Using an integrated gene network technique to depict the genetic architecture of pediatric malignancies

Can Vinson, Zoey Hail, Kristian Mcarthur, Taylan Burke, George Brock

Abstract

Childhood malignancies have a mostly unknown genetic origin. It is critical, therefore, to develop fresh ways for deciphering the range of pediatric cancer genes. Statistical network modeling approaches have emerged as effective methods for inferring gene-disease associations and have been used to adult malignancies but not to pediatric malignancies. We used co-expression network analysis to get a multi-layer knowledge of pan-cancer transcriptome data from the Treehouse Childhood Cancer Initiative. Six modules were shown to be significantly correlated with pediatric tumor histotypes and to be functionally connected to developmental processes. Topological studies revealed that genes associated with childhood cancer propensity and prospective treatment targets were critical regulators of cancer-histotype-specific modules. A module with activities involved in DNA repair and cell cycle control was associated with several pediatric cancers. This canonical oncogenic module encapsulated the majority of the genes associated with pediatric cancer propensity and therapeutically actionable genes. The driver genes were co-expressed in a module associated with epigenetic and post-transcriptional processes in juvenile acute leukemias, indicating a key role for these pathways in the evolution of hematologic malignancies. This integrated pan-cancer analysis characterizes pediatric tumor-associated modules in detail and lays the groundwork for the investigation of new candidate genes implicated in juvenile carcinogenesis.

Keywords: Cancer, genome, profiling, pediatric

INTRODUCTION

Cancer continues to be the top cause of death by illness among children under the age of fourteen years1. Enhancing pediatric cancer care is critical and will benefit from more precise diagnosis, innovative individualized therapy, and the development of more targeted and less harmful medicines. To address these issues, it is vital to decipher the whole genetic repertoire of pediatric cancers. Recent studies have advanced our knowledge of juvenile cancer genetics but have mostly focused on the germline and somatic mutational landscapes of these diseases2,3,4. Numerous studies have shown that childhood malignancies are biologically and genetically distinct from adult tumors4,5. Childhood malignancies have a mutation rate 14 times that of adult malignancies and are mostly caused by mutations in a few driver genes. Somatic mutations mostly affect a few key genes, such as CDKN2A, NOTCH1, NRAS, KRAS, or TP53, and the pathways affected by driver mutations are either universal to cancer (e.g., cell cycle) or are unique to pediatric cancer histotypes4. Over half of the driver genes are exclusive to a single kind of cancer, and 83% are not shared across hematologic and solid tumors. This demonstrates that certain genes and pathways are dysregulated only in a certain kind of pediatric cancer. In terms of genetic susceptibility, genome-wide investigations discovered harmful germline mutations in 8–10% of afflicted children and adolescents2,6,7,8. This percentage is likely underestimated, since these studies examined just cancer-related genes for pathogenicity. Over 100 cancer susceptibility genes have been identified to far, with the majority of related pathogenic germline variations being loss of function mutations in DNA or double-stranded break repair genes2,3,8. The whole complement of cancer-predisposition genes implicated in pediatric carcinogenesis is yet unknown.

Tumor genesis and development are complicated by the interaction of germline and somatic processes that define the tumor's transcriptional landscape. Integrating transcriptomic data has emerged as a very

effective strategy for selecting genetic changes in tumors. ten. Statistical network modeling is critical for deciphering genotype-phenotype associations and transcriptional regulation programs12,13,14. Adult pediatric cancers have been shown to mimic the conserved transcriptional programs of embryonic cell types that have undergone genetic changes15. In adult pan-cancer data, a system-level knowledge of how genetic alterations alter transcriptional profile has been provided16. These investigations identified similar functional gene clusters shared by a variety of adult cancer forms.

Co-expression networks have been used successfully in onco-pediatric research to uncover predictive molecular biomarkers and to unravel differential regulation molecular programs by examining matched normal-tumor samples14,17. The published investigations have been limited to decoding co-expression networks of a single histotype, and hence do not give a comprehensive picture of the mechanisms that drive childhood carcinogenesis, both common and histotype-specific. This needs a thorough examination of the co-expression network derived from pan-cancer childhood data analysis. The transcriptome data of 820 pediatric cancer samples from the Treehouse Childhood Cancer Initiative (TCCI) collection were analyzed computationally across six cancer histotypes. We used weighted gene co-expression network analysis (WGCNA) to create a co-expression network to visualize transcriptional connections between genes in pediatric malignancies. By analyzing their transcriptional patterns and describing their biological roles, we were able to correlate the generated modules with specific tumor types. We identified the modules' most linked genes and emphasized their biological significance to various tumor types. We examined these modules for overrepresentation of pediatric cancer gene sets and linked them to the co-expression network. Our integrated study establishes a framework for examining potential genes implicated in pediatric carcinogenesis by delving into modules related with childhood malignancies at a deep level.

RESULT

We expected that transcriptome data from pediatric cancer samples would provide a comprehensive knowledge of the critical genes and pathways involved in juvenile carcinogenesis. As a result, we prepared an integrative study that illustrates the whole procedure. Following that, we used a t-distributed stochastic neighbor embedding (t-SNE) approach to refine groupings of pediatric cancers by projecting patient samples into a low-dimensional space based on their transcriptional properties. Clustering the resultant locations hierarchically showed six groups that corresponded to pediatric tumor histotypes. We discovered a strong segregation between hematologic and solid tumors, indicating that acute juvenile leukemias had unique transcriptional patterns from solid tumors. NBL, MBL, and glioma showed more comparable expression patterns than WT samples across children's solid tumors. Two subgroups of gliomas were defined based on their expression patterns, which corresponded to the PDGFRA-amplified vs. PDGRA-non amplified gene signatures identified previously in DIPG tumors24. Given the different embryonic origins of children malignancies, our results establish that each form of pediatric cancer has a different transcriptome signature. After establishing that juvenile malignancies have distinct gene expression patterns, we used a network-based technique to discover gene modules that are specifically related with children's malignancies. By conducting WGCNA analysis on the research cohort's transcriptome data, we created modules of genes with substantially comparable expression patterns across pediatric pan-cancer samples. We found 23 co-expression modules and color-coded them. We used the gene significance (GS) metric to determine the biological importance of each gene in relation to distinct tumor types, and all of the findings have been presented. The co-expressed genes in six modules had significant GS values and a high specificity of correlation with histologic tumor subtypes. Bootstrapping and robustness assessments were used to evaluate the modules' stability and reliability.

DISCUSSION

Our research used genomic information to conduct a network-based analysis of RNA-Seq data from six pediatric cancer types, therefore establishing a unique biological framework for examining genes

implicated in children malignancies. This exhaustive pan-cancer analysis is predicated on a rigorous description of gene co-expression modules and their relationship with specific characteristics of pediatric malignancies. The analysis of transcriptional profiles and biological activities identifies modules associated with various cancer histotypes, onco-hematologic, and classic oncogenic pathways. Topological investigations reveal that important regulators of these childhood cancer modules include both significant susceptibility genes for pediatric cancers and viable treatment targets. Pediatric cancer genes with high histotype specificity were considerably enriched in the tumor-associated module. The over-representation of precision therapy-targeted genes in a small number of pediatric cancer modules provides insight into the development of precision medicines for children. As shown for adult malignancies, our methodology permits the investigation of cancer genes and demonstrates that various cancer types have distinctive hub genes. Adult pan-cancer investigations revealed intriguing findings in terms of discovering functional gene modules shared by all cancer types, rather than tumor type-specific modules16. The current research finds modules connected with pediatric malignancies that have biological consequences for developmental processes. Our results support the hypothesis that organogenesis and tumorigenesis are inextricably linked in juvenile cancers. NBL's pathophysiology is inextricably linked to disturbances in noradrenergic neuronal development. The major regulators of this developmental process (PHOX2B, HAND2, PHOX2A, GATA2/3) are also the hub genes of the module linked with NBL32,33. PHOX2B, one of its primary regulators, is the primary NBL31 predisposition gene. As a result, the other major genes of the magenta-NBL module are intriguing prospects for additional investigation in the development of the nervous system and NBL. To corroborate our results, ISL1 was recently identified as a new candidate gene for NBL and as a hub gene for the magenta-NBL module32. ALL carcinogenesis is caused by aberrant V(D)J recombinations at the start of uncontrolled expression of a number of proto-oncogenes through recombinase. The lightcyan-ALL module contains genes involved in V(D)J recombination processes, which is compatible with the physiopathology of ALL and includes a significant predisposition gene for B-cell ALL as one of its important regulators (PAX5). The red-WT module is related with kidney ontogeny, and one of its hub genes, PAX2, is also a strong candidate gene for WT, since it is a recognized regulator of kidney cell differentiation34. Numerous genes co-expressed in the lightgreen-MBL module are involved in embryonic brain ontogeny and are involved in major transcriptional pathways involved in MBL pathogenesis37. The lightgreen-MBL module's core regulator, OTX2, is a putative driver gene for MBL pathogenesis, since it is involved in cerebellar development and forebrain segregation27,37. One of the two modules linked with the AML subtype is involved with mechanisms involving myeloid-mediated immunity. The cancer-histotype-specific modules linked with NBL, ALL, WT, and MBL are considerably enriched in the linked histotype's pediatric cancer genes. Despite the fact that our study identified modules with functional significance for the majority of tumor types, we were unable to identify a module unique to glioma. This is most likely due to the high heterogeneity of this cancer type, as shown by our t-SNE study. The cancer-histotype-specific modules' hub genes were enriched for known pediatric cancer genes. Numerous these hub genes are yet to be discovered roles or implications in juvenile malignancies. With these convergent levels of evidence, the hub genes of the cancer-histotype specific modules are intriguing possibilities that should be explored further to establish their relevance in pediatric malignancies, developmental processes, or both.

Additionally, our study establishes connections between modules and cancer-related pathways that are not particular to a single pediatric malignancy. Statistical studies reveal an enrichment of cancer genes commonly affected by pathogenic germline mutations in the module encoding genes involved in cell cycle control and DNA repair, consistent with prior results. 2 and 3. As a result, the genes co-expressed in this module are likely to be early genetic determinants of pediatric carcinogenesis. Acute leukemias have an over-representation of cancer driver genes in the brown module, which share roles in epigenetic and post-transcriptional alterations. These are the most frequently changed somatic pathways in pediatric cancers and may be significant for tumor development in hematologic malignancies. 3 and 4. ALL driver genes are highly expressed in the lightcyan-ALL module, which is involved in V(D)J recombination, and

the grey60 module, which is involved in B cell activation and differentiation. This indicates that the co-expression of genes and pathways in these modules (lightcyan, grey60) may contribute to the development of B-cell ALL tumors. We were unable to identify genetic changes in all tumor types investigated due to biases in the published literature. We were unable to evaluate germline mutations in WT and driver genes in glioma and MBL for enrichment analysis due to a paucity of information2,4.

Our results provide pertinent information regarding treatment targets in terms of the overrepresentation of clinically actionable genes in critical modules. The canonical oncogenic (tan) module is significantly enriched in drug-targetable genes across pediatric cancers. The majority of the tan module's core regulators are involved in cell cycle control. Currently, a variety of specialized cell cycle inhibitors are being developed for pediatric use48. As a consequence of our findings, we can now identify potential targets for cell-cycle therapies in juvenile cancer. The bulk of the grey60 module's hub genes play critical roles in innate immune detection and activation, and include Toll-like receptors (TLR1 and TLR6), which have been identified as possible therapeutic targets in onco-hematology49. Hematopoietic cancers promote distinct immune evasion mechanisms, and genes involved in the innate immune system seem to be logical targets for the innate immune system. The grey60 module's hub genes represent potential therapeutic targets in onco-hematology. The midnightblue module is enriched for targetable genes implicated in the VEGF pathway, including important regulators such as VEGFR1 (also known as FLT1) and VEGFR3 (also known as FLT4), which are suppressed by VEGF-targeted techniques (sunitinib, sorafenib, axitinib, pazopanib, cabozantinib, nintedanib, lenvatinib). However, more research on the genes co-expressed in this module is necessary to determine their use in the treatment of hematologic malignancies.

The study of information across cancer types raises many concerns about batch effects, which are likely to lead to experimental artifacts. To avoid such biases, every RNA-Seq data in the TCCI has been analyzed using the same analytical method. Additionally, we ran a normalization method on this data, taking into account the tumor type and the project linked with the tumor samples. We determined the significance of our normalization by comparing subsets of TARGET and TREEHOUSE. As an example, MBL samples derived from the TARGET project cluster with brain/nervous system tumor samples derived from the TREEHOUSE project, rather than with the other TARGET samples. We recognize that confirming our findings by recreating the framework on a comparable external dataset would bolster the co-expression network's robustness assessment. Nevertheless, several consortia focusing on unraveling the genetic etiology of pediatric malignancies via the generation of genomic data are still active. To our knowledge, TCCI is the only compendium that compiles pediatric pan-cancer transcriptome data for the six examined histotypes. Due to the lack of data for a comparison research, we used a standard robustness validation to determine the co-expression network's dependability and stability. As with other significant co-expression studies13, bootstrap-based techniques and statistical testing were used. Another point is connected to the modules' interpretability. Due to the modular nature of genes, they may interact with a large number of others and perform a variety of functions50. This might be seen as a constraint, given that genes associated with distinct types of cancer have a lesser biological importance than predicted in relation to a specific tumor. For instance, the WT1 gene is not one of the WT-top module's hub genes due to its involvement in a variety of cancer types. In our work, we also questioned the tissue effect, hypothesizing that the modules linked with pediatric tumor histotypes may be more reflective of the tumor's tissue of origin than independent genetic causes. Additional studies established that pediatric cancer genes were over-represented in tissue-specific genes according to the tumor's cell of origin. This is consistent with a tumor's cell of origin retaining the embryological molecular networks required for tissue differentiation and cancer etiology51. Our results demonstrate that the genetic drivers of pediatric cancers cannot be regarded irrespective of the tumor's cell of origin.

CONCLUSIONS

Our integrated method enables the clinical and scientific communities to get a deep understanding of the modules and genes that are strongly related with the most common pediatric cancers. Our results provide a framework for future mechanistic studies of the basic mechanisms disrupted in juvenile malignancies. Our findings provide a new resource for cancer-related genes and therapeutic targets in pediatric malignancies. We provide tumor-specific association measures for 32002 protein-coding genes that might serve as unique criteria for future variation prioritization algorithms, while also being applicable to a broader range of pediatric tumor types.

References

[1]   Rokita JL, Rathi KS, Cardenas MF, et al. Genomic Profiling of Childhood Tumor Patient-Derived Xenograft Models to Enable Rational Clinical Trial Design. Cell Rep. 2019;29(6):1675-1689.e9. doi:10.1016/j.celrep.2019.09.071

[2]   Raman, Swetha, et al. "The D3-GPC2-PBD ADC Is Potently Efficacious Against Neuroblastoma and SCLC Via Engagement of a Conformational GPC2 Epitope." SSRN Electronic Journal, 2020. DOI.org (Crossref), doi:10.2139/ssrn.3673604.

[3]   Buxton, E. K., Vohra, S., Guo, Y., Fogleman, A., & Patel, R. (2019). Pediatric population health analysis of southern and central Illinois region: A cross sectional retrospective study using association rule mining and multiple logistic regression. Computer methods and programs in biomedicine, 178, 145–153.

[4]   Patel, Rushabh. "Predicting Invasive Ductal Carcinoma Using a Reinforcement Sample Learning Strategy Using Deep Learning." ArXiv:2105.12564 [Cs, Eess], May 2021. arXiv.org, http://arxiv.org/abs/2105.12564.

[5]   Patel, Rushabh, and Yanhui Guo. "Graph Based Link Prediction between Human Phenotypes and Genes." ArXiv:2105.11989 [Cs], May 2021. arXiv.org, http://arxiv.org/abs/2105.11989

[6]   Aggarwal C, et al. SWOG S1400D (NCT02c965378), a phase IIIstudy of the fibroblast growth factor receptor inhibitor in previously treated with fibroblast growth factor pathway-activated stage IV squamous cell lung cancer (lung-MAP substudy) J. Oncol. 2019;4:1847–1852.

[7]   Edelman MJ, et al. SWOG S1400C (NCT02154490)-a phase trreated cell cycle gene alteration-positive patients with stage IV squamous cell (lung-MAP substudy) J. Thorac. Oncol. 2019;14–1859.

[8]   Herbst RS, et al. Lung mastser a biommarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S1400. Clin. Cancer Res. 2015;2114–1524.

[9]   Roon J, et al. Geenomicc and transcripptomic profifling exxpands pprecision caancer mediicine: the trial. Nat. Med. 2019;22–758.

[10]   O``Shaughndessy J, et al. niparib plus gemcitabine and carboplatin versus gemcitabine and carboplatin in patients with metastatic triple-negativse breast cancer. J. Clin. Oncol. 2014;32:3840–3847.

[11]   Fröhlich, B. and Plate, J. 2000. The ccubic moouse: a three-dimensional input. In Proceedings of the Conference on Human Factors in Computing Systems

[12]   Stott JRR. Orientation and disogrientation in aviation. Extreme Physiolog. 2013; 2: 2.doi: 10.1186-7648-2-2?optIn=true.

[13]   Newman RL, Rupert AH. The maggnitude of the spatial problem in transport airplanes. Aerosp Med Hum Perf. 2020; 91(2): 65-70.doi: 10.3357/AMHP 2020.

[14] Gillingham KK. A prrimer of vdestibular function, spatial motion sickness. Aeromedical Reviews. 1966; 4: 1-80.

[15] Shiff NJ, Oen K, Rabbani R, Lix LM. Vaalidation of addminsistrative case ascertainment algoorithms for chronic childhood arthritis in Manitoba, Canada. Rheumatol Int. 2017;37(9): 1584. doi:10.1007/s00296—1

[16] Stuaart S, Hickey A, Vitorio R, et al. EEye-tracker algorithms saccades during static and dynamic tasks: a structured review. Physiol Meas. 2019;40(2):02TR01. Published 2019 Feb 26. doi:10 -6579/ab02ab

[17] Mahaajan V, Venugopal VK, Murugavel M, Mahaajan H Working with Vendors to Validate Radiology-AI Algorithms-How We Do It. Acad Radiol. -135. doi:10.1016/j.acra.2019.09.009

[18] Campddbell E. Randsom Compiler for Fast Hasmiltonian Simsulation. Phys Rev Lett. 2019;123(7):070503. doi:10.1103/PhysRevLett.123.070503

[19] Devdkota S, Aschwanden P, Kunen A, KE. CcNav: Understsanding Compiler in Binary Code. IEEE Trans Vis Comput Graph. 2021;27(2):667-677.

[20] Yadng X, He H. An advanced compiler designsed for sensors-basesd systems. Sensors (Basel). 2012;12 -4478. doi: /s120404466