The spectrum of mosaic mutations in megalencephaly and other growth disorders by ultra-deep targeted next-generation sequencing (NGS)

Nawal Madkhali

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2021

Committee:

Brian Shirts

Ghayda Mirzaa

Program Authorized to Offer Degree:

Laboratory Medicine

University of Washington

Abstract

The spectrum of mosaic mutations in megalencephaly and other growth disorders by ultra-deep targeted next-generation sequencing (NGS)

Nawal Madkhali

Chair of the Supervisory Committee:
Brian H. Shirts, M.D., Ph.D.

Associate Professor of Laboratory Medicine and Pathology

This is a retrospective study aimed at analyzing genetic variants and levels of mosaicism identified in a cohort of patients clinically tested for brain and body overgrowth phenotypes between 2014 and 2019 through the Megaplex multi-gene panel offered at the University of Washington. We analyzed the megaplex data to further characterize the molecular basis of overgrowth phenotypes and to optimize future interpretation and analysis of this panel.

In this study, we examined samples from 180 clinical patients diagnosed with brain and body overgrowth disorders. An additional 33 samples were collected from parents to determine the inheritance of compelling variants. The panel consisted of 37 genes known to be associated with brain and body overgrowth disorders. DNA was extracted from peripheral blood in 169 (53.8%) of the samples, 69 (22.8%) in skin fibroblast, 68 (21.8%) in tissue, 3 (0.9%) in saliva, 2 (0.6%) in cell-free DNA (cfDNA), and 2 (0.6%) in unknown samples. Capture-based Next-Generation sequencing (NGS) was performed using custom-designed SureSelect probe libraries and analyzed using short read sequencing on Illumina HiSeq 2000 or MiSeq

sequencers. Identified mutations were confirmed and analyzed using a custom in-house bioinformatics pipeline.

Of the 213 individuals tested, 128 (41.0%) had pathogenic and likely pathogenic mutations. Most of these variants were in *PIK3CA* (N =49, 38%) and *PTEN* 17 (13.3%). There were no pathogenic or likely pathogenic mutations reported in 161 cases (51.6%). In this study, variants of uncertain significance were reported in 23 (7.4%) of cases.

Ultra-deep NGS can efficiently identity mosaic mutations in megalencephaly and overgrowth disorders including detecting low levels of mosaicism, compared to Sanger sequencing and standard-depth NGS testing. Detecting mosaic mutations using deep NGS testing improves the clinical yield and provides a better understating of the spectrum of mosaic mutations underlying these phenotypes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A. The differences between macrocephaly and megalencephaly

Megalencephaly is defined as overgrowth of the brain and macrocephaly is defined as a large head measurement that is two or more standard deviations above the age-related mean.
In some cases, macrocephaly be benign without neurodevelopmental consequences. However, it could also be associated with other features such as anomalies of bone skull structures, subdural fluid collections, hydrocephalus, intracranial masses, and arteriovenous malformations. In contrast, megalencephaly is usually more severe, and could be associated with more significant neurological complications. It is characterized by increasing growth of the developing human brain due to an aberration of one the various steps of brain development such as neuronal proliferation or migration. (Pavone et al., 2017)

B. Mosaicism

Mosaicism is a biological phenomenon that describes an individual who has developed from a single fertilized egg and has two or more populations of cells with distinct genotypes (Biesecker et al., 2013). Errors that arise during chromosome segregation or DNA replication lead to chromosome aneuploidy, CNVs, genomic rearrangements, single-nucleotide variation, or repeat expansions, and microsatellite instabilities (Lupski, 2013).

Detecting mosaic mutations could provide insights into how megalencephaly and other neurodevelopmental disorders occur in a child of unaffected parents and better understand some of the biological processes that led to the occurrence of these phenotypes. Detecting mosaic mutation may vary based on phenotypes, type of molecular methods, and tissue types. Mosaic variants, including low-level ones, have been successfully detected and quantified by Next Generation Sequencing (NGS) approaches. NGS had been used to detect mosaic mutations in a variety of human genetic disorders, including autosomal dominant, autosomal recessive, and X-linked diseases (Qin et al., 2016). Low-level mosaicism seen in as low as 0.4% of cells was detected in patients with neurodevelopmental disorders using fluorescence in situ hybridization (FISH) (Oneda et al, 2017). NGS gene panels can be more sensitive than other sequencing methods. In comparison, mosaic mutations in a small fraction of blood cells in patients with brain malformations were detectable at allele frequencies of 1% or with read depths as higher as 1000× by NGS (Jamuar et al., 2014). The mosaicism level varied depending on sample types and may vary in the same tissue samples (Qin et al., 2016). In cohorts of megalencephaly-capillary malformation-polymicrogyria

syndrome (MCAP) patients and *PIK3CA*-related overgrowth spectrum (PROS) patients, *PIK3CA* variants were often detected at VAF lower than 10% using NGS (Park et al., 2020) (Hucthagowder et al., 2017).

C. Germline variants

Germline variants refer to DNA changes that are inherited or occur in gametes. They can be passed onto offspring or arise *de novo* during early embryogenesis. These variants affect every cell in the entire organism, while somatic mutations that occur later in development affect some parts of the body. Germline variants are easier to detect compared to somatic mutations. They can be detected in peripheral blood or any tissue of an affected individual.

Postzygotic mutations or germline variants in *AKT3, PIK3CA,* or *PIK3R2* were found to cause a pre-and postnatal overgrowth of the brain and other body parts in 37 out of 50 patients with these symptoms. (Riviere et al., 2012) While germline variants in *PTEN* gene have been found in patients with various clinical phenotypes, including Bannayan-Riley-Rubalcaba syndrome (BRRS), *PTEN* hamartoma tumor syndrome (PHTS), Cowden syndrome, Proteus syndrome (PS), and Proteus-like syndrome. (Yehia et al., 2019). Another study has found that germline *PTEN* mutations were found in patients with autistic behavior and extreme macrocephaly in the existence or absence of BRRS or CS features (Butler et al., 2005).

CHAPTER II
LITERATURE REVIEW FOR SELECTED GENES

A. *PIK3CA, PTEN,* and *AKT3*

Mirzaa et al. (2016) identified 29 *PIK3CA* mutations using targeted next-generation sequencing, Sanger sequencing, and whole-exome sequencing. These mutations are found in individuals with severe focal overgrowth, brain overgrowth (megalencephaly), and capillary and vascular malformations. Mosaic mutations account for 52% of *PIK3CA* mutations, and the level of mosaicism ranged from 5% to 42%. The study shows that the mutational spectrum in children with megalencephaly-capillary malformation syndrome (MCAP) is broader than other *PIK3CA*-related overgrowth disorders.

Mirzaa et al. (2016) performed multiplex targeted sequencing and sanger sequencing on 33 patients tested for mutations in (*PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, AKT1, AKT2, AKT3,* and *PTEN*). These patients diagnosed with dysplastic megalencephaly (DMEG), hemimegalencephaly (HMEG), and focal cortical dysplasia (FCD) and they underwent surgical resection for intractable epilepsy. Of 33, four (12%) pathogenic mutations were identified, in addition to a patient with DMEG: 3 mosaic

PIK3CA mutations, one mosaic *AKT3* mutation, and one germline *PTEN* mutation. The study suggests that MEG, HMEG, and FCD are associated with upregulation of the *PI3K*/*AKT*/*mTOR* pathway.

B. *PIK3R2*

In the Mirzaa et al. (2015) study of a cohort of 118 patients with bilateral perisylvian polymicrogyria (BPP), 38 individuals underwent targeted next-generation sequencing, and 80 underwent amplicon sequencing. The results show recurrent *PIK3R2* mutations detected in 19 patients, and a *de-novo* missense mutation identified in 1 patient. Mosaic *PIK3R2* mutations were detected in 8 patients while constitutional mutations were detected in 12 patients. Phenotypes were found to differ in patients as some of them were found to have BPP with average occipitofrontal circumference (OFC) and other patients with MPPH. *PIK3R2* mutations cause the Megalencephaly-Polymicrogyria-Polydactyly-Hydrocephalus syndrome (MPPH) syndrome, and this study suggested that *PIK3R2* mutations are likely associated with BPP, as seen in 15% of patients.

C. *AKT3*

Alcantara et al. (2017) reported 25 *AKT3* mutations: 14 were new *AKT3* mutations, and 11 were previously published variants. Patients with mosaic *AKT3* mutations (20%) were mainly diagnosed with DMEG/HMEG, while patients with constitutional *AKT3* mutations (80%) were diagnosed with megalencephaly- polymicrogyria MEG-PMG. The authors suggested that the mosaic p.E17K *AKT3* mutation is associated with segmental brain malformations, while constitutional *AKT3* mutations are associated with bilateral brain malformations.

CHAPTER III
METHODS

Cohort. We conducted a retrospective review of clinical cases submitted for megaplex panel testing, single genes, or known mutations in families with MEG at the University of Washington Genetics and Solid Tumors laboratory. Cases had clinical reporting between November 24, 2014, and December 23, 2020. These often consisted of samples of different tissues and/or samples from relatives of the patient. The participant IDs were anonymized, with only age, sex, clinical phenotype, and mutations detected included in the analysis.

Molecular diagnostic testing. Samples were accessioned for clinical testing; DNA extracted using standard QiaSymphony protocols, and genomic DNA sequenced using a custom targeted capture protocol. Genes on the Megaplex panel with reported clinical phenotypes are listed in Table 5.

Gene capture and sequencing. After extracting DNA from samples (peripheral blood, tissue, skin, saliva, or cfDNA) and preparing libraries with an average insert size of 150 bp, the Megaplex genes were captured by overnight hybridization solution with biotinylated RNA baits (Agilent SureSelect). The RNA baits were used to enrich targeted regions. After targeted capture, samples were sequenced using massively parallel sequencing on the Illumina HiSeq instrument

Capture-based panel sequencing pipeline and data analysis

After transferring raw data and generating unmapped sequencing reads, the Illumina BCL file converted to DNA primary sequencing strings and stored as a text-based FASTQ format, separated one for each sample (demultiplexing). The average length of sequencing strings was between 50 to 300 cycles, and the number of the strings was varied from ~10 million to ~2.5 billion. FASTQ files aligned to the reference human genome for identifying variants and comparison.

For variant identification, we used Genome Analysis Toolkit. For variant annotations, multiple tools were used, including ExAC and 1000 Genomes project for population frequency, GERP, CADD, Polyphen, and MutationTaster for assessments of variant severity, ClinVar database for evidence strength of variants, and VCF for gene prediction. All variant calls were analyzed by Lab Medicine Genetics and Seattle Children's Hospital lab directors, and a consensus was reached about variant calls before case sign-off. For data visualization, the following R packages were used: ggplot2, ggbeeswarm and cowplot.

The Megalencephaly Panel

The Megalencephaly gene panel has been developed over the years to help with detecting mosaic mutations that are contributing to megalencephaly and relative overgrowth disorders. Some of the overgrowth disorders are rare and difficult to diagnose. The first version of the MEGPX panel was used between 2014 to 2016, and it consisted of five genes: *AKT1,* AKT3, *PTEN*, *PIK3CA,* and *PIK3R2*. In June 2016, the panel was updated, and 32 new genes were added, with a total of 37 genes (Appendix A: Table 1).

CHAPTER IV
RESULTS

A. Sample summary

In a total of 312 samples, 228 (73.1%) samples were tested by whole gene panel testing (MEGPX), 55 (17.6%) for a single gene, and 29(9.3%) for a known mutation (KMU). We identified 161 (51.6%) negative cases, 128 (41.0%) positive cases including those with pathogenic and likely pathogenic variants, and 23 (7.4 %) cases with variants of uncertain significance (Table 2 and 3).

Table 2: Summary of tests requested and findings.

Results	Whole panel testing (%)	Single gene testing (%)	KMU (%)	Total (%)
Negative	105 (33.7)	33 (10.6)	23 (7.4)	161 (51.6)
Positive	106 (34.0)	20 (6.4)	2 (0.6)	128 (41.0)
VUS	17 (5.4)	2 (0.6)	4 (1.3)	23 (7.4)
Total	228 (73.1)	55 (17.6)	29 (9.3)	312 (100)

Table 3: Summary of test results by sample and test type.

Type of testing	Results	Sample type						
		Peripheral blood (%)	Skin (%)	Tissue (%)	Saliva (%)	cfDNA (%)	Unknown (%)	Total (%)
Whole panel testing (%)	NEG	68 (21.8)	18 (5.8)	17 (5.4)	0 (0)	1 (0.3)	1 (0.3)	105 (33.7)
	POS	46 (14.7)	34 (10.9)	25 (8.0)	0 (0)	1 (0.3)	0 (0)	106 (34.0)
	VUS	10 (3.2)	3 (1.0)	3 (1.0)	0 (0)	0 (0)	1 (0.3)	17 (5.4)
Single gene testing (%)	NEG	15 (4.8)	7 (2.2)	10 (3.2)	1 (0.3)	0 (0)	0 (0)	33 (10.6)
	POS	2 (0.6)	7 (2.2)	11 (3.5)	0 (0)	0 (0)	0 (0)	20 (6.4)
	VUS	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (0.6)
KMU (%)	NEG	21 (6.7)	0 (0)	0 (0)	2 (0.6)	0 (0)	0 (0)	23 (7.4)
	POS	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)	0 (0)	2 (0.6)
	VUS	4 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (1.3)
Total (%)		169 (53.8)	69 (22.8)	68 (21.8)	3 (0.9)	2 (0.6)	2 (0.6)	312 (100)

Abbreviations: KMU, known mutation; NEG, mutation-negative; POS, mutation-positive; VUS, variant of uncertain significance.

B. Pathogenic and Likely pathogenic variants (PATH/LP)

Out of 312 samples that were tested by deep targeted sequencing, 128 (41.0%) pathogenic and likely pathogenic variants were identified: 49 (38.3%) were detected in peripheral blood, 41 (32.0%) in skin fibroblasts, 37 (28.9%) in tissue samples, and 1 (0.8%) in cell-free DNA from CSF (cfDNA).

The analysis showed that *PIK3CA* variants were the most common among all the phenotypes except in patients with malformations of cortical development (MCD). *PIK3CA* variants accounted for 83 (64.8%) of total variants identified, and they were widely seen among patients with megalencephaly (MEG) and somatic overgrowth (OVG) phenotypes. In comparison, 17 (13.3%) *PTEN* variants were observed among patients with megalencephaly (MEG), malformations of cortical development (MCD), and somatic overgrowth (OVG) phenotypes. The other identified genes in MEG phenotypes beside *PIK3CA* and *PTEN* were *AKT3, PIK3R2, MTOR, CCND2, NSD1, BRWD3, PTCH1,* and *BRWD3*.

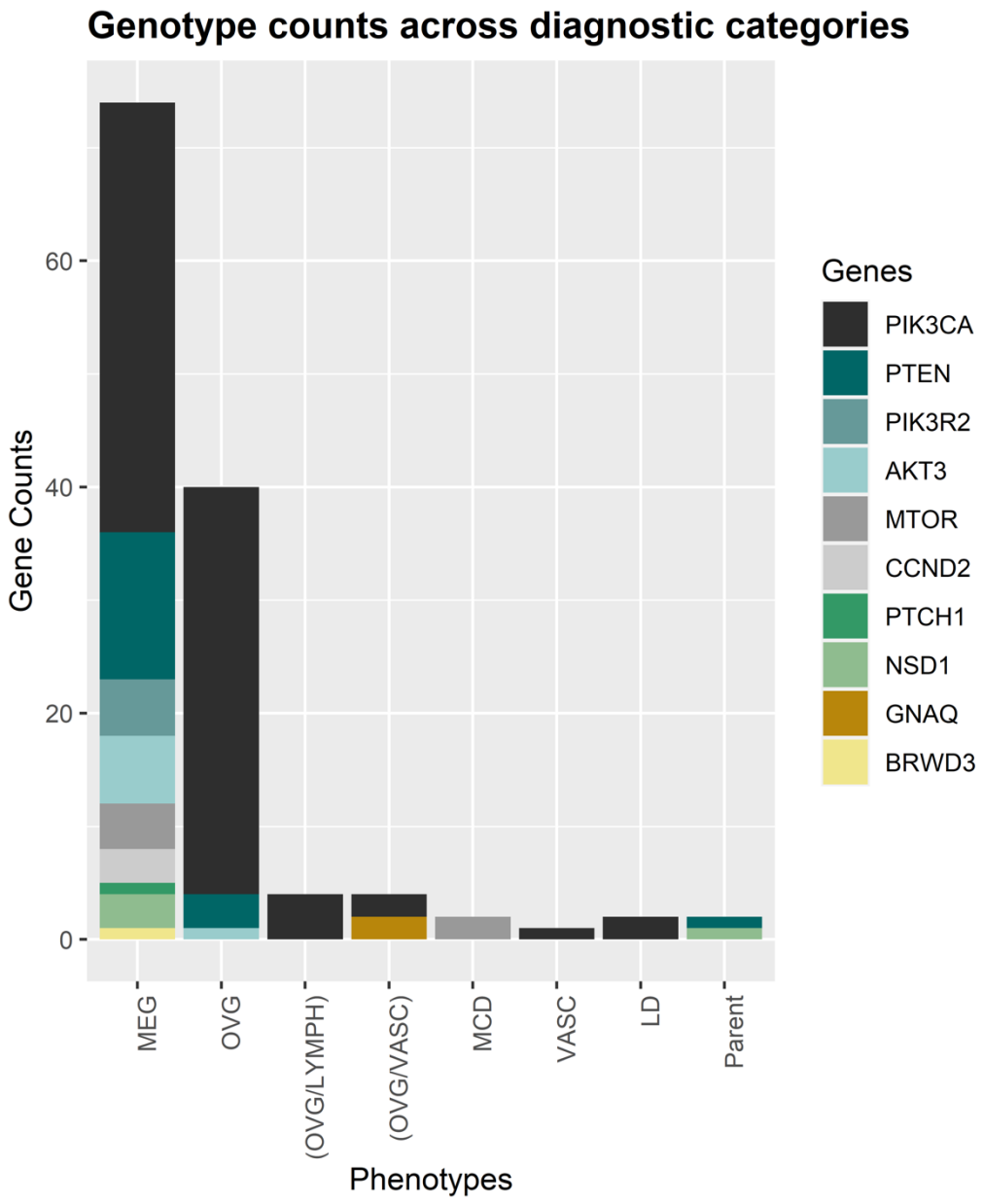
Other less frequent genes , such as *PTCH1* and *GNAQ* were observed only in certain phenotypes: MEG and somatic vascular malformations overgrowth (OVG/VASC) phenotypes, respectively. In other phenotypes such as lymphatic malformations, vascular malformations and learning delays only *PIK3CA* variants were identified and in MCD phenotypes only *MTOR* variants were observed (Figure 1).

The most common positive variants found in this panel associated with brain and somatic overgrowth phenotypes were in *PIK3CA* and *PTEN*. The distribution of variant allele fractions (VAFs) estimated for every case in both genes is shown in Figure 2. Most of the variant allele fractions in *PTEN* were between 40% to 50%, which are more likely to be heterozygous germline variants, whereas the majority of *PIK3CA* variant allele fractions were below 40% and are thus mostly mosaic. *PIK3CA* variants were more frequently detected in skin and tissue samples than in peripheral blood, while *PTEN* positive variants were mostly observed in peripheral blood (Table 4).

Additionally, we compared VAFs of all pathogenic variants based on the sample types. VAFs range was different in blood than skin and tissue samples as blood accounted for the most commonly tested sample. The VAF ranges in skin and tissues sample were relatively similar, as shown in Figure 3. We were able to identify one mosaic pathogenic variant in a cfDNA sample with a VAF of 3%.

In a total of 128 (41.1%) positive variants, 49 (38.3%) were detected in peripheral blood samples, 41 (32.0%) in skin, 37 (28.9) in tissue, and 1 (0.8%) in cfDNA. No positive variants were detected in any of the saliva samples.

Figure 1**:** Positive gene counts by phenotypes.



*Abbreviations***:**
MEG: Megalencephaly (brain overgrowth phenotypes)
OVG: Somatic overgrowth
LYMPH: Lymphatic malformations
VASC: Vascular malformations
MCD: Malformations of Cortical Development (e.g. cortical dysplasia)
LD: Learning Delays / non-specific developmental delays)

Figure 2: Variant allele fractions (VAFs) in PTEN and PIK3CA.

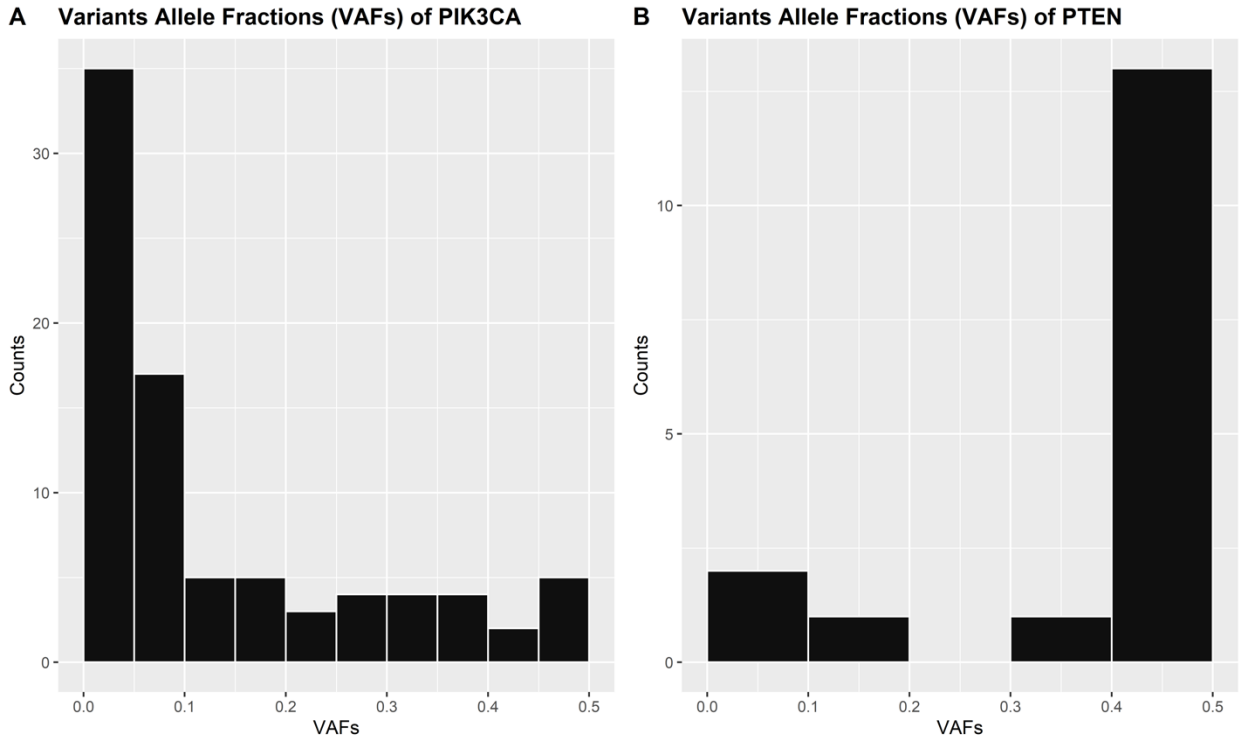
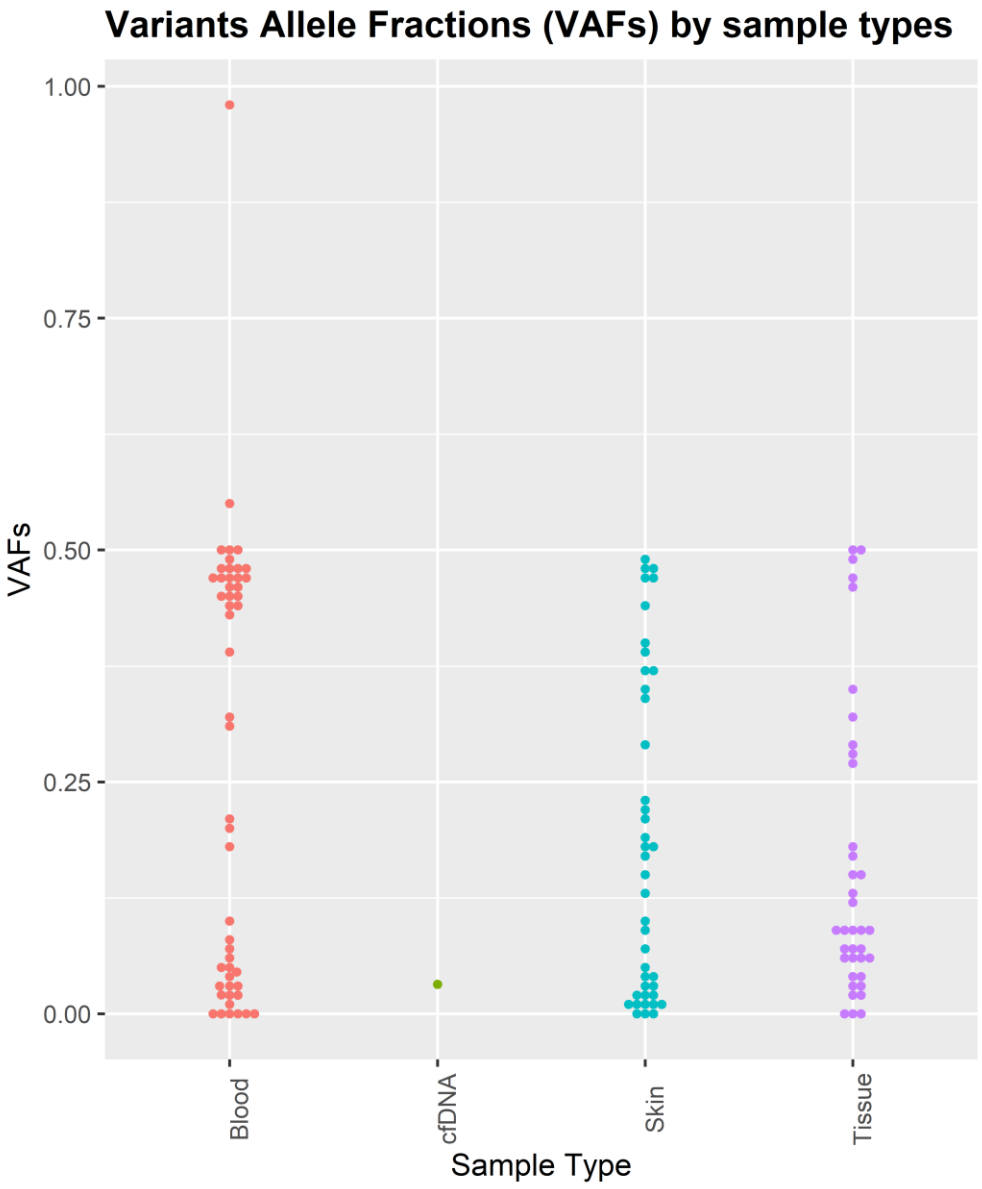


Table 4: Summary of pathogenic variants by sample type.

Genes	Peripheral blood (%)	Skin (%)	Tissue (%)	cfDNA (%)	Total (%)
PIK3CA	22 (17.2)	30 (23.4)	30 (23.4)	1 (0.8)	83 (64.8)
PTEN	12 (9.4)	1 (0.8)	4 (3.1)	0 (0)	17 (13.3)
AKT3	3 (2.3)	4 (3.1)	0 (0)	0 (0)	7 (5.5)
MTOR	3 (2.3)	2 (1.6)	1 (0.8)	0 (0)	6 (4.7)
PIK3R2	3 (2.3)	2 (1.6)	0 (0)	0 (0)	5 (3.9)
NSD1	2 (1.6)	0 (0)	2 (1.6)	0 (0)	4 (3.1)
CCND2	2 (1.6)	1 (0.8)	0 (0)	0 (0)	3 (2.3)
PTCH1	1 (0.8)	0 (0)	0 (0)	0 (0)	1 (0.8)
BRWD3	1 (0.8)	0 (0)	0 (0)	0 (0)	1 (0.8)
GNAQ	0 (0)	1 (0.8)	0 (0)	0 (0)	1 (0.8)
Total	49 (38.3)	41 (32.0)	37 (28.9)	1 (0.8)	128 (100)

Figure 3: Variant allele fractions (VAFs) by sample type.



C. Variant of uncertain significance (VUS)

Of the 213 samples, 23 (7.4%) had variants of uncertain significance (VUS), 18 in patients with megalencephaly and other growth disorders, and 5 in parents. Only one VUS was found in *PIK3CA* and one in *PTEN*, and the other VUS variants were mostly in minor genes such as *KIF7*, *GLI3*, *GPC3*, *RIN2* and *TSC2*. VUS were mostly identified in blood samples. The VAFs ranged from 0.1%to 100%.

D. Multi-sample testing

Out of 265 cases, one sample was submitted for 226 cases. In 34 cases, two samples were submitted, and in 5 cases three- samples were submitted for analysis (Figure 4A). Some patients submitted two similar sample types, and some submitted different types of samples. The total number of samples was 48 out of 312, 26 patients had blood samples in addition to tissue or skin samples.

From cases with multiple samples, there were 50 case with pathogenic variants (59.5%). The highest number of pathogenic variants were observed in tissue samples 20 (23%) followed by 15 (17.9%) in peripheral blood, 14 (16%) in skin fibroblasts, and 1 (1.2%) in cfDNA. Out of the number of pathogenic variants, 42 were in *PIK3CA*, the others were in *MTOR*, *CCND2* and *PTEN*. The VAFs of pathogenic variants ranged from 0.1% to 50%. In multi-sample cases two VUS were seen in tissue samples, one in skin fibroblasts, and 1 in blood samples, in total of 4 (4.8%). No positive variants detected in 30 (35.7%) of samples from multi-sample cases (Table 5).

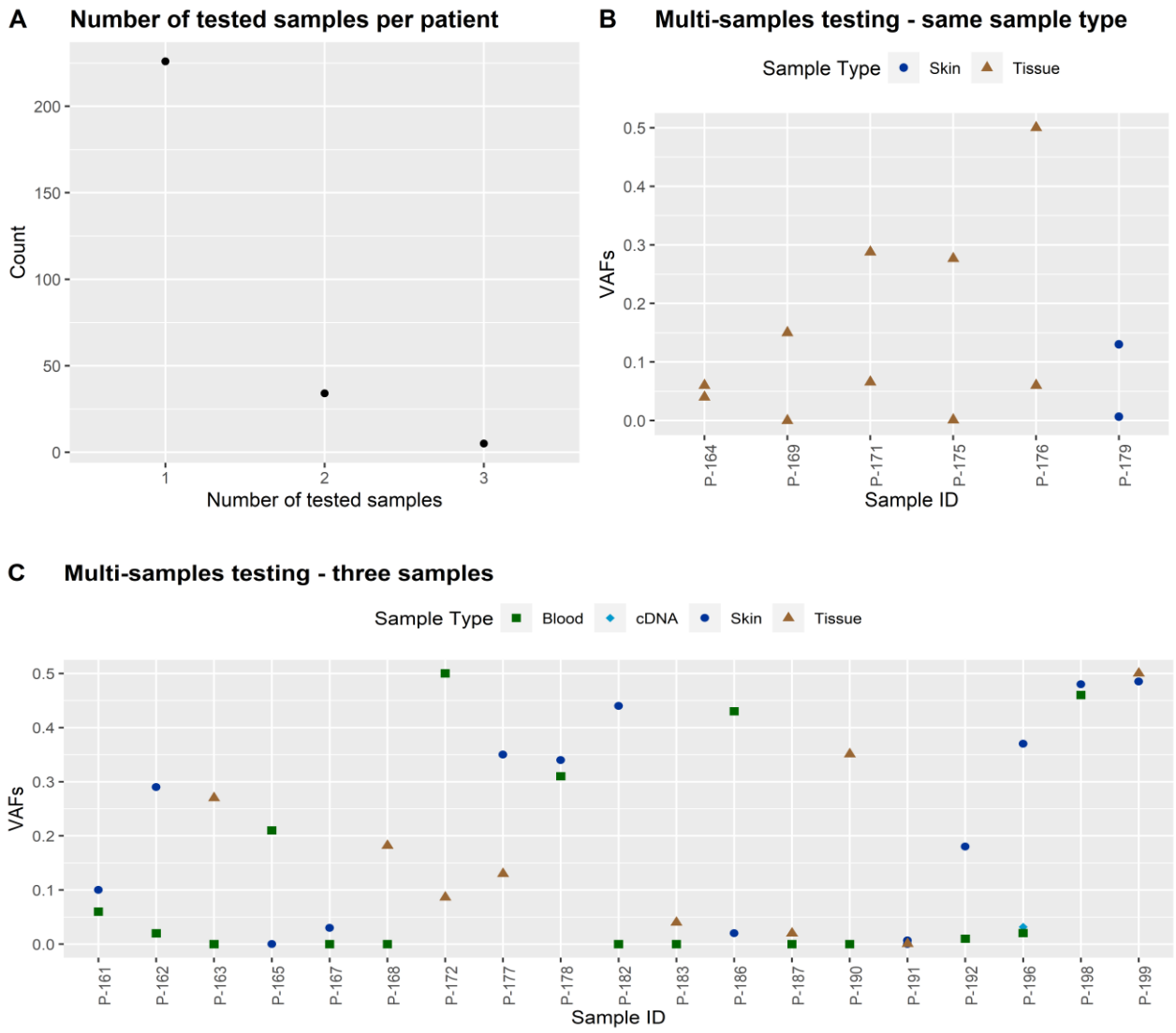
Table 5: Summary of multi-sample testing

Abbreviations: NEG, mutation-negative; POS, mutation-positive; VUS, variant of uncertain significance.

Sample type	POS (%)	VUS (%)	NEG (%)	Total (%)
Peripheral blood	15 (17.9)	1 (1.2)	12 (14.3)	28 (33.3)
Tissues	20 (23.8)	2(2.4)	13 (15.5)	35 (41.7)
Skin fibroblasts	14 (16.7)	1 (2)	5 (6.0)	20 (23.8)
Cell-free DNA	1(1.2)	0 (0)	0 (0)	1(1.2)
Total	50 (59.5)	4 (4.8)	30 (35.7)	84 (100)

The data presented in Figure 4B show that the VAF in most cases differ among similar samples sources . However, The VAFs presented in Figure 4C showed that in most cases, the VAFs in blood samples were lower than in skin and tissue samples; however , in three cases the VAFs in blood samples were higher than in VAFs in skin and tissue.

Figure 4: ***A:*** number of tested samples per patient. **B:** positive variants among patients with two similar types of samples (Skin and Tissue). **C**: positive variants among patients with two or three different types of samples.



E. Novel and rare mutations

We also identified 18 (5.7%) novel variants that were not reported in any database: 15 variants were identified in patients megalencephaly, while 3 variants were in patients with somatic overgrowth phenotype. These variants were found in *PIK3CA*, *PTEN*, *AKT3*, *MTOR*, *CCND2*, *NSD1*, *BRWD3*, and *SETD2*. Most of them were detected in blood 10 (3.2%) then in skin 5 (1.6%), and finally in 3 (0.9%) in tissue. The VAFs

ranged from 0.1% to 98%. All patients with novel variants were tested for multi-gene panel (MEGPX) except for one patient who was tested for single gene (*PIK3CA*).

F. Parental testing

In this cohort, we tested 33 samples from parents and other family members for 18 individuals to follow up or to resolve interpretation of results. Both parents were tested in 10 instances, and for the rest we tested at least one parent, siblings, or relatives. Most relative samples were tested for known mutation or for a single gene and only one parent tested for the whole gene panel (MEGPX). Two pathogenic heterozygous variants were identified in parents; one each in *PTEN* and *NSD1*. Five VUS were identified in unaffected parents allowing updated classification of these variants. No reported variants were detected in any siblings or other relatives.

CHAPTER V
DISCUSSION

A. The applications and utility of the Megaplex panel for the detection of mosaic mutation underlying megalencephaly and other overgrowth phenotypes

The megaplex gene panel identified causative genes in patients with MEG and other similar phenotypes. As a result, it facilitated the diagnosis of complex genetic disorders and detected both somatic and germline mutations in different types of tissue samples. Our findings also emphasize that there is benefit in testing parents of affected children to determine if pathogenic variants were inherited or *de novo*.

The advances in NGS testing technology allowed us to detect mosaic mutations at read frequencies above about 1% in various types of samples. Some of the mosaic mutations were only detectable in affected tissues and some were detectable in both blood and tissue. This could be because mosaic mutations are often tissue-specific and manifest in affected tissues such as lesional skin fibroblasts, and brain. The most common genes identified in patients with megalencephaly related phenotypes were *PIK3CA*, *PTEN*, *PIK3R2*, *AKT3*, and *MTOR*. These genes account for 92% of pathogenic mutations. While other genes, including *CCND2*, *GNAQ*, *MED12*, *NSD1*, and *PTCH1*, account for 8% of pathogenic mutations.

B. Comparing mosaic multi-gene panels versus other

Some of the MEGPX key features that other panels may lack are high sensitivity and high depth of coverage, which ranged between 500x to more than 1000x, and the read length ranged between 2 × 150 to

2 × 300 base pairs (bp). These features allowed targeted NGS to detect low level mosaic variants, whereas other sequencing techniques could not detect them.

C. The variation in detecting mosaic mutations based on sample types

Although the highest number of positive variants were detected in blood, the VAF ranges were inconsistent, and the highest portion of variants was between 0% to 10% and 40% to 50%, and few variants in between. Underlying possibilities for this distribution include the nature and types of phenotypes that are primarily targeted by this panel – as *PIK3CA*-related disorders, for example, are typically associated with very low-level mosaic mutations, whereas other disorders such as *PTEN*-related disorders and others are associated with germline variants. The number of positive variants in skin and other tissues were less than the ones in blood, but the VAFs in skin and tissue were more consistent, and a high number of mosaic variants were between 0.1% to around 20% (Figure 3). The variation in VAFs distribution probably explains that most of the variants in peripheral blood tend to be germline at high percentages, while VAFs in skin and tissue were expected to be mainly mosaic. The mosaic variants with low VAFs (0%-10%) in peripheral blood are not always expected to detect, especially *PIK3CA* variants (Moog et al.,2020). The severity of phenotypes can possibly contribute to the variation of VAFs distribution in peripheral blood.

D. *PIK3CA* and *PTEN*

This was one of the largest studies of clinically identified *PIK3CA* variants. *PIK3CA* mosaic variants were most likely to manifest in skin and tissue as the highest number of *PIK3CA* mosaic variants were not detected in blood, whereas positive *PTEN*, *MTOR*, *PIK3R2*, *CCND2*, *PTCH1*, *BRWD3* variants were highly seen in blood samples. One of the genetic variant in (*GNAQ*) was only seen in a skin sample.

The *PIK3CA* variants allele fractions ranged from 0.1% to 50%. Only one patient had a likely-pathogenic *PIK3CA* mutation with VAF 50%. This variant was found in an affected tissue sample, so it may not be heterozygous in other tissues. This was a novel variant that has not been reported before. The samples that tested positive for *PIK3CA* mutations were mostly from different types of tissue. This is in contrast to sample positive for *PTEN* mutations, where the samples were mostly from blood. Although there have been larger studies of *PTEN*, the variation in phenotypic findings observed with *PTEN* is still surprising. The *PTEN* variants allele fractions ranged from 0.3% to 50%. There were two patients with VAFs 50%, both of which were from blood samples. For *PTEN*, we observed similar phenotypes in patients with mosaic and heterozygous variants. Of 17 patients with *PTEN* mutations there were thirteen with MEG, three with somatic phenotype, and one in a parent.

E. Novel Genotype-Phenotype Correlation

MEG and overgrowth were the most common features present in individuals with novel or rare *PIK3CA* variants. In one case with the p.Gly914Arg *PIK3CA* variant, the patient had more affected organs beside brain including liver, and limbs, and facial characteristic. Three *PTEN* rare variants reported in patients with macrocephaly, CLOVES syndrome and MCAP. Another rare likely pathogenic *PTEN (*p.Phe81Cys) variant was observed in a patient with polymicrogyria *(*PMG) while it was reported as a pathogenic variant in an exome sequencing project and in the human gene mutation database associated with breast cancer (Stenson et al., 2014). Previous studies have shown that germline and mosaic *PIK3CA* variants are associated with polymicrogyria (Mirzaa et al., 2015). Although *MTOR* variants were highly observed in patients with FCD, in one case a novel pathogenic *MOTR* (p.Ile1417Thr) variant was seen in patient with syndromic MEG and had hydrocephalus, hemangioma, and butterfly vertebrae.

Most of *AKT3* variants in this cohort were seen in MEG cases and few in patients with overgrowth disorders. A novel *AKT3* pathogenic variant (p.Asp322Gly) was detected in a fetal blood sample at age 32 weeks and 7 days, and the fetus had PMG, ventriculomegaly, and MEG. During pregnancy the amount of amniotic fluid around the fetus exceeds the normal level and the fetus had a thick corpus callosum and it found to be associated with other brain abnormalities such as macrocephaly (Lerman-Sagie et al., 2009).

F. Comparisons with targeted ddPCR versus exome and genome sequencing

Next-generation sequencing (NGS) requires target enrichment, and it is focused on a specific set of genes while whole-genome sequencing and whole-exome sequencing are not. MEGPX gene panels targeted specific genes at a higher sequencing depth and lower cost per sample than whole exome and whole genome sequencing. Although NGS and Digital polymerase chain reaction (ddPCR) have high sensitivity and can detect low-level mosaic variants, ddPCR is not a high throughput tool and is more expensive when testing one sample for multiple probes. ddPCR is ideal to use as a confirmation test or to test samples for specific variants. Therefore, the targeted gene capture technique is a powerful tool for mosaic variants detection and the most cost-effective technique compared to others.

CHAPTER VI
LIMITATIONS

Limitations of our study include incomplete information for some patients, mainly earlier cases. Family history and phenotype-related data available to the laboratory were often limited, as we were

dependent on reports shared by providers with laboratory order forms. Another limitation is the uncertainty in testing tissues; we are not assured whether they were affected or not. This depends on the provider taking the biopsy. Other tested tissue samples did not indicate if they were taken from affected or unaffected tissue. We tested a limited number of saliva and cfDNA samples. Therefore, we could not compare the VAF percentages to the ones in blood, tissue, and skin samples. Additionally, our retrospective observational study was not designed specifically to determine which source is better, and indeed we cannot tell which is optimal from our data. Some genes were identified as associated with a broad spectrum of megalencephaly, and other overgrowth disorders are not available on our panel. Therefore, the future plan is to expand the panel by adding new genes.

CHAPTER VII
CONCLUSION

Our results demonstrate deep next-generation sequencing using a targeted capture panel allows identifying a broad spectrum of mosaic mutations in patients with megalencephaly and other related phenotypes, including detecting low levels of mosaicism in a diverse type of samples. These mosaic mutations might not be detected using Sanger sequencing. Also, it showed that NGS could detect mosaic variants in peripheral blood besides skin or tissue sample; the only difference was in VAF percentage.

APPENDIX A

Table 1: Megalencephaly Gene Panel.

Gene	Disease	Reference (PMID)
ABCC9	Cantu syndrome	22610116
AKT1	Proteus syndrome	21793738,22876373
AKT2	Asymmetric overgrowth with hypoglycemia	21979934
AKT3	Hemimegalencephaly, Megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome, focal cortical dysplasia	22729224,25722288
BRWD3	X-linked intellectual disability and macrocephaly	17668385
CCND2	Megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome	24705253
CDKN1C	IMAGe syndrome, Beckwith Wiedemann	20301568,24624461
CUL4B	Syndromic X linked mental retardation	17236139,25385192
DEPDC5	Focal epilepsy with or without focal cortical dysplasia, familial focal epilepsy with variable foci	23542701,23542697
DNMT3A	Tatton-Brown-Rahman Syndrome	24614070
EED	Overgrowth and macrocephaly	25787343
EZH2	Weaver syndrome	23865096
GLI3	Greig cephalosyndactyly, Acrocallosal syndrome	20301619,12414818
GNAQ	Sturge Weber Syndrome, capillary malformation (port-wine)	23656586
GNAS	Fibrous Dysplasia/McCune-Albright Syndrome, Pseudopseudohypoparathyroidism, pseudohypoparathyroidism 1A, 1B, and progressive osseous heterotopia	25719192
GPC3	Simpson-Golabi-Behmel syndrome	20301398
HEPACAM	Megalencephalic Leukoncephalopathy with Subcortical Cysts	20301707
KCNJ8	Cantu syndrome	24700710
KIF7	Macrocephaly, multiple epiphyseal dysplasia and distinctive facies	22587682
MED12	Opitz-Kaveggia syndrome	20301719
MLC1	Megalencephalic Leukoncephalopathy with Subcortical Cysts	20301707
MTOR	Megalencephaly, hemimegalencephaly, focal cortical dysplasia	25799227
NFIA	Macrocephaly and intellectual disability	19763616,26997977
NFIX	Overgrowth and macrocephaly	25118028
NSD1	Sotos syndrome	20301652

PIK3CA	PIK3CA-related overgrowth syndromes (PROS), Megalencephaly with capillary malformation (MCAP), CLOVES syndrome, megalencephaly, hemimegalencephaly, focal cortical dysplasia, Klippel Trenauney syndrome, isolated lymphatic malformations, isolated venous malformations	23946963,25681199, 22729224,26637981
PIK3R2	Megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome, polymicrogyria	22729224,26520804
PTCH1	Nevoid basal cell carcinoma syndrome, Basal Cell Nevus Syndrome, Gorlin syndrome	20301330
PTEN	Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome (BRRS) *PTEN* related hamartoma syndrome, autism, focal cortical dysplasia	20301661
RAB39B	X-linked intellectual disability and macrocephaly with autism	20159109
RIN2	MACS syndrome (macrocephaly, alopecia, cutis laxa and scoliosis)	19631308
RNF135	Overgrowth and macrocephaly	17632510
SETD2	Overgrowth and macrocephaly	24852293
STRADA	PMSE syndrome (polyhydramnios, megalencephaly, symptomatic epilepsy)	17522105
TBC1D7	Megalencephaly with intellectual disability, autosomal recessive	23687350,24515783
TSC1	Tuberous Sclerosis	20301399
TSC2	Tuberous Sclerosis	20301399

REFERENCES

1. Alcantara, Diana, Timms, Andrew E, Gripp, Karen, Baker, Laura, Park, Kaylee, Collins, Sarah, . . . Mirzaa, Ghayda M. (2017). Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly. *Brain (London, England : 1878), 140*(10), 2610-2622.

2. Biesecker, Leslie G, & Spinner, Nancy B. (2013). A genomic view of mosaicism and human disease. Nature Reviews. *Genetics, 14*(5), 307-320.

3. Butler, M G, Dasouki, M J, Zhou, X-P, Talebizadeh, Z, Brown, M, Takahashi, T N, . . . Eng, C. (2005). Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *Journal of Medical Genetics, 42*(4), 318-321.

4. Hucthagowder, V, Shenoy, A, Corliss, M, Vigh-Conrad, K.A, Storer, C, Grange, D.K, & Cottrell, C.E. (2017). Utility of clinical high-depth next generation sequencing for somatic variant detection in the PIK3CA-related overgrowth spectrum. *Clinical Genetics, 91*(1), 79-85.

5. Jamuar, Saumya S, Lam, Anh-Thu N, Kircher, Martin, D'Gama, Alissa M, Wang, Jian, Barry, Brenda J, . . . Walsh, Christopher A. (2014). Somatic Mutations in Cerebral Cortical Malformations. *The New England Journal of Medicine, 371*(8), 733-743.

6. Lerman-Sagie, T, Ben-Sira, L, Achiron, R, Schreiber, L, Hermann, G, Lev, D, . . . Malinger, G. (2009). Thick fetal corpus callosum: An ominous sign? *Ultrasound in Obstetrics & Gynecology, 34*(1), 55-61.

7. Lupski, J. R. (2013). Genome Mosaicism--One Human, Multiple Genomes. *Science (American Association for the Advancement of Science), 341*(6144), 358-359.

8. Mirzaa, Ghayda, Timms, Anew E, Conti, Valerio, Boyle, Evan August, Girisha, Katta M, Martin, Beth, . . . Dobyns, William B. (2016). PIK3CA-associated developmental disorders exhibit distinct classes of mutations with variable expression and tissue distribution. *JCI Insight, 1*(9), JCI insight, 2016, Vol.1 (9)

9. Mirzaa, Ghayda M, Dr, Conti, Valerio, PhD, Timms, Andrew E, PhD, Smyser, Christopher D, MD, Ahmed, Sarah, MD, Carter, Melissa, MD, . . . Guerrini, Renzo, Prof. (2015). Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2 , in perisylvian polymicrogyria: A next-generation sequencing study. *Lancet Neurology, 14*(12), 1182-1195.

10. Moog, Ute, Felbor, Ute, Has, Cristina, & Zirn, Birgit. (2020). Disorders Caused by Genetic Mosaicism. *Deutsches Ärzteblatt International, 116*(8), 119-125.

11. Oneda, Beatrice, Asadollahi, Reza, Azzarello-Burri, Silvia, Niedrist, Dunja, Baldinger, Rosa, Masood, Rahim, . . . Rauch, Anita. (2017). Low-Level Chromosomal Mosaicism in Neurodevelopmental Disorders. *Molecular Syndromology, 8*(5), 266-271.

12. Park, Hyun Jin, Shin, Chang Ho, Yoo, Won Joon, Cho, Tae-Joon, Kim, Man Jin, Seong, Moon-Woo, . . . Ko, Jung Min. (2020). Detailed analysis of phenotypes and genotypes in megalencephaly-capillary malformation-polymicrogyria syndrome caused by somatic mosaicism of PIK3CA mutations. O*rphanet Journal of Rare Diseases, 15*(1), 205.

13. Pavone, Piero, Praticò, Andrea Domenico, Rizzo, Renata, Corsello, Giovanni, Ruggieri, Martino, Parano, Enrico, & Falsaperla, Raffaele. (2017). A clinical review on megalencephaly: A large brain as a possible sign of cerebral impairment. *Medicine (Baltimore), 96*(26), E6814.

14. Qin, Lan, Wang, Jing, Tian, Xia, Yu, Hui, Truong, Cavatina, Mitchell, John J, . . . Wong, Lee-Jun C. (2016). Detection and Quantification of Mosaic Mutations in Disease Genes by Next-Generation Sequencing. *The Journal of Molecular Diagnostics : JMD, 18*(3), 446-453.

15. Riviere, Jean-Baptiste, mirzaa, Ghayda M, Worthylake, Thea, Sullivan, Christopher T, Ward, Thomas R, Butler, Hailly E, . . . Nikkel, Sarah M. (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.

16. Stenson, Peter D, Mort, Matthew, Ball, Edward V, Shaw, Katy, Phillips, Andrew D, & Cooper, David N. (2014). The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics*, 133(1), 1-9.

17. Spinner, Nancy B, & Conlin, Laura K. (2014). Mosaicism and clinical genetics. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics, 166*(4), 397-405.

18. Yehia, Lamis, Ngeow, Joanne, & Eng, Charis. (2019). PTEN-opathies: From biological insights to evidence-based precision medicine. *The Journal of Clinical Investigation, 129*(2), 452-464.