



Published in final edited form as:

*Hum Genet.* 2014 October ; 133(10): 1289–1297. doi:10.1007/s00439-014-1463-z.

## Characterization of T gene sequence variants and germline duplications in familial and sporadic chordoma

MJ Kelley<sup>1</sup>, J Shi<sup>2</sup>, B Ballew<sup>2</sup>, PL Hyland<sup>2</sup>, WQ Li<sup>2</sup>, M Rotunno<sup>2</sup>, DA Alcorta<sup>3</sup>, NJ Liebsch<sup>4</sup>, J Mitchell<sup>2</sup>, S Bass<sup>2</sup>, D Roberson<sup>2</sup>, J Boland<sup>2</sup>, M Cullen<sup>2</sup>, J He<sup>2</sup>, L Burdette<sup>2</sup>, M Yeager<sup>2</sup>, SJ Chanock<sup>2</sup>, DM Parry<sup>2</sup>, AM Goldstein<sup>2</sup>, XR Yang<sup>2</sup>

<sup>1</sup>Department of Medicine, Duke University Medical Center, Durham, NC, 27705, USA.

<sup>2</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA.

<sup>3</sup>Department of Medicine, Duke University, Durham, North Carolina 27710, United States.

<sup>4</sup>Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.

### Abstract

Chordoma is a rare bone cancer that is believed to originate from notochordal remnants. We previously identified germline *T* duplication as a major susceptibility mechanism in several chordoma families. Recently, a common genetic variant in *T*(rs2305089) was significantly associated with the risk of sporadic chordoma. We sequenced all *T* exons in 24 familial cases and 54 unaffected family members from eight chordoma families (three with *T* duplications), 103 sporadic cases, and 160 unrelated controls. We also measured *T* copy number variation in all sporadic cases. We confirmed the association between the previously reported variant rs2305089 and risk of familial [odds ratio (OR) = 2.6, 95 % confidence interval (CI) = 0.93, 7.25,  $P = 0.067$ ] and sporadic chordoma (OR = 2.85, 95 % CI = 1.89, 4.29,  $P < 0.0001$ ). We also identified a second common variant, rs1056048, that was strongly associated with chordoma in families (OR = 4.14, 95 % CI = 1.43, 11.92,  $P = 0.0086$ ). Among sporadic cases, another common variant (rs3816300) was significantly associated with risk when jointly analyzed with rs2305089. The association with rs3816300 was significantly stronger in cases with early age onset. In addition, we identified three rare variants that were only observed among sporadic chordoma cases, all of which have potential functional relevance based on in silico predictions. Finally, we did not observe *T* duplication in any sporadic chordoma case. Our findings further highlight the importance of the *T* gene in the pathogenesis of both familial and sporadic chordoma and suggest a complex susceptibility related to *T*.

### Introduction

Chordoma (MIM 215400) is a rare bone cancer, with an age-adjusted incidence rate of less than 0.1 per 100,000 in the United States (Smoll et al. 2013). It is thought to originate from notochord remnants and occurs almost exclusively in the axial skeleton where it is distributed nearly equally among cranial, vertebral and sacral sites (McMaster et al. 2001). Chordoma occurs more frequently in males than females (1.6:1) and in Caucasians than

African-Americans (4:1), and is diagnosed at a median age of 58 years with a range from infancy to over 90 (McMaster et al. 2001; Smoll et al. 2013). Treatment of chordomas is predominantly surgical, followed by radiotherapy (Healey and Lane 1989; Rich et al. 1985; Tai et al. 1995). Based on the analysis using data up to 2009 from the National Cancer Institute's Surveillance Epidemiology, and End Results (SEER) 18 registries, the median survival of chordoma patients was 7.7 years and the age-standardized 5- and 10-year survival rates were 72 and 48 %, respectively (Smoll et al. 2013).

Although etiologic factors for this rare cancer are largely unknown, the *T* gene, which encodes the protein brachyury, has been implicated in its pathogenesis (Vujovic et al. 2006; Presneau et al. 2011). Brachyury is a tissue-specific transcription factor expressed in the nucleus of notochord cells (Kispert et al. 1995) and is essential for proper notochord development and maintenance (Kispert and Herrmann 1994). We previously identified germline *T* duplication as a major susceptibility mechanism in several chordoma families (Yang et al. 2009). Recently, Pillay et al. (2012) identified a common genetic variant in *T* (rs2305089) that was significantly associated with the risk of sporadic chordoma, further highlighting the importance of *T* in chordoma pathogenesis. However, the exact role of this common variant (~50 % among Caucasians) in such a rare tumor remains unclear. In a recent study investigating skull-based chordoma patients among Chinese, rs2305089 was not significantly associated with chordoma risk (Wu et al. 2013), suggesting that the risk associated with this variant may vary by tumor site or ethnicity.

To further characterize *T* germline variants and their associations with chordoma risk, we sequenced all *T* exons and measured *T* copy number variations (CNVs) in the largest collection of chordoma cases with germline DNA available to date, consisting of 24 familial cases from eight chordoma families and 103 sporadic chordoma cases.

## Methods

### Study population

We have ascertained nine chordoma families with two or more affected members in each family and included eight of them in this examination based on DNA availability. Six families were previously described (Yang et al. 2009) and the two new families included in this study (Families 5 and 9) had three and two chordoma cases, respectively (Table 1, Supplementary Fig. 1). In Family 5, there is only one confirmed chordoma case (the proband's father). The proband had a notochord remnant of the clivus (ecchordosis physaliphora) confirmed by histopathology at autopsy. A third relative (the proband's nephew) had a bilobed lesion associated with the clivus on magnetic resonance images, which is suspected to be a chordoma. We considered these two suspected cases as chordoma cases in this analysis. One previously reported family had a *T* duplication [Family 3 from Yang et al. (2009)] but was not included in the current study because DNA was not available.

Sporadic cases ( $N=103$ ) were recruited from the United States or Canada and a subset of them was evaluated at the NIH Clinical Center. Controls were healthy unrelated individuals who were participants in family studies unrelated to chordoma. All study subjects were

Caucasians. Both the family and sporadic chordoma studies were approved by institutional review boards at the National Cancer Institute and all participants provided written informed consent.

### Quantitative PCR

We performed quantitative real-time PCR (qPCR) using the TaqMan assay with an ABI 7000 sequence detection system to measure *T* copy number variations (CNVs) in blood-derived DNA from five cases in Families 5 and 9 and 11 young sporadic cases examined at the National Cancer Institute (NCI) Clinical Center, as well as saliva DNA from 92 sporadic cases collected using Oragene™ DNA Self-Collection Kit (DNA Genotek). Custom TaqMan probes targeted to the 3' region of exon 6 of the *T* gene (forward PCR primer, GTACTCCCAATGTACGGTTTGTTG; reverse PCR primer, TCAGCAAGTCTAGTCCCGATGAC; TaqMan MGB probe, CTCTGTTCATGTCATTCTG) were designed with Primer express v2.0 software (ABI). RNaseP probe (ABI) was used as the endogenous “normal copy number reference” for each sample. Multiplex qPCR reactions using *T* and RNaseP probes were performed for each sample in triplicate. Thirty ng of DNA per individual was used as the template in each replicate. Cycling parameters were: 95 °C × 10 min, followed by 40 cycles of 95 °C × 15 s, and 60° × 1 min. Cycle time (CT), defined as the threshold number of PCR cycles at which fluorescence from the TaqMan reaction was first detected, was determined for each replicate. An average of the replicate CT values was calculated for the *T* gene and RNaseP probes for each sample. The differences between the average CT of the *T* gene and RNaseP were calculated ( $\Delta CT = CT_{\text{target}} - CT_{\text{reference}}$ ). An unrelated spouse was used as the calibrator (reference for determining Brachyury copy number) for all tested samples. The average CT value for the calibrator was subtracted from each sample to calculate the CT value ( $\Delta CT = CT_{\text{test sample}} - CT_{\text{calibrator sample}}$ ). The fold difference (reflecting copy number status) was calculated using the comparative CT method;  $2^{-\Delta CT}$  (ABI application note part #4371095). Efficiency testing of the *T* gene and RNaseP probes was performed, and the slope of  $\Delta CT$  vs. log input was <0.1, validating the use of the comparative CT method.

### Array CGH

We designed a custom-made fine-tiling array specifically targeting the 6q27 region (average probe spacing, 220 bp) using Agilent CGH arrays (8 × 60 k). For this analysis, we selected five samples with either suggestive positive or inconsistent CNV calls by qPCR as well as a positive control sample (a familial chordoma case with *T* duplication). The reference sample was a pool of six male individuals (Promega, Madison, WI) provided by Oxford Gene Technology, Inc. (UK). 500 ng of test and reference DNA was labeled with Cy3 and Cy5, respectively, and co-hybridized to the array slides. All experimental procedures and data analyses were described previously (Yang et al. 2009).

### *T* gene sequencing

A custom set of Ampliseq™ primers from Life Technologies was used to amplify the coding region of the *T* gene. The primers were created using the primer design engine at <http://www.ampliseq.com>. 40 ng of genomic DNA was used for each sample. The Ampliseq™ process was performed according to the official process documentation found

on the Ion Community (<http://www.ioncommunity.lifetechnologies.com>). The samples were quality checked for proper amplicon length and quantity on the Agilent Bioanalyzer and were then sequenced according to the current Ion Torrent template and sequencing preparation documentation. Each PGM Ion 316 sequencing run consisted of 96 barcoded samples with randomized platform position. The resulting sequencing data yielded coverage for each amplicon >300X in depth. Single nucleotide substitutions were identified and quality filtered with the Genome Analysis Toolkit (GATK) (DePristo et al. 2011). To ensure the accuracy of the variant call, we included only variants that were called by both GATK and the torrent variant caller.

### Variant annotation

We used computational tools including PolyPhen-2, SIFT, Provean, MutationAssessor, and MutationTaster to predict the potential impact of sequence variants on protein function. We also obtained conservation scores using GERP, PhastCons, and PhyloP, as well as ProPhyLER to predict mutation impact based on evolutionary constraint analyses.

### Statistical analyses

For the analysis of chordoma families, odds ratios (ORs) and 95 % confidence intervals (95 % CIs) were obtained from conditional logistic regression models by comparing cases to all unaffected subjects (spouses and unaffected family members combined), conditioning on family to account for the ascertainment. Although this approach ignores residual correlation among family members, it gives estimates that are attenuated toward the null and thus is considered conservative (Pfeiffer et al. 2001). We also used an independent approach, a generalized estimating equation with the independence working correlation matrix, to account for familial correlation and observed similar results, and we therefore only present the results obtained from the conditional logistic regression analysis. Unconditional logistic regression was used for the analysis comparing sporadic cases to unrelated controls and Mantel–Haenszel was used for the analysis comparing sporadic cases to ESP EA controls. For each SNP, we calculated the  $P_{\text{trend}}$  based on the three-level ordinal genotype variable (0, 1, 2) that counted the number of minor alleles using the homozygous common allele genotype as the referent group. Adjusting for age and gender did not substantially alter the results (data not shown). We used SAS (version 9.3, SAS Institute, Inc., Cary, NC, USA) software for all analyses.

## Results

We included eight chordoma families in this examination based on DNA availability. Six families were previously described (three with *T* germline duplications and three without) (Yang et al. 2009). The two new families (Families 5 and 9) had three and two chordoma cases, respectively (Table 1, Supplementary Fig. 1). We also evaluated 103 sporadic chordoma cases. Patient characteristics for these cases are summarized in Tables 1, 2.

### Germline *T* copy number variations (CNVs)

We evaluated CNVs in the *T* gene in chordoma cases from the two new families and all sporadic cases using a quantitative PCR (qPCR) assay. While the positive control sample (a

chordoma case from a family with known *T* duplication) clearly carried the duplication, none of the familial or sporadic cases examined had the duplication (data not shown). Samples with potentially suggestive positive signals were confirmed to be negative using Agilent CGH arrays (8 × 60 k) specifically targeting the 6q27 region (average probe spacing, 220 bp).

### Single nucleotide variants (SNVs) in *T* in chordoma families

We sequenced all *T* exons in 24 familial chordoma patients plus 54 unaffected family members with DNA available in the eight families, 103 sporadic chordoma cases, and 160 unrelated healthy controls. Table 3 lists the five *T* exonic variants observed in the eight families, all of which were common variants (>5 % in The 1000 Genomes Project [1,092 subjects] and NHLBI GO Exome Sequencing Project [ESP, including exomes from up to 4,300 subjects of European ancestry]). The T allele of rs2305089, which was previously associated with increased risk for sporadic chordoma (Pillay et al. 2012), was observed in all chordoma cases in the eight families. Compared to all controls (unaffected family members and spouses combined), chordoma cases were more likely to carry the variant allele after correction for familial correlation, although the association was not statistically significant [odds ratio (OR) = 2.6, 95 % confidence interval (CI) = 0.93, 7.25,  $P = 0.067$ ]. The frequency of this variant was similar in families with and without a *T* duplication (Supplementary Table 1). Another common variant, rs1056048, demonstrated a higher risk estimate than rs2305089 in these families (OR = 4.14, 95 % CI = 1.43, 11.92,  $P = 0.0086$ ). This variant showed only weak linkage disequilibrium (LD) with rs2305089 ( $r^2 = 0.18$ ). When including both variants in the regression model, the association for rs1056048 remained significant (OR = 4.33, 95 % CI = 1.27, 14.8,  $P = 0.019$ ), whereas the association for rs2305089 was substantially attenuated (OR = 1.49, 95 % CI = 0.45, 4.93,  $P = 0.51$ ). Further, interestingly, the association for rs1056048 was only seen in families with a *T* duplication (Supplementary Table 1).

### Single nucleotide variants (SNVs) in *T* in sporadic chordoma cases

All common *T* variants seen in the families were also observed in the sporadic cases and controls that were sequenced (Table 4). Four variants showed significant associations after controlling for the false discovery rate (FDR) at 5 % (nominal  $P < 0.033$ ) when comparing chordoma cases to the controls from this study or ESP [4,300 subjects of European ancestry (EA)]. Rs2305089 showed the strongest association (OR = 2.85, 95 % CI = 1.89, 4.29). Associations for the other three significant variants, including rs1056048, became non-significant when the analysis was conducted conditioning on rs2305089 (Table 4). Interestingly, rs3816300, which was not significant in the univariate analysis, became significant when analyzed jointly with rs2305089 (Table 4). Carrying one risk allele for both variants conferred a similar risk (OR = 11.4, 95 % CI = 3.31, 39.29,  $P = 0.0001$ ) to having two risk alleles for rs2305089 (OR = 11.52, 95 % CI = 3.83, 34.62,  $P < 0.0001$ ; Table 5). Interestingly, while the risks associated with rs2305089 were similar in all age groups, the association with rs3816300 was significantly stronger in cases with early age onset ( $P = 0.017$ ), particularly in cases with age at diagnosis younger than 20 years (Table 6). Rs3816300 also showed a stronger association in cases with skull-base chordoma compared

to those with chordoma in other sites (Table 6). Adjusting for age and gender did not substantially alter the results (data not shown).

A less common synonymous variant [rs920961, observed 28 times in ESP EA (MAF = 0.3 %)] showed a significant association in the case–ESP comparison and a suggestive association in the case–control comparison when conditioning on rs2305089 (Table 4); the lack of significance is likely due to the limited power. The crystal structure of the T domain (DNA-binding domain) from *Xenopus laevis* showed that amino acids homologous to rs2305089, rs1056048, and rs920961 are located in loop regions near a beta-sheet and affect the T domain (Muller and Herrmann 1997) (Supplementary Fig. 2).

In addition to the common variants, we also identified two rare missense variants (g.6:166571981 C>T, p.R377Q; and g.6:166576021 G>A, p.T273 M; NP\_003172) and one rare variant (g.6:166581095 C>T) located in the 5′ untranslated region (UTR); each variant was seen in a single sporadic case and no controls (Supplementary Table 2). The three variants were not reported in 1,000 Genomes Project among Europeans or ESP EA. Both missense variants were predicted to be deleterious by most of the prediction programs evaluated and are located in moderately conserved regions (Supplementary Table 2). The 5′-UTR variant (rs3734509) is located in a DNaseI site in ENCODE cell lines and overlaps with a bivalent chromatin domain in H1 cells and numerous transcription factor binding motifs (Chadwick 2012; Rosenbloom et al. 2013). In addition, MutationTaster predicted that this variant could alter splicing which may lead to the loss of the T-box domain. Carriers for the three rare variants were all <35 years at diagnosis and the carrier of p.R377Q had multicentric chordoma (skull-base, lumbar spine, and sacrum). In contrast, among 160 sequenced controls, only one rare variant (rs369239526) was observed in a single subject, but this variant was synonymous and seen three times in ESP EA.

## Discussion

In this study, we measured CNVs in the *T* gene and sequenced all *T* exons in the largest collection of chordoma families and sporadic chordoma case series reported to date. We identified several variants in the *T* gene in addition to the previously reported variant rs2305089 that are related to disease risk. Our results also demonstrated that, although germline *T* duplication is fairly common in chordoma families (44 %), it is extremely rare among sporadic cases, which is consistent with data from previous reports (Presneau et al. 2011; Pillay et al. 2012).

The risk estimates for rs2305089 were similar in our familial and sporadic case–control comparisons (OR = 2.5–3), although they were much lower compared to what was reported by Pillay et al. (OR = 5.3, 95 % CI = 3.1, 8.9) (Pillay et al. 2012). The frequency of the risk allele (T) was similar among controls (53 %) in the two studies, whereas the frequencies were higher among chordoma cases in the study of Pillay et al. (2012) (87.5 % in the discovery set and 82.5 % in the replication set) compared to those of the current study (72 and 76 % among familial and sporadic cases, respectively). It is unlikely that the difference is driven by the variation in patient characteristics, since the association for rs2305089 did not vary by age at diagnosis or site of chordoma in our analysis. Small sample size in both



studies may at least partially contribute to the variation, although the number of cases in our study doubled that in the study of Pillay et al. Interestingly, in a recent investigation of 65 skull-base chordoma cases and 120 healthy controls among Chinese, rs2305089 was not significantly associated with chordoma risk (Wu et al. 2013). Given the strong association with this variant observed in our study in which the majority of cases had skull-base chordoma, the lack of association among Chinese is more likely driven by the population difference rather than the site of chordoma. In fact, the frequency of the T allele is much lower among Asian (36 %) compared to Caucasians (~50 %). These findings highlight the need of a more comprehensive characterization of *T* variants rather than genotyping a single variant in different populations.

In the family analysis, both rs1056048 and rs2305089 were associated with chordoma risk. In contrast to what was seen in the sporadic case–control comparisons in which the association for rs1056048 became non-significant after adjusting for rs2305089, analyzing both variants together in families resulted in a significant association for rs1056048 and an attenuated non-significant association for rs2305089. Interestingly, whereas the association for rs2305089 did not vary by family *T* duplication status, the association for rs1056048 was only seen in families with *T* duplications. Although rs1056048 is a synonymous variant, according to the prediction by the program MutationTaster (Schwarz et al. 2010), it has the potential to influence splicing which may in turn affect sumoylation and nuclear localization of the protein. In addition, the region containing rs1056048 overlaps with a strong repressive H3K27me3 mark and a binding region for SUZ12, which is essential for polycomb repressive complex 2 (PRC2)-mediated gene silencing. The SNP region also maps to a DNase I site, a poised transcriptional start site and a large CpG island in a number of cell lines such as human embryonic stem cells (H1), which express *T*, based on ENCODE (Rosenbloom et al. 2013) and NIH Epigenomics Roadmap data (Chadwick 2012), indicating that this variant may have an important regulatory role(s).

In the sporadic case–control comparisons, although three other common variants in addition to rs2305089 (rs35819705, rs3127328, and rs1056048) showed associations in the univariate analysis, they became non-significant when the analysis was conducted conditioning on rs2305089. However, our data also revealed a novel association with another variant, rs3816300, which was not significant in the univariate analysis but became significant when analyzed jointly with rs2305089. This can be explained by the negative correlation ( $r = -0.40$ ,  $P < 0.0001$ ) between the two risk alleles (not in LD,  $r^2 < 0.1$ ), under which only the joint analysis can detect both associations (Yang et al. 2011). Because of the strong negative correlation, none of the cases or controls examined were homozygous for both risk alleles. Individuals carrying one risk allele for each variant were associated with a more than tenfold risk, similar to those who were homozygous for rs2305089. These data suggest that the joint analysis of the two variants may provide a more accurate risk estimate. Moreover, the association for rs3816300 showed a significant interaction with age at diagnosis, with association being significantly stronger among cases with earlier age onset. We also observed a stronger association among patients with skull-base chordoma; however, skull-base chordomas are known to have earlier age onset (Chambers et al. 2013) and therefore this finding may be driven by age. Consistent with these results, the two familial cases who

carried the variant allele both had skull-base chordoma and were diagnosed at young ages (5 and 46 years, respectively).

Our study is not powered for finding significant associations for less common or rare variants. However, the one less frequent variant observed among sporadic cases, rs920961, showed a suggestive association in the case–ESP comparison as well as in the case–control analysis conditioning on rs2305089. Together with rs2305089 and rs1056048, this variant is located in the T-box of brachyury, which is the DNA-binding domain. Although these variants do not appear to be in direct contact with DNA, rs2305089 is predicted to decrease the stability of protein structure using support vector machines leveraging both sequence and structural information (Cheng et al. 2006). Our analysis also identified three rare variants that were not reported in 1,000 Genomes Project among Europeans or ESP EA and may have potential functional relevance based on in silico predictions. Carriers for all three variants were <35 years old at diagnosis, 23 years less than the median age at chordoma diagnosis in the US (Smoll et al. 2013). In particular, the carrier of p.R377Q had multicentric chordoma, which is extremely rare, suggesting that the etiology of chordoma in these patients is likely to be influenced by genetic factors.

In summary, our findings provide more evidence for the importance of genetic variations in the *T* gene in the pathogenesis of both familial and sporadic chordoma. The susceptibility related to *T*, however, appears to be complex, involving multiple mechanisms including *T* duplication (essentially seen only in families) and multiple common and rare variants. Risks associated with some of the variants appear to vary by patient characteristics such as age onset and/or site of chordoma. Future work should focus on functional analyses in cell lines or experimental animals to elucidate the molecular mechanisms underlying these associations and interactions across different variants. Analysis of a much larger chordoma dataset is needed to obtain a more precise risk estimate for each identified *T* variant as well as their interactions and to identify additional variants in *T* and other susceptibility genes. This will require the establishment of national and international collaborations to study this rare cancer.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Chadwick LH (2012) The NIH Roadmap Epigenomics Program data resource. *Epigenomics* 4(3): 317–324. doi:10.2217/epi.12.18 [PubMed: 22690667]
2. Chambers KJ, Lin DT, Meier J, Remenschneider A, Herr M, Gray ST (2013) Incidence and survival patterns of cranial chordoma in the United States. *Laryngoscope*. doi:10.1002/lary.24420
3. Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 62(4):1125–1132. doi:10.1002/prot.20810 [PubMed: 16372356]
4. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytzky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping



- using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498. doi:10.1038/ng.806 [PubMed: 21478889]
5. Healey JH, Lane JM (1989) Chordoma: a critical review of diagnosis and treatment. *Orthop Clin North Am* 20(3):417–426 [PubMed: 2662114]
  6. Kispert A, Herrmann BG (1994) Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev Biol* 161(1):179–193. doi:10.1006/dbio.1994.1019 [PubMed: 8293872]
  7. Kispert A, Koschorz B, Herrmann BG (1995) The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J* 14(19):4763–4772 [PubMed: 7588606]
  8. McMaster ML, Goldstein AM, Bromley CM, Ishibe N, Parry DM (2001) Chordoma: incidence and survival patterns in the United States, 1973–1995. *Cancer Causes Control* 12(1):1–11 [PubMed: 11227920]
  9. Muller CW, Herrmann BG (1997) Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 389(6653):884–888. doi:10.1038/39929 [PubMed: 9349824]
  10. Pfeiffer RM, Gail MH, Pee D (2001) Inference for covariates that accounts for ascertainment and random genetic effects in family studies. *Biometrika* 88:16
  11. Pillay N, Plagnol V, Tarpey PS, Lobo SB, Presneau N, Szuhai K, Halai D, Berisha F, Cannon SR, Mead S, Kasperaviciute D, Palmen J, Talmud PJ, Kindblom LG, Amary MF, Tirabosco R, Flanagan AM (2012) A common single-nucleotide variant in T is strongly associated with chordoma. *Nat Genet* 44(11):1185–1187. doi:10.1038/ng.2419 [PubMed: 23064415]
  12. Presneau N, Shalaby A, Ye H, Pillay N, Halai D, Idowu B, Tirabosco R, Whitwell D, Jacques TS, Kindblom LG, Bruderlein S, Moller P, Leithner A, Liegl B, Amary FM, Athanasou NN, Hogendoorn PC, Mertens F, Szuhai K, Flanagan AM (2011) Role of the transcription factor T (brachyury) in the pathogenesis of sporadic chordoma: a genetic and functional-based study. *J Pathol* 223(3):327–335. doi:10.1002/path.2816 [PubMed: 21171078]
  13. Rich TA, Schiller A, Suit HD, Mankin HJ (1985) Clinical and pathologic review of 48 cases of chordoma. *Cancer* 56(1):182–187 [PubMed: 2408725]
  14. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* 41 (Database issue):D56–63. doi:10.1093/nar/gks1172
  15. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8):575–576. doi:10.1038/nmeth0810-575 [PubMed: 20676075]
  16. Smoll NR, Gautschi OP, Radovanovic I, Schaller K, Weber DC (2013) Incidence and relative survival of chordomas: the standardized mortality ratio and the impact of chordomas on a population. *Cancer* 119(11):2029–2037. doi:10.1002/cncr.28032 [PubMed: 23504991]
  17. Tai PT, Craighead P, Bagdon F (1995) Optimization of radiotherapy for patients with cranial chordoma. A review of dose-response ratios for photon techniques. *Cancer* 75(3):749–756 [PubMed: 7828124]
  18. Vujovic S, Henderson S, Presneau N, Odell E, Jacques TS, Tirabosco R, Boshoff C, Flanagan AM (2006) Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J Pathol* 209(2):157–165. doi:10.1002/path.1969 [PubMed: 16538613]
  19. Wu Z, Wang K, Wang L, Feng J, Hao S, Tian K, Zhang L, Jia G, Wan H, Zhang J (2013) The Brachyury Gly177Asp SNP is not associated with a risk of skull base chordoma in the Chinese population. *Int J Mol Sci* 14(11):21258–21265. doi:10.3390/ijms141121258 [PubMed: 24232574]
  20. Yang XHR, Ng D, Alcorta DA, Liebsch NJ, Sheridan E, Li SF, Goldstein AM, Parry DM, Kelley MJ (2009) T (brachyury) gene duplication confers major susceptibility to familial chordoma. *Nat Genet* 41(11):1176–1178. doi:10.1038/Ng.454 [PubMed: 19801981]
  21. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM (2011)

Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43(6): 519–525. doi:10.1038/ng.823 [PubMed: 21552263]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**  
**Distribution of characteristics among familial chordoma cases and family members included in this study**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

Family	#Cases	#Cases with DNA	#Unaffected family members with DNA	#Spouses with DNA	Median age at diagnosis <sup>a</sup>	Tumor site	<i>T</i> dup
1	11	11	15	8	31	1 Sacrum, 10 skull-base	Yes
2	2	2	6	4	29	Skull-base	No
4	3	2	0	1	17	Skull-base	Yes
5 <sup>b</sup>	3	3	8	3	40	1 Sacrum, 2 skull-base	No
6	2	1	6	0	46	Skull-base	No
7	2	1	0	0	37	1 Sacrum, 1 skull-base	No
8	3	3	1	0	8	Skull-base	Yes
9	2	1	2	0	26	1 Mobile spine, 1 skull-base	No

<sup>a</sup>Median age at diagnosis among chordoma cases with DNA

<sup>b</sup>There is only one confirmed chordoma case in Family 5. Chordomas in the other 2 potential cases were not confirmed; one had a notochord rest at autopsy and another had notochord remnant by imaging

**Table 2**  
**Distribution of patient characteristics among sporadic chordoma cases included in this study**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

	<i>N</i>	%
Age at diagnosis		
<10	3	2.9
10–20	10	9.7
20–30	10	9.7
30–40	19	18.4
40–50	18	17.5
50–60	32	31.1
60	11	10.7
Gender		
Male	50	40.7
Female	73	59.3
Tumor site		
Skull	61	59.2
Mobile spine	19	17.5
Sacrum/coccyx	23	22.3

**Table 3**  
**All exonic *T* sequence variants in familial chordoma cases**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

Location	dbSNP	Ref	Var	Variant allele frequency					Case-control comparison		
				1 k genome Eur <sup>a</sup>	ESP EA <sup>b</sup>	Spouses	Unaffected family members	Cases	OR <sup>c</sup>	95 % CI <sup>c</sup>	P <sup>c</sup>
166571935	rs35819705	C	T	0.25	0.30	0.44	0.26	0.19	0.64	0.26, 1.54	0.31
166572005	rs3816300	T	C	0.10	0.094	0	0	0.04	N/A	N/A	N/A
166572045	rs3127328	C	T	0.12	0.12	0.16	0.18	0.10	0.51	0.15, 1.71	0.28
166579270	rs2305089	C	T	0.47	0.51	0.66	0.56	0.72	2.60	0.93, 7.25	0.067
166580188	rs1056048	G	A	0.22	0.21	0.22	0.31	0.52	4.14	1.43, 11.92	0.009

<sup>a</sup>The 1,000 Genomes Project, European sub-population

<sup>b</sup>NHLBI GO Exome Sequencing Project (ESP), including exomes from up to 4,300 subjects of European ancestry

<sup>c</sup>ORs, 95 % CIs, and *P* values were obtained using conditional logistic regression (log-additive) models comparing cases to all unaffected subjects (spouses and unaffected family members combined) conditioning on families to account for familial correlation. The variant C allele was observed in two cases and zero controls and the logistic regression model for this SNP did not converge because of small number

**Table 4**  
**All common (reported in ESP EA) exonic *T* sequence variants in sporadic chordoma cases and controls**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

Location	dbSNP	Ref	Var	Variant allele freq		Case-ESP EA <sup>a</sup>			Case-control <sup>b</sup>			Conditioning on rs2305089 <sup>c</sup>		
				Controls	Cases	OR	95 % CI	P	OR	95 % CI	P	OR	95 % CI	P
166571935	rs35819705	C	T	0.32	0.42	1.64	1.24, 2.17	0.0006	1.51	1.05, 2.17	0.025	0.81	0.52, 1.26	0.35
166572005	rs3816300	T	C	0.1	0.092	0.98	0.61, 1.58	0.94	0.89	0.49, 1.60	0.69	2.17	1.08, 4.39	0.03
166572045	rs3127328	C	T	0.12	0.058	0.45	0.25, 0.81	0.007	0.46	0.24, 0.91	0.003	0.92	0.43, 1.95	0.82
166579270	rs2305089	C	T	0.53	0.76	3.00	2.17, 4.13	<0.0001	2.85	1.89, 4.29	<0.0001	N/A	N/A	N/A
166580188	rs1056048	G	A	0.22	0.35	2.02	1.51, 2.70	<0.0001	1.9	1.27, 2.83	0.002	1.31	0.84, 2.02	0.23
166580257	rs920961	G	A	0.0093	0.014	4.53	1.36, 15.15	0.014	1.56	0.31, 7.90	0.59	3.14	0.57, 17.13	0.19

<sup>a</sup>ORs, 95 % CIs, and *P* values were obtained using Mantel-Haenszel for the analysis comparing sporadic cases to ESP EA controls in log-additive models

<sup>b</sup>ORs, 95 % CIs, and *P* values were obtained using unconditional regression models comparing sporadic cases to unrelated controls in log-additive models. Each SNP was analyzed separately in the model

<sup>c</sup>ORs, 95 % CIs, and *P* values were obtained using unconditional regression models comparing sporadic cases to unrelated controls in log-additive models. Each SNP was adjusted for rs2305089 in the model



**Table 5**  
**Combined genotypes for rs2305089 and rs3816300 in sporadic chordoma cases and controls**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

Genotype	Control	Case	OR <sup>a</sup>	95 % CI <sup>a</sup>	P <sup>a</sup>
rs2305089 CC, rs3816300 TT	22	2	Ref <sup>b</sup>		
rs2305089 CC, rs3816300 TC	12	1			
rs2305089 CC, rs3816300 CC	3	0			
rs2305089 CT, rs3816300 TT	61	25	3.89	1.26,12.06	0.018
rs2305089 CT, rs3816300 TC	15	18	11.4	3.31, 39.29	0.0001
rs2305089 CT, rs3816300 TC	0	0			
rs2305089 TT, rs3816300 TT	47	57	11.52	3.83, 34.62	<0.0001
rs2305089 TT, rs3816300 TC	0	0			
rs2305089 TT, rs3816300 CC	0	0			

<sup>a</sup>ORs, 95 % CIs, and *P* values were obtained using unconditional regression models comparing sporadic cases to unrelated controls in log-additive models

<sup>b</sup>The three categories (rs2305089 CC/rs3816300TT, rs2305089 CC/rs3816300 TC, and rs2305089 CC/rs3816300 CC) were combined as the reference category because of small numbers

**Table 6**  
**Associations for rs2305089 and rs3810300 by age at diagnosis and tumor site**

From: Characterization of *T* gene sequence variants and germline duplications in familial and sporadic chordoma

SNP	Age at diagnosis among cases									Tumor site					
	<20 years			20–40 years			40 years			Skull-base			Others		
	OR	95 % CI	<i>P</i>	OR	95 % CI	<i>P</i>	OR	95 % CI	<i>P</i>	OR	95 % CI	<i>P</i>	OR	95 % CI	<i>P</i>
rs3816300	7.80	1.88, 32.42	0.005	2.00	0.63, 6.34	0.24	1.42	0.54, 3.70	0.48	2.88	1.24, 6.67	0.014	1.51	0.54, 4.21	0.43
rs2305089	3.61	1.02, 12.79	0.047	3.54	1.64, 7.64	0.0013	3.63	2.05, 6.41	<0.0001	4.07	2.22, 7.46	<0.0001	3.06	1.62, 5.76	0.0005

ORs, 95 % CIs, and *P* values were obtained using unconditional regression models comparing sporadic cases in each age or site stratum to unrelated controls in log-additive models including both rs3816300 and rs2305089 in the model

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript