Editor-in-Chief,

European Urology

European Association of Urology

Meester E.N. van Kleffensstraat 5

6842 CV Arnhem, Netherlands

10th January 2018

Dear Professor Catto,

I am writing with regard to the following manuscript that is currently under review with European Urology: **"Large-scale sequencing of testicular germ cell tumour (TGCT) cases excludes major TGCT predisposition gene"**.

We would like to thank the European Urology Statistical Editor for their comments (received 1st January 2018). We have now addressed these comments and enclose a point-by-point response and a set of revised manuscript files.

We look forward to your response.

Yours faithfully,

Clare Turnbull.

Dr Clare Turnbull MD PhD MA MSc MRCP MFPH

- Professor in Genomic Medicine, **Queen Mary University of London,** Charterhouse Square, London, EC1M 6BQ
- Senior Researcher, **Institute of Cancer Research,** Institute of Cancer Research, 15 Cotswold Rd, Sutton, Surrey, SM2 5NG
- Clinical Lead for Cancer Genomics, **100,000 Genomes Project,** Genomics England, Queen Mary University of London, EC1M 6BQ
- Honorary Consultant in Clinical Cancer Genetics, **Guys and St Thomas NHS Trust,** SE1 9RT
- Honorary Consultant in Public Health Medicine, **Public Health England,** SE1 8UG

T: +44 (0)20 8722 4145
F: +44 (0)20 8722 4359
E clare.turnbull@icr.ac.uk

Division of Genetics & Epidemiology
15 Cotswold Road
Sutton SM2 5NG
UNITED KINGDOM

Royal Cancer Hospital
123 Old Brompton Road
London SW7 3RP
UNITED KINGDOM

TITLE PAGE

Large-scale sequencing of testicular germ cell tumour (TGCT) cases excludes major TGCT predisposition gene

Kevin Litchfield[a,†,], Chey Loveday[a,], Max Levy[a], Darshna Dudakia[a], Elizabeth Rapley[a], Jeremie Nsengimana[b], D Tim Bishop[b], Alison Reid[c], Robert Huddart[c], Peter Broderick[a], Richard S Houlston[a,d], Clare Turnbull[a,e,f,g,*]

[a]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK; [b]Section of Epidemiology & Biostatistics, Leeds Institute of Cancer and Pathology, Leeds, LS9 7TF, UK; [c]Academic Uro-oncology Unit, The Royal Marsden NHS Foundation Trust, Sutton, Surrey; [d]Division of Molecular Pathology, The Institute of Cancer Research, London, SM2 5NG, UK; [e]William Harvey Research Institute, Queen Mary University, London, EC1M 6BQ, UK;[f] Department of Clinical Genetics, Guys and St Thomas NHS Foundation Trust, London, SE1 9RT, UK; [g]Public Health England, National Cancer Registration and Analysis Service, London, SE1 8UG, UK.

[†]Current address: Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK.

These authors shared co-first authorship.

*Corresponding author. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK; Tel: +44 (0) 208 722 4485; E-mail: clare.turnbull@icr.ac.uk (C. Turnbull).

Keywords: Testicular germ cell tumour, whole-exome sequencing, cancer susceptibility, heritability, genetics.

Word count of text: 1122 **Word count of abstract:** 201

ABSTRACT

Testicular Germ Cell Tumour (TGCT), the most common cancer in young men, has a significant heritable basis, which has long raised the question of existence of underlying major high-penetrance susceptibility gene(s). To determine the contribution of rare gene mutations to the inherited risk of TGCT, we analysed germline whole-exome data from 919 TGCT cases and 1,609 cancer-free controls. We compared frequencies between TGCT cases and controls of rare (<1%) and low frequency (1-5%) coding variants (i) individually and (ii) collapsed at gene level by burden testing (T1 disruptive, T2 all deleterious, T3 all non-synonymous), using Fisher's exact test with Bonferroni correction of significance thresholds. No individual variant or individual gene demonstrated significant association with TGCT after correction for multiple testing. In the largest whole-exome sequencing study of testicular cancer reported to date, our findings do not support existence of a major high-penetrance TGCT susceptibility gene (of odds ratio > 10 and allele frequency (combined) > 0.01%). Due to power, this study cannot exclude existence of susceptibility genes responsible for occasional TGCT families, or rare mutations which confer very modest relative risks. In concert with findings of GWAS, our data support inherited susceptibility largely being polygenic with substantial contribution from common variation.

PATIENT SUMMARY

In the largest study of its kind, sequencing ~20,000 genes in 919 men with TGCT and 1609 TGCT-free individuals, we find no evidence for a single major gene underlying predisposition to TGCT (in manner of *BRCA1* for breast cancer). Instead, familial risk of TGCT is likely to be due to varying dosages of hundreds of minor genetic factors.

MAIN TEXT

Testicular germ cell tumour (TGCT), the most common cancer affecting young men, has a strong heritable basis as evidenced by the 4-8 fold increased risk of TGCT seen in brothers of TGCT patients [[1],[2]]. High heritability and observation of multiplex TGCT families have long fuelled anticipation that there may exist a 'major' TGCT-susceptibility gene suitable for clinical testing, analogous to BRCA1/BRCA2 in breast cancer. Early genetic linkage studies however proved unfruitful but were very much limited in power by the modest size and low frequency of multiplex TGCT families [3].

We previously reported whole exome case-control segregation analysis of 150 TGCT families focusing just on gene-level analysis of rare (<1%) disruptive (truncating) mutations [4]. Although, after correcting for exome-wide analysis, no gene was significant based on segregation analysis alone, using a range of functional analyses we demonstrated association with familial TGCT of the strongest candidate, *DNAAF1,* and related cilia-microtubule genes (CMGs). However, mutations in each of those genes were infrequent and none would constitute a 'major' TGCT susceptibility gene [4]. Here we present more comprehensive germline WES analysis of the contribution of rare variants to TGCT susceptibility, examining *multiple* different types of rare coding alleles (disruptive, damaging and otherwise) and presenting primary analysis of the *full cohort* of 919 TGCT cases

(comprising 613 unselected TGCT cases in addition to the previously reported 306 familial TGCT cases), comparing these data to 1,609 healthy controls (see Supplementary Methods).

We first examined individual nonsynonymous coding variants, both rare (MAF < 1%) and low-frequency (MAF 1-5%), for association with TGCT. A total 966,695 rare and 4,994 low frequency variants were detected, and a Bonferroni-corrected threshold of $p < 5 \times 10^{-8}$ (i.e. $p < 0.05$/~1 M variants) was imposed. No variant demonstrated an association with TGCT above this significance threshold (**Table 1**).

We next analysed rare nonsynonymous variants collapsed at gene level and organised into three groups: (T1 disruptive, T2 all deleterious, and T3 all non-synonymous). A Bonferroni-corrected threshold of $p < 8 \times 10^{-7}$ (i.e. $p < 0.05$/(20,000 genes, 3 tiers)) was imposed. No gene was significant within any variant group at the exome-wide level (**Table 2**). We assessed the distribution of test statistics compared to a null model using quantile-quantile plots **(Supplementary Fig. 1**) and found inflation statistics to be in the range of $\lambda = 0.75$ to 1.0, suggesting the data fitted a null distribution overall.

Whilst TGCT has yet to be implicated in any established cancer susceptibility syndrome, it is feasible that association of TGCT with a cancer susceptibility gene (CSG) may have gone undetected due to its rarity. Gene burden testing of 114 established high/moderate penetrance CSGs provided no evidence for association between any CSG and TGCT, after correcting for multiple testing (i.e. $p < 0.0001$ ($p < 0.05$/(114*3))) (**Supplementary Table 1**) [5]. Across all 114 CSGs, 4.9% of TGCT cases carried a T1 mutation, compared to 5.5% of controls ($p > 0.5$), providing no evidence of global enrichment. Similarly, there was no difference in the proportion of cases and controls carrying a T2 or T3 mutation.

Causal variants responsible for GWAS signals largely map to non-coding regulatory regions, presumed to be influencing gene expression. However, non-synonymous variants in these same genes have also been shown to influence cancer risk [6]. Gene burden testing of 64 genes positioned within the 49 established TGCT loci provided no evidence for association with TGCT of any GWAS-associated gene, after correction for multiple testing (i.e. $p < 0.0003$ ($p < 0.05/(64*3)$)), and there was no difference in the proportion of cases and controls with a T1/T2/T3 mutation (**Supplementary Table 2**) [[7],[8],[9]].

We then proceeded to undertake additional analyses of the 150 TGCT families. First we focused on the 12 'large' pedigrees (eleven three-case families and one four-case family), on the presumption that family clustering of a rare cancer is unlikely to occur by chance. There were no genes for which rare T1 variants were found to segregate with TGCT in more than one family. For the singular four case pedigree (PED-269), the only rare variant segregating fully with TGCT was the missense variant *BOLL* c.62C>A;p.Ser21Tyr (MAF = 0.001). Given the role of *BOLL* in germ cell development [10], we then genotyped c.62C>A;p.Ser21Tyr in 3,999 unselected TGCT cases and 4,011 controls, but found no evidence of association ($p > 0.5$, **Supplementary Table 3**). Assessing low-frequency variants (MAF = 1-5%, T1/T2/T3) across the full set of families for evidence of segregation, we found strongest evidence for a haplotype in cilia gene *DNAH7* (c.1895C>G;p.Ser632Cys and c.6340A>G;p.Thr2114Ala (R^2 = 0.9, D' = 1.0, population MAF = 0.04)). This haplotype showed full segregation in two 'large' pedigrees PED-269, PED-251 and in 8/138 two-case pedigrees (**Supplementary Table 4**) but analysis in the full case-control series did not support association with disease (Case MAF = 0.05, control MAF = 0.04, $p > 0.1$).

Finally, we undertook simulations to evaluate the power our analyses had to detect TGCT predisposition genes. We modelled a hypothetical TGCT predisposition gene for which the frequency of the summed pathogenic mutations ($MAF_{combined}$) ranged from 0.01% to 1% and for which the effect size (odds ratio, OR) ranged from 2 to 10 (**Supplementary Fig. 2**). Our study was shown to be well-powered (> 0.9) to detect a high risk 'major TGCT gene" (OR > 10, $MAF_{combined}$ > 0.01%) had one existed. For intermediate effect size (OR = 5) power remains good down to $MAF_{combined}$ = 0.05% but plummets to zero at 0.01%. For lower risk effect size (OR = 2) the power is far more limited, dropping below 0.3 for $MAF_{combined}$ < 0.5%. These power analyses are therefore consistent with existence of multiple additional 'undiscovered' rare variants/susceptibility genes, of modest effect size and/or very low frequency. However, in the current analyses, the signal of association for such variants/genes would be indistinguishable above noise from innocuous background variation. To improve power to > 0.9 for detection of such variants down to $MAF_{combined}$ < 0.01% for medium (OR > 5) or low (OR > 2) effect sizes, sequencing of > 10,000 TGCT cases would be required (along with a comparable number of controls). Thus, as for other common cancers, to significantly advance rare variant discovery for TGCT, studies at least ten-fold larger in scale are required, in conjunction with advances in in-silico prediction tools to better predict disease pathogenicity of non-synonymous variants. Functional analyses can also assist in identifying the true biologically impactful genomic variants, as demonstrated in our CMG studies [[4],[8],[9]].

In summary, our findings indicate that there is unlikely to exist a 'major' high penetrance TGCT susceptibility gene suitable for clinical pre-symptomatic testing. In conjunction with

the 49 TGCT-associated common alleles identified through recent GWAS [7], these analyses serve to underscore a highly polygenic model of genomic TGCT susceptibility.

REFERENCES

[1] Litchfield, K., Thomsen, H., Mitchell, J.S., et al. Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Sci Rep* 2015;5:13889.

[2] Hemminki, K., Li, X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *Br J Cancer* 2004;90:1765-1770.

[3] Crockford, G.P., Linger, R., Hockley, S., et al. Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. *Hum Mol Genet* 2006;15:443-451.

[4] Litchfield, K., Levy, M., Dudakia, D., et al. Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. *Nat Commun* 2016;7:13840.

[5] Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* 2014;505:302-308.

[6] Sud, A., Kinnersley, B., Houlston, R.S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer.* 2017;17:692-704.

[7] Litchfield, K., Levy, M., Huddart, R.A., et al. The genomic landscape of testicular germ cell tumours: from susceptibility to treatment. *Nat Rev Urol* 2016;13:409-419.

[8] Litchfield, K., Levy, M., Orlando, G., et al. Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. *Nat Genet* 2017;49:1133-40.

[9] Wang, Z., McGlynn, K.A., Rajpert-De Meyts, E., et al. Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat Genet* 2017;49:1141-7.

[10] Fu, X.F., Cheng, S.F., Wang, L.Q., et al. DAZ Family Proteins, Key Players for Germ Cell Development. *Int J Biol Sci* 2015;11:1226-1235.

LIST OF ABBREVIATIONS

CI – confidence interval

CMG - cilia-microtubule gene

CSG – cancer susceptibility gene

FET – Fisher's Exact Test

GATK – Genome Analysis Toolkit

GWAS – genome wide association study

HWE – Hardy-Weinberg Equilibrium

ICR – Institute of Cancer Research

Indel – insertion-deletion

MAF – minor allele frequency

OR – odds ratio

SNV – single nucleotide variants

TGCT – testicular germ cell tumour

VEP – variant effect predictor

WES – whole exome sequencing

TABLES

Table 1 – Top 25 most significant individual variants

				Case Alleles			Controls Alleles				
Gene	**cDNA**	**Protein**	**Group**	**Alt no.**	**Total**	**MAF**	**Alt no.**	**Total**	**MAF**	**OR (95% CI)**	***p* value***
PGP	c.859G>A	p.Gly287Arg	T3	98	1718	0.057	97	3088	0.031	1.87 (1.40-2.49)	3×10^{-5}
FAM160A2	c.2195C>T	p.Pro732Leu	T3	11	1692	0.007	55	2594	0.021	0.30 (0.16-0.58)	7×10^{-5}
MLXIP	c.1240G>A	p.Asp414Asn	T3	84	1786	0.047	71	2844	0.025	1.93 (1.40-2.66)	7×10^{-5}
OR1N2	c.709C>T	p.Arg237Cys	T3	34	1846	0.018	20	3162	0.006	2.95 (1.69-5.14)	1×10^{-4}
OR10C1	c.169C>T	p.Pro57Ser	T2	11	1846	0.006	1	3138	0.000	18.8 (2.43-145)	1×10^{-4}
PLK1	c.1388T>A	p.Leu463His	T3	37	1782	0.021	24	3058	0.008	2.68 (1.60-4.50)	2×10^{-4}
ADAMTS18	c.3565G>A	p.Val1189Ile	T3	2	1706	0.001	27	2732	0.010	0.12 (0.03-0.50)	2×10^{-4}
EPB41L5	c.82C>T	p.Arg28Cys	T2	18	1864	0.010	6	3126	0.002	5.07 (2.01-12.8)	2×10^{-4}
JMJD4	c.1024T>C	p.Phe342Leu	T3	103	1848	0.056	101	3036	0.033	1.72 (1.30-2.27)	2×10^{-4}
SKIV2L	c.2749G>A	p.Val917Met	T2	118	1778	0.066	127	3044	0.042	1.63 (1.26-2.11)	2×10^{-4}
ARHGEF17	c.1571C>T	p.Ala524Val	T3	12	1806	0.007	2	3000	0.001	10.0 (2.24-44.8)	3×10^{-4}
SH3TC1	c.2429C>T	p.Thr810Met	T3	2	1640	0.001	27	2702	0.010	0.12 (0.03-0.51)	3×10^{-4}
MPDZ	c.2194T>A	p.Ser732Thr	T2	10	1656	0.006	1	2840	0.000	17.3 (2.21-134)	3×10^{-4}
PALB2	c.2014G>C	p.Glu672Gln	T3	80	1662	0.048	85	3066	0.028	1.77 (1.30-2.42)	3×10^{-4}
EHBP1L1	c.2683A>T	p.Ser895Cys	T2	15	1836	0.008	4	3152	0.001	6.48 (2.15-19.6)	4×10^{-4}
ABCC4	c.1141G>A	p.Val381Ile	T3	11	1676	0.007	2	3060	0.001	10.1 (2.24-45.6)	4×10^{-4}
R3HCC1	c.500C>T	p.Thr167Ile	T2	36	1798	0.020	111	2906	0.038	0.51 (0.35-0.75)	4×10^{-4}
VPS16	c.1561G>A	p.Asp521Asn	T3	10	1802	0.006	1	2970	0.000	16.6 (2.12-129)	4×10^{-4}
P2RX7	c.827G>A	p.Arg276His	T3	56	1792	0.031	50	3160	0.016	2.01 (1.36-2.95)	4×10^{-4}
OR1N1	c.680G>A	p.Arg227Gln	T3	33	1800	0.018	22	3112	0.007	2.62 (1.52-4.51)	5×10^{-4}
DEFB132	c.277G>A	p.Val93Ile	T3	37	1832	0.020	27	3166	0.009	2.40 (1.45-3.95)	6×10^{-4}
ALPK1	c.2042G>A	p.Gly681Asp	T3	47	1870	0.025	38	3164	0.012	2.12 (1.38-3.27)	6×10^{-4}
IYD	c.794G>A	p.Cys265Tyr	T3	78	1824	0.043	77	3108	0.025	1.76 (1.28-2.42)	7×10^{-4}
ZYG11A	c.1027A>G	p.Met343Val	T3	15	1846	0.008	5	3154	0.002	5.16 (1.87-14.2)	7×10^{-4}
POLI	c.1595T>C	p.Phe532Ser	T2	78	1816	0.043	79	3164	0.025	1.75 (1.27-2.41)	7×10^{-4}

* Bonferroni-corrected threshold of $p < 5 \times 10^{-8}$ for significance. MAF, minor allele frequency. OR, odds ratio.

Table 2 – Top ranked genes by variant group

T1 - Disruptive					T2 - Deleterious					T3 – Non-synonymous				
Gene	**Ca**	**Co**	**OR (95% CI)**	***p* value***	**Gene**	**Ca**	**Co**	**OR (95% CI)**	***p* value***	**Gene**	**Ca**	**Co**	**OR (95% CI)**	***p* value***
THADA	6	0	NA	1×10^{-3}	*CDHR4*	11	3	6.40 (1.80-23.1)	5×10^{-3}	*GFAP*	10	1	17.5 (2.24-136)	9×10^{-5}
PLA2G3	16	10	2.80 (1.27-6.20)	2×10^{-3}	*CSF2RB*	9	2	7.90 (1.70-36.5)	1×10^{-3}	*XPO6*	10	1	17.5 (2.24-136)	9×10^{-5}
OR5AU1	7	2	6.13 (1.27-29.6)	7×10^{-3}	*FCHSD1*	1	25	0.10 (0.00-0.50)	1×10^{-3}	*ZKSCAN3*	16	6	4.67 (1.82-12.0)	1×10^{-4}
IQGAP3	0	14	NA	7×10^{-3}	*GFAP*	6	0	NA	1×10^{-3}	*RNF213*	80	100	1.40 (1.03-1.90)	3×10^{-4}
MYO1A	15	11	2.39 (1.09-5.22)	1×10^{-2}	*NTRK3*	6	0	NA	1×10^{-3}	*GTF3A*	3	36	0.15 (0.04-0.48)	4×10^{-4}
C4orf21	4	0	NA	1×10^{-2}	*FOXM1*	10	3	5.84 (1.60-21.3)	1×10^{-3}	*MAN2B2*	44	44	1.75 (1.14-2.68)	4×10^{-4}
EIF2A	4	0	NA	1×10^{-2}	*LLGL1*	10	3	5.84 (1.60-21.3)	1×10^{-3}	*KCNA7*	11	3	6.42 (1.79-23.1)	5×10^{-4}
FASTKD1	4	0	NA	1×10^{-2}	*COL9A1*	15	8	3.28 (1.39-7.77)	1×10^{-3}	*KRTAP13-2*	19	11	3.02 (1.43-6.38)	5×10^{-4}
MIS12	4	0	NA	1×10^{-2}	*KARS*	24	19	2.21 (1.20-4.06)	1×10^{-3}	*C10orf12*	31	27	2.01 (1.19-3.39)	9×10^{-4}
NDUFV3	4	0	NA	1×10^{-2}	*KDM5A*	13	6	3.79 (1.44-10.0)	2×10^{-3}	*ENO3*	14	6	4.09 (1.56-10.7)	1×10^{-3}

* Bonferroni-corrected threshold of $p < 8 \times 10^{-7}$ for significance. Ca, number of cases with a variant. Co, number of controls with a variant. OR, odds ratio. CI, confidence interval.

DECLARATIONS

Ethics approval and consent to participate

Collection of blood samples and clinical information from subjects was undertaken with informed written consent and relevant ethical review board approval at respective institutions.

Consent for publication

No identifiable persons' data was used in this study.

Availability of data and materials

Case WES data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI); accession number EGAS00001001789.

Competing interests

The authors declare that they have no competing interests.

Funding

We acknowledge the National Health Service funding to the National Institute for Health Research Biomedical Research Centre. This study was supported by the Movember foundation and the Institute of Cancer Research. K. Litchfield was 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).

Authors' contributions

C.T, and K.L. designed the study. 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, M.L and P.B designed and conducted laboratory experiments. K.L., C.L., C.T. and R.S.H. designed bioinformatics and statistical analyses. K. L. and C.L. conducted bioinformatics and statistical analyses. C.L, C.T., and K.L. drafted the manuscript with assistance from R.S.H., and M.L. and K. L generated images for publication. All authors reviewed and contributed to the manuscript.

Acknowledgements

We thank the patients and their clinicians within for participation in this study. We acknowledge the International Testicular Cancer Linkage Consortium and the UK Testicular Cancer Collaboration. 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.