



**HAL**  
open science

## Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours

Kevin Litchfield, Brenda Summersgill, Shawn Yost, Razvan Sultana, Karim Labreche, Darshna Dudakia, Anthony Renwick, Sheila Seal, Reem Al-Saadi, Peter Broderick, et al.

### ► To cite this version:

Kevin Litchfield, Brenda Summersgill, Shawn Yost, Razvan Sultana, Karim Labreche, et al.. Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nature Communications*, 2015, 6, pp.5973. 10.1038/ncomms6973 . hal-01116429

**HAL Id: hal-01116429**

**<https://hal.sorbonne-universite.fr/hal-01116429>**

Submitted on 13 Feb 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARTICLE

Received 26 Sep 2014 | Accepted 25 Nov 2014 | Published 22 Jan 2015

DOI: 10.1038/ncomms6973

OPEN

# Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours

Kevin Litchfield<sup>1</sup>, Brenda Summersgill<sup>2</sup>, Shawn Yost<sup>1</sup>, Razvan Sultana<sup>1</sup>, Karim Labreche<sup>1,3</sup>, Darshna Dudakia<sup>1</sup>, Anthony Renwick<sup>1</sup>, Sheila Seal<sup>1</sup>, Reem Al-Saadi<sup>2</sup>, Peter Broderick<sup>1</sup>, Nicholas C. Turner<sup>4</sup>, Richard S. Houlston<sup>1</sup>, Robert Huddart<sup>5</sup>, Janet Shipley<sup>2</sup> & Clare Turnbull<sup>1,6</sup>

Testicular germ cell tumours (TGCTs) are the most common cancer in young men. Here we perform whole-exome sequencing (WES) of 42 TGCTs to comprehensively study the cancer's mutational profile. The mutation rate is uniformly low in all of the tumours (mean 0.5 mutations per Mb) as compared with common cancers, consistent with the embryological origin of TGCT. In addition to expected copy number gain of chromosome 12p and mutation of *KIT*, we identify recurrent mutations in the tumour suppressor gene *CDC27* (11.9%). Copy number analysis reveals recurring amplification of the spermatocyte development gene *FSIP2* (15.3%) and a 0.4 Mb region at Xq28 (15.3%). Two treatment-refractory patients are shown to harbour *XRCC2* mutations, a gene strongly implicated in defining cisplatin resistance. Our findings provide further insights into genes involved in the development and progression of TGCT.

<sup>1</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. <sup>2</sup>Divisions of Molecular Pathology and Cancer Therapeutics, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. <sup>3</sup>Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75019, Paris, France. <sup>4</sup>The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. <sup>5</sup>Academic Radiotherapy Unit, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. <sup>6</sup>William Harvey Research Institute, Queen Mary University London, Charterhouse Square, London EC1M 6BQ, UK. Correspondence and requests for materials should be addressed to C.T. (email: clare.turnbull@icr.ac.uk).

**T**GCTs are the most common cancer affecting young men, with a mean age at diagnosis of 36 years<sup>1,2</sup>. The main TGCT histologies are seminomas, which resemble undifferentiated primary germ cells, and non-seminomas, which show differing degrees of differentiation. Cure rates for TGCTs are generally high, due to the sensitivity of malignant testicular germ cells to platinum-based chemotherapies, however this is at the cost of an increased risk of metabolic syndrome, infertility and secondary cancer<sup>3–5</sup>. Furthermore, there are limited options for the patients who are platinum resistant, a group for whom the long-term survival rate is poor<sup>6</sup>.

Overall, TGCTs are markedly aneuploid with recurring gain of chromosomes 7, 8, 21, 22 and X<sup>7–13</sup>. In addition, gain of chromosomal material from 12p is noted in virtually all cases<sup>7–9</sup>, with genomic amplification and overexpression of genes in the 12p11.2-p12.1 region reported in ~10% of TGCTs<sup>14</sup>. *KRAS* is located in this region and has been proposed as the candidate driver<sup>14</sup>. Focused studies of TGCTs have identified somatic missense mutations and amplifications of the oncogene *KIT*, present in ~25% of seminomas<sup>15,16</sup>. These reported mutations are clustered in the juxta membrane and kinase encoding domains of *KIT*<sup>15,16</sup>. However, a study of 518 other protein kinase encoding genes failed to conclusively identify any new driver mutations<sup>17</sup>. Beyond these focused interrogations of specific genes, no systematic mutational analysis across all genes in a large series of TGCT samples has been reported to our knowledge.

Here we perform WES of a series of 42 TGCTs to characterize the mutational signature of these tumours and to search for additional driver mutations and pathways disrupted. Our analyses demonstrate these tumours to be relatively homogeneous in profile with a markedly low rate of non-synonymous mutations and provide some novel insights into the genomic architecture of this biologically interesting tumour type.

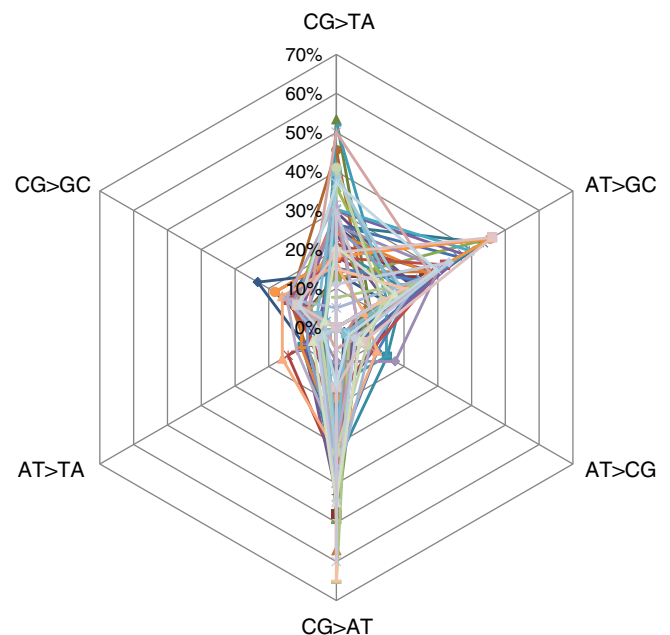
## Results

**Overview of TGCT mutational landscape.** The 42 TGCT cases comprised 16 seminomas, 18 non-seminomas, 4 mixed seminoma/non-seminoma histology and 4 tumours of indeterminate classification. Fresh frozen tumour tissue and matched germline blood samples were obtained from each patient and WES was performed on extracted DNA, achieving mean coverage of  $72 \times$  across targeted bases with 86% of targeted bases being covered at  $\geq 20$ . Sequencing was conducted using Illumina technology, with subsequent alignment, mapping and variant calling performed using Burrows–Wheeler Aligner (BWA)/Stampy/GATK/MuTect software. Across all 42 cases a total of 1,168 somatic single nucleotide variants (SNVs), and 111 small scale somatic insertion–deletions (indels) were identified, resulting in a combined total of 795 non-synonymous mutations, equating to a mean rate of 0.51 somatic mutations per Mb. By comparison, recent large-scale analysis across 27 cancer types recorded mean rates as high as  $11.0 \text{ Mb}^{-1}$  in melanoma and  $8.0 \text{ Mb}^{-1}$  in lung cancers with a mean rate across all tumour types of  $4.0 \text{ Mb}^{-1}$ , some eight times higher than that seen here in TGCT (ref. 18). Indeed the mutation rate in TGCT is within the second lowest decile, only marginally greater than paediatric cancers such as Ewing sarcoma ( $0.3 \text{ Mb}^{-1}$ ) and Rhabdoid tumour ( $0.15 \text{ Mb}^{-1}$ ). This observation is entirely consistent with oncogenic origins of TGCT arising during embryonic development<sup>19</sup>. Of additional note is the high intra-patient homogeneity in mutation rate present in our data, with a s.d. of just 0.24 across the 42 tumours and the extreme lowest to extreme highest mutation rate varying by only 1 order of magnitude. This variation is low compared with the 3 orders of magnitude inter-sample variation observed for acute

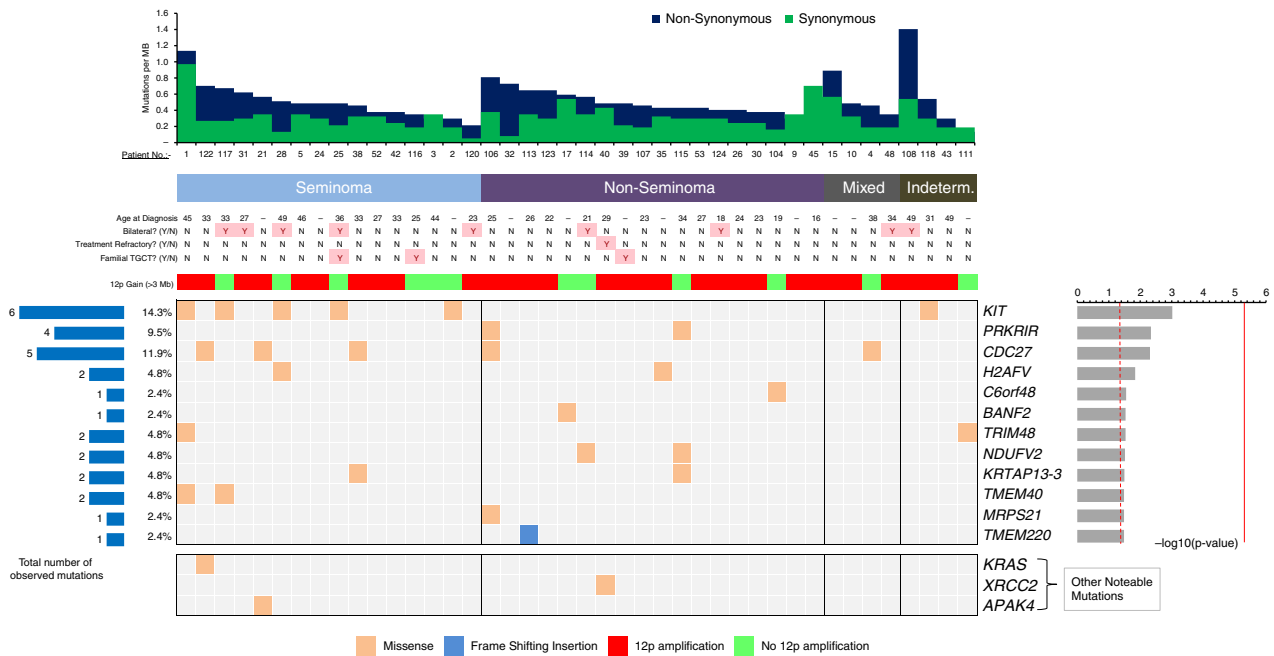
myeloid leukaemia, which has a comparable mutation rate<sup>18</sup>. Of note, there were no genes that were recurrently mutated or structural variants shared between the tumours in which the mutational rate was  $> 2$  s.d. above the mean (two tumours). The mutational spectrum of SNVs in the TGCTs was typified by an excess of CG>TA transitions (27% of SNVs), as observed in most solid tumours<sup>18,20</sup> (Fig. 1). In addition, TA>CG transitions (23%) as well as CG>AT transversions (31%, of which the majority were C>A) were also over-represented. While C>A transversions are observed at higher proportion in lung cancers postulated to be due to exposure to tobacco carcinogens<sup>18</sup>, this pattern is also has also been reported in melanoma, neuroblastoma and chronic lymphocytic leukaemia<sup>21</sup>.

**Driver genes.** We used MutSigCV version 1.4 to identify genes harbouring more non-synonymous mutations than expected by chance given gene size, sequence context and gene-specific background mutation rates<sup>18</sup>. *KIT* was identified as the most significantly mutated gene (Fig. 2), with mutations seen in 14.3% across all TGC tumours, but predominantly found in seminomas (31.3%); a result consistent with previously reported observations<sup>16,22</sup>. All of the six *KIT* mutations we identified were in hotspot domains—five non-synonymous SNVs in exon 17 (kinase encoding domain) and one in exon 11 (juxta membrane domain). The absence of another gene ranked above *KIT* is a notable result, given our study assesses an exome-wide complement of genes. In addition to *KIT*, a non-synonymous SNV was also observed in previously proposed TGCT driver gene *KRAS*. While *p53* mutations have been suggested to be a feature of TGCT<sup>23</sup>, none were observed in our data set, consistent with most recent studies<sup>17,24,25</sup>. We validated all *KIT/KRAS* mutations called by next generation sequencing (NGS) using Sanger sequencing of the respective exons across all samples and to ensure no additional mutations were missed. In all cases, Sanger sequencing was 100% concordant with NGS.

In addition to *KIT* and *KRAS*, there was an over-representation of mutations in cell division cycle 27 (*CDC27*) (11.9%;



**Figure 1 | TGCT somatic SNV spectrum exome wide.** Proportions are displayed for all 12 possible SNV alterations, collapsed by strand complementarity. Each line represents one of the 42 tumours.



**Figure 2 | Mutated genes in testicular germ cell tumour by histological subtype.** The top bars represent somatic mutation rate per sample for the 42 samples (synonymous and non-synonymous (including small-scale indels)). The genes listed on the right are mutated genes as prioritized by MutSigCV, ranked by  $-\log_{10}(P \text{ value})$  (far right), with the dotted red line denoting a significance threshold of  $P=0.05$  and the solid red line a genome-wide significance threshold of  $5 \times 10^{-6}$  (see Methods). Below the top ranked genes in a separate box are other notable but non-significant mutations. Mutations by sample are depicted in the central box, with colour indicating mutation type as per the legend. The far left bars represent the absolute number of mutations observed per gene across all samples and adjacent to this is the % of samples this represents.

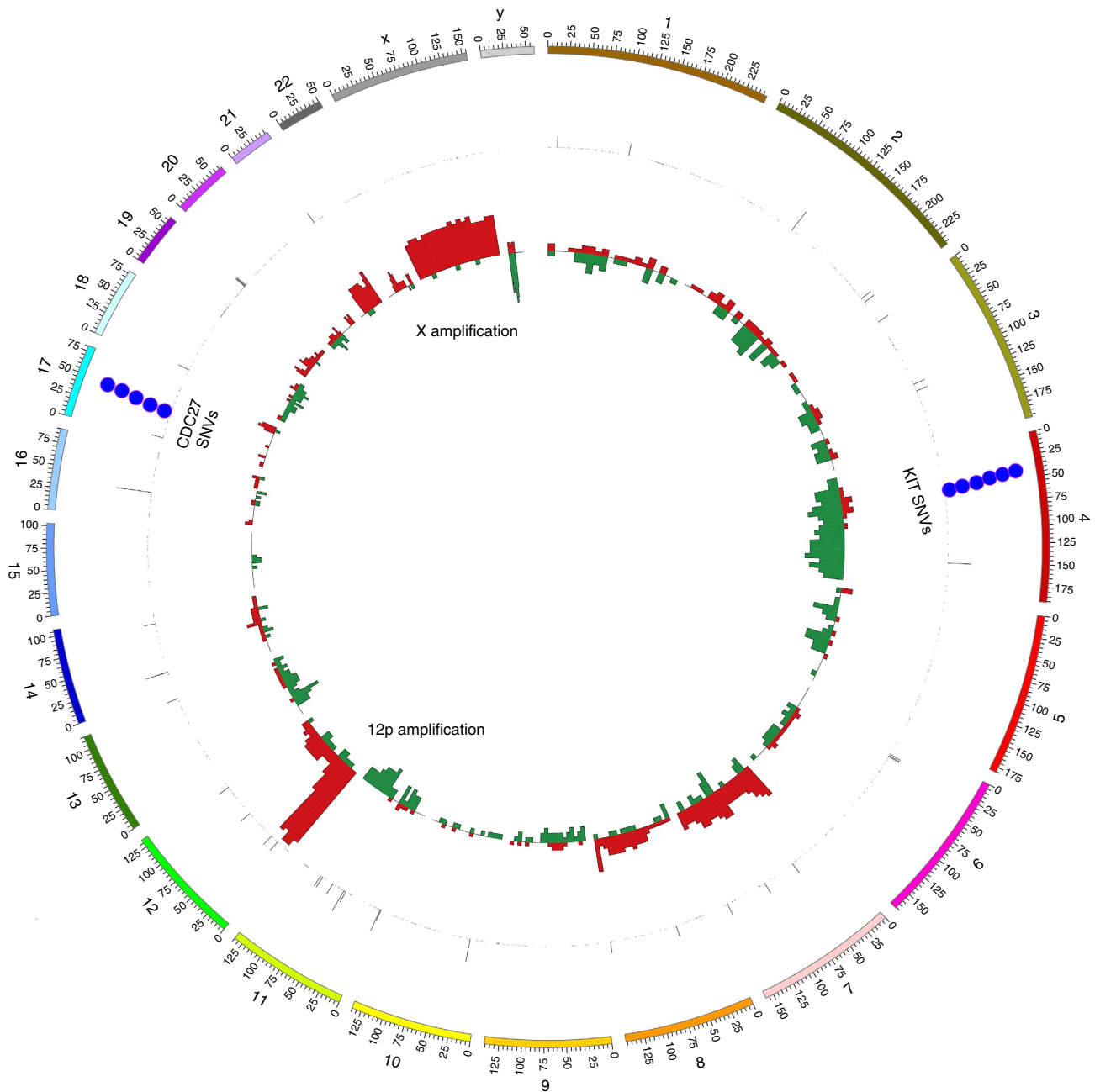
5 mutations, 5 tumours) and *PRKRIR* (4 mutations, 2 tumours), neither of which have been previously reported as TGCT drivers. *CDC27* is a core component of the anaphase-promoting complex/cyclosome, a multi-subunit E3 ubiquitin ligase that governs cell cycle progression, through ubiquitination and degradation of G1/mitotic checkpoint regulators<sup>26</sup>. Anaphase-promoting complex/cyclosome recruits its substrates via one of the two adaptor proteins CDC20 or CDH1, overexpression of which have been linked to multiple tumours<sup>27–29</sup>. *CDC27* is downregulated in breast cancer and *CDC27* is postulated to be a tumour suppressor<sup>30</sup>. All of the *CDC27* mutations we identified were missense variants, characterized by a consistently low frequency of mutant allelic reads (8–14%), consistent with *CDC27* mutation being present only in a subclone of each tumour sample. Intriguingly subclonal low frequency of *CDC27* mutation has also recently been demonstrated in a colonic adenocarcinoma<sup>31</sup>.

**Pathway analysis.** To increase our ability to identify cancer drivers and delineate associated oncogenic pathways for TGCT, we incorporated mutation data from multiple tumour types using OncoPrint-fm<sup>32</sup> as implemented within the IntOGen-mutations platform<sup>33</sup>. The most frequently mutated pathways were those involved in metabolism (mutated in 93%), pathways in cancer (54%), endocytosis (54%) and PI3K–Akt signalling (54%). The most significantly mutated pathway was RNA degradation (14.6%), with a biased accumulation of functional mutations (fm-bias,  $P=3.8 \times 10^{-3}$ ), observed across six different genes (see methods and Supplementary Table 2).

**Copy number variation.** The 42 tumours were analyzed for copy number variation (CNV) using software package ExomeCNV<sup>34</sup>. Focal CNVs (up to 3 Mb) were identified in all tumours and large-scale CNVs ( $\geq 3$  Mb) were detected in 35 (83%) tumours,

(Fig. 3). Across all 42 cases the proportion of the tumour genome showing CNV ranged from 0.1 to 48.4% per genome (mean 10.8%). The most frequent large-scale chromosome abnormality was 12p copy number gain, present in 30 of the 42 tumours (71%), of which 25 were 12p isochromosomes, a result consistent with previous experimental observations<sup>7–9</sup>. The remaining 12 cases without large-scale 12p gain all showed evidence of focal copy number amplification of 12p, however, detailed analysis of these sub-regions did not reveal any recurring hotspots. Other recurring large-scale copy number changes included gain of chromosome X (16 cases, 38%) as well as gains of chromosomes 7 ( $n=15$ ; 36%), 21 ( $n=12$ ; 29%) and 22 ( $n=11$ ; 26%), findings again consistent with previous studies<sup>7–13</sup>. In addition, we observed large-scale copy number deletion of chromosome Y (10 cases, 24%). We used previously generated chromosomal comparative genomic hybridization (CGH) data for 24 of the tumours<sup>12,35,36</sup> to validate our large-scale CNVs for the known mutational event at 12p; concordance between NGS/CGH was 92%.

In terms of focal events three tumours (patients 115, 53 and 43) exhibited a high degree of chromosomal instability, with a 19-fold increase in focal alterations compared with the others. We assessed these cases for evidence of chromothripsis, which we defined as  $>20$  CNVs on a chromosome single arm. While this technical definition was met for several loci, the majority of events were spread uniformly across the genome with no common hotspots across the three tumours. Excluding these three tumours we undertook an analysis of the focal alterations seen in the remaining 39 tumours to identify any recurrent patterns. Mapping the coordinates of all focal copy number events to genes, all possible gene alterations were assessed, quality filtered and ranked by frequency (Table 1 and methods). The highest ranking gene from this analysis was fibrous sheath interacting protein 2 (*FSIP2*) at 2q32.1, with seven recurring amplifications observed across six (15.3%) tumours. *FSIP2*



**Figure 3 | Circos Plot showing the count of SNV variants and copy number changes in the 42 tumours.** Outer ring marks the count of SNV variants across all 42 samples with proposed driver SNVs as blue dots and other SNVs as black lines; inner ring marks large-scale copy number gains (red) and losses (green).

amplifications were all 8–9 kb in length spanning a sub-region of the gene coding sequence, encompassing exons 16–17. Recent functional evidence has demonstrated that part-gene amplifications do affect gene expression levels, with an effect size comparable to that of full-gene amplification<sup>37</sup>. Our finding of recurrent *FSIP2* amplification is corroborated by recent high resolution SNP array data on an independent series of seminomas<sup>38</sup>, which documented *FSIP2* amplification in 22% of tumours. Across both studies *FSIP2* is the only gene consistently observed with focal amplification in >10% of cases. There is a strong biological basis for abnormalities of *FSIP2* being a feature of TGCTs *a priori*. The fibrous sheath is a cytoskeletal structure located in the principle piece region of the sperm flagellum. Transcription of *FSIP2* begins in late spermatocyte development

with mouse model data demonstrating it to be expressed exclusively in the testis<sup>39</sup>. Furthermore, *FSIP2* also binds to another fibrous sheath enzyme A kinase (PRKA) anchor protein 4 (AKAP4), which has been linked to male infertility<sup>40</sup>. Interestingly the tumour from patient 21, which harboured a *FSIP2* amplification, also carried a missense mutation in *AKAP4*.

Other focal events observed included a 0.4 Mb region at Xq28, with amplification in six cases. This region contains 18 genes, including testis expressed 28 (*TEX28*) and transketolase like gene 1 (*TKTL1*), both of which are overexpressed in the human testis<sup>41</sup>. *TKTL1* is hypothesized to play a role in tumour response to hypoxia with increased *TKTL1* expression correlating with poor patient outcome in many solid tumours<sup>42</sup>.

**Table 1 | Genes with five or more recurrent copy number gains/losses.**

Gene (s)	Region	Losses	Gains	Total CNVs
<i>FSIP2</i>	2q32.1	2	7	9
<i>AK2</i>	1p35.1	0	7	7
<i>ZNF644</i>	1p22.2	0	7	7
<i>ENPP3</i>	6q23.2	0	7	7
<i>MUC12</i>	7q22.1	0	7	7
<i>AHNAK2</i>	14q32.33	0	7	7
<i>TSPEAR</i>	21q22.3	0	7	7
<i>FLG</i>	1q21.3	1	6	7
<i>AK056431</i>	1q21.3	1	6	7
<i>HCFC1, TMEM187, MIR3202-1, IRAK1, MIR718, MECP2, OPN1LW, TEX28, OPN1MW, TKTL1, FLNA, EMD, AK307233, RPL10, SNORA70, DQ570720, DNASE1L1, TAZ</i>	Xq28	0	6	6
<i>CHRND</i>	2q37.1	0	6	6
<i>CTAGE9</i>	6q23.2	0	6	6
<i>MUC5B</i>	11p15.5	0	6	6

CNV, copy number variation.  
Focal CNVs included are defined as <3 Mb in length. See methods for further details on quality filters applied.

**Clinicopathological-molecular associations.** SNV/indel somatic mutation rates between seminoma and non-seminoma cases were almost identical; 0.50 mutations per Mb and 0.49 mutations per Mb respectively. *KIT* mutations were observed predominantly in seminoma cases, as previously reported. The proportion of the genome showing CNV was elevated (+47%) in non-seminoma tumours. A correlation between somatic mutational rate and patient age was seen ( $r=0.36$ ), with the mean rate for patients aged >40 years being 0.69 compared with 0.48 for cases <40 ( $P=0.05$ , two-sided Student's *t*-test). This is consistent with a model in which the majority of mutations are passenger mutations that accumulate with patient age following the early *in utero* oncogenic transformation of germ cells. Of particular clinical interest is the mutational profile of treatment-refractory TGCT, a rare subset of ~3% of patients in whom there is disease progression despite platinum-based chemotherapy. Within our cohort only one such patient, 40, had this profile of therapeutic response, so any conclusions are speculative. Accepting this caveat the mutational rate for this tumour was  $0.49 \text{ Mb}^{-1}$ , a rate comparable to the overall cohort, and of the 18 SNVs identified in this patient (see Supplementary Table 1), a mutation in gene *XRCC2* (c.6T>Gp.Cys2Trp) is of particular note. *XRCC2* encodes a member of the RecA/Rad51-related protein family, which participates in homologous recombination maintaining chromosome stability and repair of DNA damage. Importantly *XRCC2* mutant animal clones show increased resistance to cisplatin through enhanced DNA repair activity<sup>43</sup>, and *XRCC2* germline variants have been shown to significantly associate with cytotoxic resistance in breast cancer<sup>44</sup>. In addition to the treatment-refractory patient in our main cohort, we also performed exome sequencing of tumour DNA from one additional platinum refractory case (germline DNA was not available, patient 109), identifying a further mutation in *XRCC2* (c.2T>Gp.Met1Arg). This additional variant had alternative allele frequency of only 4%, making it difficult to validate by Sanger. Both *XRCC2* mutations are predicted to be pathogenic on the basis of *in silico* analysis using the CONDEL algorithm (CONsensus DELEteriousness (CONDEL) score of non-synonymous SNVs, <http://bg.upf.edu/fannsdh/help>)<sup>45,46</sup>.

## Discussion

Our exome analysis has confirmed mutation of *KIT* and recurrent copy number gain of 12p as archetypical features of TGCT. We have also characterized the mutational signature of TGCTs, demonstrating a homogeneous profile with a markedly low SNV mutation rate, consistent with the embryonic origins of the disease. This low rate of point mutations (that is, SNVs) is contrasted, however, by frequent large-scale copy number gains, of not only 12p but also chromosomes 7, 21, 22 and X. Since our study was empowered to identify recurrent mutations having frequency of >15% (84% power), we can conclude that it is unlikely that additional high frequency driver mutations will exist.

We did, however, identify novel mutations in the probable tumour suppressor gene *CDC27*, implicating *CDC27* mutation as a potential oncogenic factor in a subset of TGCTs. Functionally *CDC27* interacts with spindle checkpoint proteins encoded by *MAD2* (ref. 47) and *TEX14* (ref. 48) genes, the latter of which resides in a linkage disequilibrium block associated through recent genome-wide association study (GWAS) with germline TGCT predisposition<sup>49</sup>. Interestingly three of the other TGCT GWAS risk loci contain genes also related to mitotic spindle assembly—*MAD1L1*, *CENPE* and *PMF1* (refs 49,50). Collectively, such observations provide further evidence of commonality between germline and somatic TGCT pathways, a notable result given the previous precedent that *KITLG*, the ligand which binds *KIT*, is the only gene within the linkage disequilibrium block at the strongest existing TGCT GWAS risk locus (odds ratio ~2.5)<sup>51</sup>. Aside from *CDC27*, we also observed mutations in several other genes at a frequency of <10%; at this lower frequency our study was not sufficiently powered to comprehensively evaluate the genetic mutational profile (our power to detect mutations with frequencies of 10% and 5% was only 14%).

Previous CGH studies have characterized the aneuploidy nature of TGCTs, and our findings are consistent with these analyses. We hypothesized that NGS exome data, with average probe lengths of ~200 bp, would allow identification of novel small-scale CNVs below the level detectable by CGH. We performed this analysis and identified recurring focal copy number alterations in the spermatocyte development gene *FSIP2*, a finding corroborated by previous independent orthologous study. Meta-analysis of the two experiments shows this to be significant at  $P=6.8 \times 10^{-9}$ . *FSIP2* is shown to be unique to spermatogenic cells and is hypothesized to act as a linker protein, binding *AKAP4* to the fibrous sheath<sup>59</sup>. Dysplasia of the fibrous sheath and mutations in *AKAP4* have both been linked to male infertility<sup>40,52</sup>, an established risk factor for TGCT<sup>53</sup>. The additional observation of an *AKAP4* missense mutation further implicates this pathway, although the exact mechanisms facilitating tumorigenesis remain to be elucidated. Furthermore, we observed recurrent deletion of chromosome Y, a finding that also has interesting resonance with the germline as chromosome Y 'gr/gr' germline deletions are linked to both TGCT predisposition and male infertility<sup>54,55</sup>. In addition, we identified a recurring focal amplification of 0.4 Mb in length at Xq28, a region encompassing 18 genes, several of which may plausibly link to TGCT. Several observations implicate chromosome X in germ cell oncogenesis, with family studies suggesting a possible X-linked model of inheritance for TGCT genetic susceptibility<sup>56</sup>. In addition, patients with Klinefelter syndrome (47XXY constitutional karyotype) have a 67-fold elevated risk of developing mediastinal germ cell tumours<sup>57</sup>.

We found no significant difference observed in the mutational rate between seminoma and non-seminoma cases. This is consistent with findings from germline genetic studies of TGCT,

where no differential genotype risk has been observed between histological sub-groups<sup>49,51,58</sup>. This supports a hypothesis of commonality in the oncogenic pathways activated, with differentiation occurring later in the tumour formation. This hypothesis is further supported by the observation of TGCT cases with mixed pathology<sup>59</sup>, as well as bilateral and familial cases displaying tumours with inconsistent histological types<sup>60,61</sup>. Descriptive analysis of a single treatment-refractory patient in our cohort revealed a *XRCC2* mutation, a DNA repair gene which has been demonstrated to promote cisplatin resistance in animal studies<sup>43</sup>. Further analysis of one additional treatment-refractory tumour sample revealed some evidence for a second *XRCC2* mutation. Cell line studies suggest that the exceptional sensitivity of TGCTs to cisplatin is due to their inability to repair treatment-induced DNA damage, due to the low expression of DNA repair genes such as *ERCC1* (ref. 62). In addition, cisplatin-resistant embryonal carcinoma cell lines show sensitivity to poly(ADP-ribose) polymerase (PARP) inhibition, through blocking their acquired ability to repair DNA<sup>63</sup>. The observation of *XRCC2* mutations in our patient tumour data expands on these previous animal and cell line studies, further supporting an important role for this pathway.

To our knowledge this study represents the largest comprehensive sequencing study of TGCT conducted to date. While we have implemented strategies to accurately identify the mutational landscape of this tumour, we were only well powered to identify genes with high mutational frequency. Hence further insights into the biology of TGCT should be forthcoming through additional sequencing initiatives and meta-analyses of such data. This is likely to be especially important given the importance of probable histological subtype-specific changes, the subclonal architecture of TGCT and differences that are likely to be seen in platinum-resistant tumours.

## Methods

**Sample description.** Samples were collected from TGCT patients at the Royal Marsden Hospital NHS Trust, UK. Informed consent was obtained from all participants and the study was approved by the Institute of Cancer Research/Royal Marsden Hospital Committee for Clinical Research (study number CCR2014). The samples have been previously reported in other studies<sup>10,12,36,61,64</sup>. Surgical specimens were snap frozen within 30 min of surgery and matched blood samples were collected at the time of surgery. Tumour samples were trimmed to remove surrounding normal tissue, and tumour cells were confirmed by histological assessment. Tumour and matched lymphocyte DNA were extracted by standard techniques<sup>65,66</sup>. Tumour samples from patients 26 and 9 were obtained post chemotherapy. Clinical characteristics of our sample cohort were representative of the broader patient population, in terms of histological sub-types, patient age, familial TGCT and response to treatment. Our series was, however, enriched for cases with bilateral disease (9/42 cases in our series compared with a frequency of ~5% in the broader patient population).

**Whole-exome sequencing.** Samples were quantified using Qubit technology (Invitrogen, Carlsbad, CA, USA) and sequencing libraries constructed from 50 ng of respective normal/tumour DNA. Library preparation was performed using 37 Mb Nextera Rapid Capture Exome kits (Illumina, San Diego, CA, USA), with enzymatic tagmentation, indexing PCR, clean-up, pooling, target enrichment and post-capture PCR amplification/quality control performed in-house, following standardized protocols as per manufacturer guidelines. Samples underwent paired-end sequencing using the Illumina HiSeq2500 platform with a 100-bp read length. Mean coverage of 73.6 × and 69.0 × were achieved across targeted bases for tumour and normal samples, respectively. FASTQ files were generated using Illumina CASAVA software (v.1.8.1, Illumina) and aligned to the human reference genome (b37/hg19) using BWA (v. 0.5.10, <http://bio-bwa.sourceforge.net/>) Stampy (v.1.0.23) packages. PCR duplicates were removed and coverage metrics were calculated using Picard-tools (v.1.48, <http://picard.sourceforge.net/>). Coverage metrics demonstrated a mean of 95% of target bases achieved >10 × coverage and 86% >20 ×. The Genome Analysis Toolkit (GATK, v. 3.1-1, <http://www.broadinstitute.org/gatk/>) was used for local indel realignment/base quality score recalibration and SNVs were called using MuTect (v. 1.1.4). Data was quality filtered using in-house FoxoG software to remove potential artefactual variants introduced through DNA oxidation<sup>21</sup>. FoxoG ensured variants were supported by a minimum of one alternative read in each strand direction, a mean Phred base

quality score of >26, mean mapping quality ≥50 and an alignability site score of 1.0. Small-scale insertion/deletions (indels) were called using GATK.

We used MutSigCV (v.1.4) to identify genes that somatically mutated more often than would be expected by chance<sup>18</sup>, after first excluding common germline SNPs with minor allele frequency >25% as recorded in either dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), 1000 genomes (<http://www.1000genomes.org/>) or in our in-house data from exome sequencing of the UK 1958 birth cohort (Houlston *et al.*, personal communication). In total, 33 common germline SNP variants were removed across all samples. MutSigCV was run using the standard genomic covariates of (i) global gene expression data, (ii) DNA replication time and (iii) HiC statistic of open versus closed chromatin states. We used Oncodrive-fm<sup>32</sup> as implemented within the IntOGen-mutations platform<sup>67</sup>, using data mutation data from multiple tumour studies (<http://bg.upf.edu/group/projects/oncodrive-fm.php>; <http://www.intogen.org/analysis/mutations/>)

**Confirmation sequencing.** Confirmation sequencing was performed with bidirectional Sanger sequencing of *KIT* (exons 11 and 17) and *KRAS* (exon 2) across all 84 tumour/normal samples. Primer sequences are shown in Supplementary Table 3. Mutational analysis was conducted using Mutation Surveyor (v.3.97, SoftGenetics, State College, PA, USA).

**CNV analysis.** CNV analysis was conducted using the CRAN package ExomeCNV<sup>34</sup>, a statistical algorithm designed to detect CNV, and loss of heterozygosity (LOH) events using depth-of-coverage and B-allele frequencies ([https://secure.genome.ucla.edu/index.php/ExomeCNV\\_User\\_Guide](https://secure.genome.ucla.edu/index.php/ExomeCNV_User_Guide)). ExomeCNV is calibrated to achieve high levels of sensitivity and specificity, with a power to detect 95% for CNVs down to 500 bp in length<sup>34</sup>. When recently tested using a matched tumour/normal exome data set with ~40 × coverage, ExomeCNV achieved 97% specificity and 86% sensitivity compared with results from Illumina Omni-1 SNP array<sup>34</sup>. To calculate CNVs, we first generated coverage files using GATK, and then used ExomeCNV to calculate log coverage ratios between matched tumour/normal samples and make CNV calls per exon. Exonic CNV calls were combined into segments using circular binary segmentation. LOH calls were made by first identifying all heterozygous germline positions per case, using Platypus (v.0.5.2) for germline variant calling. GATK was then used to create BAF files per case and ExomeCNV used to call LOH at heterozygous positions individually and at combined LOH segments.

CNV results were classified as large-scale (>3 Mb in length) or focal (<3 Mb) and filtered by coverage ratio selecting copy number gain >1.3 or loss <0.7, retaining calls with a specificity confidence score of 1.0. Focal events were analyzed by gene, mapping the coordinates of all events to gene coding start and end points to assess all possible gene alterations. Small-scale regions showing susceptibility to variable levels of coverage, that is, exact same probes frequently altered and with both copy number gain and loss, were removed to avoid false-positive associations.

**Pathway analysis.** Pathway analysis was performed using Oncodrive-fm<sup>32</sup> as implemented within the IntOGen-mutations platform<sup>67</sup>, using the 1,168 SNVs and 111 indel mutations called across the 42 tumours.

**Statistical analyses.** Statistical significance of mutations were determined by testing whether the observed mutation counts in a gene significantly exceeded the expected counts based on a gene-specific background mutation rate, as implemented in MutSigCV (v.1.4). Plotted in the far section of Fig. 2 are the resulting  $-\log_{10}(P \text{ values})$ , with the dotted red line denoting a significance threshold of  $P=0.05$  and the solid red line a genome-wide significance threshold of  $P=5 \times 10^{-6}$ . Due to the overall low frequency of mutations observed in our data set, and the way such tumour types are treated by MutSigCV, no genes were significant at the genome-wide level, not even previously known TGCT driver gene *KIT*. Power analysis was conducted using a binomial power model, based on recent methods published by the Cancer Genome Analysis group at the Broad Institute<sup>68</sup>, incorporating the average background somatic mutation rate specifically observed for TGCT, sample size and assuming a genome-wide significance level of  $P \leq 5 \times 10^{-6}$ . Significance of focal copy number events by gene was calculated under a binomial distribution. Meta-analysis was conducted using the Fisher method of combining  $P$  values from independent tests. Statistical analysis were carried out using R3.0.2 (<http://www.r-project.org/>) and Stata12 (StataCorp, Lakeway Drive College Station, TX, USA) software. Continuous variables were analyzed using Student's  $t$ -tests. We considered a  $P$  value of 0.05 (two sided) as being statistically significant.

## 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**, e31–e36 (2014).

3. de Haas, E. C. *et al.* Early development of the metabolic syndrome after chemotherapy for testicular cancer. *Ann. Oncol.* **24**, 749–755 (2013).
4. 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).
5. 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).
6. Nitzsche, B. *et al.* Anti-tumour activity of two novel compounds in cisplatin-resistant testicular germ cell cancer. *Br. J. Cancer* **107**, 1853–1863 (2012).
7. Sandberg, A. A., Meloni, A. M. & Suijkerbuijk, R. F. Reviews of chromosome studies in urological tumors.3. Cytogenetics and genes in testicular tumors. *J. Urol.* **155**, 1531–1556 (1996).
8. Atkin, N. B. & Baker, M. C. Specific chromosome change, I(12p), in testicular-tumors. *Lancet* **2**, 1349–1349 (1982).
9. Atkin, N. B. & Baker, M. C. I(12p)—specific chromosomal marker in seminoma and malignant teratoma of the testis. *Cancer. Genet. Cytogenet.* **10**, 199–204 (1983).
10. Henegariu, O., Vance, G. H., Heiber, D., Pera, M. & Heerema, N. A. Triple-color FISH analysis of 12p amplification in testicular germ-cell tumors using 12p band-specific painting probes. *J. Mol. Med.* **76**, 648–655 (1998).
11. Roelofs, H. *et al.* Restricted 12p amplification and RAS mutation in human germ cell tumors of the adult testis. *Am. J. Pathol.* **157**, 1155–1166 (2000).
12. Summersgill, B. *et al.* Molecular cytogenetic analysis of adult testicular germ cell tumours and identification of regions of consensus copy number change. *Br. J. Cancer* **77**, 305–313 (1998).
13. Zafarana, G. *et al.* 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. *Oncogene* **22**, 7695–7701 (2003).
14. Rodriguez, S. *et al.* Expression profile of genes from 12p in testicular germ cell tumors of adolescents and adults associated with i(12p) and amplification at 12p11.2-p12.1. *Oncogene* **22**, 1880–1891 (2003).
15. McIntyre, A. *et al.* Amplification and overexpression of the KIT gene is associated with progression in the seminoma subtype of testicular germ cell tumors of adolescents and adults. *Cancer Res.* **65**, 8085–8089 (2005).
16. Kemmer, K. *et al.* KIT mutations are common in testicular seminomas. *Am. J. Pathol.* **164**, 305–313 (2004).
17. Bignell, G. *et al.* Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosomes Cancer* **45**, 42–46 (2006).
18. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
19. Kristensen, D. M. *et al.* Origin of pluripotent germ cell tumours: the role of microenvironment during embryonic development. *Mol. Cell Endocrinol.* **288**, 111–118 (2008).
20. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
21. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
22. Nakai, Y. *et al.* KIT (c-kit oncogene product) pathway is constitutively activated in human testicular germ cell tumors. *Biochem. Biophys. Res. Commun.* **337**, 289–296 (2005).
23. Ye, D. W. *et al.* P53 gene-mutations in chinese human testicular seminoma. *J. Urol.* **150**, 884–886 (1993).
24. Heidenreich, A. *et al.* Immunohistochemical and mutational analysis of the p53 tumour suppressor gene aml the bcl-2 oncogene in primary testicular germ cell, tumours. *APMIS* **106**, 90–99 (1998).
25. Laumann, R., Jucker, M. & Tesch, H. Point mutations in the conserved regions of the P53 tumor suppressor gene do not account for the transforming process in the jurkat acute lymphoblastic-leukemia T-cells. *Leukemia* **6**, 227–228 (1992).
26. Zhang, J. F., Wan, L. X., Dai, X. P., Sun, Y. & Wei, W. Y. Functional characterization of Anaphase Promoting Complex/Cyclosome (APC/C) E3 ubiquitin ligases in tumorigenesis. *Biochim. Biophys. Acta* **1845**, 277–293 (2014).
27. Kato, T. *et al.* Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. *J. Surg. Oncol.* **106**, 423–430 (2012).
28. Chang, D. Z. *et al.* Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. *J. Hematol. Oncol.* **5**, 15 (2012).
29. Marucci, G. *et al.* Gene expression profiling in glioblastoma and immunohistochemical evaluation of IGFBP-2 and CDC20. *Virchows Arch.* **453**, 599–609 (2008).
30. Pawar, S. A. *et al.* C/EBP delta targets cyclin D1 for proteasome-mediated degradation via induction of CDC27/APC3 expression. *Proc. Natl Acad. Sci. USA* **107**, 9210–9215 (2010).
31. Yu, C. *et al.* Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing. *Cell Res.* **24**, 701–712 (2014).
32. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169 (2012).
33. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
34. Sathirapongsasuti, J. F. *et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* **27**, 2648–2654 (2011).
35. Gilbert, D. C. *et al.* Minimum regions of genomic imbalance in stage I testicular embryonal carcinoma and association of 22q loss with relapse. *Genes Chromosomes Cancer* **50**, 186–195 (2011).
36. Summersgill, B., Osin, P. S., Lu, Y. J., Huddart, R. & Shipley, J. Chromosomal imbalances associated with carcinoma in situ and associated testicular germ cell tumours of adolescents and adults. *Br. J. Cancer* **85**, 213–219 (2001).
37. Chao, H. H., He, X., Parker, J. S., Zhao, W. & Perou, C. M. Micro-scale genomic DNA copy number aberrations as another means of mutagenesis in breast cancer. *PLoS ONE* **7**, e51719 (2012).
38. LeBron, C. *et al.* Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays. *Genomics* **97**, 341–349 (2011).
39. Brown, P. R., Miki, K., Harper, D. B. & Eddy, E. M. A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. *Biol. Reprod.* **68**, 2241–2248 (2003).
40. Miki, K. *et al.* Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. *Dev. Biol.* **248**, 331–342 (2002).
41. Coy, J. F., Dressler, D., Wilde, J. & Schubert, P. Mutations in the transketolase-like gene TKTL1: clinical implications for neurodegenerative diseases, diabetes and cancer. *Clin. Lab.* **51**, 257–273 (2005).
42. Schwaab, J. *et al.* Expression of Transketolase like gene 1 (TKTL1) predicts disease-free survival in patients with locally advanced rectal cancer receiving neoadjuvant chemoradiotherapy. *BMC Cancer* **11**, 363 (2011).
43. Danoy, P., Sonoda, E., Lathrop, M., Takeda, S. & Matsuda, F. A naturally occurring genetic variant of human XRCC2 (R188H) confers increased resistance to cisplatin-induced DNA damage. *Biochem. Biophys. Res. Commun.* **352**, 763–768 (2007).
44. Lin, W. Y. *et al.* A role for XRCC2 gene polymorphisms in breast cancer risk and survival. *J. Med. Genet.* **48**, 477–484 (2011).
45. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001).
46. Sunyaev, S. *et al.* Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591–597 (2001).
47. Poddar, A., Stukenberg, P. T. & Burke, D. J. Two complexes of spindle checkpoint proteins containing Cdc20 and Mad2 assemble during mitosis independently of the kinetochore in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**, 867–878 (2005).
48. Monda, G., Ohashi, A., Yang, L., Rowley, M. & Couch, F. J. Tex14, a Plk1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the spindle assembly checkpoint. *Mol. Cell* **45**, 680–695 (2012).
49. 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).
50. Chung, C. C. *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* **45**, 680–685 (2013).
51. Rapley, E. A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* **41**, 807–810 (2009).
52. Chemes, H. E., Brugo, S., Zanchetti, F., Carrere, C. & Lavieri, J. C. Dysplasia of the fibrous sheath: an ultrastructural defect of human spermatozoa associated with sperm immotility and primary sterility. *Fertil. Steril.* **48**, 664–669 (1987).
53. Hotaling, J. M. & Walsh, T. J. Male infertility: a risk factor for testicular cancer. *Nat. Rev. Urol.* **6**, 550–556 (2009).
54. Nathanson, K. L. *et al.* The Y deletion gr/gr and susceptibility to testicular germ cell tumor. *Am. J. Hum. Genet.* **77**, 1034–1043 (2005).
55. Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**, 279–286 (2001).
56. 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).
57. Hasle, H., Mellempgaard, A., Nielsen, J. & Hansen, J. Cancer incidence in men with Klinefelter syndrome. *Br. J. Cancer* **71**, 416–420 (1995).
58. Turnbull, C. *et al.* Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat. Genet.* **42**, 604–607 (2010).
59. Gori, S. *et al.* Germ cell tumours of the testis. *Crit. Rev. Oncol. Hematol.* **53**, 141–164 (2005).
60. Mai, P. L. *et al.* The international testicular cancer linkage consortium: a clinicopathologic descriptive analysis of 461 familial malignant testicular germ cell tumor kindred. *Urol. Oncol.* **28**, 492–499 (2010).
61. Forman, D. *et al.* Familial testicular cancer: a report of the UK family register, estimation of risk and an HLA class I sib-pair analysis. *Br. J. Cancer* **65**, 255–262 (1992).



62. Usanova, S. *et al.* Cisplatin sensitivity of testis tumour cells is due to deficiency in interstrand-crosslink repair and low ERCC1-XPF expression. *Mol. Cancer* **9**, 248 (2010).
63. Cavallo, F. *et al.* Reduced proficiency in homologous recombination underlies the high sensitivity of embryonal carcinoma testicular germ cell tumors to cisplatin and poly (ADP-ribose) polymerase inhibition. *PLoS ONE* **7**, e51563 (2012).
64. Huddart, R. A., Wooster, R., Horwich, A. & Cooper, C. S. Microsatellite instability in human testicular germ cell tumours. *Br. J. Cancer* **72**, 642–645 (1995).
65. Orkin, S. Molecular-cloning—a laboratory manual, 2nd edn (Sambrook, J., Fritsch, E.F., Maniatis, T.) *Nature* **343**, 604–605 (1990).
66. Lahiri, D. K. & Nurnberger, J. I. A Rapid Nonenzymatic Method for the Preparation of Hmw DNA from Blood for Rflp Studies. *Nucleic Acids Res.* **19**, 5444 (1991).
67. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
68. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

### Acknowledgements

We thank the patients and their clinicians for participation in this study. We acknowledge the National Health Service funding to the National Institute for Health Research Biomedical Research Centre. We acknowledge the facilities and expertise of the Cancer Genetics Core Laboratory Facility and the Cancer Genetics Sequencing Facility made available at the Institute of Cancer Research by Professor Nazneen Rahman. This study was supported by the Movember foundation and the Institute of Cancer Research. K. Litchfield is supported by a PhD fellowship 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).

### Author contributions

C.T. designed the study. J.S. and R.A.H. provided the samples. D.D., B.S. and R.A.-S. coordinated sample administration and tracking. K. Litchfield and S.S. coordinated sample management. J.S. and B.S. provided CGH validation data. C.T., R.S.H., P.B. A.R. and K. Litchfield designed laboratory experiments. K. Litchfield and A.R. conducted laboratory experiments. K. Litchfield, R.S.H., K.L., N.C.T., S.Y. and R.S. designed bioinformatic analyses. K. Litchfield, S.Y. and R.S. carried out bioinformatics analyses. K. Litchfield performed statistical analyses. K. Litchfield drafted the manuscript with assistance from C.T., R.S.H. and J.S. All authors reviewed and contributed to the manuscript.

### Additional information

**Accession codes:** Whole-exome sequencing have been deposited in the European Genome–phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI), under the accession code EGAS00001001084.

**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.* Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nat. Commun.* **6**:5973 doi: 10.1038/ncomms6973 (2015).



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/>