Runs of Homozygosity and Testicular Cancer Risk

Chey Loveday[1], Amit Sud[1], Kevin Litchfield[2], Max Levy[1], Amy Holroyd[1], Peter Broderick[1], Zsofia Kote-Jarai[1], Alison M Dunning[3], Kenneth Muir[4,5], Julian Peto[6], Rosalind Eeles[1,7], Douglas F Easton[3,8], Darshna Dudakia[1], Nick Orr[9], Nora Pashayan[10], UK Testicular Cancer Collaboration, The PRACTICAL Consortium, Alison Reid[11], Robert A Huddart[12], Richard S Houlston[1] and Clare Turnbull[1,13,14,15*]

1. Division of Genetics & Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK.
2. Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK.
3. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
4. Division of Health Sciences, Warwick Medical School, Warwick University, Warwick, UK.
5. Institute of Population Health, University of Manchester, Manchester, UK.
6. Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.
7. Royal Marsden NHS Foundation Trust, London, UK.
8. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
9. The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK.
10. Department of Applied Health Research, University College London, London, UK.
11. Academic Uro-oncology Unit, The Royal Marsden NHS Foundation Trust, Sutton, Surrey, UK.
12. Academic Radiotherapy Unit, Institute of Cancer Research, Sutton, Surrey, UK.
13. William Harvey Research Institute, Queen Mary University, London, EC1M 6BQ, UK.
14. Guys and St Thomas Foundation NHS Trust, Great Maze Pond, London, SE1 9RT, UK.

15. Public Health England, National Cancer Registration and Analysis Service, London, SE1 8UG, UK.

*clare.turnbull@icr.ac.uk

Abstract (200 words)

Runs of homozygosity (RoH) are observed in the outbred population and have been associated with increased cancer risk, being a signature indicating underlying recessively acting alleles. To examine whether RoH are associated with TGCT risk, we performed a genome-wide RoH analysis using GWAS data from 3,206 TGCT cases and 7,422 controls uniformly genotyped using the OncoArray platform. Global measures of homozygosity were not significantly different between cases and controls, and the frequency of individual consensus RoH were not significantly different between cases and controls, following correction for multiple testing. RoH at three regions, 11p13-11p14.3, 5q14.1-5q22.3 and 13q14.11-13q.14.13, were nominally statistically significant at $p < 0.01$, indicating potential selective enrichment in TGCT cases. Interestingly, RoH200 at 11p13-11p14.3 encompasses *Wilms tumour 1* (*WT1*), a recognised cancer susceptibility gene with roles in sex determination and developmental transcriptional regulation, processes repeatedly implicated to date in TGCT etiology. Overall, our data does not support a major role in the risk of TGCT for recessively-acting alleles acting through homozygosity, as measured by RoH in an outbred population.

Introduction

Testicular germ cell tumour (TGCT) is the most common cancer in men aged 18-45, with over 52,000 new cases diagnosed annually worldwide[1]. TGCT has a strong heritable basis, as evidenced by the 4-8 fold increased risk of TGCT seen in first degree male relatives of TGCT patients[2-6]. Heritability analyses estimate that genetic factors contribute to approximately half of all disease risk[2].

Early linkage analyses of TGCT did not indicate existence of a major Mendelian TGCT susceptibility locus, albeit that these studies were limited in power on account of modest sample sizes and low prevalence of multiplex TGCT pedigrees[7]. Recent familial exome sequencing studies had improved power to identify rare susceptibility alleles of a frequency/effect size profile not tractable by the initial linkage analyses. Nevertheless, these exome sequencing analyses demonstrated that susceptibility genes/rare alleles for TGCT of moderately 'major' frequency *and* effect size (MAF>0.001 and OR>10) are also unlikely to exist. Very rare alleles in ciliary microtubule genes have been implicated in TGCT susceptibility only through extensive functional validation with segregation analysis in multiplex families; these are typical exemplars of the susceptibility alleles of very low frequency likely to contribute to TGCT susceptibility, with each allele/gene only accounting for a tiny fraction of disease heritability[8]. These findings are consistent with advanced analyses of TGCT heritability, which have indicated that the genetic component of TGCT heritability is largely constituted by common variants. Recent genome-wide association studies (GWAS) have made substantial progress in exposition of this partition of heritability with 49 independent TGCT risk loci identified, together accounting for ~37% of the excess genetic risk of disease[9-20]. These TGCT susceptibility loci have provided invaluable insight into the biology of TGCT susceptibility, implicating as underlying mechanisms, widespread transcriptional dysregulation linked to developmental arrest of primordial germ cells, aberrant KIT-MAPK signaling and defective microtubule function[19]. From these GWAS loci approximately half of the genetic component of TGCT heritability has been accounted for, with heritability analysis indicating that the outstanding 'missing heritability' of TGCT is likely polygenic, with substantial contribution from common variation[2, 19, 21]. GWAS analysis has likewise made substantial impact in delineating the genetic architecture of many common cancers but almost uniformly the reported susceptibility loci have been identified through analysis based on a log additive model of inheritance, with little evidence generated for alleles acting recessively[22]. This observation may be a reflection that GWAS is suboptimal in its ability to detect these alleles rather than

an observation truly reflective of the underlying biology. In principle, it is entirely plausible that there may be an association between recessively acting disease alleles and susceptibility to cancer. Such a hypothesis is supported by observations reporting an increased burden of cancer in the offspring of consanguineous unions and in populations with a high degree of inbreeding[23-27]. Furthermore, experimental inbreeding (e.g. backcrossing mice) has also been shown to increase tumour burden in mice[28]. In addition, uniparental disomy through dysregulated imprinting is a specific situation in which homozygosity can be directly associated with cancer[29]. Of note, for TGCT, there has been a long-standing hypothesis that recessive (or X-linked) susceptibility factors are highly likely to be important, based on repeated epidemiological observation of a higher familial relative risk to brothers of cases than that to fathers/sons. Whilst it could be hypothesized that such an observation might be an artefact common on account of biases in age-related case ascertainment, from multi-generational analysis of cancer incidence from >10 million records over 40 years from the Swedish cancer registry, this phenomenon was observed for very few tumour types [3, 6].

Homozygosity mapping is undertaken in order to identify potential recessive components of inheritance. It has been demonstrated that, on account of selective pressure, runs of homozygosity (RoH) occur at high frequency in outbred populations, the result of autozygosity (i.e. the co-location of two alleles at a given locus originating from a common ancestor by way of non-random mating) [30, 31]. These RoH can be enriched for rare deleterious variants in homozygous form; multiple disease-associated susceptibility loci have been reported, identified through genome-wide analyses for RoH of SNP array data (reviewed by Ceballos et al., 2018)[32], [33].

Here, we sought to identify associations between homozygosity and TGCT risk through the characterization and comparison of genome-wide homozygosity measures and specific loci identified

through consensus mapping of recurrent RoH in 3,206 TGCT cases vs 7,422 controls directly genotyped for 371,504 SNPs.

Methods

Sample description

TGCT cases (n=3,206) were ascertained via two UK studies: (1) a UK study of familial testicular cancer and (2) a systematic collection of UK TGCT cases. Case recruitment was via the UK Testicular Cancer Collaboration, a group of oncologists and surgeons treating TGCT in the UK (Supplementary note). The studies were coordinated at the Institute of Cancer Research (ICR). Samples and information were obtained with full informed consent and Medical Research and Ethics Committee approval (MREC02/06/66 and 06/MRE06/41).

Control samples for the primary GWAS were all taken from within the UK. Specifically 2,976 cancer-free, male controls were recruited through two studies within the PRACTICAL Consortium (Supplementary note): (1) the UK Genetic Prostate Cancer Study (UKGPCS) (age <65), a study conducted through the Royal Marsden NHS Foundation Trust and (2) SEARCH (Study of Epidemiology & Risk Factors in Cancer), recruited via GP practices in East Anglia (2003-2009). 4,446 cancer-free female controls from across the UK were recruited via the Breast Cancer Association Consortium (BCAC).

GWAS

Genotyping was conducted using a custom Infinium OncoArray-500K BeadChip (OncoArray) from Illumina (Illumina, San Diego, CA, USA), comprising a 250K SNP genome-wide backbone and 250K SNP custom content selected across multiple consortia within COGS (Collaborative Oncological Gene-environment Study). OncoArray genotyping was conducted in accordance with the manufacturer's

recommendations by the Edinburgh Clinical Research Facility, Wellcome Trust CRF, Western General Hospital, Edinburgh EH4 2XU.

The UK TGCT OncoArray dataset was filtered as follows: we excluded individuals with low call rate (<95%), with abnormal autosomal heterozygosity (>3 SD above the mean) or with >10% non-European ancestry (based on multi-dimensional scaling); we excluded SNPs with minor allele frequency <1%, a call rate of <95% in cases or controls or with a minor allele frequency of 1–5% and a call rate of <99%; and those deviating from Hardy-Weinberg equilibrium ($P > 10^{-12}$ in controls and 10^{-5} in cases). The final number of SNPs passing quality control filters was 371,504. Case data are deposited at European Genome–phenome Archive [EGA] under accession code EGAS00001001836.

Bioinformatic and statistical analysis

Bioinformatic and statistical analyses were performed as previously described[34]. In brief, we detected RoH using PLINK v1.90[35], which moves a sliding window of SNPs across the entire genome. To allow for genotyping error or other sources of artificial heterozygosity (such as paralogous sequences) within a stretch of truly homozygous SNPs, 2% heterozygous SNPs were allowed in each window. This measure was implemented to prevent underestimation of the number and size of RoH. Default parameter values were employed (including allowing 5 missing calls per window), with the exception that we varied the parameter homozyg-snp according to our heuristic preferences for defining the RoH as detailed below. Subsequent statistical analyses were performed using packages available in R (version 3.1.2) and custom written Perl code. Comparisons of global homozygosity measures between cases and controls were made using the Student t-test. Adjustment for multiple testing was based on the Bonferroni correction.

We used three metrics to investigate the selection pressure on each RoH. Integrated Haplotype Score (iHS) is based on LD surrounding a positively selected allele compared to background, providing evidence of recent positive selection at a locus[36]. An iHS score >2.0 reflects that haplotypes on the ancestral background are longer compared to the derived allelic background. Episodes of selection tend to skew SNP frequencies in different directions and Tajima's D is based on the frequencies of SNPs segregating in the region of interest[37]. Fixation index (Fst) measures the degree of population differentiation at a locus, taking values from 0 to 1.0[38]. iHS, D and Fst metrics were obtained from dbPSHP[39].

Identification of Consensus RoH

In order to focus on commonly occurring ROH and to empower our analysis to identify meaningful associations, only RoH in which 10 or more individuals shared the same RoH were retained for these analyses. The initial search for RoH was performed using PLINK[35] with a specified length of 68 consecutive SNPs (homozyg-snp parameter). This ROH length was chosen (i) to be more than an order of magnitude larger than the mean haploblock size in the human genome (ii) without being so large as to be very rare. The likelihood of observing 68 consecutive chance events can be calculated as follows[40]. Mean heterozygosity in the samples was calculated to be 42%. Thus, given 371,504 SNPs and 10,628 individuals, a minimum length of 47 would be required to produce <5% randomly generated ROH across all subjects ($[1 - 0.42]^{47} \times 371{,}504 \times 10{,}628 < 0.05$). A consequence of LD is that the SNP genotypes are not always independent, thereby inflating the probability of chance occurrences of biologically meaningless ROH. Analysis based on PLINK's pairwise LD SNP pruning function showed an approximate reduction of information compared to the original number of SNPs of 25%. Thus RoH of length 68 SNPs were used to approximate the degrees of freedom of 47 independent SNP calls.

Once all RoH of at least 68 SNPs in length were identified, these were pruned to only those RoH that occurred in more than 10 individuals. To ensure that a minimum length and minimum number of SNPs in each RoH was maintained, each individual's SNP data were recoded as one if the SNP was in an RoH for that individual and zero otherwise. Then, for each SNP, those SNPs with fewer than 10 individuals coded as one were recoded to zero before removing any ROH that due to this recoding were now less than the required number of SNPs in length. This process therefore resulted in a list of "consensus" ROH having a minimum of 68 consecutive homozygous SNP calls across 10 or more samples.

Results and Discussion

We have previously implemented rigorous quality control measures to the UK TGCT OncoArray GWAS dataset[19], excluding samples and SNPs with poor call rates, SNPs with significant departure from HWE, and samples of non-European ancestry or with a sex discrepancy as inferred from the data. The final dataset included 10,628 individuals from the UK and of European ancestry, comprising 3,206 TGCT cases and 7,422 controls, all genotyped on the same platform. The final number of SNPs passing quality control filters was 371,504.

Across all samples (n=10,628), the total number of discrete autosomal RoH >1000 kb and comprising at least 68 consecutive SNPs as identified by PLINK was 137,833, with an average number of 12.97 RoH per individual, an average size of 1630.17 Kb per RoH per individual, and an average total length of the genome covered by RoH of 21,216.01 Kb per individual. These results are broadly similar to other studies using similar methodologies[34, 41-43]. There was no significant difference in the average number, length per RoH, or total length of RoH per individual between TGCT cases and controls when compared using Student's T test (Table 1). Likewise, the cumulative distribution of RoH was broadly similar for TGCT cases and controls (Figure 1).

Data indicate two different types of RoH shaped by different selective pressures, with the different types characterised by different run length[44]. Small/intermediate sized RoH (<1.6 Mb) are shaped via serial migration as a result of decreasing population size, generating LD, reducing haplotype diversity and increasing chance pairing of identical haplotypes. Conversely, long RoH (>1.6 Mb) are generated through inbreeding. There was no difference in global homozygosity measures between TGCT cases and controls when performing these analyses on RoH separated into these size categories (Table 1).

We next identified a set of 319 consensus RoH (Supplementary Table 1), that is RoH that are present in at least 10 individuals. Eight of these consensus RoH had a frequency of greater than 0.25 across all individuals (Table 2). The vast majority of these common consensus RoH has been previously reported in other studies of RoH. For these RoH, selective pressure metrics are indicative of positive selection in Caucasian populations, and their locations are within genomic regions characterised by reduced numbers of structural variants and low recombination rates. The most frequently occurring RoH in our dataset (RoH164) has previously been identified as a site of selective sweep in multiple studies and is frequently identified in studies of common consensus RoH. Importantly, previous reports of these RoH provide further validity of our approach.

Figure 2 shows the correlation between the frequency of consensus RoH in TGCT cases and controls. No consensus RoH was exclusive to either group and none was significantly associated with TGCT risk after correcting for multiple testing (i.e. $p < 0.0001$). Three consensus RoH demonstrated nominal associations with TGCT at a suggestive significance level ($P < 0.01$) (Table 3). Each of these regions showed highly significant values for three estimates of selective pressure (iHS_{max}, Tajimas' D_{max} , and Fay Wu's H), indicating that these regions may have been generated as the result of a selective sweep.

The RoH with the strongest evidence of association, RoH200, was identified in 5% of TGCT cases (n=148) and 3% of controls (n=243) (p= 0.0009; Table 3). It comprises 866 SNPs spanning 9 Mb of chromosome 11 and encompasses 52 genes/predicted transcripts, including *Wilms Tumor 1* (*WT1*), a developmental transcription factor involved in sex determination and establishment of the urogenital system, and with established oncogenic and proto-oncogenic roles in tumour formation. To further investigate a potential link between *WT1* and TGCT risk, we performed an association analysis of individual SNPs within 25 kb of WT1, considering only those with an info score > 0.8 and MAF > 0.01 (n=432). The strongest putative association was for a directly genotyped SNP, rs11031783, which maps to the non-coding *WT1 antisense RNA* (*WT1-AS*), OR = 1.18, p = 0.0003). These data are suggestive and warrant additional validation.

Conclusions

In conclusion, our analyses demonstrate that levels of homozygosity are unlikely to play a substantial role in defining the risk of TGCT. Moreover, our findings suggest that existence of large numbers of recessive alleles that predispose to TGCT when unmasked by autozygosity is unlikely in outbred populations such as that of the UK. Therefore, from these analyses we are unable to provide explanation for epidemiological observation of the higher risks to siblings of cases than to other male family members. However, due to genome-wide testing and requisite correction for multiple-testing, this analysis certainly does not preclude existence of recessively acting disease alleles in TGCT risk; alternative analytic strategies will be needed to identify such alleles if they do exist. Though not statistically significant, the possible link between TGCT and an RoH hotspot that encompasses 11p.13 is an interesting observation that warrants further investigation.

References

Acknowledgements

We thank the patients with TGCT and the clinicians involved in their care for participation in this study. We thank the patients and all clinicians forming part of the UK Testicular Cancer Collaboration (UKTCC) for their participation in this study. This study would not have been possible without the contributions of M.K. Bolla (Breast Cancer Association Consortium (BCAC)), Q. Wang (BCAC), K. Michailido (BCAC), J. Dennis (BCAC), P. Hall (Collaborative Oncological Gene-environment Study (COGS)); D.F. Easton (BCAC), A. Berchuck (Ovarian Cancer Association Consortium), R. Eeles (PRACTICAL), G. Chenevix-Trench (The Consortium of Investigators of Modifiers of BRCA1/2), J. Dennis, P. Pharoah, A. Dunning, K. Muir, J. Peto, A. Lee and E. Dicks. We also thank the following for their contributions to this project: J. Simard, P. Kraft, C. Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory; and K.F. Doheny and the staff of the Center for Inherited Disease Research (CIDR) genotyping facility. The results published here are based in part on data generated by the TCGA Research Network. This study makes use of data generated by the Wellcome Trust Case Control Consortium 2 (WTCCC2). We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out.

Author Contributions Statement

C.T., C.L., A.S. and R.S.H. designed the study. Case samples were recruited by A.R., R.A.H. and through UKTCC. R.E., A.M.D., K.M., J.P., Z.K.-J., N.P. and D.F.E. supplied OncoArray control data. N.O. administrated genotyping of OncoArray case samples. D.D. coordinated all case sample administration and tracking. K.L., M.L., A.H. and P.B. prepared samples for genotyping experiments. C.T., R.S.H., A.S. and C.L. designed bioinformatics and statistical analyses. C.L., K.L. and M.L. performed bioinformatics and statistical analyses. C.L. drafted the manuscript with assistance from C.T., A.S., and R.S.H. All authors reviewed and contributed to the manuscript.

Additional Information

Competing financial interests

The authors declare that they have no competing financial interests.

Figures

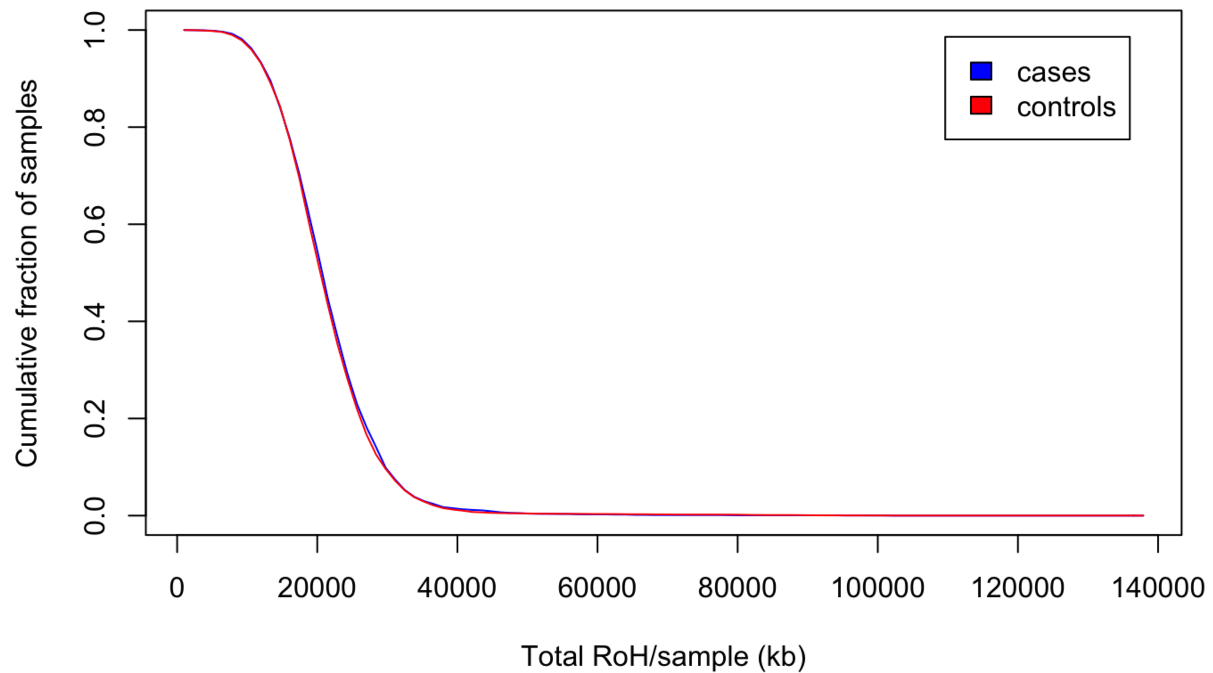

Figure 1. Cumulative distribution of runs of homozygosity (RoH) in TGCT cases and controls. Data is presented in such a way that each data point represents the cumulative fraction (y-axis) of the samples with the corresponding minimum total length of the genome covered by RoH (x-axis), as determined from PLINK.

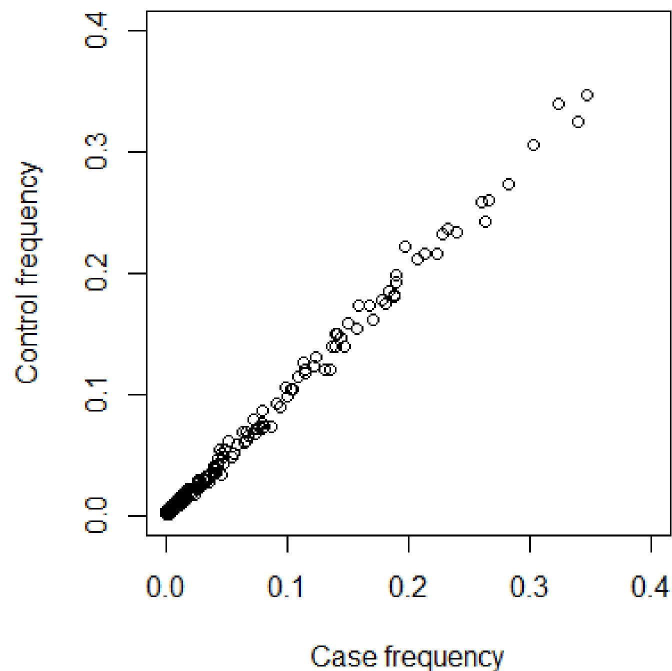

Figure 2. Frequency of consensus runs of homozygosity (RoH) in TGCT cases versus controls. Consensus RoH were defined on the basis of being present in 10 or more individuals.

References

1. Le Cornet C, Lortet-Tieulent J, Forman D, 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* 2014, 50:831-839.
2. Litchfield K, Thomsen H, Mitchell JS, et al: Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Sci Rep* 2015, 5:13889.
3. Hemminki K, Li X: Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *British Journal of Cancer* 2004, 90:1765-1770.
4. Swerdlow AJ, De Stavola BL, Swanwick MA, et al: Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology. *Lancet* 1997, 350:1723-1728.
5. McGlynn KA, Devesa SS, Graubard BI, et al: Increasing incidence of testicular germ cell tumors among black men in the United States. *J Clin Oncol* 2005, 23:5757-5761.
6. Kharazmi E, Hemminki K, Pukkala 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* 2015, 68:283-289.
7. Crockford GP, Linger R, Hockley S, et al: Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. *Hum Mol Genet* 2006, 15:443-451.

8. 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.
9. Rapley EA, Turnbull C, Al Olama AA, et al: A genome-wide association study of testicular germ cell tumor. *Nat Genet* 2009, 41:807-810.
10. Kanetsky PA, Mitra N, Vardhanabhuti S, et al: Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nat Genet* 2009, 41:811-815.
11. Turnbull C, Rapley EA, Seal S, et al: Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat Genet* 2010, 42:604-607.
12. Kanetsky PA, Mitra N, Vardhanabhuti S, et al: A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. *Hum Mol Genet* 2011, 20:3109-3117.
13. Turnbull C, Rahman N: Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. *Int J Androl* 2011, 34:e86-96; discussion e96-87.
14. Chung CC, Kanetsky PA, Wang Z, et al: Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat Genet* 2013, 45:680-685.
15. Ruark E, Seal S, McDonald H, et al: Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat Genet* 2013, 45:686-689.
16. Kristiansen W, Karlsson R, Rounge TB, 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* 2015, 24:4138-4146.
17. Litchfield K, Sultana R, Renwick A, et al: Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. *Hum Mol Genet* 2015, 24:1169-1176.
18. Litchfield K, Holroyd A, Lloyd A, et al: Identification of four new susceptibility loci for testicular germ cell tumour. *Nat Commun* 2015, 6:8690.
19. 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, 49:1133-1140.
20. Wang Z, McGlynn KA, 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, 49:1141-1147.
21. Litchfield K, Mitchell JS, Shipley J, et al: Polygenic susceptibility to testicular cancer: implications for personalised health care. *Br J Cancer* 2015, 113:1512-1518.
22. Sud A, Kinnersley B, Houlston RS: Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer* 2017.
23. Bener A, El Ayoubi HR, Chouchane L, et al: Impact of consanguinity on cancer in a highly endogamous population. *Asian Pac J Cancer Prev* 2009, 10:35-40.
24. Lebel RR, Gallagher WB: Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. *Am J Med Genet* 1989, 33:1-6.
25. Shami SA, Qaisar R, Bittles AH: Consanguinity and adult morbidity in Pakistan. *Lancet* 1991, 338:954.
26. Simpson JL, Martin AO, Elias S, et al: Cancers of the breast and female genital system: search for recessive genetic factors through analysis of human isolate. *Am J Obstet Gynecol* 1981, 141:629-636.
27. Assie G, LaFramboise T, Platzer P, et al: Frequency of germline genomic homozygosity associated with cancer cases. *JAMA* 2008, 299:1437-1445.
28. Demant P: Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat Rev Genet* 2003, 4:721-734.

29. Henry I, Bonaiti-Pellie C, Chehensse V, et al: Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature* 1991, 351:665-667.
30. McQuillan R, Leutenegger AL, Abdel-Rahman R, et al: Runs of homozygosity in European populations. *Am J Hum Genet* 2008, 83:359-372.
31. Ku CS, Naidoo N, Teo SM, et al: Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet* 2011, 129:1-15.
32. Szpiech ZA, Xu J, Pemberton TJ, et al: Long runs of homozygosity are enriched for deleterious variation. *Am J Hum Genet* 2013, 93:90-102.
33. Ceballos FC, Joshi PK, Clark DW, et al: Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet* 2018.
34. Sud A, Cooke R, Swerdlow AJ, et al: Genome-wide homozygosity signature and risk of Hodgkin lymphoma. *Sci Rep* 2015, 5:14315.
35. Purcell S, Neale B, Todd-Brown K, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007, 81:559-575.
36. Voight BF, Kudaravalli S, Wen X, et al: A map of recent positive selection in the human genome. *PLoS Biol* 2006, 4:e72.
37. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, 123:585-595.
38. Holsinger KE, Weir BS: Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet* 2009, 10:639-650.
39. Li MJ, Wang LY, Xia Z, et al: dbPSHP: a database of recent positive selection across human populations. *Nucleic Acids Res* 2014, 42:D910-916.
40. Lencz T, Lambert C, DeRosse P, et al: Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A* 2007, 104:19942-19947.
41. Hosking FJ, Papaemmanuil E, Sheridan E, et al: Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood* 2010, 115:4472-4477.
42. Thomsen H, Filho MI, Woltmann A, et al: Inbreeding and homozygosity in breast cancer survival. *Sci Rep* 2015, 5:16467.
43. Thomsen H, Chen B, Figlioli G, et al: Runs of homozygosity and inbreeding in thyroid cancer. *BMC Cancer* 2016, 16:227.
44. Pemberton TJ, Absher D, Feldman MW, et al: Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 2012, 91:275-292.