Title

Comparative RNA-seq analysis aids in diagnosis of a rare pediatric tumor

Journal

Molecular Case Studies, 5(5)

ISSN

2373-2865

Authors

Sanders, Lauren M

Rangaswami, Arun

Bjork, Isabel

et al.

Publication Date

2019-10-01

DOI

10.1101/mcs.a004317

Peer reviewed

COLD SPRING HARBOR
Molecular Case Studies

RESEARCH REPORT

Comparative RNA-seq analysis aids in diagnosis of a rare pediatric tumor

Lauren M. Sanders,[1] Arun Rangaswami,[2] Isabel Bjork,[1] Du Linh Lam,[1] Holly C. Beale,[3] Ellen Towle Kephart,[1] Ann Durbin,[1] Katrina Learned,[1] Rob Currie,[1] A. Geoffrey Lyle,[3] Jacob Pfeil,[1] Avanthi Tayi Shah,[4] Alex G. Lee,[4] Stanley G. Leung,[4] Inge H. Behroozfard,[4] Marcus R. Breese,[4] Jennifer Peralez,[2] Florette K. Hazard,[2] Norman Lacayo,[2] Sheri L. Spunt,[2] David Haussler,[1,5] Sofie R. Salama,[1,5] E. Alejandro Sweet-Cordero,[4] and Olena M. Vaske[3]

[1]Department of Biomolecular Engineering, UC Santa Cruz Genomics Institute, Santa Cruz, California 95064, USA; [2]Stanford University School of Medicine and Stanford Cancer Institute, Stanford, California 94305, USA; [3]Department of Molecular, Cell and Developmental Biology, UC Santa Cruz Genomics Institute, Santa Cruz, California 95064, USA; [4]Department of Pediatrics, Division of Hematology and Oncology, University of California San Francisco, San Francisco, California 94143, USA; [5]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA

Abstract Gliomatosis peritonei is a rare pathologic finding that is associated with ovarian teratomas and malignant mixed germ cell tumors. The occurrence of gliomatosis as a mature glial implant can impart an improved prognosis to patients with immature ovarian teratoma, making prompt and accurate diagnosis important. We describe a case of recurrent immature teratoma in a 10-yr-old female patient, in which comparative analysis of the RNA sequencing gene expression data from the patient's tumor was used effectively to aid in the diagnosis of gliomatosis peritonei.

[Supplemental material is available for this article.]

INTRODUCTION

Immature ovarian teratomas are malignant tumors of germ cell origin (Gheorghisan-Galateanu et al. 2013). Teratomas are the most common germ cell tumor, but, in rare cases, immature teratoma can occur with gliomatosis peritonei, which is characterized by mature glial tissue in the peritoneum (Liang et al. 2015). The presence of mature glial tissue implants can indicate a favorable prognosis in patients with immature ovarian teratoma (Marwah et al. 2016). However, all lesions must be sampled to confirm mature histological status, and full excision is important. Additionally, recurrence potential is high, requiring careful follow-up and monitoring.

Gliomatosis peritonei can be difficult to identify through histopathological analysis alone. Recent studies have shown that molecular analysis can aid cancer type classification (Cancer Genome Atlas Research Network 2015). In many cases, DNA sequencing and variant identification can help subtype cancers by grade and outcome. However, the paucity of recurrent DNA mutations in rare pediatric cancers can make variant-based disease classification difficult. RNA sequencing (RNA-seq) of tumor gene expression can provide additional classification information, through a comparative analysis of RNA-seq from the pediatric tumor and with a compendium of RNA-seq data from known cancer types (Newton et al. 2018). In this study, we describe the use of comparative RNA-seq analysis in a case of relapsed

Corresponding author: olena@ucsc.edu

Ontology term: neoplasm of the nervous system

Published by Cold Spring Harbor Laboratory Press

doi:10.1101/mcs.a004317

pediatric immature teratoma cooccurring with gliomatosis peritonei with no informative DNA mutations.

RESULTS

Clinical Presentation and Family History

A 10-yr-old female patient was diagnosed with immature teratoma, relapsed to the pericardium and diaphragm. Treatment history included laparotomy with resection of ovarian and fallopian tube mass, video-assisted thoracoscopic resection of lung nodule, and resection of diaphragmatic and pericardial lesions. The pericardial lesion was submitted for RNA sequencing. Foundation Medicine DNA testing of the diaphragmatic lesion identified only one variant: MLL3 p.C310S. This mutation was not informative for diagnosis or subtyping. Whole-genome sequencing (WGS) detected three somatic coding variants in genes *FCGR1A*, *ANKRD36C*, and *HLA-DRB1* at hg38 Chr 1:g.[149790230C>T], Chr 2:g.[95855406C>G], and Chr 6:g.[32584172C>G] (Table 1). Interestingly, *ANKRD36* has been previously characterized by TCGA as a significantly mutated gene in adult glioblastoma (Brennan et al. 2013).

Histologic sections of the pericardium (Fig. 1A) showed fibroadipose tissue and mature glial cells, with lymphovascular invasion. Histologic sections of the right diaphragm (Fig. 1B) showed extensive involvement by teratoma, with no malignant elements present.

Genomic Analyses

The patient was enrolled in the "Clinical Implementation of Genomic Analysis in Pediatric Malignancies" study at Stanford University, and through this trial, her tumor RNA-seq data set was analyzed. This analysis uses an *N*-of-1 analysis approach, which compares an individual pediatric tumor to a cancer compendium of uniformly processed RNA-seq data from 11,456 other tumors (https://treehousegenomics.ucsc.edu/public-data/). This approach aids in the molecular classification of the pediatric tumor through the identification of most similar tumors in the Treehouse cancer compendium.

We calculated pairwise Spearman correlation scores between the patient's tumor and all tumors in the Treehouse cancer compendium. For 232 samples, the pairwise correlation scores with the focus sample exceeded the 95th percentile correlation score in the cancer compendium (0.875); 228 of these (98%) were glioma or glioblastoma multiforme samples, and the remaining four were various brain tumors (Supplemental Fig. S1). The top 6 most correlated tumors are shown in Table 2. In addition, glioma samples had a significantly higher correlation to the patient's tumor than to other tumor types in the cancer compendium (Supplemental Fig. S2). A neural network classification approach (Abadi et al. 2016) also classified the patient's tumor as most similar to glioma (Supplemental Fig. S3). Overall this indicates a strong gene expression similarity between the patient's pericardial lesion and high-grade adult glioma tumors.

Table 1. Variants detected in the pericardial lesion by whole-genome sequencing

Gene	Chr	HGVS DNA reference	HGVS protein reference	Variant type	Predicted effect (substitution, deletion, etc.)	dbSNP/ dbVar ID	Genotype (heterozygous/ homozygous)
FCGR1A	1	GRCh38	NP_000557.1	Missense	Substitution	rs637882	Heterozygous
ANKRD36C	2	GRCh38	NP_001297083.1	Missense	Substitution	rs77972623	Heterozygous
HLA-DRB1	6	GRCh38	NP_001230894.1	Missense	Substitution	rs16822805	Heterozygous

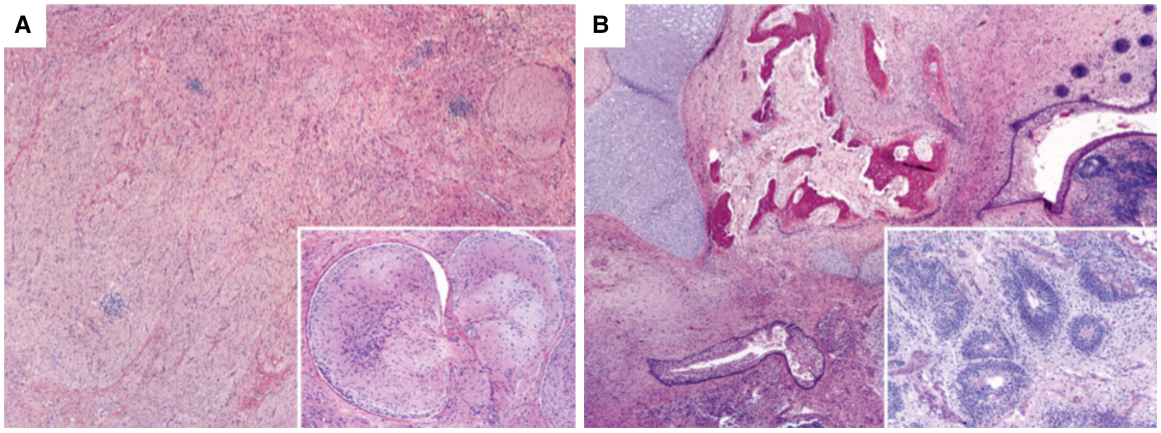

Figure 1. Teratoma involving the pericardium and diaphragm. (*A*) (H&E stain, 40×) Pericardial involvement by mature glial implant composed of mature neurons, neuropil, and schwannian stroma. (*Inset*) (H&E stain, 100×) Nodules of implants within vascular spaces. (*B*) (H&E stain, 40×) Diaphragm involvement by a mixture of mature and immature germ cell components. (*Inset*) (H&E stain, 100×) Immature neuroepithelium forming rosettes set within neuropil.

The TumorMap algorithm (see Methods) was used to visualize the top six most correlated tumors in the context of all tumors in the Treehouse cancer compendium (Newton et al. 2017). TumorMap visualization generates a two-dimensional (2D) "map" of the similarity between tumor RNA-seq samples based on pairwise Spearman correlation. The six most correlated samples to the focus tumor are indicated using red pins (Fig. 2). All six most correlated samples fall in a modular cluster that includes both adult and pediatric glioma (yellow) and glioblastoma brain tumors (green). This indicates that the patient's pericardial lesion is most transcriptionally similar to high-grade glial brain tumors.

Diagnosis of Gliomatosis Peritonei

As a result of the combined histologic and genomic analysis, the patient was subsequently diagnosed with gliomatosis peritonei. This diagnosis is consistent both with the presence of glial tissue in the pericardial lesion, and with the Treehouse genomic finding that the patient's tumor is most similar to high-grade glioma tumors. Molecular similarities between high-grade glioma and gliomatosis peritonei include high expression of the stem cell marker *SOX2* and low expression of transcription factors *OCT4* and *NANOG* (Nogales et al. 2014; Liang et al. 2015). Consistent with these characteristics, Figure 3A shows that the patient's

Table 2. The top six most correlated RNA-seq samples to the patient's RNA-seq sample belong to older patients with glioma or glioblastoma

Sample ID	Diagnosis (grade)	Histology	Patient age (yr)	Spearman correlation
TCGA-DU-7012-01	Glioma (3)	Astrocytoma	74	0.93
THR14_0312_S01	Glioma (3)	Astrocytoma	18	0.92
TCGA-CS-4941-01	Glioma (3)	Astrocytoma	67	0.91
TCGA-HT-7680-01	Glioma (2)	Astrocytoma	32	0.91
TCGA-DU-8158-01	Glioma (3)	Astrocytoma	57	0.91
TCGA-28-5215-01	Glioblastoma (4)	Astrocytoma	62	0.91

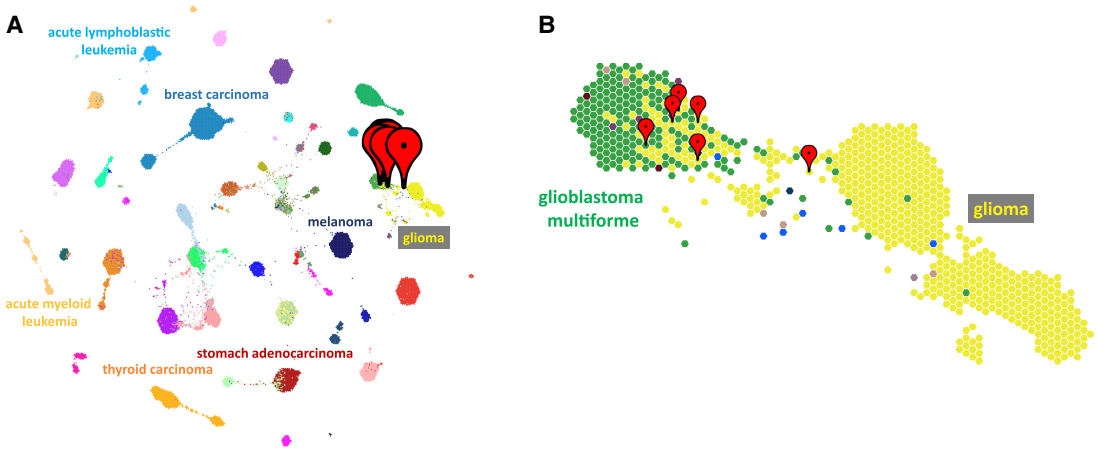

Figure 2. TumorMap clustering visualization of the Treehouse cancer compendium. (*A*) Treehouse cancer compendium v8 shown visualized with the TumorMap tool. Each colored dot represents an individual patient's tumor RNA-seq data. Tumors are grouped based on RNA-seq similarity and selected tumor types are labeled. The top six most similar tumors to the patient's pericardial lesion are indicated with red pins. (*B*) Zoomed-in image of the location of the six most correlated tumors on the TumorMap. Five out of six of the most correlated tumors fall in the leftmost brain tumor cluster, which contains the majority of glioblastoma samples and high-grade glioma samples.

tumor expresses *SOX2* at very high levels, comparable to glioma and glioblastoma tumors in the Treehouse cancer compendium. Figure 3B,C shows that the patient's tumor also expresses very low levels of *OCT4* and *NANOG*, similar to gliomas.

The patient underwent resection of the pericardial gliomatosis implant as well as the diaphragmatic immature teratoma implant. Two years post-resection, she was healthy and was discharged from oncology.

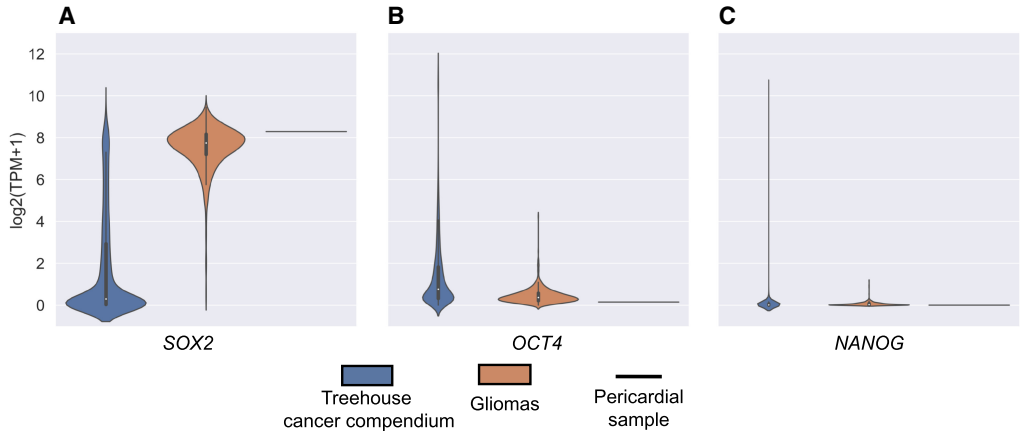

Figure 3. Expression levels of *SOX2*, *OCT4*, and *NANOG*. (*A*) *SOX2* expression. The majority of the tumors in the Treehouse cancer compendium express *SOX2* at very low levels. However, the glioma tumors exhibit exceptionally high *SOX2* expression. The patient's pericardial tumor expresses *SOX2* at a level comparable with the glioma tumor group. (*B*,*C*) *OCT4*, *NANOG* expression. The majority of Treehouse cancer compendium tumors express *OCT4* and *NANOG* at very low levels, including most gliomas. The patient's pericardial tumor also expresses *OCT4* and *NANOG* at very low levels, comparable with the glioma tumor group.

DISCUSSION

We describe here the utility of comparative RNA-seq analysis using the TumorMap method for molecular classification and diagnosis of a rare pediatric tumor. The TumorMap algorithm has been used previously to describe the global similarity between tumor types and to discover novel subtypes (Ceccarelli et al. 2016; Farshidfar et al. 2017), but this is the first published use of TumorMap for *N*-of-1 tumor classification.

We demonstrate the utility of the publicly available Treehouse cancer compendium, an extensive database of thousands of tumor RNA-seq samples. *N*-of-1 comparison of a pediatric pericardial tumor aided in molecular classification by identifying other tumors with similar gene expression profiles, all of which were glioma or glioblastoma brain tumors. The diagnoses of the most similar tumors were clinically meaningful and consistent with the subsequent diagnosis of gliomatosis peritonei in this pediatric patient.

The methods described here are widely applicable for enabling precision molecular classification or diagnosis in cases of rare or difficult-to-diagnose cancer. Beyond the application described here, the Treehouse cancer compendium and TumorMap clustering analysis can also be used to accurately identify molecular subtypes of cancer, a useful application for cancer types with the subtype-dependent outcome and treatment differences (Newton et al. 2017). Additionally, in some cases it can be impossible to determine cancer tissue of origin using clinical or radiologic data, making it difficult to design a treatment regimen (Park et al. 2018). These methods could aid diagnosis and treatment strategies for both childhood and adult cancers with unknown tissue of origin, by clustering a tumor tissue RNA-seq sample in the TumorMap and identifying other tumors with most similar molecular features. Overall, the comparative RNA sequencing analysis presented here is a powerful tool for precision molecular subtype classification and diagnosis of cancer.

METHODS

Tissue Source and Processing

A sample of the pericardial lesion was flash frozen, embedded into OCT, sectioned to a depth of 5 µm, and stained with H&E. The sample was evaluated for tumor content by a certified pathologist. The tumor was macro-dissected from the OCT block to a depth of up to 5 mm, disrupted with a mortar and pestle under liquid nitrogen, and homogenized with a QIAshredder (QIAGEN, 79654). Nucleic acids were extracted using the AllPrep DNA/RNA kit (QIAGEN, 80204). The RNA integrity was quantified using the RNA 6000 Pico kit (Agilent, 5067-1513) on the Bioanalyzer (Agilent).

Whole-Genome Sequencing

WGS was performed on the pericardial lesion. Average WGS depth was 60.67× (tumor), 29.10× (germline). The read length was 2 × 150 bp (paired-end). The read depths for reported somatic variants are as follows: Chr 1:g.[149790230C>T] Tumor ref,alt: 45, 6; Germline ref,alt: 31,0. Chr 2:g.[95855406C>G] Tumor ref,alt: 59, 9; Germline ref,alt: 35, 1. Chr 6:g.[32584172C>G] Tumor ref,alt: 11, 5; Germline ref, alt: 19, 0 (Table 3).

RNA Sequencing

Libraries were prepared using the TruSeq Stranded mRNA kit (Illumina, RS-122-2101) with an input of 400 ng in accordance with manufacturer's instructions. All manufacturer controls were used in preparation. Libraries were quantified using the High Sensitivity DNA kit (Agilent, 5067-4626) on the BioAnalyzer (Agilent). Sequencing was performed on the

Table 3. Sequencing coverage table for somatic variants detected in the pericardial lesion

Gene	Chr	Pos	Ref allele	Alt allele	Avg depth (tumor)	Ref (tumor)	Alt (tumor)	Avg depth (germline)	Ref (germline)	Alt (germline)
FCGR1A	1	149790230	C	T	60.67	45	6	29.10	31	0
ANKRD36C	2	95855406	C	G	60.67	59	9	29.10	35	1
HLA-DRB1	6	32584172	C	G	60.67	11	5	29.10	19	0

(Avg depth) average read depth, (Ref) number of reads for reference allele, (Alt) number of reads for alternative allele.

Illumina HiSeq 4000 with PE75 chemistry at the Stanford Functional Genomics Facility. The total sequence depth for this sample was 97,983,221 reads.

Comparative RNA-seq Analysis

The RNA-seq data from the patient's pericardial lesion was processed at UC Santa Cruz and gene expression quantification was performed using the TOIL RNA-seq pipeline (Vivian et al. 2017). Genome alignment was performed with genome assembly hg38. RSEM quantification TPM measurements were used as input to normalization and compendium building. The Treehouse comparative RNA-seq analysis is designed to compare an *N*-of-1 sample against a larger background cohort of RNA-seq data from many cancer samples. Pairwise Spearman correlation scores are calculated between the gene expression vector from the focus sample and all other samples in the background cohort. The top six most correlated samples are used to identify tumor types with gene expression are most similar to the focus sample.

TumorMap

TumorMap (tumormap.ucsc.edu) is a hexagonal 2D representation of similarity between samples based on gene expression (Newton et al. 2017). The spatial representation of samples in the TumorMap is based on vector similarity using a multidimensional scaling tool called OpenOrd. Sample information, such as patient age, tumor grade, or clinical outcome, can be displayed as attributes on the map. TumorMap also has built-in correlation analysis tools for discovering correlations between attributes.

Neural Network Classification

As validation, we trained a fully connected neural network to classify disease type. The network was comprised of an input layer, a batch normalization layer, two hidden layers of size 32 with a dropout of 0.5 and relu activation, and a one-hot output layer, one per disease type with sigmoid activation. We trained the network on 80% of the data using binary cross entropy as a loss function and tested it using the remaining 20% of the data. The train and test were stratified by disease to ensure an equal representation of each disease type. The network was implemented using the Keras library in a Jupyter notebook. All code is available here: https://nbviewer.jupyter.org/github/rcurrie/pancan-gtex/tree/cf249e64f7ac5da95c264f946f2ae5fd69410f63/

ADDITIONAL INFORMATION

Data Deposition and Access

All processed RNA sequencing data is publicly available in the Treehouse cancer compendium: https://treehousegenomics.soe.ucsc.edu/public-data/. The interpreted variants were

submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV000994650–SCV000994652.

Ethics Statement

The patient was enrolled on a Stanford protocol "Clinical implementation of genomic analysis in pediatric malignancies" (IRB#34383). The UCSC Treehouse protocol was approved by the institutional review board at the University of California Santa Cruz (No. HS2648).

Acknowledgments

We thank the patient and her family for their support. We thank the California Initiative to Advance Precision Medicine, St Baldrick's Foundation, Alex's Lemonade Stand Foundation, Unravel Pediatric Cancer, Team G Childhood Cancer Foundation, Live for Others Foundation, and Lucile Packard Children's Hospital Stanford.

Author Contributions

L.M.S., O.M.V., I.B., S.R.S., S.L.S., E.A.S.-C., A.R., and A.T.S. were responsible for conception and design. D.L.L., H.C.B., E.T.K., A.D., K.L., R.C., A.G.L., J.P., A.G.L., S.G.L., I.H.B., M.R.B., F.K.H., D.H., and S.R.S. collected and assembled the data. L.M.S., O.M.V., D.L.L., H.C.B., E.T.K., A.D., K.L., R.C., A.G.L., J.P., and S.R.S. analyzed and interpreted the data. A.R., A.T.S., E.A.S.-C., N.L., and S.L.S. provided study material or patients. I.B., J.P., and A.D. provided administrative support. All authors wrote the manuscript.

Competing Interest Statement

The authors have declared no competing interest.

Received May 10, 2019; accepted in revised form August 15, 2019.

Funding

This work was supported by the California Initiative to Advance Precision Medicine, St Baldrick's Foundation, Alex's Lemonade Stand Foundation, Unravel Pediatric Cancer, Team G Childhood Cancer Foundation, Live for Others Foundation, and Lucile Packard Children's Hospital Stanford.

REFERENCES

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, et al. 2016. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467.

Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, et al. 2013. The somatic genomic landscape of glioblastoma. *Cell* **155:** 462–477. doi:10.1016/j.cell.2013.09.034

Cancer Genome Atlas Research Network. 2015. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* **372:** 2481–2498. doi:10.1056/NEJMoa1402121

Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM, et al. 2016. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164:** 550–563. doi:10.1016/j.cell.2015.12.028

Farshidfar F, Zheng S, Gingras M-C, Newton Y, Shih J, Robertson AG, Hinoue T, Hoadley KA, Gibb EA, Roszik J, et al. 2017. Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep* **18:** 2780–2794. doi:10.1016/j.celrep.2017.02.033

Gheorghisan-Galateanu A, Terzea DC, Carsote M, Poiana C. 2013. Immature ovarian teratoma with unusual gliomatosis. *J Ovarian Res* **6:** 28. doi:10.1186/1757-2215-6-28

Liang L, Zhang Y, Malpica A, Ramalingam P, Euscher ED, Fuller GN, Liu J. 2015. Gliomatosis peritonei: a clinicopathologic and immunohistochemical study of 21 cases. *Mod Pathol* **28:** 1613–1620. doi:10.1038/modpathol.2015.116

Marwah N, Batra A, Gupta S, Singhal SR, Sen R. 2016. Gliomatosis peritonei arising in setting of immature teratoma of ovary: a case report and review of literature. *J Obstet Gynecol India* **66:** 192–195. doi:10.1007/s13224-015-0708-7

Newton Y, Novak AM, Swatloski T, McColl DC, Chopra S, Graim K, Weinstein AS, Baertsch R, Salama SR, Ellrott K, et al. 2017. TumorMap: exploring the molecular similarities of cancer samples in an interactive portal. *Cancer Res* **77:** e111–e114. doi:10.1158/0008-5472.CAN-17-0580

Newton Y, Rassekh SR, Deyell RJ, Shen Y, Jones MR, Dunham C, Yip S, Leelakumari S, Zhu J, McColl D, et al. 2018. Comparative RNA-sequencing analysis benefits a pediatric patient with relapsed cancer. *JCO Precis Oncol* **2:** 1–16. doi:10.1200/PO.17.00198

Nogales FF, Dulcey I, Preda O. 2014. Germ cell tumors of the ovary: an update. *Arch Pathol Lab Med* **138:** 351–362. doi:10.5858/arpa.2012-0547-RA

Park CK, Malinowski DP, Cho NH. 2018. Diagnostic algorithm for determining primary tumor sites using peritoneal fluid. *PLoS One* **13:** e0199715. doi:10.1371/journal.pone.0199715

Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, Pfeil J, Narkizian J, Deran AD, Musselman-Brown A, et al. 2017. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol* **35:** 314–316. doi:10.1038/nbt.3772