RESEARCH PROJECT CALL
'TODOS SOMOS RAROS'
TECHNICAL PROPOSAL

File number:_____

TITLE: *PIK3CA* Overgrowth Syndromes: Diagnosis, Phenotype and Clinical Guidelines

PRINCIPAL INVESTIGATOR: Martinez Gonzalez, Victor Manuel

TYPE OF PROJECT (please underline as appropriate): **INDIVIDUAL** **COOPERATIVE**

PROJECT COORDINATOR:

DURATION (please underline as appropriate): **1 Year** **2 Years**

SUMMARY (Objectives and Methodology) **(Maximum of 1 page, Arial 10)**

RESUMEN EN ESPAÑOL

La presencia de múltiples clones de células con varios genotipos en un mismo individuo se conoce como "mosaicismo somático", un mecanismo patogénico en el que las mutaciones no están presentes en la línea germinal, sino que surgen como un evento post-cigótico, pudiendo dar origen tanto a neoplasias como a síndromes genéticos del desarrollo muy variados. Recientemente, se ha descrito un grupo de síndromes de sobrecrecimiento segmentario causado por mutaciones somáticas en el gen *PIK3CA*. Este grupo incluye síndromes previamente considerados independientes como la Megalencefalia-Malformación Capilar (MCAP) o el CLOVES, junto con algunas variantes como la fibrodisplasia adiposa aislada o la hemi-megalencefalia. Sin embargo, ya que los mosaicismos somáticos producen una expresión clínica variable tanto en severidad como en localización, el espectro fenotípico de este grupo de sobrecrecimientos segmentarios debe todavía ser bien definido para evitar que pacientes que presenten manifestaciones fenotípicas no habituales sean excluíos del diagnóstico. Otra de las complicaciones asociadas con el diagnóstico molecular de mutaciones somáticas es que éstas se pueden presentar en mosaicos bajos, haciéndolas difícil de detectar con técnicas moleculares clásicas.

Este proyecto utilizará Secuenciación Masiva [Next Generation Sequencing (NGS)] para desarrollar un protocolo experimental y bio-informático, robusto y aplicable a la práctica clínica, para el diagnóstico de pacientes con síndromes de sobrecrecimiento asociado a alteraciones en *PIK3CA*, lo que además servirá como modelo para realizar un panel diagnóstico por NGS para la detección de mosaicos bajos en otros síndromes genéticos del desarrollo. Como parte del proceso, evaluaremos las características clínicas que podrían sugerir el diagnóstico en un paciente a pesar de no cumplir con los criterios estrictos establecidos para MCAP y CLOVES. Esto permitirá aumentar el porcentaje de pacientes diagnosticados. También desarrollaremos guías clínicas tanto para los profesionales de la salud como para afectados y familiares, dándoles un fácil acceso a esta (muchas veces escasa) información. Así mismo daremos apoyo a la creación de una asociación de pacientes específica para éstas patologías.

RESEARCH PROJECT CALL
'TODOS SOMOS RAROS'
TECHNICAL PROPOSAL

File number:_____

TITLE: *PIK3CA* Overgrowth Syndromes: Diagnosis, Phenotype and Clinical Guidelines

PRINCIPAL INVESTIGATOR: Martinez Gonzalez, Victor Manuel

TYPE OF PROJECT (please underline as appropriate): **INDIVIDUAL** **COOPERATIVE**

PROJECT COORDINATOR:

DURATION (please underline as appropriate): **1 Year** **2 Years**

SUMMARY (Objectives and Methodology) **(Maximum of 1 page, Arial 10)**

The presence of multiple clones of cells with various genotypes in the same individual is known as "somatic mosaicism", a pathogenic mechanism in which the mutations are not present in the germ-line but arise as a post-zygotic event, causing both cancer and highly variable developmental genetic syndromes. Recently, a group of segmental overgrowth syndromes caused by somatic mutations in the *PIK3CA* gene have been described. The group includes previously considered separate syndromes as Megalencephaly-Capillary malformation (MCAP) and CLOVES, along with some other variants as isolated adipose fibrodysplasia or hemimegalencephaly. However, as somatic mosaicism makes clinical expression variable in severity and location, the phenotypic spectrum of this group of segmental overgrowth syndromes is still to be elucidated, in order to prevent patients with unusual phenotypic manifestations to be excluded from the diagnosis. Another complication associated with the diagnosis of somatic mutations is that they can occur in low mosaics, making them difficult to detect by standard molecular techniques.

This project will use Next Generation Sequencing (NGS) to perform an experimental and bioinformatic protocol, reliable and applicable to clinical practice, for the diagnosis of patients with *PIK3CA* segmental overgrowth syndromes, as a model for an NGS diagnostic panel to detect low mosaic mutations in developmental syndromes. As part of the clinical diagnostic process, we will evaluate those key clinical features that could suggest the diagnosis of a patient despite not meeting the strict diagnostic criteria established for MCAP and CLOVES. This will allow increasing the percentage of diagnosed patients. We will also develop clinical guidelines for both health professionals and patients and their families to give them access to this generally limited information, and we will promote the creation of a specific association of patients for these diseases.

RESEARCH PROJECT TECHNICAL PROPOSAL

BACKGROUND AND CURRENT STATUS

Aim of the project, background and current status of scientific-technological knowledge, national and international research groups working in this specific line or related
Cite references in the section: Relevant Bibliography **(Maximum of 3 pages, Arial 10)**

Somatic cells, from its first post-zygotic division, may accumulate genetic changes, so that cells from different tissues or even within the same tissue differ genetically. The presence of multiple clones of cells with various genotypes in the same individual is known as "somatic mosaicism". In recent years, the search for the genetic causes of many developmental disorders has enabled define somatic mosaicism as a pathogenic mechanism in which the mutations are not present in the germ-line but arise as a post-zygotic event, causing highly variable phenotypes sometimes difficult to correlate clinically [1-8].

The first description of a genetic disorder caused by this type of mechanism was achieved in Proteus syndrome, caused by the activation of somatic mutations in the *AKT1* gene, a kinase acting downstream PI3K and therefore involved in proliferation and survival [7]. Mutations in this gene are likely to be only tolerated in mosaic state. Since this first finding, the list of syndromes associated to somatic mutations present in mosaic have increased and include multiple genes in different molecular pathways. Among them we can find the PI3K/PTEN/AKT/mTOR [2; 7; 8] and the RAS/MAPK [1; 9] pathways, mainly associated with overgrowth syndromes, but also well known for their involvement in a variety of oncogenic processes.

Mutations in some of these genes cause well-defined genetic syndromes when present in the germ-line, but also can be found as somatic mosaicism in more than one phenotypically distinct segmental disorder. As an example we have mutations in the *HRAS* gene. When they are present in germinal state cause Costello syndrome, but in somatic mosaicism can cause different entities as the keratinocytic epidermal nevus syndrome or the Schimmelpenning syndrome [1]. Thus, the clinical features of the disorder are determined by the time at which the mutation arises during the embryonic development, the level of activation of the possible molecular pathways affected, the cell type involved, and the genetic background of each individual [10].

Recently, a group of segmental overgrowth syndromes caused by mutations in the *PIK3CA* gene have been described [2-5]. *PIK3CA* is an oncogene already known because it is frequently implicated in various types of neoplasms due to somatic mutations associated to gain of function. The group of *PIK3CA* segmental overgrowths includes previously considered separate syndromes as Megalencephaly-Capillary malformation (MCAP) and CLOVES.

Common clinical features comprise overgrowth and vascular malformations. However, the differential features between them seem to depend on the affected tissues. The overgrowth in MCAP is associated to progressive hemimegalencephaly or megalencephaly, while in CLOVES is more frequent the lipomatous and bone overgrowth. The main vascular anomaly in MCAP is the cutaneous capillary malformation, particularly striking on facial mid-line, while in CLOVES are frequent the low-flow capillary, venous and lymphatic malformations, or even the high-flow arterio-venous malformations. Connective tissue dysplasia is more common in MCAP and

macrodactyly in CLOVES. Some variants of these two syndromes such as the isolated adipose fibrodysplasia or the hemimegalencephaly are also associated with segmental overgrowth caused by somatic mutations in *PIK3CA* [11].

With the publication in recent years of a large number of cases in the literature, it has been possible to outline the clinical characteristics that define MCAP and CLOVES [11; 12]. However, these clinical criteria could be leaving out a large group of patients with low levels of somatic mosaicism for mutations in the *PIK3CA* gene. As somatic mosaicism makes clinical expression variable in severity and location, the phenotypic spectrum of this group of segmental overgrowth syndromes is still to be elucidated.

In 2008, two of the researchers of this project described the CLAPO syndrome [13], characterized by Capillary malformation (CM) of the lower lip; Lymphatic malformation (LM) of the face/neck; Asymmetry of face and limbs, and Partial/generalized Overgrowth. This syndrome, whose cause is unknown to date, shares many of the features present in overgrowth syndromes associated PIK3CA, and therefore could be caused by mutations in this gene, or genes within the same pathway. Because of its clinical features is also possible that the alteration could be present as somatic mosaic. This shows that the phenotypic spectrum of the segmental overgrowth remains to be defined.

Both MCAP and CLOVES are diagnosed in less than 5 cases per 10,000 inhabitants, so they are considered rare diseases. This low prevalence together with its recent description, the limited current knowledge about their pathogenic mechanisms, and the generalized low institutional visibility of the rare diseases, leads to low diagnosis of patients. Affected individuals often need several years to find physicians able to recognize and diagnose these syndromes, so they cannot provide them adequate follow-up or give the necessary genetic counseling required by both patients and relatives.

Although the genetic cause of the *PIK3CA* associated overgrowths has been described, the molecular diagnostic in clinical practice remains elusive. Unlike other genetic syndromes in which the detection of mutations may be performed by conventional methods by using DNA obtained from any source, as saliva or peripheral blood, in the case of *PIK3CA* associated overgrowths there are two obstacles to overcome: 1) the mutation is usually present only in the affected tissue, so a biopsy is required, and 2) mutations within the same tissue can be found in varying percentages of cells, so that conventional methods may not detect low mosaicism.

Next generation sequencing (NGS) or massive parallel sequencing, defines a series of new technologies that have revolutionized the fields of basic and clinical research. NGS is a powerful tool for the discovery of genetic variation by providing rapid and complete sequencing of a set of candidate genes, the entire exome or even the whole genome [14; 15]. During the next few years we are likely to see public databases filled with hundreds of thousands of terabases of human sequence data. Anticipating that most resequencing studies will produce data for multiple individuals, we realize that there is available information about sample allele frequency and nucleotide-read error in these data. Consequently, genotype-calling algorithms could be improved by incorporating this information into the algorithm. In this project, the expertise in this specific technique and technology will be offered by INGEMM, where the group already has experience in using 3 different technologies (Roche, MySeq and IonTorrent). In addition, INGEMM has a Section of Bioinformatics since recent technological advances in NGS are producing an unprecedented volume of sequencing data that need mathematical and biocomputation analysis.

Next Generation Sequencing technique has already been used to solve the detection of low mosaics in some developmental syndromes. Specifically in *PIK3CA*, in a work studying 6 individuals clinically diagnosed with CLOVES, mutant alleles were detected at frequencies ranging from 3 to 30%. To detect and classify low frequency variants, first eliminating false positive sequencing errors, authors ranked the variants according to the number of reads of the mutant allele, without a minimum threshold of read depth [4]. In a later work, using similar criteria, the authors used as threshold of at least 4 mutated reads, which allowed detecting low mosaics in 5 out of 6 patients with MCAP [2]. However, in these two studies the overall read depth used —the number of times each nucleotide is read— was relatively low. This is an effective strategy to find heterozygous mutations (50%) but not as good to detect low mosaics. Using another approach, a more recent study used bioinformatics specific criteria (allele frequency spectrum and variant calling) to detect low mosaics in 3 patients with megalencephaly syndromes [16].

Despite these efforts, there is still no standardized method for the detection of such low mosaic mutations, and the published parameters for detection of low mosaics so far, are more a description made after the study, than a real study designed to establish a real detection protocol. It is needed to design studies to establish protocols that take into account the read depth, the type of tissue studied, and the bioinformatic parameters needed to detect low mosaics, avoiding the occurrence of false positives.

Therefore, this project will take a dual approach to the study of patients with overgrowth associated to *PIK3CA*. On the one hand, the development of a protocol for experimental and bioinformatic detection of low mosaics in patients with clear clinical diagnosis, which may also be used in other syndromes associated to somatic mosaicism; and secondly, a review of clinical criteria on the phenotypic spectrum of these diseases, thus reducing the proportion of clinically undiagnosed individuals, who naturally do not undergo molecular diagnosis. We will also elaborate clinical and diagnostic guidelines to allow patients, relatives and healthcare professionals to obtain complete information about these diseases, and will support the creation of a Spanish association of patients with segmental overgrowth caused by mutations in *PIK3CA*.

RESEARCH PROJECT TECHNICAL PROPOSAL

BACKGROUND AND CURRENT STATUS

Cite references included in the previous section: Background and Current Status (Maximum of 1 page, Arial 10)

1. Hafner, C., and Groesser, L. (2013). Mosaic RASopathies. Cell Cycle 12, 43-50.
2. Riviere, J.B., Mirzaa, G.M., O'Roak, B.J., Beddaoui, M., Alcantara, D., Conway, R.L., St-Onge, J., Schwartzentruber, J.A., Gripp, K.W., Nikkel, S.M., et al. (2012). De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. Nat Genet 44, 934-940.
3. Lindhurst, M.J., Parker, V.E., Payne, F., Sapp, J.C., Rudge, S., Harris, J., Witkowski, A.M., Zhang, Q., Groeneveld, M.P., Scott, C.E., et al. (2012). Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. Nat Genet 44, 928-933.
4. Kurek, K.C., Luks, V.L., Ayturk, U.M., Alomari, A.I., Fishman, S.J., Spencer, S.A., Mulliken, J.B., Bowen, M.E., Yamamoto, G.L., Kozakewich, H.P., et al. (2012). Somatic mosaic activating mutations in PIK3CA cause CLOVES syndrome. Am J Hum Genet 90, 1108-1115.
5. Baradaran-Heravi, A. (2012). PIK3CA, a hotspot for postzygotic mutations in nonhereditary overgrowth syndromes. Clinical genetics 82, 523-525.
6. Pansuriya, T.C., van Eijk, R., d'Adamo, P., van Ruler, M.A., Kuijjer, M.L., Oosting, J., Cleton-Jansen, A.M., van Oosterwijk, J.G., Verbeke, S.L., Meijer, D., et al. (2011). Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. Nat Genet 43, 1256-1261.
7. Lindhurst, M.J., Sapp, J.C., Teer, J.K., Johnston, J.J., Finn, E.M., Peters, K., Turner, J., Cannons, J.L., Bick, D., Blakemore, L., et al. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. The New England journal of medicine 365, 611-619.
8. Zhou, X.P., Marsh, D.J., Hampel, H., Mulliken, J.B., Gimm, O., and Eng, C. (2000). Germline and germline mosaic PTEN mutations associated with a Proteus-like syndrome of hemihypertrophy, lower limb asymmetry, arteriovenous malformations and lipomatosis. Human molecular genetics 9, 765-768.
9. Sun, B.K., Saggini, A., Sarin, K.Y., Kim, J., Benjamin, L., LeBoit, P.E., and Khavari, P.A. (2013). Mosaic activating RAS mutations in nevus sebaceus and nevus sebaceus syndrome. The Journal of investigative dermatology 133, 824-827.
10. De, S. (2011). Somatic mosaicism in healthy human tissues. Trends in genetics : TIG 27, 217-223.
11. Keppler-Noreuil, K.M., Sapp, J.C., Lindhurst, M.J., Parker, V.E., Blumhorst, C., Darling, T., Tosi, L.L., Huson, S.M., Whitehouse, R.W., Jakkula, E., et al. (2014). Clinical delineation and natural history of the PIK3CA-related overgrowth spectrum. Am J Med Genet A.
12. Mirzaa, G.M., Riviere, J.B., and Dobyns, W.B. (2013). Megalencephaly syndromes and activating mutations in the PI3K-AKT pathway: MPPH and MCAP. Am J Med Genet C Semin Med Genet 163C, 122-130.
13. Lopez-Gutierrez, J.C., and Lapunzina, P. (2008). Capillary malformation of the lower lip, lymphatic malformation of the face and neck, asymmetry and partial/generalized overgrowth (CLAPO): report of six cases of a new syndrome/association. Am J Med Genet A 146A, 2583-2588.
14. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42, 30-35.
15. Tucker, T., Marra, M., and Friedman, J.M. (2009). Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 85, 142-154.
16. Tapper, W.J., Foulds, N., Cross, N.C., Aranaz, P., Score, J., Hidalgo-Curtis, C., Robinson, D.O., Gibson, J., Ennis, S., Temple, I.K., et al. (2014). Megalencephaly syndromes: exome pipeline strategies for detecting low-level mosaic mutations. PloS one 9, e86940.

RESEARCH PROJECT TECHNICAL PROPOSAL

HYPOTHESIS AND OBJECTIVES

(Maximum of 1 page, Arial 10)

HYPOTHESIS

1. The evaluation of NGS experimental conditions *(differential detection in tissues, NGS coverage and read depth, etc.)* and bioinformatic algorithms *(pipeline strategies, allele frequency spectrum, variant calling tools, etc.)* needed to detect low somatic mutations, will allow us to develop a molecular diagnostic protocol in patients with segmental overgrowth syndromes associated to mosaic *PIK3CA* mutations, which will serve as a model to be applied in clinical practice to other mosaic developmental syndromes.

2. The magnitude of the phenotypic spectrum of mutations in *PIK3CA* is not yet well defined. Mutational analysis of this gene in patients with clinical manifestations similar to those of MCAP and CLOVES, but who do not strictly meet the diagnostic criteria (including CLAPO syndrome), will allow us to open the range of clinical manifestations associated with this type of segmental overgrowth.

3. Besides the development of a diagnostic test for low mosaic mutations, the elaboration of clinical guidelines and protocols for diagnosis and monitoring will allow different health professionals to have the appropriate tools to provide better care for patients and families.

4. The elaboration of comprehensive and informative guidelines aimed at patients will improve their knowledge of the disease, allowing them to make better decisions and help them to rationalize the emotional facets of the disease.

5. The creation of patient organizations that provide guidance and support to those affected and their families, will allow them to occupy the center of the healthcare process and research efforts, which they seldom conquer when talking about rare diseases.

OBJECTIVES

1. **Technological development:** To develop an experimental and bioinformatic NGS protocol, reliable and applicable to clinical practice, for the diagnosis of patients with *PIK3CA* segmental overgrowth syndromes, as a model for an NGS diagnostic panel to detect low mosaic mutations in developmental syndromes.

2. **Expansion of the phenotype:** To expand the phenotypic spectrum of the overgrowth syndromes associated with *PIK3CA* by including patients who do not completely meet the actual clinical criteria associated with MCAP or CLOVES, thus reducing the number of clinically and molecularly undiagnosed patients.

3. **Clinical care:** To develop clinical guidelines and protocols for diagnosis and monitoring of patients with segmental overgrowth syndromes associated to *PIK3CA*, intended for physicians and health care professionals.

4. **Information for patients:** To develop publicly accessible guidelines/fact sheets for patients with segmental overgrowth syndromes associated to *PIK3CA*.

5. **Association of patients:** To provide scientific and institutional support needed to create a spanish association of patients with segmental overgrowth syndromes associated to *PIK3CA*.

Design, study subjects, variables, data collection and analysis and limitations of the study.

(Maximum of 3 pages, Arial 10)

The project has been divided in 4 working packages (WP) detailed below.

WP		Methodology
WP1	**Patients, samples, experiments and data analysis**	
	Patients and databases	The INGEMM at La Paz Hospital in Madrid is a referral service nationwide for patients with overgrowth syndromes and vascular malformations. Until today we have seen and follow up in the clinics 17 patients with clinical diagnosis of MCAP, 13 patients with CLOVES and a significant number of patients with clinical features included in the *PIK3CA* spectrum, but that do not meet the clinical criteria. The clinical team at the INGEMM will select at least 50 patients (belonging to the entire national territory) with segmental overgrowth syndromes associated *PIK3CA,* and interested in participating in the study (informed consent will be requested). Clinical team will collect all clinical data necessary to carry out the project. We will also create a specific database for the project, and patients and families will be offered the opportunity to enter in the Spanish Overgrowth Syndromes Registry, located in the INGEMM.
	Sample Collection	We will collect samples of peripheral blood (2x3ml EDTA) and saliva (oragene) from patients and their parents, and biopsy of the affected tissue from the patients. Affected tissue biopsies may be performed at the Department of Dermatology, Hospital Infantil La Paz, or when it is not possible for the patient to move to Madrid, we will provide the means to collect samples in their reference hospital.
	Sample processing and DNA extraction	Extracting DNA from peripheral blood and saliva will be performed by conventional methods already standardized in our Institute. Biopsies will be immediately collected for primary culture and subsequent DNA extraction by conventional methods in the INGEMM.
	NGS	The experiment is designed to obtain an initial read depth of 500x, which further allows to define the minimum read depth threshold necessary to detect low mosaic mutations. The preparation of the libraries will be performed by using Roche Nimblegen NimbleDesign program (https://design.nimblegen.com/) to select genes of interest. Capture will be performed with the SeqCap EZ Developer Library. SeqCap EZ Oligo pool is made against target regions in the genome. Standard shot-gun sequencing library is made from genomic DNA. The sequencing library is hybridized to the SeqCap EZ Oligo pool. Capture beads are used to pull down the complex of capture oligos and genomic DNA fragments. Unbound fragments are removed by washing. Enriched fragment pool is amplified by PCR. The success of enrichment is measured by qPCR at control loci. The end product is a sequencing library enriched for target regions, ready for high throughput sequencing. Run will be performed with the MiSeq platform which allow parallel sequencing of DNA molecules, reaching 15Gb of output with 25 M sequencing reads and 2x300 bp read lengths, by performing automated clonal amplification and sequencing followed by a primary data analysis (quality, basecalling). Raw data goes then to the Bioinformatic analysis.
	Bioinformatics	Bioinformatic analysis will be developed in the INGEMM with the direct support of the Bioinformatics Section. The sequenced samples will be separated into 2 groups. The first to develop the tool, and a second for subsequent validation. The analysis involves the most popular open-source bioinformatic tools in the NGS field and custom methods. First, the data is aligned using the Bowtie2 software [SourceForge.net, Johns Hopkins University, USA] and the alignments are revised for candidate indels (re-alignment) and the quality of the nucleotides is recalibrated. In a second step, the variant calling is performed using GATK suite of tools [Broad Institute] for detecting SNVs and Indels. The changes detected are filtered and recalibrated to remove false positives. The analysis process included tests and controls for checking the integrity of the process. A quality control is performed to be sure the nucleotides are reliable sequenced. Only nucleotides with a phred quality score above 28 are selected. In addition, the alignments are checked and

		only those reads that are paired are included for the variant calling. The histograms of coverage are also calculated. The variants are evaluated as well. The number of novel/known variants has to be concordant to the expected and the ratio tiTv obtained shows the validity of the analysis. Detection, quality, depth and validation parameters are evaluated according to the different percentages of mosaics in order to obtain the best pipeline for detection and subsequent application in the control group.
	Validation and analysis	To validate mosaic levels found by NGS, pyrosequencing for NGS-detected mutations of selected samples in different tissues will be designed. Pyrosequencing will be performed in the Pyromark Q96 MD – QIAGEN, available in the INGEMM.
WP2	**Expansion of the phenotype**	
	Defining the phenotypic spectrum	Since the discovery of the involvement of *PIK3CA* in MCAP and CLOVES, other patients not classified in these two diseases but sharing common manifestations have been reported. The different clinical variables will be analyzed together with molecular results (presence or absence of mutation in *PIK3CA*) to establish recognizable patterns associated with this group of diseases, to allow us to establish criteria for inclusion of patients who do not meet the classic diagnostic criteria for CLOVES and MCAP.
	Genotype-phenotype correlations	Statistical correlations will be performed between clinical phenotype and the specific mutations detected.
WP3	**Guidelines and patients association**	
	Clinical guidelines and protocols	The guide will contain a complete description of the clinical features and molecular alterations, as well as protocols for diagnosis and monitoring of patients with these syndromes. It will also include a sheet with the minimum information necessary to refer patients between different specialists. INGEMM geneticists are authors of the only clinical guide in spanish about MCAP, published in 2012 before the discovery of the causative gene for this syndrome.
	Guidelines/fact sheets for patients	The Guidelines/fact sheets will be directed to provide a complete and understandable information to patients and families.
	Association of patients	We will contact parents of children affected to offer our support for the creation of a specific patient association of these pathologies. The research group will provide scientific support and the institutional support will be provided from the CIBERER through direct collaboration by one of its scientific managers.
WP4	**Training and dissemination**	
	Publication and dissemination of guidelines	The guideline will be published in the Orphanet portal, and on the CIBERER, GT-CSGP and AEGH webpages, and will be available to different patient groups or organizations interested in promoting it.
	Communication of results	The results will be presented at conferences, and will be published in scientific journals, preferably in open acces format.
	Training	The student we propose (see attached CV), will perform the activities in the INGEMM where she will develop and refine the molecular techniques and bioinformatics required for the project. She will attend meetings, training seminars and conferences, and will be stress upon the acquisition of critical thinking skills necessary for the development of a research career.

RESEARCH PROJECT TECHNICAL PROPOSAL

WORK PLAN

Stages of development and task distribution of the research team, allocations provided for the technical staff requested. Indicate the place of realization of the project. (Maximum of 1 page, Arial 10)

WP	Objectives	Participants	Working Plan	Deliverables
WP1	**Patients, samples, experiments and data analysis**			
	Patients and databases	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Clinical team will evaluate the clinical history of the patients to decide who will be included. Patients will be cited and included until obtaining the expected number for the project.	Patient database. NGS and data analysis protocol for detection of low mosaic mutations in clinical practice.
	Sample Collection	Gordo G; Martínez-Glez V	Patients and parents will be cited to the clinics (INGEMM) to collect samples. For patients outside Madrid, when necessary, we will provide the means to collect samples in their reference hospital.	
	Sample processing and DNA extraction	Gordo G	The samples will be processed by the student in the INGEMM using established protocols.	
	NGS	Gordo G; Martínez-Glez V	All NGS experiments will be performed at the INGEMM using the equipment and infrastructure already available.	
	Bioinformatics	Gordo G; Martínez-Glez V; Section of Bioinformatics INGEMM	In close collaboration with the Section of Bioinformatics, the proposed analysis will be developed to obtain the protocol for detection of low mosaics.	
	Validation and analysis	Gordo G; Martínez-Glez V	After obtaining the NGS data and bioinformatic analysis, data on detected mosaic levels will be validated by Pyrosequencing. Draw of conclusions.	
WP2	**Expansion of the phenotype**			
	Phenotypic spectrum	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	All studies involving correlations between clinical and molecular variables will be performed at the INGEMM. Analysis will be assisted by the Biostatistics Service at Hospital La Paz.	Diagnostic Criteria
	Genotype-phenotype correlations			
WP3	**Guidelines and patients association**			
	Clinical guidelines and protocols	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	The guide will be elaborated by the clinical team. The group will review the current scientific evidence and will contribute their professional expertise to develop a clinical guide intended for health professionals.	Clinical guidelines, protocols and fact sheets.
	Guidelines/fact sheets for patients		From the clinical guide, team members will draft a reduced but complete version in plain language for the general public.	
	Association of patients	Martínez-Glez V; Guillén-Navarro E; Gómez B	The experience of the research group and its relationship with other associations or federations of patients, will be used to help the creation of a specific association of this disease. This will be proposed to families of children affected in the clinic when explainig the project and collecting the samples.	
WP4	**Training and dissemination**			
	Publication and dissemination of guidelines	Martínez-Glez V; Gómez B	In the last months of the project the guidelines and fact sheets will be graphically edited for subsequent publication on internet: Orphanet, CIBERER, AEGH, GT-CSGP. All scientific results will be published on indexed journals.	Publication of clinical guidelines
	Communication of results	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G		Results publication in scientific journals and conferences
	Training	Gordo G	Learning activities of the pre-doctoral student will be strengthened. The student will be physically located at the INGEMM at La Paz Hospital. Two years of the project may not be sufficient to achieve the experiemntal and training milestones necessary for a PhD. However, we will seek, and is expected, the continuation of this research line to ensure that the student obtains the PhD degree.	PhD degree

TIMELINE

*Indicate the activities / tasks and assign staff from the research team indicating the time necessary to accomplish them.

WP		Tasks	Participants	Timeline (1-24 months)
WP1	**Patients, samples, experiments and data analysis**			
	Patients and databases	Select a representative group of patients with MCAP, CLOVES, and other patients with clinical features included in the *PIK3CA* spectrum. Create a specific clinical/molecular database.	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Months 1 to 6
	Sample Collection	Collect samples of peripheral blood, saliva, and a biopsy of the affected tissue, from each patient.	Gordo G; Martínez-Glez V	Months 3 to 9
	Sample processing and DNA extraction	Obtain DNA from different tissues for each patient	Gordo G	Months 3 to 9
	NGS	Establish an experimental protocol to obtain the data necessary to allow subsequent computer processing in order to perform diagnosis in tissue specific low mosaic mutations.	Gordo G; Martínez-Glez V	Months 7 to 11
	Bioinformatics and validation	Establish bioinformatic algorithms (pipeline strategies, allele frequency spectrum, variant calling tools, etc.) needed to detect low somatic mutations. Perform pyrosequencing experiments to validate NGS results.	Gordo G; Martínez-Glez V; Section of Bioinformatics INGEMM	Months 10 to 20
WP2	**Expansion of the phenotype**			
	Defining the phenotypic spectrum	Expand the phenotypic spectrum of these pathologies by detecting *PIK3CA* mutations in patients who do not meet all the criteria for CLOVES and MCAP.	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J	Months 10 to 16
	Genotype-phenotype correlations	Obtain correlations between the type of mutation, the affected tissue and the clinical manifestations.	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Months 16 to 18
WP3	**Guidelines and patients association**			
	Clinical guidelines and protocols	Develop an objective and updated clinical guideline aimed at healthcare professionals, including protocols for clinical diagnosis and monitoring of patients with segmental overgrowth associated with *PIK3CA*.	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Months 8 to 20
	Guidelines/fact sheets for patients	Develop publicly accessible guidelines/fact sheets for patients with segmental overgrowth syndromes associated *PIK3CA*.	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Months 8 to 20
	Association of patients	Provide scientific and institutional support needed to create a Spanish association of patients with segmental overgrowth syndromes associated *PIK3CA*.	Martínez-Glez V; Guillén-Navarro E; Gómez B	Months 1 to 24
WP4	**Training and dissemination**			
	Publication and dissemination of guidelines	Edit and publish the clinical guideline, making it free and accessible online	Martínez-Glez V; Gómez B	Months 20 to 24
	Communication of results	Publication of results in scientific journals and conferences	Martínez-Glez V; López-Gutierrez JC; Lapunzina P; Guillén E; Rosell J; Gordo G	Months 18 to 24
	Educational and training	The pre-doctoral student will conduct its activities in the INGEMM, where she will learn the molecular and bioinformatic techniques required for the Project, and will attend meetings, training seminars and conferences.	Gordo G	Months 1 to 24

RESEARCH PROJECT TECHNICAL PROPOSAL

WORK PLAN

A timeline image may be attached.

	Timeline (months)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
WP1	**Patients, samples, experiments and data analysis**																								
	Patients and databases	X	X	X	X	X	X																		
	Sample Collection			X	X	X	X	X	X	X															
	Sample processing and DNA extraction			X	X	X	X	X	X	X															
	NGS							X	X	X	X	X													
	Bioinformatics										X	X	X	X	X	X	X								
	Validation																X	X	X	X	X				
WP2	**Expansion of the phenotype**																								
	Defining the phenotypic spectrum										X	X	X	X	X	X	X	X	X	X	X	X			
	Genotype-phenotype correlations																X	X	X	X	X	X			
WP3	**Guidelines and patients association**																								
	Clinical guidelines and protocols								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
	Guidelines/fact sheets for patients								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
	Association of patients	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
WP4	**Training and dissemination**																								
	Publication and dissemination of guidelines																					X	X	X	X
	Communication of results																					X	X	X	X
	Educational and training	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

RESEARCH PROJECT TECHNICAL PROPOSAL
EXPERIENCE OF THE TEAM IN THIS SPECIFIC RESEARCH LINE
(Maximum of 1 page, Arial 10)

The team involved in this project consists of professionals nationally and internationally recognized in the study and in the clinical and translational approach to rare diseases and especially in overgrowth syndromes. The PI of this project (Dr. Martínez-Glez) is the author of numerous scientific papers in indexed journals related to overgrowth syndromes, including pathologies studied in this project. One publication is specifically about MCAP, in which the PI, along with the associated researcher Dr. Lapunzina, performed a review of the syndrome and proposed diagnostic criteria. Similarly, Dr. Martinez- Glez and Dr. Lapunzina are authors of the only clinical guide written in Spanish on MCAP, published in 2011. Dr. Martinez- Glez is also the Coordinator of the Working Group on Cancer within Genetic Polimalformative Syndromes, in whose area of study are included the diseases addressed in this project, and on this matter he is the PI of a national project dedicated to the study of the relationship between tumors and developmental syndromes. The latest article published by Dr. Martínez-Glez concerns the use of NGS in the study of diseases such as those associated with mutations in *PIK3CA*.

Dr. Lapunzina is also the author of numerous scientific articles on overgrowth syndromes, field in which he is internationally recognized as an expert. Dr. Lapunzina is the coordinator of the Spanish Registry of Overgrowth Syndromes and has directed numerous research projects on this topic. Dr. Lopez-Gutierrez, a vascular surgeon, has a large experience in vascular disorders associated with overgrowth and, along with Dr. Lapunzina, described the CLAPO syndrome (Capillary malformation of the lower lip, lymphatic malformation of the face and neck, asymmetry and partial/generalized overgrowth), whose clinical features may overlap with MCAP and CLOVES. Dr. Guillen (President of the Spanish Association of Clinical Genetics and Dysmorphology) and Dr. Jordi Rosell, are clinical geneticists with extensive experience in the pathologies studied in this project, and are part of the clinical groups linked to CIBERER.

All team members have experience in the clinical practice, as well as in the development of projects in basic and translational research. Overall, team researchers of this project total 305 scientific publications (131 in the last 5 years), 38 research projects, and 8 clinical guidelines.

RESEARCH PROJECT TECHNICAL PROPOSAL

DISSEMINATION PLAN AND AVAILABLE MEANS

DISSEMINATION PLAN:
1. Relevance of the project with regards to clinical / medical care impact and/or technological development.
2. Relevance of the Project with regards to bibliometric impact. (Maximum of 1 page, Arial 10)

1. The relevance of this project with regards to clinical / medical care impact and/or technological development is a direct consequence of its objectives.

Technological development: The development of an experimental and bioinformatic NGS protocol for the diagnosis of patients with *PIK3CA* segmental overgrowth syndromes can be extrapolated and used as a model for an NGS diagnostic panel able to detect low mosaic mutations in developmental syndromes in the clinical practice.

Clinical care: Expanding the phenotypic spectrum of the overgrowth syndromes associated with *PIK3CA* by including patients who do not completely meet the actual clinical criteria associated with MCAP or CLOVES, will allow to reduce the number of clinically and molecularly undiagnosed patients

The development of clinical guidelines and protocols for diagnosis and monitoring of patients intended for physicians and health care professionals, will ostensibly improve the care they can provide to patients, improving the diagnosis and monitoring, as well as reducing the time until a patients is clinical or molecularly diagnosed, and allowing to offer better genetic counseling. Additionally, publicly accessible guidelines/fact sheets for patients is an essential part of the comprehensive management of patients, improving their knowledge of the disease, allowing them to make better decisions and helping to rationalize the emotional facets of the disease.

2. **Bibliometric impact:** The results of this project will be useful to the international community, so likely will be widely cited; besides published in high impact scientific journals.

AVAILABLE MEANS FOR THE DEVELOPMENT OF THE PROJECT (Maximum of 1 page, Arial 10)

The Institute of Medical and Molecular Genetics (INGEMM), at HULP, is part of the IdiPAZ Medical Research Institute and the CIBERER (Unit 753). INGEMM is a multidisciplinary research institute with extensive experience in diagnosis, research and teaching in human medical and molecular genetics, with a particular interest in clinical genetics and rare diseases. INGEMM has the necessary infrastructure, equipment and proven experience in molecular genetic techniques including next-generation sequencing. INGEMM is organized in 10 different Sections or Units, 4 of which will contribute to this project: 1) The Structural and Functional Genomics Unit (Dr. Víctor Martínez-Glez), the Molecular Endocrinology and overgrowth Unit, (Dr. Pablo Lapunzina), the Bio-informatics Unit, and the Clinical Genetics Unit (Dr. Pablo Lapunzina and Dr. Víctor Martínez-Glez). The INGEMM at La Paz is the only public hospital having the three massive benchtop sequencing platforms on the market: Junior from Roche (http://www.gsjunior.com/) , MiSeq from Illumina (http://www.illumina.com/systems/miseq.ilmn) and Ion Torrent from Life Technologies (http://www.invitrogen.com/site/us/en/home/brands/Ion-Torrent.html?cid=fl-iontorrent), all of which allow parallel sequencing of DNA molecules, reaching 15Gb per run in the case of MiSeq Illumina, the platform that will be used to carry out this project. In terms of computational infrastructure for genome analysis, the Bio-informatics Unit owns its computational cluster with the following characteristics: Head node: 64Gb RAM, Dual CPU Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz. 40 Tb redundant disk array for storage. LTO6 Tape backup library with capacity for 24 labeled media. Four Compute Nodes: 64Gb RAM. Dual CPU Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz. All compute nodes and head node are connected through a 10Gb Ethernet network. The INGEMM also have the necessary equipment and facilities for cell cultures and molecular genetics, and the Pyromark Q96 MD – QIAGEN for validation by pyrosequencing, all of which are necessary for the proper development of the project.

RESEARCH PROJECT TECHNICAL PROPOSAL

BUDGET

1. **Include itemized information regarding the amount requested for each eligible cost: human resources, equipment, consumables, travel and subsistence, subcontracting, overheads.**
2. **Detailed justification of eligible costs requested.**

(Maximum of 2 pages, Arial 10)

Item		Justification	Budget
	Patients, samples, experiments and data analysis		
	Patients and databases	--	- €
	Sample Collection	EDTA tubes for blood collection, Oragene kits for saliva collection, samples transport for patients out of Madrid.	2.100,00 €
WP1	Sample processing and DNA extraction	Extraction kits, cultures, reagents needed for DNA extraction from different tissues.	2.800,00 €
	NGS	Capture and sequencing kits and reagents needed for NGS experiments.	31.000,00 €
	Bioinformatics	--	- €
	Validation	Pyrosequencing reagents (Oligonucleotides and reagents) needed for validation of results.	2.250,00 €
		SUBTOTAL	**38.150,00 €**
	Expansion of the phenotype		
WP2	Defining the phenotypic spectrum	--	- €
	Genotype-phenotype correlations	--	- €
		SUBTOTAL	**- €**
	Guidelines and patients association		
WP3	Clinical guidelines and protocols	--	- €
	Guidelines/fact sheets for patients	--	- €
	Association of patients	--	- €
		SUBTOTAL	**- €**
	Training, dissemination, meetings		
	Team members meetings	Biannual meeting in Madrid of all team members. Travel expenses from Murcia (Guillen E) and Majorca (Rosell J).	1.500,00 €
WP4	Publication and dissemination of guidelines	External graphic editing and printing of some copies of the guidelines.	2.000,00 €
	Communication of results	Indexed journals (open access), attendance at conferences.	2.000,00 €
	Educational and training	Personnel: Pre-doctoral student (2 years) needed to perform experiments.	44.165,02 €
		SUBTOTAL	**49.665,02 €**
		TOTAL	**87.815,02 €**