

Request for INRG Data Analysis

Proposal Title: Identifying neuroblastoma drivers and bringing them to the clinic

Principle Investigators: Stephen Skapek, MD, Lin Xu, PhD, Susan Cohn, MD, Mark Applebaum, MD

Institution: University of Texas Southwestern Medical Center and University of Chicago
Email addresses: Stephen.Skapek@UTSouthwestern.edu; Lin.Xu@UTSouthwestern.edu; scohn@peds.bsd.uchicago.edu; mapplebaum@peds.bsd.uchicago.edu;

Specific Aims:

We are exploring the potential for new computational and functional genomics approaches to reveal oncogenic drivers of neuroblastoma. Based on our previous rhabdomyosarcoma (RMS) studies,¹ we expect this innovative analytic strategy will provide an increased understanding of the biologic underpinnings of high-risk neuroblastoma and also lead to the identification of new therapeutic targets, and biomarkers for high-risk neuroblastoma.

In general, the most frequent genetic alterations in cancer are single-nucleotide and gene copy-number variants (SNV and CNV, respectively). In neuroblastoma, next-generation sequencing efforts have only revealed approximately 18 somatic protein-altering SNVs per tumor.^{2,3} With a notable exception of *ALK*,²⁻⁵ few genes are recurrently mutated, and that poses challenges to targeting SNVs. Copy-number gains in *MYCN* have long been linked to a more aggressive disease,⁶ but *MYCN*-directed therapies have also been difficult to develop.⁷ Other CNVs, including segmental alterations in 1p-, 11q-, 17q+ and sub-segmental alterations have been described.⁸⁻¹² While serving as additional biomarkers for higher risk disease, how they drive less favorable outcome is not yet clear.^{13,14}

Our team recently conceived of a plan to explore the hypotheses that a) relevant CNVs in neuroblastoma are those in which gene copy-number alteration correlates with gene expression, and b) identifying those genes will reveal new therapeutic targets and better biomarkers for risk stratification. In this project, we are employing a new computational algorithm in which Bayesian methodology provides an integrative analysis of gene Expression and Copy-Number (iExCN), an approach that has already met with some success in RMS.¹ Our three Specific Aims are:

Specific Aim 1: To uncover oncogenic drivers and tumor suppressors in neuroblastoma by using iExCN, a Bayesian-based, integrative analysis algorithm

Specific Aim 2: To experimentally validate iExCN-predicted drivers and tumor suppressors in a panel of neuroblastoma cell lines and PDX models

Specific Aim 3: To explore how CNVs and expression of iExCN genes correlates with clinical variables to improve risk stratification models and guide therapy for children with neuroblastoma

Hypothesis:

We hypothesize that relevant CNVs in neuroblastoma are those in which gene copy-number alteration correlates with gene expression, and further that identifying those genes will reveal new therapeutic targets and better biomarkers for risk stratification.

Patient Cohort (Eligibility Criteria):

The Gabriella Miller Kids First (GMKF) Pediatric Research Program was established by the NIH to enable researchers, clinicians, and patients to work together to accelerate research and promote discoveries for children affected with cancer and structural birth defects.¹⁵ GMKF

datasets provide a variety of next-gen sequencing data, including somatic and germline DNA and somatic RNA analyses. Similar types of data are also provided by the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) initiative.¹⁶

While GMKF and TARGET are unparalleled repositories of genomics data, they were not designed to house rich clinical annotation. To address that gap, we plan to link GMKF data to the International Neuroblastoma Research Group (INRG) Data Commons. Importantly, the GMKF and TARGET datasets are connected to additional sources of data by a Universal Specimen Identifier (USI) which forms the backbone of patient identifiers for our studies. The USI numbers for the GMKF and TARGET data are publicly available. Thus, eligible patients for this proposal will be those neuroblastoma patients with sequencing data in the GMKF and TARGET databases as identified by USI numbers.

Background:

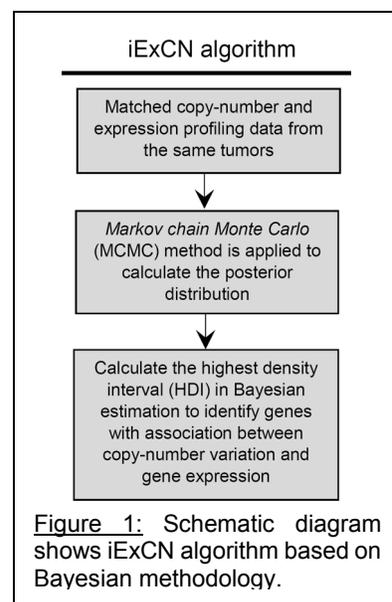
MYCN status is one of the criteria used to assign patient risk, and the presence of *MYCN* amplification stratifies patients to the high-risk group. Although *MYCN* amplified high-risk tumors are biologically distinct from non-*MYCN* amplified high-risk neuroblastoma and display different transcriptional networks,⁹ among high-risk patients treated with modern therapy, *MYCN* status is not prognostic of outcome.¹⁷ Other CNVs, including segmental and subsegmental chromosomal aberrations are described and appear to have prognostic value,⁸ but *CHD5* (chromosome 1p) appears to be the one of only tumor suppressor gene consistently lost¹⁸ and *MYCN* remains the only known oncogene consistently gained.³

Gene mutation and expression are also important in cell biology. *ALK* is one of the most commonly mutated gene in neuroblastoma, but it is detected in only 8-10% of diagnostic tumors.¹⁹ *ALK* inhibitors are being tested in the small subset of children with tumors that harbor *ALK* mutations.²⁰ Multiple gene expression signatures have also been developed as potential prognostic biomarkers,²¹⁻²⁴ but none have made it to widespread clinical use or identified putative oncogenic drivers.

As the Skapek lab recently showed for RMS,¹ the gene expression changes likely to be most important in driving tumor phenotype can be “hard-wired” into the genome by local CNVs. In previous studies analyzing two independent RMS cohorts, we have demonstrated that patients with tumors that harbor a greater number of iExCN genes with any level of CNVs have poor outcome. We expect that applying the biologically-based iExCN analysis and functional studies to neuroblastoma will reveal the CNVs and expression changes likely to be actual disease drivers (Figure 1).

Additionally, our analysis algorithm is coupled to a CRISPR-based “mini-pool” screen to quickly provide functional validation of the candidate disease genes in a panel of human neuroblastoma models. This *tandem, computationally-guided, functional screen* could become a new standard for helping to move genomics data into the clinic.

The Skapek lab applied this tandem analysis to RMS. The iExCN algorithm predicted 25 oncogenic drivers and 4 tumor suppressors, based on copy-number/expression gain and loss, respectively. Fifteen (52%) of 29 iExCN-identified genes were shown by CRISPR/Cas9 to be functionally important in two RMS models.¹ This level of enrichment approximated the significant



changes in 57% of the guides targeting “positive control” genes; only 4% of vectors targeting randomly selected negative controls changed significantly (Figure 2). A number of RMS genes are targetable, including EZH2 and RIPK2,¹ or highlight targetable pathways, such as IGF2 signaling downstream of PLAG1.²⁵ We are confident this approach will also work in neuroblastoma.

Significance:

In many cases, next generation sequencing has not transformed childhood cancer care the way most had hoped. The promise of

“precision medicine” has fallen short because most childhood cancers harbor few, if any, targetable mutations detectable by SNV analyses, and the mere presence of a mutant, highly expressed, or amplified gene does not necessarily signify that gene as an oncogenic driver. Our proposal is founded on the following premises:

- 1) Childhood cancers, especially neuroblastoma, harboring few recurrent mutations are better viewed as diseases in which otherwise normal cells are reprogrammed based on altered gene expression, and many of the important expression changes are likely to be “hard-wired” by gene copy-number gain or loss.
- 2) A biologically-based integrative analysis of coordinate changes in gene expression and gene copy-number can identify candidate oncogenic drivers and tumor suppressors that have escaped notice thus far.
- 3) Functional screens leveraging CRISPR/Cas9-based high throughput approaches can provide functional validation to help prioritize targetable oncogenic drivers and tumor suppressors.
- 4) The presence of CNVs in *bona fide* oncogenic drivers and tumor suppressors will represent a new biomarker for clinically relevant features and illuminate potential targetable vulnerabilities.

Proposal Description:

The GMKF project includes genomics data generated from almost 500 children with high-risk neuroblastoma. For our project, we will use the matched DNA and RNA from 175 diagnostic biopsies with whole-genome sequencing (WGS) and RNA-seq (<https://portal.kidsfirstdrc.org>). The FASTQ files for both data types will be downloaded from the GMKF portal (dbGaP phs001436, approval pending). With CNVs and gene expression data as input, we will run the iExCN algorithm to identify potential cancer genes. As in our previous application of iExCN,¹ we will use a stringent criterion in iExCN to define statistically significant cancer genes: the probability that group difference is greater than zero must be more than 99.9% in Bayesian estimation. That approach led to a high validation rate of iExCN-predicted cancer genes in our previous study (Figure 2). We will first analyze the entire cohort of 175 cases. Recognizing the potential for biological differences based on MYCN amplification, expected to occur in ~50% of the cases, we will carry out separate iExCN analyses in cases with and without this feature.

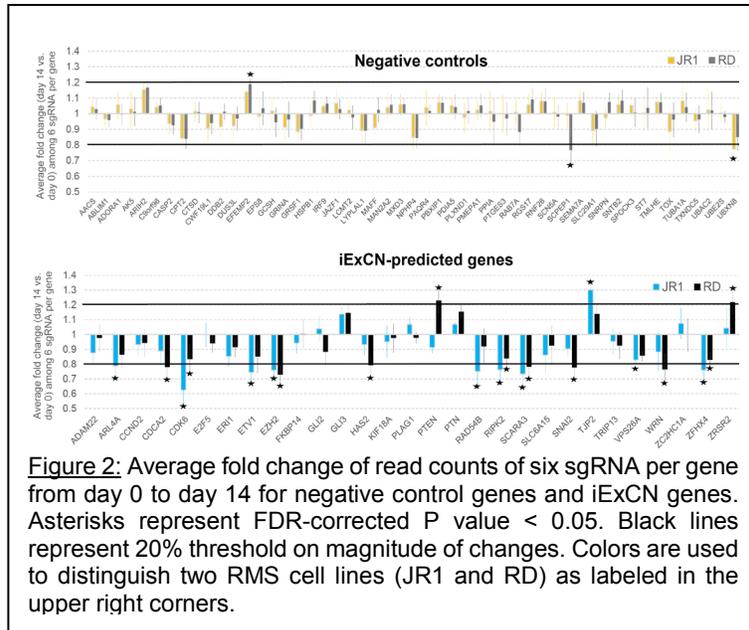


Figure 2: Average fold change of read counts of six sgRNA per gene from day 0 to day 14 for negative control genes and iExCN genes. Asterisks represent FDR-corrected P value < 0.05. Black lines represent 20% threshold on magnitude of changes. Colors are used to distinguish two RMS cell lines (JR1 and RD) as labeled in the upper right corners.

Findings will be validated using matched whole exome sequencing and RNAseq data from 97 non-overlapping patients in the TARGET dataset.

We will use USI numbers to link the genomic data from GMKF and TARGET with the demographic, treatment, and outcomes information in the INRG data commons to create genomic data with robust clinical annotation, and we will analyze these two cohorts independently. With each cohort, we will conduct univariate analyses with Cox's proportional hazards regression model to assess CNVs in individual iExCN genes failure-free and overall survival. We will also consider iExCN gene status in multivariate analyses with other known risk factors, including *MYCN* amplification status, age, and stage. Association with event-free and overall survival will be plotted by Kaplan–Meier analysis with P-value calculated by the log-rank test. We will use Holm procedure or the bootstrap-based step-down method to perform multiple testing correction. Adjusted P-values < 0.05 will be considered as statistically significant.

Next, we will determine whether the number of iExCN genes harboring CNVs correlates with survival. If it does, we will also test whether cases with copy-number gains in a greater number of iExCN genes also have significantly higher average expression of those genes. Because increased chromosomal instability is also associated with worse prognosis in neuroblastoma,⁸ we will compare the chromosomal instability (CIN) index²⁶ subsets of cases with more or fewer iExCN genes involved.

Finally, we will use complementary approaches to prove the functional importance of iExCN-predicted neuroblastoma disease genes identified using high-throughput CRISPR/Cas9-based gene editing and molecular biology approaches to interrogate cells for growth, apoptosis, tumorigenicity, and cell differentiation.

Data Requested: All patient characteristics, clinical, histopathologic, and biology variables for patients with data in the GMKF and TARGET databases (USI numbers to be provided).

Collaborating Biostatistician: Lin Xu

References:

1. Xu L, Zheng Y, Liu J, et al. Integrative Bayesian Analysis Identifies Rhabdomyosarcoma Disease Genes. *Cell Rep*. 2018;24(1):238-251.
2. Molenaar JJ, Domingo-Fernandez R, Ebus ME, et al. LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nature genetics*. 2012;44(11):1199-1206.
3. Pugh TJ, Morozova O, Attiyeh EF, et al. The genetic landscape of high-risk neuroblastoma. *Nat Genet*. 2013;45(3):279-284.
4. Bellini A, Bernard V, Leroy Q, et al. Deep Sequencing Reveals Occurrence of Subclonal ALK Mutations in Neuroblastoma at Diagnosis. *Clin Cancer Res*. 2015;21(21):4913-4921.
5. Mosse YP, Laudenslager M, Longo L, et al. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. 2008;455(7215):930-935.
6. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science*. 1984;224:1121-1124.
7. Barone G, Anderson J, Pearson AD, Petrie K, Chesler L. New strategies in neuroblastoma: Therapeutic targeting of MYCN and ALK. *Clin Cancer Res*. 2013;19(21):5814-5821.

8. Schleiermacher G, Mosseri V, London WB, et al. Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *Br J Cancer*. 2012;107(8):1418-1422.
9. Pinto NR, Applebaum MA, Volchenbom SL, et al. Advances in Risk Classification and Treatment Strategies for Neuroblastoma. *J Clin Oncol*. 2015;33(27):3008-3017.
10. Chicard M, Colmet-Daage L, Clement N, et al. Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-spatial Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma. *Clin Cancer Res*. 2018;24(4):939-949.
11. Attiyeh EF, London WB, Mosse YP, et al. Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N Engl J Med*. 2005;353(21):2243-2253.
12. Janoueix-Lerosey I, Schleiermacher G, Michels E, et al. Overall genomic pattern is a predictor of outcome in neuroblastoma. *J Clin Oncol*. 2009;27(7):1026-1033.
13. Chang X, Zhao Y, Hou C, et al. Common variants in MMP20 at 11q22.2 predispose to 11q deletion and neuroblastoma risk. *Nature communications*. 2017;8(1):569.
14. Okawa ER, Gotoh T, Manne J, et al. Expression and sequence analysis of candidates for the 1p36.31 tumor suppressor gene deleted in neuroblastomas. *Oncogene*. 2008;27(6):803-810.
15. Major A, Cox SM, Volchenbom SL. Using big data in pediatric oncology: Current applications and future directions. *Semin Oncol*. 2020;47(1):56-64.
16. Volchenbom SL, Cox SM, Heath A, Resnick A, Cohn SL, Grossman R. Data Commons to Support Pediatric Cancer Research. *Am Soc Clin Oncol Educ Book*. 2017;37:746-752.
17. Park JR, Kreissman SG, London WB, et al. Effect of Tandem Autologous Stem Cell Transplant vs Single Transplant on Event-Free Survival in Patients With High-Risk Neuroblastoma: A Randomized Clinical Trial. *JAMA*. 2019;322(8):746-755.
18. Koyama H, Zhuang T, Light JE, et al. Mechanisms of CHD5 Inactivation in neuroblastomas. *Clin Cancer Res*. 2012;18(6):1588-1597.
19. Bresler SC, Weiser DA, Huwe PJ, et al. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. *Cancer Cell*. 2014;26(5):682-694.
20. Mosse YP, Lim MS, Voss SD, et al. Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *Lancet Oncol*. 2013;14(6):472-480.
21. Stricker TP, Morales La Madrid A, Chlenski A, et al. Validation of a prognostic multi-gene signature in high-risk neuroblastoma using the high throughput digital NanoString nCounter system. *Mol Oncol*. 2014;8(3):669-678.
22. Barbieri E, De Preter K, Capasso M, et al. A p53 drug response signature identifies prognostic genes in high-risk neuroblastoma. *PLoS One*. 2013;8(11):e79843.
23. Fardin P, Barla A, Mosci S, et al. A biology-driven approach identifies the hypoxia gene signature as a predictor of the outcome of neuroblastoma patients. *Mol Cancer*. 2010;9:185.
24. Hallett RM, Seong AB, Kaplan DR, Irwin MS. Transcript signatures that predict outcome and identify targetable pathways in MYCN-amplified neuroblastoma. *Mol Oncol*. 2016;10(9):1461-1472.
25. Zheng Y, Xu L, Hassan M, et al. Bayesian Modeling Identifies PLAG1 as a Key Regulator of Proliferation and Survival in Rhabdomyosarcoma Cells. *Mol Cancer Res*. 2019.
26. Wang K, Lim HY, Shi S, et al. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology*. 2013;58(2):706-717.