MS222 or ice bath. Sections for IHC were fixed overnight in 4% paraformaldehyde and 2% acetic acid, embedded in paraffin and 6  $\mu\text{m}$  obtained. Tissue for morphological analysis was fixed using 4% glutaraldehyde and embedded in glycol methacrylate (Technovit 7100, Hereaus Kulzer), and 4- $\mu\text{m}$  sections stained with toluidine blue. Images were captured using a Nikon Eclipse E800 equipped with a Nikon

DXM1200 digital camera and Plan Apo  $\times 2/0.1$ ,  $\times 10/0.45$ ,  $\times 20/0.75$  and  $\times 40/0.95$  NA objectives.

**Electron microscopy.** Testes tissue was fixed in Karnovsky fixative (2% paraformaldehyde, 2.5% glutaraldehyde, 0.08M Na-cacodylate (pH 7.4), 0.25 mM calcium chloride and 0.5 mM magnesium chloride (pH 7.4)) for at least 24 h at 4 °C. Samples were postfixed in 1% osmiumtetroxide and embedded in Epon 812. Ultrathin sections (60nm) were contrasted with 3% uranyl magnesium acetate and lead citrate and viewed with a Jeol (<http://www.jeol.com/>) JEM 1010 or a Philips (Eindhoven, The Netherlands) CM10 transmission electron microscope.

**Analysis of The Cancer Genome Atlas data.** Copy number status, RNA-Seq expression data (RPKM counts) and gene methylation data were extracted for 'Testicular Germ Cell Tumours' from the TCGA Broad Firehose pipeline run on 28 January 2016. Normal testicular tissue RNA-Seq expression data (RPKM counts) were downloaded from GTEx<sup>54</sup>. Associations between differential methylation and expression levels across seminoma, non-seminoma and normal testicular tissue were quantified using the Kruskal–Wallis trend test.

#### Web addresses.

Genome Analysis Tool Kit (GATKv3): <https://www.broadinstitute.org/gatk>  
 Online Mendelian Inheritance in Man (OMIM): <http://omim.org/>  
 Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA, USA: <http://evs.gs.washington.edu/EVS> (accessed January 2016)  
 Exome Aggregation Consortium (EXAC): <http://exac.broadinstitute.org/>  
 Gene set enrichment analysis (GSEA): <http://software.broadinstitute.org/gsea/index.jsp>  
 The cancer genome atlas (TCGA): <http://cancergenome.nih.gov/>  
 Broad Firehose: <https://gdac.broadinstitute.org/>

**Data availability.** The WES data that supports this study have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI); accession numbers EGAS00001001789, EGAD00001001021 and EGAS00001001667. The TCGA data is available from the database of Genotypes and Phenotypes (dbGaP), Study Accession: phs000178.v9.p8. The remaining data are available within the article and its Supplementary Information files or available from the authors upon request.

#### References

- Bray, F., Ferlay, J., Devesa, S. S., McGlynn, K. A. & Moller, H. Interpreting the international trends in testicular seminoma and nonseminoma incidence. *Nat. Clin. Pract. Urol.* **3**, 532–543 (2006).
- Ruf, C. G. *et al.* Changes in epidemiologic features of testicular germ cell cancer: age at diagnosis and relative frequency of seminoma are constantly and significantly increasing. *Urol. Oncol.* **32**, e1–e6 (2014).
- Le Cornet, C. *et al.* Testicular cancer incidence to rise by 25% by 2025 in Europe? Model-based predictions in 40 countries using population-based registry data. *Eur. J. Cancer.* **50**, 831–839 (2014).
- de Haas, E. C. *et al.* Early development of the metabolic syndrome after chemotherapy for testicular cancer. *Ann. Oncol.* **24**, 749–755 (2013).
- Bujan, L. *et al.* Impact of chemotherapy and radiotherapy for testicular germ cell tumors on spermatogenesis and sperm DNA: a multicenter prospective study from the CECOS network. *Fertil. Steril.* **100**, 673–680 (2013).
- Rusner, C. *et al.* Risk of second primary cancers after testicular cancer in East and West Germany: a focus on contralateral testicular cancers. *Asian J. Androl.* **16**, 285–289 (2014).
- Swerdlow, A. J., De Stavola, B. L., Swanwick, M. A. & Maconochie, N. E. Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology. *Lancet* **350**, 1723–1728 (1997).
- McGlynn, K. A., Devesa, S. S., Graubard, B. I. & Castle, P. E. Increasing incidence of testicular germ cell tumors among black men in the United States. *J. Clin. Oncol.* **23**, 5757–5761 (2005).
- Hemminki, K. & Li, X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *Br. J. Cancer* **90**, 1765–1770 (2004).
- Kharazmi, E. *et al.* Cancer risk in relatives of testicular cancer patients by histology type and age at diagnosis: a joint study from five nordic countries. *Eur. Urol.* **68**, 283–289 (2015).
- Crockford, G. P. *et al.* Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. *Hum. Mol. Genet.* **15**, 443–451 (2006).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- van Rooijen, E. *et al.* LRRc50, a conserved ciliary protein implicated in polycystic kidney disease. *J. Am. Soc. Nephrol.* **19**, 1128–1138 (2008).
- Loges, N. T. *et al.* Deletions and point mutations of *Lrrc50* cause primary ciliary dyskinesia due to dynein arm defects. *Am. J. Hum. Genet.* **85**, 883–889 (2009).

15. McGlynn, K. A. & Trabert, B. Adolescent and adult risk factors for testicular cancer. *Nat. Rev. Urol.* **9**, 339–349 (2012).
16. Lucas, J. S. *et al.* Diagnosis and management of primary ciliary dyskinesia. *Arch. Dis. Child.* **99**, 850–856 (2014).
17. Rattei, T. *et al.* SIMAP: the similarity matrix of proteins. *Nucleic Acids Res.* **34**, D252–D256 (2006).
18. Aken, B. L. *et al.* The Ensembl gene annotation system. *Database.* <http://dx.doi.org/10.1093/database/baw093> (2016).
19. Wirschell, M. *et al.* The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. *Nat. Genet.* **45**, 262–268 (2013).
20. Dagoneau, N. *et al.* *DYNC2H1* mutations cause asphyxiating thoracic dystrophy and short rib-polydactyly syndrome, type III. *Am. J. Hum. Genet.* **84**, 706–711 (2009).
21. Brancati, F. *et al.* *CEP290* mutations are frequently identified in the oculo-renal form of Joubert syndrome-related disorders. *Am. J. Hum. Genet.* **81**, 104–113 (2007).
22. Sayer, J. A. *et al.* The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. *Nat. Genet.* **38**, 674–681 (2006).
23. Bizet, A. A. *et al.* Mutations in TRAF3IP1/IFT54 reveal a new role for IFT proteins in microtubule stabilization. *Nat. Commun.* **6**, 8666 (2015).
24. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
25. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
26. Basten, S. G. *et al.* Mutations in *LRRC50* predispose zebrafish and humans to seminomas. *PLoS. Genet.* **9**, e1003384 (2013).
27. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).
28. Meyer, C. H. On the combination of Kartagener syndrome and malignant embryonal testis teratoma. *Frankf. Z. Pathol.* **76**, 21–31 (1966).
29. Barselo, E. R. *et al.* Testicular seminoma in a patient with Kartagener's syndrome. *Arch. Esp. Urol.* **61**, 431–434 (2008).
30. Schraml, P. *et al.* Sporadic clear cell renal cell carcinoma but not the papillary type is characterized by severely reduced frequency of primary cilia. *Mod. Pathol.* **22**, 31–36 (2009).
31. Wong, S. Y. *et al.* Primary cilia can both mediate and suppress Hedgehog pathway-dependent tumorigenesis. *Nat. Med.* **15**, 1055–U109 (2009).
32. Yuan, K. *et al.* Primary cilia are decreased in breast cancer: analysis of a collection of human breast cancer cell lines and tissues. *J. Histochem. Cytochem.* **58**, 857–870 (2010).
33. Basten, S. G. & Giles, R. H. Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis. *Cilia* **2**, 6 (2013).
34. Litchfield, K. *et al.* Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Sci. Rep.* **5**, 13889 (2015).
35. Rapley, E. A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* **41**, 807–810 (2009).
36. Turnbull, C. & Rahman, N. Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. *Int. J. Androl.* **34**, e86–e96 (2011).
37. Kanetsky, P. A. *et al.* Common variation in *KITLG* and at 5q31.3 predisposes to testicular germ cell cancer. *Nat. Genet.* **41**, 811–815 (2009).
38. Turnbull, C. *et al.* Variants near *DMRT1*, *TERT* and *ATF7IP* are associated with testicular germ cell cancer. *Nat. Genet.* **42**, 604–607 (2010).
39. Kanetsky, P. A. *et al.* A second independent locus within *DMRT1* is associated with testicular germ cell tumor susceptibility. *Hum. Mol. Genet.* **20**, 3109–3117 (2011).
40. Ruark, E. *et al.* Identification of nine new susceptibility loci for testicular cancer, including variants near *DAZL* and *PRDM14*. *Nat. Genet.* **45**, 686–689 (2013).
41. Bojesen, S. E. *et al.* Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45**, 371–384 (2013).
42. Chung, C. C. *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* **45**, 680–685 (2013).
43. Litchfield, K. *et al.* Polygenic susceptibility to testicular cancer: implications for personalised health care. *Br. J. Cancer* **113**, 1512–1518 (2015).
44. Litchfield, K. *et al.* Identification of four new susceptibility loci for testicular germ cell tumour. *Nat. Commun.* **6**, 8690 (2015).
45. Litchfield, K. *et al.* Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. *Hum. Mol. Genet.* **24**, 1169–1176 (2015).
46. Litchfield, K., Shipley, J. & Turnbull, C. Common variants identified in genome-wide association studies of testicular germ cell tumour: an update, biological insights and clinical application. *Andrology* **3**, 34–46 (2015).
47. Kristiansen, W. *et al.* Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. *Hum. Mol. Genet.* **24**, 4138–4146 (2015).
48. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* **35**, 34–41 (2006).
49. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
50. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
51. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
52. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
53. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
54. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

## Acknowledgements

We thank the patients and their clinicians for participation in this study. We acknowledge the International Testicular Cancer Linkage Consortium and the UK Testicular Cancer Collaboration (members defined in Supplementary Notes 1 and 2). We acknowledge the National Health Service funding to the National Institute for Health Research Biomedical Research Centre. We acknowledge Michael Stratton, Gillian Crockford and Rachel Linger for coordination of family recruitment and all members of the ITCLC. We acknowledge the facilities and expertise of the Cancer Genetics Core Laboratory Facility and Cancer Genetics Sequencing Facility at The Institute of Cancer Research directed by Nazneen Rahman, supported by Sheila Seal, Anthony Renwick, Emma Ramsey and Elise Ruark. This study was supported by the Movember foundation and the Institute of Cancer Research. K.L. is supported by a PhD Studentship from Cancer Research UK. R.S.H. and P.B. are supported by Cancer Research UK (C1298/A8362 Bobby Moore Fund for Cancer Research UK). We thank Glenn van de Hoek for assistance with zebrafish experiments, and R.H.G. acknowledges funding from the Dutch Kidney Foundation KOUNCIL program CP11.18.

## Author contributions

C.T. and K.L. designed the study with input from R.S.H. E.R., R.H., A.R. and D.T.B. lead patient recruitment. D.D., M.L. and K.L. coordinated assembly and tracking of blood and tumour samples. K.L., D.G.d.C., P.P. and C.S. coordinated management of tumour tissue and DNA extraction. K.L., M.L. and P.B. designed and conducted laboratory experiments. K.L., C.T. and R.S.H. designed bioinformatic analyses. K.L. conducted bioinformatics and statistical analyses. S.O.'C. undertook histological examination and IHC analysis of tumour tissue. R.H.G. and S.B. designed and conducted zebrafish analyses. K.L. drafted the manuscript with assistance from C.T., R.S.H., R.H.G. and M.L., K.L. generated images for publication, supported by M.L. All authors reviewed and contributed to the 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:** Litchfield K. *et al.* Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. *Nat. Commun.* **7**, 13840 doi: 10.1038/ncomms13840 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016