The Potential Genetic Contributions of Novel Variants in *GATAD2B*, *BCL11A*, and *PIK3R2* to Cerebral Palsy

Vedika Jha

2021 Laidlaw Scholar

Sick Kids Summer Research Student

Abstract

*Background***:** Cerebral Palsy (CP) is a group of neurodevelopmental disorders characterized by injury to the developing brain. CP is caused by environmental and genetic factors; however, the genetic contributions are poorly understood.

*Objectives***:** To better understand genetic factors, The Centre For Applied Genomics conducted whole genome sequencing on DNA from CP patients and their parents and identified potential clinically relevant variants. Of these variants, I confirmed the *de novo* predictive loss-of-function and missense variants using Sanger Sequencing.

Here, I highlight three *de novo* variants in genes that have been previously linked to neurodevelopmental disorders: *BCL11A* (c.A200G:p.(K67R)), *GATAD2B* (c.T1321C:p.(C441R)) and *PIK3R2* (c.G1117A:p.(G373R)). I investigated each gene's functions, previously identified gene variants, and impacts of those variants in animal models, cell lines, and clinical studies to understand each genes' potential link to CP.

*Results***:** *BCL11A* encodes a transcription regulating protein that suppresses fetal hemoglobin after birth and promotes axonal growth and branching. *GATAD2B* encodes a subunit of the nucleosome remodeling and histone deacetylase complex which regulates transcription and is linked to neural development. *PIK3R2* encodes a regulatory subunit of a kinase which is involved in growth signaling and transcription regulation. All genes are highly expressed in the brain, and all variants are in conserved regions of their respective gene. Other *de novo* missense alterations in the same exon of each gene have been identified in patients with other neurodevelopmental disorders such as intellectual disability. Knockdown studies in mice and *Drosophila* as well as molecular pathway analyses show evidence for pathogenicity of similar variants in all genes.

*Conclusions***:** The validation of sequence level variants adds to the growing database of genes and variants connected to CP, including three variants, two of which are completely novel, on genes that have not previously been linked to CP. Further investigation can be done on other variants and how they impact gene function, adding to the overall understanding of potential causes of CP.

Introduction

Cerebral Palsy, or CP, is a group of developmental conditions characterized by injury to the developing brain causing motor, speech, vision, or hearing impairments, and intellectual disabilities (CPnet, 2021). CP is the most common childhood physical disability, and in most industrialized countries the prevalence is approximately 1 in 500 live births (Colledge, 2001).

CP is thought to be caused by both environmental and genetic factors creating a brain injury before or during birth, or within the first few years of life (Nelson, 2008). Environmental risk factors include prematurity, birth asphyxia, and exposure to infection (CPnet, 2021). Population studies have also shown evidence for a genetic contribution to risk (Tollånes Mette et al., 2014). However, genetic contributions to CP are still poorly understood. The objective of my research is to better understand the genetic factors of cerebral palsy. To do so, I validated 72 genetic variants from the DNA of CP patients, and conducted in-depth literature searches on genes which contain 3 of the validated variants.

Methods

Whole Genome Sequencing

Prior to the start of my research, The Centre for Applied Genomics (TCAG) conducted whole genome sequencing (WGS) on DNA from over 200 CP patients and their families and identified copy number variants, single nucleotide variants, insertions, and deletions in each participants' genomes. Variants were evaluated based on the 5-tier pathogenicity classification through which likely pathogenic and thus potential clinically relevant variants were identified. Of the identified variants, I validated and analyzed 72 *de novo*[1], exonic[2], missense and loss of function[3] variants.

Primer Picking, PCR, and Cleaning

I picked primers for each variant using Primer 3. I used the UCSC In-Silico PCR tool to check for single nucleotide polymorphisms (SNPs) on the primer sequence, and I used the Human BLAT Search tool to check for primer overlap with other DNA segments. I then tested the primers through PCR, and subsequently conducted PCR on the DNA samples themselves. After checking that the PCR successfully amplified the DNA sequences using gel electrophoresis, I cleaned the PCR products using the Invitrogen PCR Purification Kit. Cleaned products were sent to the TCAG Sanger Sequencing facility. Sanger sequencing results were analyzed using FinchTV software (Figure 1) and were compared against the UCSC Genome Browser.

---

[1] *De novo* variants are new mutations that arise first in germ cells; they are not inherited from parents.
[2] Exonic variants are found in protein coding regions of the gene.
[3] Missense and loss of function variants are changes in the DNA sequence that affect the amino acid sequence of a given protein.

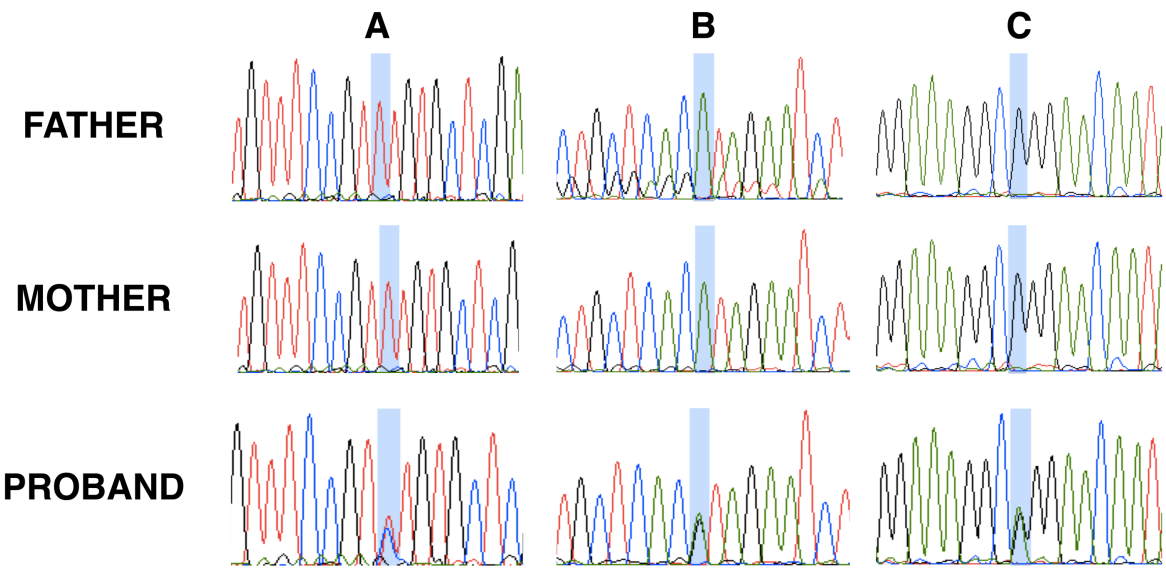

Figure 1: Sanger Sequencing results read through FinchTV for sequences on genes *BCL11A* (A), *GATAD2B* (B), and *PIK3R2* (C). *De* novo SNP variants can be seen in proband sequences for each gene, indicated by two overlapping arcs.

Literature Searches

I first conducted an overview literature search on each variant to see if the variant was already listed in existing genomic databases such as OMIM, ClinVar, ClinGen, and gnomAD. Majority of the variants had not been previously identified. However, some variants appeared in genes that have been linked to other neurodevelopmental conditions and literature investigates similar missense and loss of function variants nearby on the given gene. For these genes, I conducted in depth literature searches on each gene's function, other variants identified in the gene, and impacts of those variants in animal models, cell lines, and clinical studies. While many genes linked to other neurodevelopmental disorders still returned little to no information during the literature search, three genes were selected as "high impact" genes on which there is abundant existing literature. This report explores these three genes: GATA Zinc Finger Domain Containing 2B (*GATAD2B*), B-cell lymphoma 11A (*BCL11A*), and Phosphoinositide-3-kinase regulatory subunit 2 (*PIK3R2*).

Results and Discussion

Variants

Analysis of the Sanger Sequencing results showed that 67 out of 72 variants (93%) were successfully validated. 71% (51) of the variants were SNPs, and 22% (16) were insertions or deletions (Table 1).

Table 1: Summary of Variant Validations.

Type of Variant	Number of Variants	Percentage of total (N=76)
SNPs	51	71%
Insertions/Deletions	16	22%
Unsuccessful Validation	5	7%

Literature searches showed that majority of the identified variants are novel and have not been previously identified in existing genomic databases. However, existing literature can still inform us on the potential pathogenicity of novel variants, as will be explored further with *GATAD2B*, *BCL11A*, and *PIK3R2* (Figure 1).

GATA Zinc Finger Domain Containing 2B (GATAD2B)

Gene Function

GATAD2B encodes a subunit of the nucleosome remodeling and histone deacetylase complex (NURD complex) (OMIM, 2020). The NURD complex regulates transcription through binding DNA and histones, deacetylation of histone tails, and playing a role in nucleosome sliding and positioning (Pierson et al., 2019). It is highly conserved across species (Luo et al., 2017). The complex is linked to both neural development as well as embryonic development and is highly expressed in the brain (OMIM, 2020) (Pierson et al., 2019), which may be why *GATAD2B* variants have been linked to neurodevelopmental disorders. *GATAD2B's* known role in the complex is to recruit and localize other subunits (Feng et al., 2002) (Figure 2).

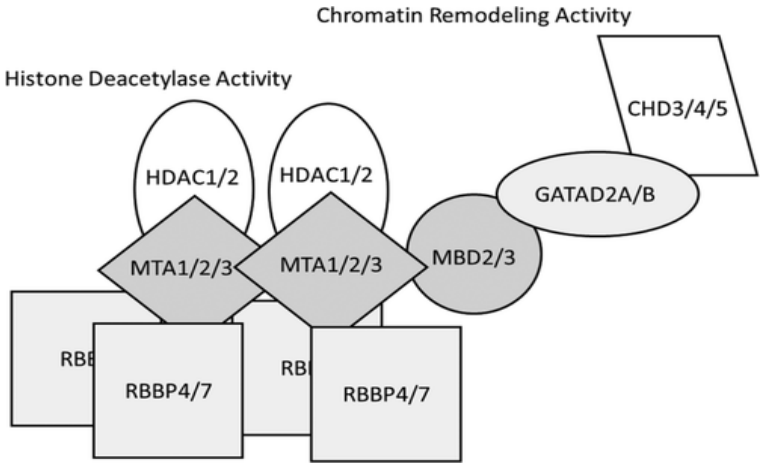

Figure 2: Illustration of the nucleosome remodeling and histone deacetylase (NURD) complex. Image from Pierson et al., 2019.

Previously Recorded Variants

While our identified variant (c.T1321C:p.(C441R)) has not previously been found, other *GATAD2B* variants have been studied and linked to macrocephaly (abnormally large head size), intellectual disability, developmental delay, and crossed eyes (Shieh et al., 2020). This phenotype has been called *GATAD2B* associated neurodevelopmental disorder (GAND). GAND shares many common characteristics with CP phenotypes, thus it is possible that variants similar to those that cause GAND may cause CP. Over 50 cases of GAND have been previously identified (Shieh et al., 2020) (Pierson et al., 2019) (Figure 3). Most cases are caused by missense and loss of function variants. Many variants are found in one of two conserved regions of the gene, including our identified variant, which is in conserved region 2 (CR2)[4].

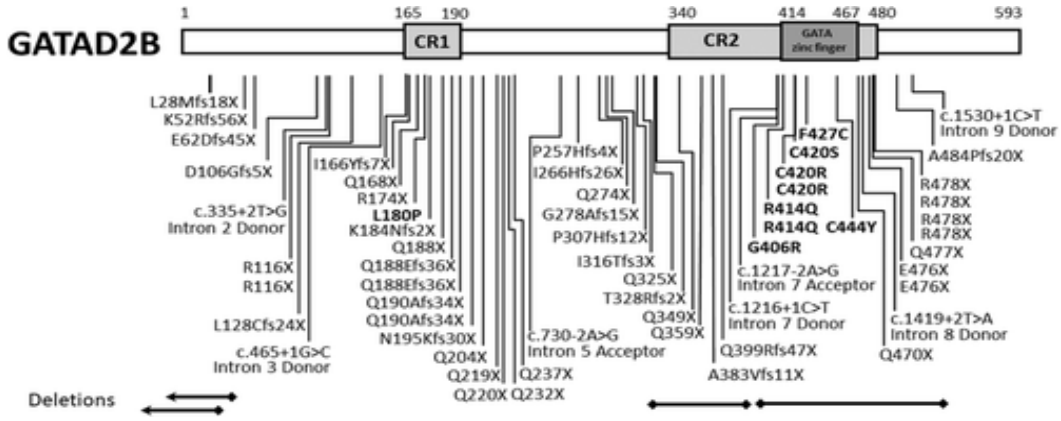

Figure 3: Genomic Map of previously recorded *GATAD2B* variants, many of which are found in conserved region 2 along with our identified variant (c.T1321C:p.(C441R)). Image from Pierson et al., 2019.

Functional Impacts of Variants

Due to the large number of variants found in CR2 of *GATAD2B*, functional studies have investigated how variants may impact the function of this region. One study used immunoprecipitation assays to show that CR2 helps direct the *GATAD2B* protein to specific nuclear loci, and variants in this region may disrupt that pathway (Feng et al., 2002). Another study used immunoprecipitation assays to show how missense variants in CR2 disrupted interaction of the protein with binding partners *CHD3*, *CHD4*, and *CHD5* (Shieh et al, 2020) (Figure 4). By disrupting the interaction of *GATAD2B* with binding partners, the variants may disrupt the overall structure of this region of the complex, inhibiting it from functioning properly.

[4] Conserved regions are regions of DNA that are similar or identical across many species, showing that the region contains important DNA that has been preserved through natural selection.

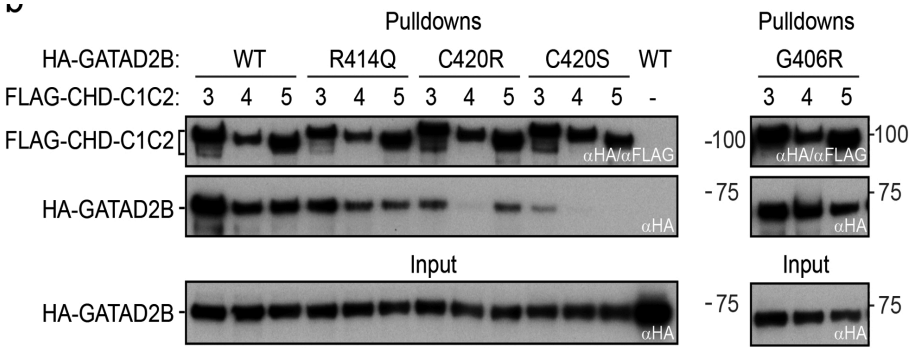

Figure 4: In-vitro transcription-translation of *GATAD2B* with NURD components (detected by Western blot) show that missense variants disrupt binding of *GATAD2B* to other components of the NURD complex. Image from Shieh et al., 2020.

Additionally, a knockdown of a *GATAD2B* ortholog[5] in fruit flies led to impaired learning and synaptic development (Willemsen et al., 2013) (Figure 5). The fruit flies' learning was measured through a timed jump response to a stimulus, which quickly improved in controls as they became habituated but didn't improve in the knocked down fruit flies. This result provides strong evidence for pathogenicity. However, since this knockdown used a microdeletion of the gene, deleting 240kb of DNA, instead of a SNP, this evidence is only partially applicable to our variant of interest, which may have similar but reduced effects when knocked down.

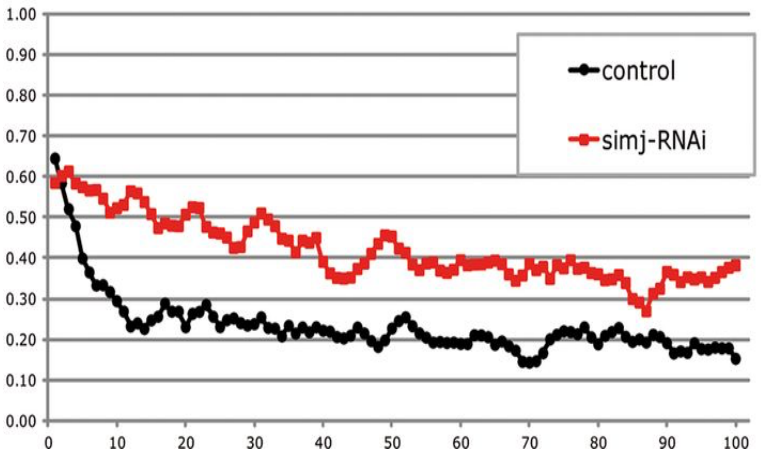

Figure 5: Knockdown of *GATAD2B* orthologue in fruit flies led to slower jump responses to a stimulus and decreased habituation Image from Willemsen et al., 2013.

B-cell lymphoma 11A

Gene Function and Conservation

BCL11A encodes a transcription regulating protein that suppresses fetal hemoglobin after birth and promotes axonal growth and branching (Kuo et al., 2010). Like the NURD complex

[5] An orthologue is an alternate version of a given gene found in a different species

which hosts GATAD2B, it is also involved in chromatin remodeling for gene suppression, and is highly expressed in the lymph nodes and brain (Liu et al., 2006). Studies have also found that it promotes young neuron outgrowth (Kuo et al., 2010), and plays a role in the polarity of upper neurons (Wiegreffe et al., 2015). To function, the BCL11A protein forms a dimer with other BCL11A isoforms to localize to nucleus and bind to chromatin (Kuo et al., 2010).

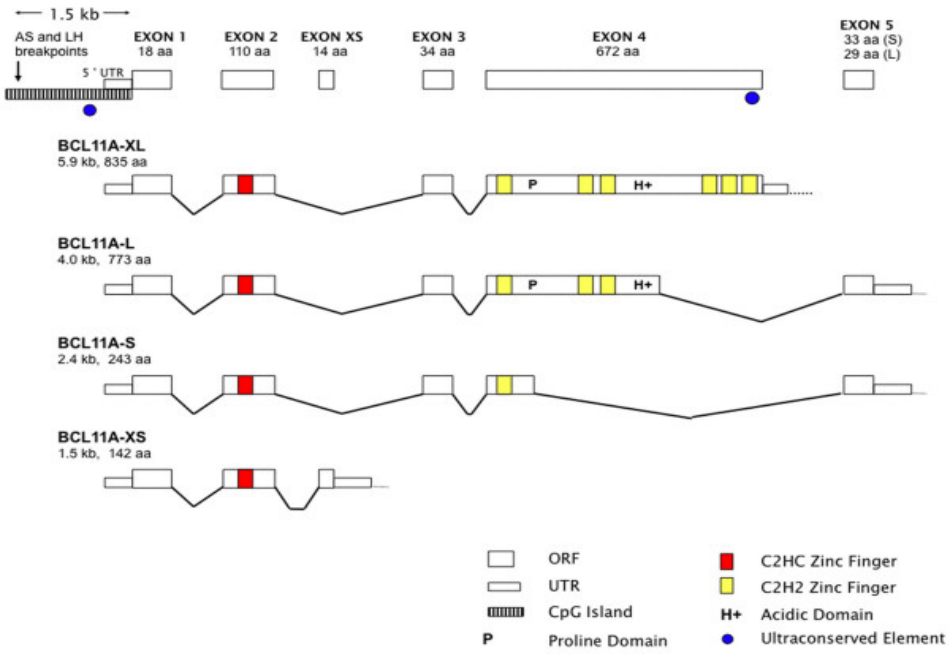

Figure 6: Isoforms of the *BCL11A* protein. Exon 2 is conserved in each isoform of the protein, thus a mutation in this exon would affect each isoform of the protein. Image from Liu et al., 2006.

BCL11A is largely conserved between mice chicken, clawed frogs, and humans, suggesting the importance of its function (Satterwhite et al., 2001). Additionally, exon 2 of the gene, which contains the identified variant (c.A200G:p.(K67R)), is found in each isoform of the protein the gene produces (Liu et al., 2006) (Figure 6). Thus, if the identified variant is pathogenic, it would affect the function of several resulting proteins.

Previously Identified Variants

While *BCL11A* has not been as extensively studied as *GATAD2B*, at least 7 missense or loss of function variants have been found in exon 2 in patients with neurodevelopmental disorders, specifically intellectual disability and speech and language development delay (Dias et al., 2016) (Korenke et al., 2020). In all individuals with *BCL11A* missense variants, fetal hemoglobin levels were significantly elevated. Since *BCL11A* suppresses fetal hemoglobin, this suggests that the *BCL11A* variant had a functional impact on the body.

Functional Impacts of Variants

Molecular assays investigated missense variants on exon 2 of *BCL11A* by measuring its interaction with *NONO* (a related protein) and with wild-type isoforms of the *BCL11A* protein (Figure 7). These studies found that the variants disrupt protein dimerization and localization (Dias et al., 2016). Because the *BCL11A* protein must form a dimer to function properly, disrupting dimerization stopped the protein from proper functioning and decreased its transcriptional regulatory activity (Dias et al., 2016).

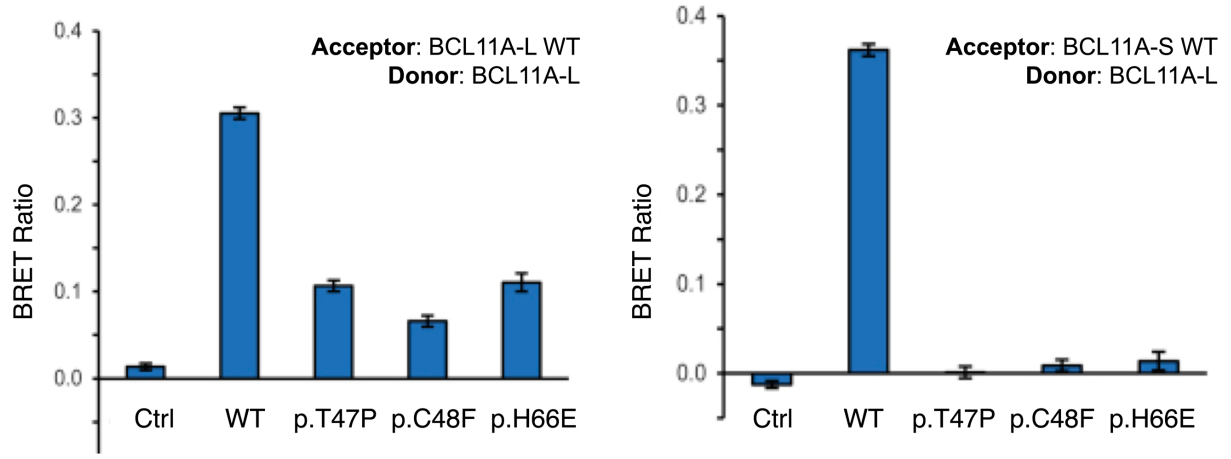

Figure 7: Missense variants on exon 2 of *BCL11A-L* reduce its ability to form a dimer with with wild type isoforms of *BCL11A* (*BCL11A-S* and *BCL11A-L)*. Image from Dias et al., 2016.

Researchers also investigated *BCL11A* exon 2 missense variants with mice models, as mice *BCL11A* is highly homologous to human *BCL11A*. Mice models showed that mice with *BCL11A* missense variants showed phenotypical features of intellectual disability, such as decreased brain volume and social recognition (Dias et al., 2016). Additionally, RNA sequencing of the mouse cortex and hippocampus showed that missense variants lead to transcriptional deregulation and thus significantly altered gene expression, strongly suggesting pathogenicity (Dias et al., 2016).

Phosphoinositide-3-kinase regulatory subunit 2 (PIK3R2)

Gene Function and Conservation

PIK3R2 encodes the regulatory subunit of the protein phosphatidylinositol 3-kinase (*PI3K*) (NCBI Gene, 2021). PI3K is a heterodimer protein with a regulatory subunit and a catalytic one, and performs a variety of functions by phosphorylating phosphatidylinositol and similar compounds that are involved in cell signaling (NCBI Gene, 2021). This creates secondary messengers that are important in growth signaling pathways, including the mTOR pathway, which is vital for cell proliferation and maintenance of brain tissues (Negishi et al., 2017). *PIK3R2* along with *PIK3R1* also regulate the function of another protein: *XBP1S*. This protein is a transcription factor that initiates protein unfolding in response to endoplasmic reticulum stress (Park et al., 2010).

The *PI3K* pathway is highly conserved in eukaryotic evolution, showing its important role in cellular function and suggesting that variants disrupting this pathway may have grave impacts on the body (Rivière et al., 2012). Additionally, this pathway is especially important for brain development; *PIK3R2* is expressed widely throughout the body but is especially highly expressed in the brain (NCBI Gene, 2021).

Variants

Unlike *GATAD2B* and *BCL11A*, the identified variant on *PIK3R2* has been previously identified several times in the literature: over 40 individuals have been identified with the c.G1117A:p.(G373R) variant (Terrone et al., 2016) (Mirzaa et al., 2016) (Rivière et al., 2012). Phenotypic features associated with this variant include megalencephaly (large head size), overgrowth and asymmetry, intellectual disability, developmental delay, and other neurologic issues such as seizures and polymicrogyria[6] (Negishi et al., 2017) (Terrone et al., 2016) (Mirzaa et al., 2016). While this variant has not previously been identified in a CP patient, many of the identified phenotypic features relate to CP and are shared by CP patients.

Functional Impacts of Variants

Functional analyses of the *PIK3R2* gene, and the *G373R* variant specifically, found that aberrations on *PIK3R2* alter activity of the mTOR pathway by preventing PI3K from entering its inactive conformation, thus keeping the pathway permanently in a high activity state (Rivière et al., 2012) (Negishi et al., 2017). Since the mTOR pathway is important for growth and cell proliferation, this may lead to the large head size, overgrowth, and other kinds of growth asymmetry observed in individuals with this variant (Negishi et al., 2017).

Conclusion

By validating 72 variants among CP patients, many of which are novel variants, this research adds to the growing database of genes and variants connected to CP. Such data can help other researchers recognize the significance of a given variant. Additionally, these variants can be further studied, both through literature searches and through experimental work, to better understand their pathogenicity and specific functional impacts.

I conducted in-depth literature searches on three genes that have previously been linked to neurodevelopmental conditions similar to CP and whose functional impacts have been studied. Literature searches show that all the genes are involved in neural development and highly expressed in the brain. Additionally, the variants are either found in conserved regions of their gene or the gene itself is highly conserved.

In the case of *GATAD2B* and *BCL11A*, our identified variants are completely novel. However, nearby variants on each gene have been previously studied and linked to other

[6] Polymicrogyria is a condition in which the brain develops abnormally to have too many ridges or folds, which are often too small as well.

neurodevelopmental disorders such as intellectual disability and developmental delay. Investigations of gene function through molecular analysis as well as animal models have provided evidence for the pathogenicity of these nearby variants. Additionally, the cellular explanation for the variants' pathogenicity, i.e. how it affects the body, can be linked to CP. In the case of *PIK3R2*, our identified variant has previously been seen in patients with other neurodevelopmental disorders, but has not been linked to CP. However, the molecular analysis of the variant shows how it may be pathogenic and linked to CP. Thus, while we cannot definitively declare our identified variants as pathogenic, the literature provides strong evidence for pathogenicity of each of these variants among CP patients.

Future Directions

Further investigation is needed to explore if other variants found in CP patients lie on genes that have been linked to neurodevelopmental conditions, as this may provide evidence to suggest pathogenicity of these variants even without experimental work on the CP variants themselves. However, this may prove difficult as there is little information about many genes and how they relate to neural development, as discovered with many of our identified variants and their genes. Thus, experiments with cell lines, molecular assays, and animal models should also be conducted to confirm pathogenicity on variants that are highly likely to be pathogenic, such as the variants on *GATAD2B*, *BCL11A*, and *PIK3R2* explored in this paper. Overall, such investigation expands the overall genetic understanding of CP, prompting further research on potential prevention and treatment methods.

Acknowledgements

I would like to thank The Centre For Applied Genomics and the Sick Kids Summer Student Research for supporting my research this summer. Thanks to Stephen W. Scherer, Richard Wintle, and Mehdi Zarrei for overseeing and guiding me this summer. Thanks to Sylvia Lamoreaux and Jill De Rijke for guiding me through the experimental procedures of the lab. Thanks to Lord Laidlaw and the Laidlaw Foundation for their generous funding of this project as well as their continual support and advice through the Laidlaw Scholars Network. Thanks to Shraddha Prasad and Flo Zeng for their help and support throughout this project and beyond.

References

CP-Net. (2021). About Cerebral Palsy. *CanChild*. Retrieved from http://cpnet.canchild.ca/en/about-cp/about-cerebral-palsy

Colledge N. (2001). A Guide To Cerebral Palsy. *The Ontario Federation for Cerebral Palsy*. Retrieved from https://www.ofcp.ca/pdf/Web-Guide-To-CP.pdf

Dias C, Estruch SB, Graham SA, McRae J, Sawiak SJ, Hurst JA, ... Logan DW. (2016). *BCL11A* Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *American Journal of Human Genetics, 99*(2); 253–274. https://doi.org/10.1016/j.ajhg.2016.05.030

Feng Q, Cao R, Xia L, Erdjument-Bromage H, Tempst P, Zhang Y. (2002). Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Molecular and cellular biology*, *22*(2), 536–546. https://doi.org/10.1128/MCB.22.2.536-546.2002

Korenke GC, Schulte B, Biskup S, Neidhardt J, Owczarek-Lipska M. (2020). A Novel de novo Frameshift Mutation in the *BCL11A* Gene in a Patient with Intellectual Disability Syndrome and Epilepsy. *Mol Syndromol*, *11*;135-140. doi: 10.1159/000508566

Kuo TY, Hong CJ, Chien HL, Hsueh YP. (2010). X-linked mental retardation gene *CASK* interacts with *Bcl11A*/*CTIP1* and regulates axon branching and outgrowth. *Journal of Neuroscience*, *88*: 2364-2373. https://doi-org.myaccess.library.utoronto.ca/10.1002/jnr.22407

Liu H, Ippolito GC, Wall JK, Niu T, Probst L, Lee BS, ... Tucker PW. (2006). Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Molecular cancer, 5*(18). https://doi.org/10.1186/1476-4598-5-18

Luo X, Zou Y, Tan B, Zhang Y, Guo J, Zeng L,…Wu L. *(2017).* Novel *GATAD2B* loss-of-function mutations cause intellectual disability in two unrelated cases. *Journal of Human Genetics,* 62, 513–516. https://doi.org.myaccess.library.utoronto.ca/10.1038/jhg.2016.164

Mirzaa GM, Conti V, Timms AE, Smyser CD, Ahmed S, Carter M, Guerrini R. (2015). Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2, in perisylvian polymicrogyria: a next-generation sequencing study. *The Lancet Neurology, 14*(12), 1182–1195. https://doi.org/10.1016/S1474-4422(15)00278-1

Negishi Y, Miya F, Hattori A, Johmura Y, Nakagawa M, Ando N, …, Saitoh S. (2017). A combination of genetic and biochemical analyses for the diagnosis of PI3K-AKT-mTOR pathway-associated megalencephaly. BMC medical genetics, 18(1), 4. https://doi.org/10.1186/s12881-016-0363-6

Nelson KB. Causative factors in cerebral palsy. (2008). *Clinical Obstetricts and Gynecology*, *51*(4):749-62. doi: 10.1097/GRF.0b013e318187087c.

Online Mendelian Inheritance in Man, OMIM®. (2020). GATA ZINC FINGER DOMAIN-CONTAINING PROTEIN 2B. Johns Hopkins University, Baltimore, MD. MIM Number: 614998.

Oskoui M, Gazzellone MJ, Thiruvahindrapuram B,...Scherer SW. (2015). Clinically relevant copy number variations detected in cerebral palsy. *Nature Communications, 6*:7949. doi: 10.1038/ncomms8949.

Park SW, Zhou Y, Lee J, Lu A, Sun C, Chung J, …, Ozcan U. (2010). The regulatory subunits of PI3K, p85α and p85β, interact with XBP-1 and increase its nuclear translocation. *Nature Medicine, 16*, 429–437

PIK3R2 phosphoinositide-3-kinase regulatory subunit 2 [Homo sapiens (human) [Internet]. (2021). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; Available from: https://www.ncbi.nlm.nih.gov/gene/5296

Rivière JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, …, Dobyns WB. (2012). De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nature genetics, 44*(8), 934–940. https://doi.org/10.1038/ng.2331

Satterwhite E, Sonoki T, Willis TG, Harder L, Nowak R, Arriola EL, … Dyer MJ. (2001). The BCL11 gene family: involvement of *BCL11A* in lymphoid malignancies. *Blood, 98*(12), 3413–3420. https://doi.org/10.1182/blood.v98.12.3413

Shieh C, Jones N, Vanle B, Au M, Huang AY, Silva A, Lee H, … Pierson TM. (2020). *GATAD2B*-associated neurodevelopmental disorder (GAND): clinical and molecular insights into a NuRD-related disorder. *Genetics in medicine : official journal of the American College of Medical Genetics*, *22*(5), 878–888. https://doi.org/10.1038/s41436-019-0747-z.

Terrone G, Voisin N, Abdullah Alfaiz A, Cappuccio G, Vitiello G, Guex N, D'Amico A, James Barkovich A, Brunetti-Pierri N, Del Giudice E, Reymond A. (2016). De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. *European journal of human genetics : EJHG, 24*(9), 1359–1362. https://doi.org/10.1038/ejhg.2016.7

Tollånes Mette C, Wilcox Allen J, Lie Rolv T, Moster Dag. (2014). Familial risk of cerebral palsy: population based cohort study. *British Medical Journal, 349*

Wiegreffe C, Simon R, Peschkes K, Kling C, Strehle M, Cheng J, …, Stefan Britsch. (2015). *Bcl11a* (Ctip1) Controls Migration of Cortical Projection Neurons through Regulation of Sema3c. *Neuron, 87*(2);311-325.

Willemsen MH, Nijhof B, Fenckova M, Millesen WM, Bongers EMHF, Castells-Nobau A, … Kleefstra T. (2013). *GATAD2B* loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. *Journal of Mendelian Genetics*, *50*:507-514.