Chordoma is a rare malignant tumor demonstrating notochordal differentiation. It is dependent on brachyury (TBXT), a hallmark notochordal gene and transcription factor, and shares histologic features and the same anatomic location as the notochord. This study involved a molecular comparison of chordoma and notochord to identify dysregulated cellular pathways. The lack of a molecular reference from appropriate control tissue limits our understanding of chordoma and its relationship to notochord. Therefore, an unbiased comparison of chordoma, human notochord, and an atlas of normal and cancerous tissue was conducted using gene expression profiling to clarify the chordoma/notochord relationship and potentially identify novel drug targets. The study found striking consistency in gene expression profiles between chordoma and notochord, supporting the hypothesis that chordoma develops from notochordal remnants. A 12-gene diagnostic chordoma signature was identified and the TBXT/transforming growth factor beta (TGF-β)/SOX6/SOX9 pathway was hyperactivated in the tumor, suggesting that pathways associated with chondrogenesis were a central driver of chordoma development. Experimental validation in chordoma cells confirmed these findings and emphasized the dependence of chordoma proliferation and survival on TGF-β. The computational and experimental evidence provided the first molecular connection between notochord and chordoma and identified core members of a chordoma regulatory pathway involving TBXT. This pathway provides new therapeutic targets for this unique malignant neoplasm and highlights TGF-β as a prime druggable candidate.
the axial skeleton, and approximately 50% of cases arise in the sacrum. The 5-year survival rate of patients with chordoma is 67%, and the median survival time is 6.3 years. Surgery is the primary treatment option, with radiotherapy used as an adjuvant. Clinical trials using cytotoxic chemotherapy have shown little benefit for the treatment of chordoma; however, initial tests of targeted therapies have shown some promising results in small cohorts of patients. The molecular mechanisms underlying chordoma are poorly understood; therefore, clinical trials based on genetic mechanisms are limited. The current study attempted to expand our understanding of genetically driven cellular pathways to provide new therapeutic targets to explore.

Current hypotheses about chordoma development rely mostly on histologic and immunohistochemical studies that show similarities between chordoma and embryonic human notochord. During higher vertebrate development, most fetal notochordal cells regress during embryogenesis; however, remnants within the adult vertebral disk and bone can occur. Benign notochordal cell tumors are relatively common lesions in the adult vertebral bodies, and their etiology is unclear. In some cases of chordoma, benign notochordal cell tumors have been found within the affected vertebral body, suggesting that benign notochordal cell tumors may be a precursor lesion.

Limited molecular examinations support the connection between chordoma and embryonic notochord vestiges. The transcription factor brachyury (TBXT), encoded by the gene \textit{TBXT} (previously named \textit{T}), has been previously identified as being important in the development of the normal notochord and appears essential for chordoma survival. The \textit{TBXT} gene has been found mutated in some chordomas, and copy number variants have been identified in familial cases. Chordoma cell lines have been shown to depend on \textit{TBXT}, where its decreased expression leads to cell cycle arrest. However, the mechanism that brachyury plays in the context of the disease requires further investigation. \textit{TBXT} is required for most chordoma tumors to survive, yet it does not have a role in most normal adult tissues, making it an ideal drug target. However, brachyury is a transcription factor, a challenging category of molecules. Currently, there are no US Food and Drug Administration–approved drugs targeting a transcription factor, but it is an active area of research. A better understanding of the molecular mechanisms active in chordoma pathogenesis is necessary to provide additional targets for therapeutic development.

Standard molecular profiling approaches require a control sample for comparison against a disease state, and they generally highlight differences but not similarities between the samples. Chordoma is problematic as its normal tissue counterpart is the notochord, which is only present in the human during the early stages of embryogenesis. To overcome these challenges, gene expression profiling of chordoma and human notochord was conducted and compared with many non-chordoma tissues to identify the best control tissue and find key pathways associated with tumor pathogenesis. We identified a chordoma diagnostic gene signature, overactive pathways within chordoma cells, and a core gene interaction network required for chordoma survival and proliferation. Finally, shRNA knockdowns and chemical inhibitor experiments support the relevance of the gene network and highlight the importance of transforming growth factor beta (TGF-\(\beta\)) in chordoma cell survival. We conclude that the pathways related to chondrogenesis are a vital driver of chordoma progression and a promising candidate for therapeutic disruption. Our findings indicate that TGF-\(\beta\) is central to chordoma and is an important druggable target.

Materials and Methods

Ethics Statement

This study used discarded tumor tissues from patients with chordoma as well as notochord cells from discarded human embryonic tissues. The Institutional Review Board at Massachusetts General Hospital (Boston, MA) approved the chordoma study protocol (2009A052093) and the notochord study protocol (2007P-002239). As both protocols used discarded material, the Institutional Review Board waived the need for written informed consent.

Cell lines U-CH1, U-CH2, U-CH12, HEK293, and 293T were purchased from ATCC (Manassas, VA), and CH22 cells were provided by F.J.H. Cells were cultured in a biosafety level 2 environment according to ATCC guidelines.

Tumor Sample Processing

Fresh chordoma specimens from five patients were obtained immediately on surgical resection. The specimens were low in cellularity with abundant extracellular matrix, presenting difficulties in capturing high-quality and adequate amounts of RNA. Tumor samples were cut into 0.5- to 1-cm–diameter pieces. Each sample was either stored in 10 mL of RNAlater (Qiagen, Germantown, MD) or frozen immediately in liquid nitrogen. Homogenization and RNA extraction were performed in the same step. RNA was extracted by placing each tissue sample in 4 mL of Trizol along with 25% (by volume) silicon carbide beads (1 mm; BioSpec, Bartlesville, OK). The tube was placed in a bead beater (BioSpec Mini-Beadbeater-96) and shaken vigorously for 1 to 5 minutes in 30-second intervals. The duration of the homogenization was determined by examining the solution for the disappearance of large masses of tissue. It was not uncommon for tumor samples to contain bone fragments embedded within the tumor mass. In such cases, the tissue was homogenized until only the bone fragments remained intact. The Trizol/tissue homogenate was then transferred to a fresh tube, to which 5 mL of Trizol was added. The tubes were centrifuged for 5 minutes at 12,000
× g to pellet the debris and silicon carbide dust. The homogenate was then transferred to a fresh tube, and 1.6 mL of chloroform was added to the extracted aqueous phase. To precipitate the RNA, 4 mL of isopropanol was immediately added. The pellet was washed in 80% ethanol, dried, and dissolved in 100 μL of water. In cases in which there was still significant visible insoluble material (protein), additional RNA isolation with Trizol was performed on the purified RNA. The RNA solution was then further purified using an RNeasy silica column (Qiagen), according to the manufacturer’s protocol. The RNA quality was assessed with a Bioanalyzer RNA Pico kit (Agilent, Santa Clara, CA), and mRNA was purified from samples that passed the quality control step.

Notochord Laser Microdissection

Human embryonic notochord specimens were obtained from discarded tissues. To obtain sufficient high-quality mRNA, it was necessary to laser microdissect the cellular vestige of the notochord from the fetal spine. Fetal spinal columns were grossly dissected and frozen to −20°C. The specimens were then mounted on a cryostat chuck using M-1 embedding matrix (ThermoFisher, Waltham, MA) and sectioned every 25 μm until the central region of the spine was approximated. At this point, the cryomicrotomy blades were changed to a new blade, and 12-μm thin cryosections were mounted on microscopy slides (Gold Seal Rite-On Micro Slides; VWR, Radnor, PA) and immediately processed for laser capture microdissection.

For the cytoarchitectural visualization of the notochord cells, each tissue section was fixed in 70% ethanol for 30 seconds. The sections were then rinsed with RNase-free distilled water and incubated in HistoGene staining solution (Arcturus; MDS Analytical Technologies, Sunnyvale, CA) for 1 minute, followed by dehydration in increasingly concentrated ethanol (75% to 100%) into xylene and subsequent laser capture microdissection. All incubations and washes were performed at room temperature. Cells representative of the degenerating notochord were clearly visible residing in central lacunae of the nucleus pulposus, and approximately 2000 of these notochord cells from each fetus were captured onto separate polyethylene collecting caps (Macro Cap, Arcturus; MDS Analytical Technologies).

RNA isolation was performed using the PicoPure RNA isolation system (Arcturus; MDS Analytical Technologies). Plastic laser capture microdissection collecting caps were incubated at 42°C for 30 minutes in 20 μL of the extraction buffer containing guanidinium thiocyanate, centrifuged briefly to collect the extracted solution, and then frozen at −80°C. Genomic DNA was removed via RNase-free DNase (Qiagen) digestion on the columns. Total cellular RNA from each column was eluted in a two-step process with 6 μL per step for a total of 12 μL of elution buffer. Isolated RNA was then stored at −80°C until further analysis. The quality of the RNA preparations at various stages was measured using an RNA 6000 Pico chip (Agilent).

RNA Amplification and Sample Preparation for Microarrays and RNA-Seq

Given the generally low RNA yield from the chordoma samples, the mRNA was amplified to produce cDNA in sufficient quantity for both microarray and RNA sequencing (RNA-Seq). Single-primer isothermal amplification (Nugen, San Carlos, CA) was used to linearly amplify the purified RNA. For the initial amplification, an Ovation Pico WTA kit (Nugen) was used. Samples destined for microarray were further processed with the Encore Biotin Module (Tecn, Männedorf, Switzerland). RNA-Seq samples were also processed with the WT-Ovation Exon Module (Nugen) to synthesize the cDNA second strand. High-throughput sequencing libraries for the GAII Illumina platform were constructed using an Illumina kit (Illumina, San Diego, CA). An RNA Amplification System V2 (Ovation) was also used, which includes a complete solution and was not available at the start of the project. In all cases, the manufacturer’s directions were followed.

Microarray Analysis

Primary normal tissue and cell type expression files were obtained from Gene Expression Omnibus18,19 and ArrayExpress.17 All samples obtained for analysis were profiled on the Affymetrix U133plus2 platform (ThermoFisher). The authors used version 14 of the custom transcript definition files provided by Brainarray.18 These files redefine Affymetrix probes by remapping individual probes to the human genome and adjusting them to the most up-to-date annotation. The data files were then normalized using the GCRMA module of the Bioconductor software library version 2.22.0 (https://www.bioconductor.org/packages/release/bioc/html/gcrma.html), and present/absent calls were calculated for each probe using the MAS5 module.19 All probes with no present calls were removed and, from the remaining probes, at least one sample was required to have an expression value larger than log2(100). Tumor expression profiles were obtained from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2109; accession number GSE2109), and cell line profiles were obtained from the Cancer Cell Line Encyclopedia.20

Gene Enrichment Profiles

The gene enrichment algorithm has previously been described21 (source code can be found at http://xavierlab2.mgh.harvard.edu/chordoma/code.html, last accessed November 7, 2022). Briefly, all normal primary cell types and tissues contained replicates, and each such group was compared with each of the other normal tissues using the Limma module of Bioconductor version 3.6.9.22 Limma
uses linear models and Bayes methods to assess differential expression. For each group, a linear model coefficient was obtained, which is a measure of differences between two cell types. The enrichment score for each probe was defined as the sum of all statistically significant (Bonferroni-adjusted $P < 0.05$) coefficients. In this study, the authors extended this method further by calculating the gene enrichment in cell lines and tumors. To avoid a bias due to the large number of tumors and cell lines, each individual sample was compared with all normal cell types. Thus, the enrichment value in each tumor or cell line sample reflects enrichment with respect to a body atlas of all normal cell types.

Assembly of BA0

To assemble BA0, the authors identified and downloaded 1499 publicly available Affymetrix gene expression microarray data files representing 127 normal human cell types and tissues. Next, the authors filtered and normalized the chordoma and notochord gene expression microarrays together with those of the assembled body atlas as a single data set (BA0). To identify enriched genes, the authors compared chordoma samples pairwise with each primary cell type using a linear model. This comparison yielded 126 linear models, each with a coefficient and an associated $P$ value for each gene. The coefficient is a measure of differences between two samples, with larger coefficient values associated with larger differences. In each comparison, the enrichment score for each gene was defined as the sum of all individual coefficients with an adjusted $P < 0.05$. Thus, a gene that was highly expressed in one cell type would produce a high enrichment score, as a result of its 126 large and statistically significant coefficient values. This score is comparable between genes within a sample, enabling ranking to identify the specific genes that define chordoma or notochord.

Assembly of BA1 and Identification of Diagnostic Signature

The authors obtained 2158 tumor profiles [Expression Project for Oncology (ExpO) data set] and 810 cell lines from the public domain and normalized and processed all samples (including BA0) as a single data set containing 4475 microarrays (BA1). The authors derived enrichment scores for each tumor and cell line by comparing it with all normal cell types, as described above, one sample at a time. The authors used two parameters to identify genes that differentiate chordoma from other tumors. First, the authors identified genes enriched within each chordoma sample with a $z$-score $\geq 5$; this stringent cutoff was used to increase specificity. Next, using the normalized gene expression data, the authors calculated a $z$-score for each gene across all the tumor samples (expression $z$-score). The authors deemed genes with a chordoma enrichment $z$-score $\geq 5$, as well as a chordoma expression $z$-score $\geq 5$, to be both specific and highly expressed in chordoma.

RNA-Seq Analysis

Next-generation sequencing data were first mapped to ribosomal RNA (12s, 16s, 18s, 28s, and 5.8s) using Bowtie version 0.12.7.23 All matching reads were discarded. The human genome sequence version HG19 and associated gene models (knownGenes) were obtained from the University of California, Santa Cruz, genome browser.24 Short reads were mapped to the genome using TopHat version 1.4.1.25 Only reads mapping at least partially to defined exons were retained and were aggregated for a single gene locus. The number of reads per sample was normalized to reads per million; and for each gene, the expression value was calculated as the normalized number of reads/observed gene length. The observed length of a gene was defined as the number of bases detected by sequencing. This adjustment was necessary because the Nugen kit applied to the samples was based on nonrandom probes and resulted in inconsistent coverage of genes with typical 3'-end enrichment (Supplemental Figure S1 provides a representative chordoma/notochord gene coverage plot). rRNA-depleted data, as well as count-based expression profiles generated using STAR version 2.7.10a,26 have been uploaded to Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE205458; accession number GSE205458).

Functional Enrichment Analysis

Pathway data were obtained from the following data sources: MetaBase from GeneGo version 6.3 (Clarivate, Chandler, AZ), Ingenuity Pathway Analysis from Ingenuity Systems (Redwood City, CA), Reactome Pathway,27 and Gene Ontology.28 To identify pathways and processes that were enriched in a gene list, a hypergeometric-based enrichment analysis was implemented. The hypergeometric $P$ value was calculated using the R program version 3.5.0 (https://www.R-project.org, last accessed November 7, 2022) with the following command: phyper(x – 1, m, n-m, k and lower. tail = FALSE), where $x$ is the number of genes from the gene list that are members of the pathway, $m$ is the number of genes in the pathway, $n$ is the total number of unique genes in all pathways, and $k$ is the number of genes from the list that were present in at least one pathway. The resulting $P$ value is indicative of the likelihood of enrichment for a specific pathway by chance given the size of the gene list.

This approach typically results in multiple significant pathways because of redundancy. To control for redundancy, this approach was employed iteratively. In each iteration, the most significantly enriched pathway and associated genes from the list were set aside, shortening the
gene list. The process was repeated until no pathway was significant \( P < 0.05 \).

**Protein Interactions and Protein Networks**

Protein interaction data were obtained from National Center for Biotechnology Information GeneRIF, MetaBase from GeneGo version 6.3, Ingenuity Pathway Analysis from Ingenuity Systems, and NetPro from Molecular Connections (Bangalore, India). Nonhuman protein interactions were mapped to human homologs using National Center for Biotechnology Information HomoloGene.\(^29\)

Data were integrated on the basis of PubMed identifier references supporting each interaction. PubMed identifiers associated with >10 interactions were discarded, and from the remaining interactions, at least two PubMed identifiers were required for the interaction to be retained. Pathway-associated interactions were defined as any of the above interactions that were present in one of the GeneGo pathway maps.

Interaction networks were simplified by including only interactions that were present in the manually curated GeneGo pathway maps, producing a high-confidence network. Networks were further subjected to functional Reactome, GeneGo, Gene Ontology, and Ingenuity pathway analyses. To eliminate database redundancy, genes were allowed to be associated with only the single-most significant pathway per database.

All graphical networks were constructed using the OmniGraffle implementation of the GraphViz Dot language (The Omni Group, Seattle, WA).

All raw and processed profiling data and source code are available for download from a supplementary web page (http://xavierlab2.mgh.harvard.edu/chordoma, last accessed November 7, 2022).

**Lentivirus Production**

shRNA clones were provided by the Broad (within either their pLKO.1 or their pLKO_TRC005 vector) along with packaging (pCMV-dR8.91) and envelope (pMD2.G) plasmids (https://www.broadinstitute.org/rna-consortium/rna-consortium-shRNA-library, last accessed November 7, 2022). Lentivirus-expressing shRNA against genes of interest was produced as follows: 293T packaging cells were seeded at 2.2 \( \times 10^5 \) cells/mL in complete Dulbecco’s modified Eagle’s medium in 6-well tissue culture plates. The cells were incubated for about 24 hours in a humidified incubator at 37°C with 5% CO\(_2\). When the cells reached 80% confluency, each well was transfected as follows: a mixture of DNA was prepared with 500 ng of packaging plasmid, 50 ng of envelope plasmid, and 500 ng of pooled shRNA vectors targeting the gene of interest (sequences in Supplemental Table S1). In addition, a reagent mix was prepared containing 3 \( \mu \)L of TransIT-LT1 (Mirus Bio, Madison, WI; catalog number MIR 2305) and 15 \( \mu \)L of Opti-Mem (ThermoFisher; 31985070). Both solutions were mixed after a 5-minute incubation at room temperature, and then the reagent mix was added, dropwise, to the DNA mix. This final mix was incubated at room temperature for 30 minutes and then transferred, dropwise, to the cells. The cells were incubated for 18 hours; then, the medium was removed and replaced with a high-growth medium (Dulbecco’s modified Eagle’s medium with 20% serum). The cells were incubated for a further 24 hours, at which point the medium was harvested, stored at 4°C, and replaced with fresh high-growth medium. After another 24 hours, the medium was harvested and pooled with that from the first harvest. The cell plate was then discarded. The pooled medium was spun at 500 \( \times \) g for 5 minutes to pellet any carryover packaging cells, and then the supernatant was aliquoted into 1-mL aliquots and stored at \(-80\)°C. Lenti-X GoStix (Takara, San Jose, CA) was used to verify that the viral concentrations were \( >5 \times 10^5 \) infectious units per mL.

**Gene Expression Knockdown**

The cells of interest were plated in T25 tissue culture flasks and grown to 80% confluency using the supplier’s recommended protocol for each cell line. For each flask, the following mixture was used for the transduction: 3 mL of media, 8 \( \mu \)L of polybrene (4 mg/mL stock), and a 1-mL virus aliquot. The transduction mix was added to the cells and incubated at 37°C. After 24 hours, the medium was changed. After a further 24 hours, the medium was changed to complete medium with 0.5 \( \mu \)g/mL puromycin to start the selection process. After 3 days, the cells were switched to 1 \( \mu \)g/mL puromycin, and split 1:2 if they were approaching confluency. The cells were monitored daily, and split at 80% confluency; at each split, the puromycin concentration was increased by 0.5 \( \mu \)g/mL until the cells were stable at 2 \( \mu \)g/mL puromycin. The cells were eventually transferred to a T75 flask and subsequently subjected to RNA isolation, bromodeoxyuridine (BrdU) treatment, and cytometric profiling.

**BrdU Proliferation Assay**

A BD Biosciences APC BrdU Flow kit (BD Biosciences, Franklin Lakes, NJ; catalog number 552598) was used to measure cell proliferation. BrdU was diluted in culture medium to a 10 \( \mu \)mol/L final concentration and added to 50% confluent cells. The cells were incubated at 37°C/5% CO\(_2\) until control cells reached near confluency (up to 1 week for chordoma cell lines), at which point the cells were lifted off the plate using TrypLE Express (ThermoFisher; 12605093). The cells were then fixed and permeabilized, then stained with anti-BrdU antibody in accordance with the manufacturer’s recommendations, and analyzed on a Sony SH800 cell sorter (Sony, Tokyo, Japan).
Annexin and Apoptosis Assay

Cells were lifted off tissue culture plates with TrypLE Express (ThermoFisher; 12605093) and split into two aliquots. One aliquot was stained with a Molecular Probes Vybrant FAM Poly Caspases Assay Kit (ThermoFisher; V35117) followed by Hoechst 33342 (ThermoFisher; 62249), according to manufacturer’s directions. The second aliquot of cells was stained for annexin V as follows: cells were resuspended in 150 μL of 1× Annexin V binding buffer (BD Biosciences; 51-66121E); 5 μL of fluorescein isothiocyanate–annexin V (BioLegend, San Diego, CA; 640905) and 1.2 μL of propidium iodide (BioLegend; 79997) were added; and cells were incubated at room temperature for 20 minutes. Both aliquots of cells (stained for annexin and caspases) were run on a NovoCyte flow cytometer (Agilent) for quantitation and analysis.

Western Blot Analysis

To isolate protein for Western blot analysis, cell lines were grown to 80% confluency. Cells were washed with phosphate-buffered saline (PBS), then lifted with TrypLE Express (ThermoFisher; 12605093), spun down at 500 × g for 5 minutes, washed in 13 mL PBS, spun down again, resuspended in 5 mL PBS, and then a 10-μL aliquot was removed and stained with acridine orange/propidium iodide dye (Logos Biosystems, Anyang, South Korea; F23011) and counted with a Luna FX7 (Logos Biosystems) in fluorescent mode. Cells were spun down once more, supernatant was aspirated, and the pellets were stored at −80°C. Lysis buffer was formulated as follows: PBS; 0.4% Triton X-100 (Sigma, St. Louis, MO; X100-100 ML); 1% Benzonase (Sigma; E1014-5KU); and manufacturer-recommended concentration of PhosSTOP (Roche, Basel, Switzerland; 04 906 837 001) and EDTA-free EASYpack (Roche; 04 693 159 001). Cell pellets were thawed and resuspended in 100 μL lysis buffer per 1 million cells, then incubated at room temperature for 15 minutes. Lysis mixture was spun down at 4°C for 10 minutes at maximum centrifuge speed, and then supernatant was removed to a new tube.

Electrophoresis was performed using precast NuPAGE gels (ThermoFisher; NP0322BOX). Protein lysate was mixed with concentrated PAGE buffer (with tris (2-carboxyethyl) phosphine) to a final 1× PAGE buffer solution, then heated to 95°C for 3 minutes. The 1× SDS running buffer was prepared from a 20× stock (ThermoFisher; NP0002). Gel running box was assembled as per manufacturer’s directions, and 20 μL of sample was loaded in each well. A total of 5 μL of ladder (Bio-Rad, Hercules, CA; 1610374) was loaded in end lanes. Electrophoresis was performed at 140 V for 90 minutes.

Protein was transferred to a low fluorescence polyvinylidene difluoride membrane using a Trans-Blot Turbo (Bio-Rad), according to manufacturer’s directions (Mixed MW MIDID Run protocol was used, at 2.5 A and 25 V for 7 minutes). Membranes were placed into blocking solution (5% milk in PBS) and incubated at room temperature for 1 hour on a gently shaking platform. Primary antibody solutions were made with 1% milk in PBS with antibody diluted at these ratios: actin (Cell Signaling Technology, Danvers, MA; 8H10D10) 1:1000; ACAN (ThermoFisher; MA3-16888) 1:750; TBXT (ThermoFisher; MA5-17185) 1:1000; WWP2 (ThermoFisher; A302-935A) 1:500; KCNK2 (ThermoFisher; PA1-16981) 1:1000; XYL (ThermoFisher; 26541-1-AP) 1:500; ENPPI (ThermoFisher; BS-4913R) 1:250; and SPDYE1 (ThermoFisher; PA5-62470) 1:125. Blocking solution was removed, and membranes were placed into primary antibody solution and left overnight at 4°C on a gently shaking platform.

The next day, membranes were washed using PBS with 0.1% Tween by rinsing the membrane twice, then performing three washes for 5 minutes each on a gently shaking platform at room temperature. Secondary antibody solutions were made with PBS with 0.1% Tween and these dilutions of antibody: goat anti-mouse (Li-Cor, Lincoln, NE; 926-32210) 1:5000; donkey anti-rabbit (Li-Cor; 926-68073) 1:5000; goat anti-rabbit (Li-Cor; 926-32211) 1:5000; and donkey anti-mouse (Li-Cor; 926-68072) 1:10,000. Membranes were placed into 10 mL of secondary staining mix, and then wrapped in aluminum foil to protect from light. Membranes were incubated on a gently shaking platform at room temperature for 1 hour. Secondary antibody solution was washed off by rinsing the membrane twice in PBS with 0.1% Tween, and then performing three washes for 5 minutes each on a gently shaking platform at room temperature. Membranes were imaged using an Odyssey DLx (Li-Cor).

Quantitative PCR and Empirical Network Construction

For laboratory-grown cells, RNA was isolated using a Qiagen RNeasy micro kit (Qiagen; catalog number 74004). The RNA was reverse transcribed into cDNA using an iScript cDNA synthesis kit (Bio-Rad; catalog number 1708891). This cDNA was diluted 1:5.5 with nuclease-free water (final volume, 110 μL), and real-time quantitative PCR was then performed with 5 μL of diluted cDNA, 0.3 μmol/L forward/reverse primer, and 10 μL of SybrGreen Supermix (Bio-Rad; catalog number 1725121) in a 20-μL reaction. The real-time quantitative PCR was thermocycled on a Bio-Rad CFX384 Touch with the following program: 95°C for 3 minutes, followed by 46 cycles of 95°C for 10 seconds and 60°C for 30 seconds. The fluorescence measurements were performed every cycle, and the threshold cycle (C_T) measurements were calculated with the Bio-Rad software (CFX Maestro 2.3 version 5.3.022.1030). All downstream calculations were performed on the average of three technical replicates. Measurements for each sample were normalized to those of glyceraldehyde-3-phosphate dehydrogenase; then, the fold change was calculated on the basis of cycle-count differences between the test and the sh-GFP control samples [with shRNA targeted against green
fluorescent protein (GFP). The empirical network was generated using the R igraph package version 1.2.4. In the network graph, nodes represent the target shRNA, and edges (arrows) point to the regulated gene. Only relationships with a fold change >1.5 or <1/1.5 were kept.

Drug Treatment Viability Assays

HEK293 or U-CH2 cells were plated into 96-well plates at a density of 15,000 or 10,000 cells per well, respectively. Wells containing U-CH2 were previously coated with collagen and fibronectin. The plates used were black tissue culture plates with clear bottoms (Corning, Corning, NY; 3904). The reagents from the RealTime-Glo assay kit (Promega, Madison, WI) were added at the time of plating, according to the manufacturer’s recommendations. Luciferase measurements were made using a Microbeta2 (Perkin Elmer, Waltham, MA) at 1 hour predosing, and the following post-dose time points: 1, 3, 19.5, 24, 27.5, 44, and 48 hours. The drug LDN-212854 (Selleck Chemicals, Houston, TX) was reconstituted in dimethyl sulfoxide at 100 mmol/L and dosed at concentrations from 1 to 100 μmol/L 24 hours after cell plating. The luciferase values for each well were normalized to the predose measurement, and dose-response curves were calculated using the DRC package version 3.0-1. The normalized luciferase values for each time point/cell line combination were fitted with a three-parameter log-logistic function using the LL.3() function call, then EC₅₀ values were calculated using the ED() function, and 95% CIs were calculated using the delta method. Relative potency was calculated using the EDcomp() function, and represents the EC₅₀ of HEK293 divided by the EC₅₀ of U-CH2. CIs were again calculated using the delta method. EC₅₀ values for time points <19.5 hours were excluded from bar plots comparing U-CH2 and HEK293 because these early time points did not exhibit dose-dependent killing of HEK293 cells, and an accurate dose-response curve could not be fitted. The results were plotted using the ggplot2 R package version 3.3.6 (https://ggplot2.tidyverse.org, last accessed November 7, 2022).

Results

Chordoma and Notochord Gene Expression Profiles Are Closest Compared with All Other Examined Profiles

A method was developed to extract high-quality RNA from chordomas and five non-clival chordoma tumors were profiled using RNA-Seq. Three of the tumor samples were also profiled using Affymetrix gene expression microarrays. Notochordal cells were obtained through the laser microdissection of three human fetal notochord tissue samples, and the cells were profiled using both RNA-Seq and Affymetrix microarrays (Gene Expression Omnibus study, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE205458; accession number GSE205458).

Comparing chordoma and notochord gene expression identified differences between the two tissues, but not similarities that could be consequential for tumor progression. For instance, TBXT has been previously identified as critical for chordoma proliferation, but is highly expressed in both chordoma and notochord tissues; from a differential expression standpoint, it is thus indistinguishable from other housekeeping genes that would also be expressed at similar levels, and would not be identified as a driver gene. Therefore, an unbiased method was developed to compare chordoma and notochord tissues, while also identifying similarities with other tissues that might enhance the mechanistic knowledge surrounding chordoma tumorigenesis. This gene enrichment method was used to compare the expression levels of each gene against those in a body atlas data set of normal primary tissues. This body atlas, BA0, was assembled from 1499 publicly available Affymetrix gene expression microarray data files that span 127 normal human cell types/tissues.

Within BA0, comparing the 20 genes most enriched in chordoma with the top 20 genes for notochord revealed five genes, including COL2A1 (collagen 2A1), TBXT (brachyury), CA3 (carbonic anhydrase 3), ACAN (aggrecan), and CD109, that were enriched in both tissue types (Figure 1A and Supplemental Figure S2, A and B). In addition, by exploring the scores of individual genes enriched in chordoma, TBXT enrichment was identified as strictly limited to chordoma and notochord, whereas COL2A1, ACAN, and KCNK2 were also highly enriched in fetal cartilage, chondrocytes, and notochord (Supplemental Figure S2C). Finally, all genes with a z-score ≥4 within chordoma or notochord were extracted (Figure 1A). Out of these high z-score genes, 28 were enriched in both tissues, 45 genes were enriched only in chordoma, and 57 genes were enriched only in notochord (Supplemental Table S2). To explore the significance of this overlap, genes were randomly selected from the chordoma and notochord lists. After 100,000 iterations, the greatest overlap observed was only six genes.

Chordoma Is Characterized by a 12-Gene Diagnostic Signature

Having identified gene expression signatures specific to notochord and chordoma, the next question addressed was whether a gene expression signature could be used to diagnose chordoma. To generate this gene expression signature, chordoma-enriched genes were compared with those enriched in other tumors to exclude general cancer-associated genes. To perform this comparison, another body atlas, BA1, was generated. BA1 included both normal and tumor samples.

Consequently, a panel of 12 genes was identified. These 12 genes were more highly expressed in all chordoma samples than in normal tissues, tumors, and cell lines, and thus formed a chordoma signature. Next, the mean
The enrichment score of the 12 genes was used as a chordoma signature score to rank tumors and cell lines. Of all the unique samples in the data set (BA0 + ExpO + cell lines), notochord, the chordoma-derived cell line UCH1, fetal and bone cartilage, and chondrocytes scored the highest (Figure 1B).

To validate the chordoma diagnostic signature, a second independent body atlas (BA2) was assembled. BA2 included the only chordoma samples available in the public domain at the time and 557 normal samples representing 126 cell types and tissues. All BA2 samples were profiled on the Affymetrix U133A platform. This data set had a
different cell type composition compared with BA1, with the most obvious difference being the larger number of different immune cell types (approximately 30% of the panel). BA2 dataset was normalized and processed for enrichment as described for BA1. Of the 12 genes identified as the chordoma signature, 8 were available on the U133A platform. Chordoma was the top-scoring group in the BA2 data set, enriched for all eight genes (Figure 1C). For further validation, the expression of a selection of these genes was checked in chordoma cell lines using Western blot analysis (Supplemental Figure S2D).

Functional Analysis of Chordoma Genes

Having defined a core chordoma gene expression signature, the molecular mechanism underlying chordoma was pursued. First, the chordoma signature genes with shared functionality were identified. A protein interaction network was generated using the 12 signature genes, allowing for a single intermediate chordoma-expressed gene to serve as a link between each of the signature genes. The resulting network of 7 signature proteins and 13 intermediate proteins is shown in Figure 2A. The network appears centered around TGF-β, SOX9, and the extracellular matrix proteins COL2A1 and ACAN, which are direct transcriptional targets of SOX9.

In addition to analyzing the chordoma signature genes, genes that had enriched expression in chordoma compared with normal cells and tissues were extracted. Initially, seven chordoma gene lists (genes with a z-score >4) were made from the three chordoma samples and four publicly available samples. Each gene list was analyzed individually for pathway enrichment using GeneGo pathway maps, and cartilage development/chondrogenesis and cell adhesion ranked as the two most significant pathways (Supplemental Figure S3A). To assess the significance of this finding, the same procedure was repeated for each of the 2158 tumor samples in the ExpO data set. This analysis yielded a similar list of pathways as identified in the chordoma samples (Supplemental Figure S3B). Further exploration of this resemblance revealed that although similar pathways were enriched in chordoma and non-chordoma tumor samples, the genes within the pathways differed. For instance, within the cartilage development pathway, ACAN, COL2A1, SOX9, SOX6, CTGF, FNI, and others were enriched in chordoma tumors but rarely enriched in other tumor types (Figure 2B). In contrast, collagen type I and III genes, enriched in some chordomas, were also frequently enriched in other tumors. In summary, chordoma signature genes participate in pathways common to many other types of cancer, yet frequently represent unique portions of the pathways.

Although the data set includes only a small number of chordoma tumors, it might indicate additional pathways that might play a role in chordoma and identify potential therapeutic targets for further validation. To this end, the interaction network analysis above was repeated to include all chordoma genes that were enriched in at least two of the three chordoma samples and expressed on average >20 reads per million in the RNA-Seq data; 62 genes met this criteria (Supplemental Table S3). Next, a protein interaction network was generated that included the proteins encoded by the 62 chordoma-enriched genes and the chordoma-expressed genes (Supplemental Table S4) that serve as a link between at least two nodes among the 62 genes. The resulting network was dense and revealed an interconnected TGF-β–SOX9 hub, an MDM2-p53 hub, and a RELA, β-catenin, and HIF-1 hub (Supplemental Figure S4A). The network was simplified by including only high-confidence interactions (Supplemental Figure S4B) and then enhanced with pathway analysis. The resulting analysis is illustrated as a network in Supplemental Figure S5, where each edge represents pathway membership, as determined through the analysis. The pathways and processes identified using this method included cartilage development, integrin binding, Wnt signaling, NF-κB activation, and angiogenesis. Taken together, all the functional analyses suggest the presence of a forward-feedback loop where the TGF-β pathway is activated by extracellular matrix via integrin signaling, leading to SOX9 activation and additional extracellular matrix generation. This integrative model is illustrated in Supplemental Figure S6.

Brachyury Core Network Members Are Highly Interconnected

The computational analysis identified a core chordoma network that encompasses members with published interactions, so a central hub potentially presenting a therapeutic target was sought. The expression of SOX5, SOX6, SOX9, TBXT, and TGFB1 was knocked down via shRNA and the effect of the knockdown on the expression of other genes of interest was examined using real-time quantitative PCR. The expression results were normalized to that of GAPDH; then, the fold induction was calculated relative to that of a control sample expressing shRNA against green fluorescent protein. The results are shown in Figure 2C. Notably, TGFB1 expression increased significantly as a result of the knockdown of any other network member (aside from TGFB1 itself). To further illustrate the interconnectedness, a network diagram representing the relationships empirically identified via real-time quantitative PCR was generated; each edge (arrow) represents a regulatory effect from the sending node (the one probed via shRNA) to the receiving node (gene with altered expression level) (Figure 2D). These experiments supported TGF-β as a central hub, exhibiting both regulation of and by other genes in the network.

Core Chordoma Network Members Are Required for Chordoma Cell Line Survival

Having established a core interconnected chordoma network, the reliance of chordoma cells on the identified
network genes was probed. The expression of SOX6, SOX9, TGFB1, and TBXT was knocked down using lentiviral shRNA constructs and cells were placed under selection for those stably expressing the relevant shRNAs. Subsequently, cell proliferation was examined by BrdU incorporation in newly synthesized DNA, as measured by flow cytometry with an anti-BrdU antibody. SOX6, SOX9, and TBXT knockdown resulted in severely decreased proliferation.
whereas TGFB1 knockdown resulted in only a modest decrease (Figure 3). However, the BrdU incorporation assay did not demonstrate the extent of visually observed cell death caused by the gene knockdowns (especially in the case of TGFB1); it only measured the proliferation in the surviving cells. To address this gap, media were collected from TGFB1 and TBXT knockdown cells and was quantitatively measured using annexin V binding and propidium iodide staining. Most floating cells were dead, regardless of which gene was knocked down. However, TGFB1 knockdown resulted in a far greater number of floating dead cells than TBXT knockdown. Furthermore, the surviving cells

**Figure 3** TGFB1 knockdown results in apoptotic phenotype. **A:** Phase-contrast microscopy images demonstrate the hyperapoptotic phenotype after TGFB1 knockdown. Cells were imaged in a time course after selection for integration of lentiviral vector. Day 1 represents the first day after antibiotic selection was finished. By day 9, the control knockdown cells expressing shRNA against green fluorescent protein (sh-GFP) exhibit proliferation consistent with U-CH2 cells, whereas brachyury (TBXT) knockdown cells have not expanded much. In contrast, TGFB1 knockdown results in extensive cell death, with remaining cells lacking the chordoma characteristic physaliferous phenotype and presenting as extremely unhealthy. **B–D:** Cytometric studies characterize the proportion of GFP knockdown cells (sh-GFP; **B**), TBXT knockdown cells (sh-TBXT; **C**), or TGFB1 knockdown cells (sh-TGFB1; **D**) that are in the early stage of apoptosis, late stage of apoptosis, or actively proliferating. **E:** Bar plot summarizing the cytometric proportions characterized in **B** through **D**. In panels **B**–**D**, the x axes of the histograms represent the reported cytometric pulse height (H), while the y axes represent the number of events. In the scatter plots, the x axes represent intensity of staining for bromodeoxyuridine (BrdU), while the y axes represent forward scatter pulse area. FITC Annexin V-H indicates the staining intensity for annexin five with the fluorescein isothiocyanate (FITC) dye, and FAM-VAD-FMDK-H represents the staining intensity for the early apoptosis caspase assay.
from shRNA knockdowns were assayed for early and late apoptosis. The cells were stained with annexin V and propidium iodide to measure late apoptosis, and TGFB1 knockdown yielded higher proportions of apoptotic cells than either TBXT knockdown or the negative control sample (shRNA against green fluorescent protein) (Figure 3). A caspase assay was used to measure early apoptosis, and TGFB1 knockdown cells had the highest proportion of dying cells (Figure 3). To verify whether the TGFB1 knockdown-induced apoptosis is a chordoma-specific phenomenon, the same annexin and early apoptosis assay was performed in HEK293 cells expressing shRNAs against either TGFB1 or green fluorescent protein. HEK293 cells did not exhibit any increased apoptosis after TGFB1 knockdown compared with the green fluorescent protein negative control (Supplemental Figure S7). In summary, all members of the core chordoma network are necessary for cell survival, with some genes exhibiting a dichotomy between proliferative and immediate apoptotic effects. TGFB1 knockdown results in moderately decreased proliferation and a strong induction of apoptosis specifically in chordoma cells, whereas TBXT knockdown results in severely reduced proliferation and mildly increased cell death during the experimental time frame.

**TGF-β Pathway Inhibition Limits the Growth of Chordoma Cell Lines**

Computational analysis and experimental data support the hypothesis that TGF-β acts as a hub for chordoma growth. Several chemical inhibitors of the TGF-β pathway are available, and LDN-212854 (which acts primarily through inhibition of ALK2 and ALK1) was used to interrogate the effect of TGF-β pathway inhibition in chordoma cell lines. HEK293 (control) and U-CH2 chordoma cells were treated with varying concentrations of LDN-212854 (Selleck Chemicals) and viability was measured using a RealTime-Glo viability assay kit (Promega). LDN-212854 had about a twofold higher relative potency in U-CH2 cells over HEK293, with U-CH2 cells exhibiting strong growth inhibition at lower drug concentrations than HEK293 (Figure 4). U-CH2 cells also exhibited characteristic dose-dependent killing at earlier time points than HEK293 cells (Figure 4). A more comprehensive survey of all TGF-β pathway inhibitors might identify more potent and practical molecules. These findings suggest that TGF-β pathway inhibition is a viable chordoma treatment strategy.

**Discussion**

This study utilized molecular profiling approaches to compare human notochord and chordoma to provide insight into alternative and potentially synergistic therapeutic targets in addition to brachyury, an established critical transcription factor.11 The tumors profiled were all non-clival in origin. Although both clival and sacral tumors are known to express and be dependent on brachyury, clival and sacral tumors have differences in other pathway dependencies.33,34 The results were generated from sacral tumors, and sacral cell lines were used to perform validation experiments; thus, the current results are specific to tumors of a sacral origin, which is the predominant anatomic location for chordomas.35 Human notochord had the closest resemblance to chordoma among the hundreds of tissues and cancers surveyed, providing strong molecular support to the close relationship between notochord and chordoma previously observed via histology and immunohistochemistry. In addition, a TGF-β/SOX6/SOX9/TBXT pathway was identified that appeared to play a key role in chordoma development and survival, with TGF-β serving as a critical member. TGF-β presents an attractive therapeutic target, as indicated by both genetic (shRNA knockdown) and chemical (TGF-β pathway inhibitor) methods.

Two challenging issues were addressed herein that have previously made the molecular profiling and analysis of chordoma difficult. First, the extraction of high-quality RNA from chordoma specimens was addressed, which is notoriously difficult because of the low cellularity and large quantity and composition of the extracellular matrix.36,37 The second issue addressed problems in identifying both similarities and differences in gene expression between chordoma and notochord. A method was developed that compared gene expression in individual samples with that in a body atlas of normal tissues and tumors, a method previously applied to the identification of cell- and tissue-specific genes.21 This method allowed identification of genes that define chordoma and notochord independently, eliminating the need to compare the two directly and rely on assumptions about proper control tissues. The composition of the body atlas might bias the results of this analysis; however, observations made in one body atlas could be validated in a second body atlas with a different sample composition and a different platform. This approach can be extended to other cancer studies and types of data, and it will be useful for cases where control tissues are unavailable, unknown, or difficult to obtain. Examples of such studies include nervous system tumors, sarcomas, other heterogeneous tumors, and tumors with an extreme mutational burden.

The comparison of chordoma and notochord in the context of the body atlas showed high molecular resemblance. The chordoma gene-expression signature was more similar to that of notochord than that of any other cancer. In addition, although some genes were different between chordoma and notochord, the diagnostic signature separating chordoma from thousands of tumors and normal cells did not distinguish chordoma from notochord (Figure 1, A and B). Of the top 10 chordoma-enriched genes (Figure 1A and Supplemental Figure S2A), 9, including TBXT, are also enriched in notochord. These data provide novel and strong
molecular evidence suggesting that chordoma is strongly related to notochord.

To date, most chordoma studies have primarily focused on brachyury, which is used clinically as a diagnostic marker and has been implicated in chordoma by several genetic studies.12,13 Functionally, the role of brachyury in cancer is unknown, although some studies have suggested that it contributes to a mesenchymal phenotype.38 The current data confirm that brachyury is a unique marker of chordoma and notochord, and is rarely expressed in other normal or cancer tissues. Knockdown experiments further confirm its essential role in underlying chordoma proliferation, although it is unclear whether brachyury’s primary role is in tumor onset or maintenance/growth. In addition, the current experiments suggest that brachyury participates in the TGF-β pathway. These observations suggest a potential role for brachyury in chordoma that needs further study.

The analysis of the molecular mechanism driving chordoma in the current study further indicated that the process of chondrogenesis was the most significant and highly enriched cellular pathway. Although this pathway is common to many cancer types,41–43 the chordoma-relevant chondrogenesis genes were not typically expressed in other cancers. Furthermore, the protein interaction network analysis indicated that the extracellular matrix proteins COL2A1 and ACAN, integrins, TGF-β, and SOX9 were all key players in the process of chondrogenesis.

Integrins are adhesion molecules that play an important role in the initiation and progression of cancer.46 Previous studies have shown that COL2A1 can bind and activate integrins on chondrocytes.47,48 Furthermore, integrins are modulators of the TGF-β pathway,49,50 which is well characterized in cancer as an inducer of epithelial-to-mesenchymal transformation.51,52 TGF-β treatment of mesenchymal stem cells can also induce a chondrogenic phenotype that induces COL2A1 and ACAN.53 A link between TGF-β and chordoma has been suggested by a study of copy number alteration in 21 chordoma specimens, in which a deletion of 1p36 was identified in 90% of chordoma samples.54 The tumor suppressor implicated in this region is RUNX3, a known repressor of TGF-β. In addition, a significant number of tumors had genomic amplifications of the TGFB1 gene, a key regulator of the pathway.54 Later studies have also found TGF-β signaling and TGFB1 expression56 associated with tumor progression, whereas down-regulation of TGFB3 has been reported to cause chordomagenesis.57 The TGF-β pathway plays a well-established role in chondrogenesis through the activation of
SOX9,58–60 a transcription factor that directly induces COL2A1 and ACAN expression.61–64 Interestingly, pathogenic germline variants in COL2A1 have recently been associated with chordoma in a cohort of patients with European or Chinese ancestry.65 Additional pathways that have been observed enriched in chordoma, such as NF-kB, can also serve to enhance chondrogenesis.56,67

Regardless of the role brachyury plays in tumorigenesis or chordoma risk, as a transcription factor it remains a challenging drug target for cancer therapy.14 The current analysis presents a new list of potential therapeutic targets, the TGF-β pathway. In vitro shRNA knockdown experiments revealed the core chordoma TGF-β/SOX6/SOX9/TBXT network to be highly interdependent, with TGF-β potentially playing a central role. Further flow cytometry analysis indicated a dichotomous relationship between proliferation and apoptosis associated with knockdown of pathway members. Specifically, during the experimental timeline of about 1 week, TBXT knockdown was associated with vastly decreased proliferation and moderately increased apoptosis, whereas TGFBI knockdown was associated with moderately decreased proliferation and vastly increased apoptosis. Although TBXT knockdown led to cell death after a significant delay, TGFBI knockdown resulted in fewer surviving chordoma cells than TBXT knockdown, presenting a therapeutic alternative or supplement to targeting TBXT.

Experiments with small-molecule inhibitors further pinpoint the hyperactivated areas within the TGF-β pathway. The tests with chemical inhibitors of various members of the TGF-β pathway identified ALK1 and ALK2 inhibitors as especially potent in chordoma cells (Figure 4), with inhibitors of other TGF-β pathway receptors showing little differential effect in chordoma cells versus HEK293 controls (data not shown). Better therapeutic molecules should be further investigated within the TGF-β pathway. More importantly, TGF-β inhibition has been proposed as a therapeutic avenue in many other cancers, making the drug development tractable for a rare tumor, such as chordoma. In support of this, there are already multiple clinical trials underway for various inhibitors of the TGF-β pathway.68

The findings of this study provide a new window into the molecular mechanism of chordoma, which, until now, was primarily centered on brachyury. Current treatment options for patients with chordoma are severely limited, and additional therapeutic targets driven by unbiased data analysis are desperately needed. This analysis opens a new avenue for chordoma research and therapy, implicating the TGF-β pathway as a promising target.

Acknowledgments

We thank Brian Seed and Ramnik J. Xavier for input, support, and discussions about this project; and the Chordoma Foundation for support.

Supplemental Data

Supplemental material for this article can be found at http://doi.org/10.1016/j.ajpath.2023.01.014.

References


64. Han Y, Lefebvre V: L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. Mol Cell Biol 2008, 28: 4999–5013