

Fondation 101 Génomes

FONDATION PRIVÉE

WHITE PAPER:

101 GENOMES FOUNDATION FONDATION PRIVÉE &

PROJECT 101 GENOMES MARFAN

(VERSION 7 – 03/12/2018)

INTRODUCTION	2
1. THE 101 GENOMES FOUNDATION – F101G.....	4
2. THE PROJECT 101 GENOMES MARFAN – P101GM	5
2.1 DESCRIPTION OF THE PROJECT: OBJECTIVE, ORIGIN AND INNOVATIONS	5
2.2 PARTNERS FOR IMPLEMENTATION OF THE PLATFORM	10
a) <i>Academic Partners</i>	10
b) <i>Biobank</i>	11
c) <i>Business Partners</i>	11
2.3 DRAFT BUDGET	16
2.4 FROM 101 TO 1000 GENOMES	18
a) <i>Extension of the original cohort</i>	18
b) <i>Reuse and/or extension of the platform</i>	20
2.5 FINANCIAL PARTNERS TO FINANCE THE PLATFORM	21
a) <i>Partners & Funding sources</i>	21
b) <i>King Baudouin Foundation</i>	23
2.6 CONSENT	27
a) <i>Protection of personal data</i>	27
b) <i>Non-medicinal interventional clinical study</i>	27
2.7 ETHICS COMMITTEE AND SCIENTIFIC COMMITTEE.....	28
3. FINAL REMARK	32
LIST OF ANNEXES.....	33

INTRODUCTION

Our son Aurélien born on 3rd of September 2015 has Marfan syndrome.

Marfan's syndrome is the result of a connective tissues¹ defect caused by a deleterious mutation of the FBN1 gene on chromosome 15, which results in the fibrillin protein encoded by this gene not being or being poorly produced and therefore does not properly fulfil its function in the organisms of affected individuals.

This dysfunction of fibrillin normally has consequences for the whole body of affected persons and usually manifests itself in skeletal, pulmonary, ocular and especially cardiovascular disorders² which in turn leads to sustained and often heavy therapeutic management. In most cases, this treatment involves major aortic surgery (often in combination with surgery of the aortic valve).

However, even within the same family with the same mutation, the intensity of the damage can greatly vary. Thus, some people affected by the syndrome have few symptoms³ while at the other end of the spectrum, the life expectancy of some Marfans is very limited⁴. Between these two extremities, we find the majority of Marfans who are sometimes severely handicapped by the disease and who must regularly monitor the dilatation of their aorta.

¹ Extra-cellular tissues that hold together the cells that make up the human body. See the page dedicated to Marfan Syndrome on the [site orphanet](#).

² DIETZ H., "Marfan Syndrome" in *GeneReviews*, 18 April 2001, last updated 2 February 2017, available at : <https://www.ncbi.nlm.nih.gov/books/NBK1335/>.

³ In exceptional cases, careers as top sportsmen and women are even achieved, for example: [Flo Hyman](#), captain of the American volleyball team that won a silver medal at the 1984 Olympic Games; [Isaiah Austin](#), a professional basketball player currently playing in the Chinese championship, and as [Jonathan Jeanne](#), both saw their careers in NBA was abruptly ended when diagnosed with Marfan. Apart from these athletes of whom the Marfan diagnosis has been confirmed, [Michael Phelps](#)' physical characteristics lead some observers to speculate - currently unconfirmed - that the most successful athlete in the history of the modern Olympic Games might also be affected by Marfan syndrome or another connective tissue disorder: <http://www.raredr.com/news/michael-phelps-marfan>.

⁴ And would be on average of about 16 months for the neonatal Marfans, or more precisely Marfan with rapid trigger "early onset":

"Marfan syndrome (MFS) (OMIM 154700) is an autosomal dominant disorder of fibrous connective tissue involving the ocular, skeletal, and cardiovascular systems. MFS patients present with clinical variability, in which the rare neonatal Marfan syndrome (nMFS) has the most severe presentation in early childhood. The prognosis of nMFS is very poor, with a mean survival age of only 16.3 months. Valvular insufficiencies and diaphragmatic hernias have been associated with shorter survival in patients diagnosed before the age of 1 year. [...] The term neonatal Marfan syndrome was first used in 1991 to describe the most severe phenotype of MFS similar to cases previously known as infantile Marfan syndrome, congenital Marfan syndrome, and severe perinatal Marfan syndrome. Recently, it has been suggested that the term neonatal MFS should be replaced by early onset and rapidly progressive MFS to represent the most severe features of MFS in early childhood" in PENG Q. et al., "A novel fibrillin-1 gene missense mutation associated with neonatal Marfan syndrome : a case report and review of the mutation spectrum", *BMC Pediatrics*, 30 avril 2016, 16:60, DOI 10.1186/s12887-016-0598-6.

At this stage in scientific knowledge, this great variability of damages and their intensity is not yet well understood (apart from certain leads relating to: the position of deleterious mutations on certain exons⁵; the type of consequences of deleterious mutations⁶ and the sex of patients⁷).

It is in this context that we have decided to set up a foundation aimed at supporting scientific research. The Foundation's first innovative project consists of providing interested researchers with a tool in the form of a digital platform containing a genomic/phenotypic cross-database of a cohort of 101 patients suffering from Marfan syndrome in order to better understand the variability and intensity of harm using the most advanced technological means.

The purpose of the Foundation is to provide a vehicle for raising the funds needed to finance the development and management of this bioinformatics tool.

Ludivine Verboogen &
Romain Alderweireldt

⁵ “Most previously identified nMFS-associated FBN1 mutations are known to cluster between exons 24 and 32, which is the neonatal region of FBN1” in PENG Q. et al., “A novel fibrillin-1 gene missense mutation associated with neonatal Marfan syndrome : a case report and review of the mutation spectrum” in BMC Pediatrics, 30 april 2016, 16:60, DOI 10.1186/s12887-016-0598-6.

See also : STHENEUR C. et al., “Prognosis Factors in Proband With an FBN1 Mutation Diagnosed Before the Age of 1 Year” in Pediatric Research, March 2011, DOI: 10.1203/PDR.0b013e3182097219 et ; MAEDA J. et al., “Variable severity of cardiovascular phenotypes in patients with an early-onset form of Marfan syndrome harboring FBN1 mutations in exons 24–32” in Heart and Vessels, January 2016, DOI: 10.1007/s00380-016-0793-2

⁶ “Haploinsufficiency” or “Dominant Negative” see for instance: LANDIS B.J. et al., “Genotype–phenotype correlations in Marfan syndrome” in Heart Online First, published on 8 June 2017 as 10.1136/heartjnl-2017-311513 et ; FRANKEN R. et al., “Beneficial Outcome of Losartan Therapy Depends on Type of FBN1 Mutation in Marfan Syndrome” in Circ Cardiovasc Genet, April 2015, DOI: 10.1161/CIRCGENETICS.114.000950

⁷ RENARD M. et al., “Sex, pregnancy and aortic disease in Marfan syndrome” in PLoS ONE, July 2017, <https://doi.org/10.1371/journal.pone.0181166>

1. THE 101 GENOMES FOUNDATION – F101G

1. **F101G.** The “101 Genomes Foundation” (hereinafter referred to as “F101G”⁸) has been established on the 10th November 2017.

The F101G *fondation privée* is registered under the company number BE0684609172. Its statutes have been published in the annexes of the *Moniteur Belge* on the 14th of November 2017. Its day-to-day management has been delegated to its Managing Director, Mrs. Ludivine Verboogen (who can be contacted at +32476871863). Its registered office is located at 6 avenue de Sumatra, 1180 Brussels. The F101G holds the IBAN account: BE15-6787-1386-2730 BIC: DELEBE22 at the Bank Delen.

The Agreement signed on the 17th of November 2017 with the King Baudouin Foundation provides that donations for the F101G are to be made to the IBAN account: BE10-0000-0000-0000-0404 (or any other account of the King Baudouin Foundation) with the mention "Funds 101 Genomes" or the structured code ***017/1730/00036*** at *Banque de la Poste*.

2. **Objective.** In accordance with its Statutes available in Annex 1, the F101G has the objective:

*“to improve the living conditions of people affected by rare diseases, in particular diseases affecting connective tissue, mainly by supporting research and development of new therapies for these people, and in particular by working, as a first step, to discover and adopt innovative therapies that will benefit children and adults with Marfan syndrome, by making available to scientists a bioinformatics platform containing phenotypic and genomic cross-data from a cohort of 101 patients with Marfan Syndrome (this is the "101 Marfans Genomes" or "101GM" project), the original cohort could be expanded and/or the phenotypic and genomic bioinformatics platform could be broadened to other rare diseases”*⁹.

Thus, F101G intends to build on progress in genomics to hope to be able to offer new therapeutic approaches to patients with rare diseases. The action of the F101G will make it possible to host on the same bioinformatics platform different projects based on cohorts of 101 patients suffering from rare diseases.

The first pilot project of the F101G is dedicated to Marfan syndrome. This is the Project "101 Marfans Genomes". This project is being carried out in order to try to improve the living conditions of people with this syndrome in general and to try to avoid them heavy surgical interventions in the future.

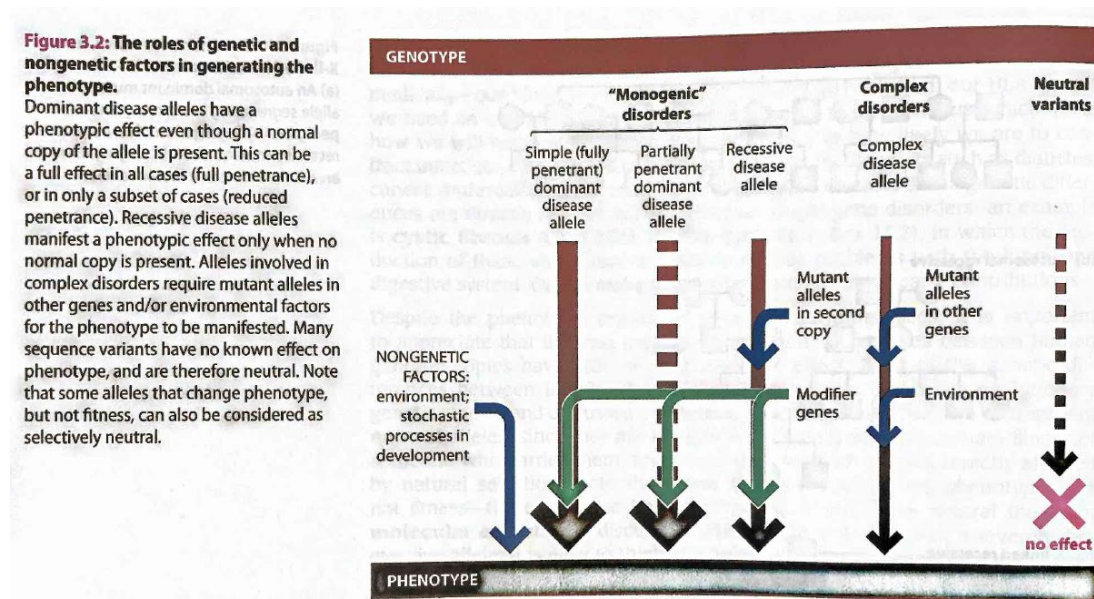
⁸ For “Fondation 101 Genomes” into French.

⁹ Annex 1: Statutes for the private foundation “*Foundation 101 Genomes*”, Article 3 (free translation).

2. THE PROJECT 101 GENOMES MARFAN – P101GM

2.1 DESCRIPTION OF THE PROJECT: OBJECTIVE, ORIGIN AND INNOVATIONS

3. **Objective.** The objective of the main F101G's project is to make available to researchers a cross-database containing genomic and phenotypic data of genetically confirmed Marfan patients. This tool, accessible through a secure computerized platform, should enable them to identify the existence of possible modifier genes^{10 11} that would protect certain Marfans against the main afflictions (mainly cardiovascular, skeletal and ocular) of the disease in order to develop on this basis treatments that replicate the effects of these protective modifier genes in Marfans in which these genes are not activated.



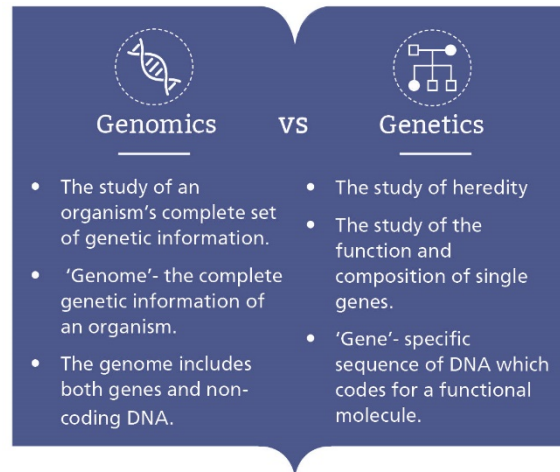
4. **Origin.** This project was born from the reading of an interview with Professor Hal Dietz, who mentioned that the intersection of genomic and phenotypic data could make it possible to understand “*how natural genetic variants can protect some people from the consequences of a fibrillin-1 mutation*” and on this basis possibly be able to “*identify drugs that can mimic nature’s successful strategy*”¹².
5. **Innovations.** At this stage, we could only note the absence of a tool that would allow us to conduct this research and we therefore wished to study the conditions for its implementation. It quickly became apparent that this project was innovative in two ways.

¹⁰ The modifier gene is a gene that affects the expression of one or more other genes. For more information about the action of modifier genes see Annex 5: RIORDAN J.D., NADEAU J. H., “*From Peas to Disease: Modifier Genes, Network Resilience and the Genetics of Health*” in *The American Journal of Human Genetics*, 101, 177–191, 3 August 2017, <http://dx.doi.org/10.1016/j.ajhg.2017.06.004>.

¹¹ See « Figure 3.2 » from TYLER-SMITH et al., *Human evolutionary genetics (second edition)*, Garland Science, 2014, ISBN 978-0-8153-4148-2, p. 46.

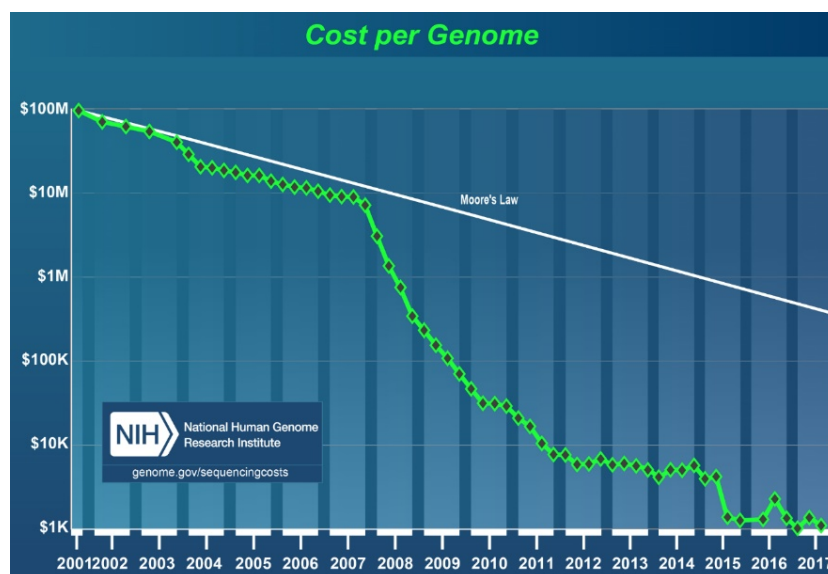
¹² WEISMAN R., “*Meet Your Gene: An Introduction to the Marfan Gene and Current Research*”, 10 janvier 2017. Available at: <http://blog.marfan.org/meet-your-gene-an-introduction-to-the-marfan-gene-and-current-research>.

On the one hand, this project is innovative because it is not a **genetic project** that concerns only the study of the expression of a single gene, but rather a **genomic project** concerning the whole genome, that is to say the sum of all the genetic information that make up the organism.



13

This difference in approach was made possible by the development of bioinformatics¹⁴ and the progressive decrease of the costs of use of next-generation sequencing tools (which dropped from \$ 100,000,000 in 2001 to \$ 1,000 in 2017)¹⁵ :



¹³ NHS, Genomics Education Programme, <https://www.genomicseducation.hee.nhs.uk/>.

¹⁴ One genome represents a volume of 300 gigabytes.

¹⁵ National Human Genome Research Institute (NHGRI), <https://www.genome.gov/sequencingcostsdata/>. See also RIORDAN, *Op. Cit.* (Annex 5).

On the other hand, this project is innovative because it is not a genomic project that identifies variations in a gene or in an association of genes that cause a rare disease or cancer (as many projects do¹⁶). On the contrary, it is a genomic project that is in a way “**inversed**”¹⁷ as the starting point is a deleterious mutation of the FBN1 gene that has already been previously identified, from which we try to identify the action of one or more modifier gene(s) that will be able to counteract this FBN1 deleterious mutation. This approach could thus provide a better understanding of the variability and severity of Marfan afflictions.

The identification of these possible modifier genes with the tool that we wish to set up could thus ultimately lead to the development of new therapeutic approaches that could prevent and/or limit the damage caused by Marfan syndrome.

This project is thus truly innovative and pioneering within the framework of Marfan syndrome¹⁸ and is in line with the current approach of the “**Genome-wide association studies**” or “**GWAS**”, which have already met with some success in the context of other rare diseases¹⁹.

It should also be noted that once the tool is in place, it could, if necessary, be reused to carry out other similar projects for other rare diseases.

The tool would thus have a “pilot” project dimension that would allow for the hosting and bringing together of other similar projects dealing with other rare diseases.

And so could be used as the starting point for a tool of another magnitude that would gradually grow as it is used and the cohorts that could join it.

¹⁶ MACARTHUR J. et al., “The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)”, *Nucleic Acids Res.*, 4 January 2017; 45(Database issue): D896–D901. Available online on 28 november 2016. doi: [10.1093/nar/gkw1133](https://doi.org/10.1093/nar/gkw1133).

¹⁷ Sometimes called: “reverse genomics”.

¹⁸ For completeness, the existence should be mentioned of an ongoing research carried out at exome level by Dr. Gabor Matyas at the “Zentrum für Kardiovaskuläre Genetik und Gendiagnostik” de Zurich in Switzerland entitled (“[Molecular basis of Marfan syndrome and related disorders: Whole-exome sequencing and targeted therapy](#)”) evoking certain possible complementarities with the main project that could be explored at a later stage.

¹⁹ See Table 1 and references in RIORDAN, *Op. cit.* (Annex 5).

6. **Combination of two strategies.** In order to identify the action of possible modifier genes that would play a protective role in relation to the action of a deleterious mutation of FBN1, two strategies can be envisaged²⁰.

The first strategy, known as “*bottom-up*”, consists in identifying a statistically representative cohort of different families affected by a previously identified deleterious mutation of FBN1, some members of which seem to be protected from the most harmful effects in an attempt to identify possible modifier gene (s) within this cohort.

The second strategy, known as “*top-down*”, is to explore the vast genomic databases that already exist and that are constantly growing and multiplying around the world²¹ in an attempt to find mutations of FBN1 previously identified as deleterious and hope to be able to identify in the genome of these individuals possible protective modifier genes.

In the present case, the main project starts with the first strategy (bottom-up) while leaving the possibility of combining later on with the second strategy (top-down)²².

7. **101 Genomes.** The main project is to build a platform on the first basis of a cohort of 101 Marfan patients. The costs of setting up the platform are calculated on the basis of this initial cohort of 101 genomes.

The starting number of 101 genomes to feed the main project may appear to be low in order to detect a possible modifier/protective gene for Marfan syndrome. However, there is at least one known precedent in which the study of a limited cohort has led to the identification of modifier genes.

²⁰ These two strategies are described in Annex 5 as follows: “*Strategies to Identify Genetic Modifiers in Humans: Several approaches can be used to study modifier genes directly in human populations, including comparative expression profiling, genome-wide association studies (GWASs), and family-based association analyses. To maximize the potential therapeutic benefit of any findings, studies are often designed to search for shared variants (candidate modifier genes) in individuals that are protected from a particular disease phenotype despite carrying a causal target allele. Two strategies are used to pursue such protective modifiers: (1) comprehensive sequence profiling of a population or family carrying the disease allele to identify variants that associate with phenotypic severity and (2) targeted sequencing of genes known to harbor disease-associated mutations in a healthy population to identify individuals who are phenotypically unaffected despite harboring causal alleles. Once these individuals are identified, mechanisms of resistance can be studied to identify candidate therapeutic targets for susceptible individuals. The targeted sequencing strategy is exemplified by the Resilience Project, an effort to identify unaffected carriers of mutations known to cause highly penetrant and severe childhood diseases. Screening for 874 distinct Mendelian disease-causing mutations across 589,306 genomes, Chen and colleagues identified 13 asymptomatic adults carrying disease alleles. [...] [T]his work provides proof-of-principle that individuals resistant to highly penetrant genetic diseases can be identified, paving the way for mechanistic studies to discover modifier genes that may be therapeutically manipulated to benefit susceptible individuals.*” in RIORDAN, *Op. cit.* (Annex 5).

²¹ As for example “*The 100.000 genomes project*” in the UK: <https://www.genomicsengland.co.uk/the-100000-genomes-project/>.

²² See *infra* paragraph 17.

Annex 5 describes another GWAS project conducted from a cohort of 91 patients that identified the existence of a modifier/protective gene that delays the onset of cystic fibrosis²³.

In any case, it is essential to adopt a “smart analysis” approach by giving priority to selecting cohort composition strategies that will maximize the chances that the analysis of 101 genomes will advance scientific knowledge.

Without forgetting that it will always be technically possible to extend this original cohort since it is explained below that the projected platform can accommodate a 10 times larger cohort²⁴.

²³ “Another search for modifiers of lung infection in CF patients used an “extreme phenotype” study design to increase efficiency of modifier gene identification. Exome sequencing was performed on individuals with CFTR mutations specifically selected from the top and bottom of the distribution for age at onset of infection. **From a set of just 91 affected subjects**, missense mutations in DCTN4 (dynactin 4 [MIM: 614758]) were linked to early infection. A validation set of 696 affected subjects provided further support for a modifier effect of DCTN4 on susceptibility to airway infection conferred by CFTR mutation” (underlining added) in RIORDAN, *Op. cit.* (Annex 5) quoting EMOND M.J. et al., “Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis” in *Nat. Genet.* 44, 886–889 (2012), doi:10.1038/ng.2344.

²⁴ See *infra* paragraph 15.

2.2 PARTNERS FOR IMPLEMENTATION OF THE PLATFORM

a) *Academic Partners*

8. **Belgium.** The “*101 Genomes Marfan Project*” (hereafter the “P101GM”) could not start in Belgium without the support of academic partners at the institutions that care for Marfan patients.

At this stage, the academic institutions and services revealing both clinical and academic expertise in the field of Marfan syndrome in Belgium are:

- the [Center for Medical Genetics Ghent](#) of the [Ghent University Hospital](#) (UZ Gent);
- the [Center of Medical Genetics](#) of the [University Hospital of Antwerp](#) (UZA);
- the Congenital Heart Disease Centre of the Saint-Luc University Clinics in Brussels (UCL);
- the Department of Cardiac Surgery at Sart-Tilman University in Liège (CHU);
- the Genetics Department of the Queen Fabiola Children’s University Hospital in Brussels (HUDERF);
- The Cardiac Department of UZ Leuven;
- The Genetics Department of the Erasme Hospital (ULB).

The partnership with several of these institutions should make it possible to identify the best candidates to compose the cohort of 101 patients.

9. **Europe / VASCERN.** It is preferable to design the P101GM so that it can, if necessary, be integrated into a European dynamic.

This is all the more so since the recognition by the Council of Member States of the Union on 15 December 2016 of the “*European Reference Networks*” (called “ERNs”). And particularly since the launch in March 2017²⁵ of the “*European Reference Network on Rare Multisystemic Vascular Diseases*” (called “VASCERN”²⁶), that includes a working group dedicated to “*Hereditary Thoracic Aorta Diseases*” (called “HTAD WG”²⁷) which brings together European specialists in Marfan Syndrome.

²⁵ The EU Commission define the ERNs as follows: “*European Reference Networks (ERNs) are virtual networks involving healthcare providers across Europe. They aim to tackle complex or rare diseases and conditions that require highly specialised treatment and concentrated knowledge and resources. Health systems in the European Union aim to provide high-quality, cost-effective care. This is particularly difficult with rare or low-prevalence complex diseases or conditions. Between 5 000 and 8 000 rare diseases affect the daily lives of around 30 million people in the EU. **How does it work?** To review a patient’s diagnosis and treatment, ERN coordinators convene ‘virtual’ advisory boards of medical specialists across different disciplines, using a dedicated IT platform and telemedicine tools. The process and criteria for establishing an ERN and for selecting its members are set in EU legislation. **24 networks:** The first ERNs were launched in March 2017, involving more than 900 highly-specialised healthcare units from over 300 hospitals in 26 Member States. 24 ERNs are working on a range of thematic issues including bone disorders, childhood cancer and immunodeficiency*” https://ec.europa.eu/health/ern/policy_en.

²⁶ The European Reference Network on Rare Multisystemic Vascular Diseases: <http://vascern.eu/>.

²⁷ <https://vascern.eu/expertise/rare-diseases-wgs/aorta-wg/#1461009962709-540360ce-6a2e>.

The launch of VASCERN means ensuring that the platform is not limited to Belgium alone and that it can be seen as part of a European perspective.

This would make it possible to ensure that the platform is known to the main European specialists in Marfan syndrome and to allow them to use it as a support for research projects or even decide to contribute to its potential expansion by proposing the addition of genomes to be sequenced or already available.

As will be explained later, the genomic data that feeds the platform are collected in international standard formats widely used in bioinformatics²⁸. This data will not pose any interoperability problems. It could therefore be ensured in consultation with VASCERN that the genomic data collected could potentially be reused one day during the actual implementation of the Clinical Patient Management System (or CPMS) as part of the ERNs deployment.

This approach would thus correspond precisely to that communicated at the 2017 Symposium organised by Eurordis²⁹, at which it was stated that the setting up of ERN registries was part of a long-term perspective and that this perspective should in no way hinder the emergence of specific national initiatives (led, for example, by patient associations) which would at a later stage fit within the framework of ERNs and which would in fact power and drive them.

b) Biobank

10. **Biobank.** The P101GM could be conducted in partnership with the “*Biothèque Wallonie Bruxelles*”³⁰, that could be in charge of organizing the conservation and referencing of samples collected as part of the platform’s implementation. As P101GM is intended to be part of a European approach, the biobank solution that will be chosen could be part of BBMRI.be³¹.

c) Business Partners

11. **Partners.** At this stage, several potential business partners have been identified for the three main stages of the platform’s operational implementation: (a) sequencing, (b) data processing/storage and (c) data crossing. They are: (a) Macrogen or BGI, (b) Bluebee or Nijmegen University and (c) Medisapiens.

²⁸ Recognized by the “*Global Alliance for Genomics and Health*” (GA4GH) et retained by the IRDiRC See the reference to R. TONDA, S. LAURIE, S. BELTRAN, AND AL. at: http://www.irdirc.org/wp-content/uploads/2017/03/PAR1_FO_05-Gut.pdf (Annex 10).

²⁹ “*EURORDIS-Rare Diseases Europe is a unique, non-profit alliance of over 700 rare disease patient organisations from more than 60 countries that work together to improve the lives of the 30 million people living with a rare disease in Europe*”. <http://www.eurordis.org/who-we-are>.

³⁰ <http://www.biotheque-wallonie-bruxelles.be/>.

³¹ <http://www.bbMRI.be/> .

i) Sequencing

12. **Option Illumina.** MacroGen³² could be responsible for sequencing the 101 genomes of the initial cohort.

The price charged by MacroGen for sequencing an entire genome (WGS) and generating a “FASTQ”³³ file containing all sequenced information using the technology developed by Illumina (in 30x)³⁴ is around 1,500 euros (excl. VAT).

This price could be reduced to around 1,200 euros per genome by sending the materials to be sequenced in groups of at least 50 units.

Other users of the Illumina technology could be approached to perform sequencing but the costs would remain around 1,000 euros (excl. VAT) per genome. Examples include the Hartwig Foundation³⁵ in the Netherlands, the UZLeuven³⁶ genetics laboratory and the BrightCore³⁷ inter-university laboratory in Brussels.

The Illumina technology for sequencing the entire genome is the reference technology³⁸.

13. **Option BGI.** A second option is to consider entrusting sequencing to the *Beijing Genomics Institute* (hereafter BGI)³⁹.

³² **MacroGen Europe** - Meibergdreef 31, 1105 AZ, Amsterdam, The Netherlands; Tel: +31 (0)20 333 7563. Website: www.macrogen.com/eng/. Contacts: Dr. HyungTae Kim, Director of MacroGen Europe et Doori Jung. Email: d.jung@macrogen-europe.com.

³³ The FASTQ format is a text file format that has established itself de facto as the standard format for storing high speed sequencer outputs (including Sanger and Illumina). See for example: COCK, P. et al. “*The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants*”, *Nucleic Acids Research*, 2010 (published online 16 December 2009), Vol. 38, No. 6 1767–1771 doi:10.1093/nar/gkp1137:

“One of the core issues of Bioinformatics is dealing with a profusion of (often poorly defined or ambiguous) file formats. Some ad hoc simple human readable formats have over time attained the status of de facto standards.[...] In the area of DNA sequencing, the FASTQ file format has emerged as another de facto common format for data exchange between tools. It provides a simple extension to the FASTA format: the ability to store a numeric quality score associated with each nucleotide in a sequence”.

³⁴ Illumina : <https://www.illumina.com/>.

³⁵ Hartwig Foundation : <https://www.hartwigmedicalfoundation.nl/>.

³⁶ UZLeuven : <https://www.uzleuven.be/en/centre-for-human-genetics>. Prof. Joris Vermeesch, joris.vermeesch@uzleuven.be, 32 16 34 5941, Campus Gasthuisberg, Herestraat 49, B - 3000 Leuven.

³⁷ BrightCore : <http://www.brightcore.be/>.

³⁸ Illumina competitors are: BGI (<http://www.genomics.cn/en/index>) Pacific Biosciences (<http://www.pacb.com/>), Oxford Nanopore (<https://nanoporetech.com/>), Complete Genomics (<http://www.completegenomics.com/>), Roswell Biotechnologies (<https://www.roswellbiotech.com/>).

³⁹ <http://www.genomics.cn/en/index>

Through the intermediary of Professor H. G. Brunner of the Radboud University in Nijmegen⁴⁰, BGI could propose a tariff of about 300 euros (excl. VAT) for a quality sequencing of the whole genome and the production of FASTQ files.

To offer such pricing conditions, the BGI uses its own “BGISEQ-500 30x” machines, which are derived from the technology developed by Complete Genomics⁴¹.

According to information gathered from several sources, the quality of the results obtained by BGI is good.

The contract between the University of Nijmegen and the BGI provides that the BGI commits itself to keep no sequenced information (and to destroy any information in its possession within a very short period of time) after its actual transmission and receipt by the F101G.

14. **Illumina vs BGI.** The confrontation of Illumina vs BGI technologies has apparently not yet been conducted on an objective basis.

The results of the only comparison exercise identified at this stage were published on April 1, 2017⁴². The fact that this exercise was conducted exclusively by Chinese researchers, some of whom are employees of BGI (as indicated in the conflict of interest statement at the end of the article) obliges one to examine its results with caution.

This comparison reports that the performance of the BGISEQ is very slightly lower than that of Illumina’s HiSeq⁴³:

Table 4: Performances of variation calling of dataset*

Variant Type	Metrics	BGISEQ-500 PE50	BGISEQ-500 PE100	HiSeq2500 PE150
SNPs	True positive	3 006 132	3 071 579	3 084 449
	False positive	15 203	6907	4318
	False negative	186 825	121 379	108 508
	Precision	99.50%	99.78%	99.86%
	Sensitivity	94.15%	96.20%	96.60%
	FPR	0.00060%	0.00020%	0.00017%
	FNR	5.85%	3.80%	3.40%
indels	True positive	261 867	326 810	355 728
	False positive	16 931	22 246	7981
	False negative	107 311	42 391	13 751
	Precision	93.93%	93.63%	97.81%
	Sensitivity	70.93%	88.52%	96.28%
	FPR	0.00067%	0.00069%	0.00032%
	FNR	29.7%	11.48%	3.72%

*Above, the first four metrics are calculated using rtg-tools software. True positive (TP) is the number of SNPs that are found in the high-confidence reference dataset, false positive (FP) is the number of SNPs that are not found in reference dataset, and false negative (FN) is the number of SNPs that are found in high-confidence reference dataset but are not found in reference dataset. Precision is $TP/(TP+FP)*100$. Sensitivity is $TP/(TP+FN)*100$. FPR is $FP/(all\ high\ confident\ region\ length-TP)*100$, where high-confident region length equals 252 9164 928 bp, which comes from GIAB released high-confidence variants datasets [19]. FNR is $FN/(FN+TP)*100$.

It remains to be seen whether this slight difference in performance could increase the risk of missing certain mutations, or focusing on false positives

⁴⁰ <http://www.ru.nl/english/people/brunner-h/>.

⁴¹ <https://www.bgi.com/us/sequencing-services/dna-sequencing/whole-genome-sequencing/>

⁴² GAO S. et al., « A reference human genome dataset of the BGISEQ-500 sequencer », *GigaScience*, Volume 6, Issue 5, 1 May 2017, Pages 1–9, <https://doi.org/10.1093/gigascience/gix024>.

⁴³ GAO S. et al., *Op. Cit.*, page 7.

15. **Observation.** The price offered by BGI is particularly attractive. The quality offered by Illumina is a guarantee of the relevance and interoperability⁴⁴ of the results obtained. The progression of the sequencer market must be monitored to identify possible alternatives⁴⁵.

ii) *Processing and storage of genomic data*

16. **Option 1.** Once the genomes are sequenced, Bluebee⁴⁶ would intervene in the following way:
- 1- repatriate and store the FASTQs for a period of one year in a secure manner for an amount of 35 euros per unit. After the first year, storage of FASTQs would cost 10 euros per year per unit;
 - 2- From the FASTQ files, generate BAM and VCF files⁴⁷ that it would actively store for one year for 175 euros per unit. After the first year, active storage of BAM and VCF files would cost 0.60 euros per year per unit;
 - 3- enable secure access⁴⁸ to the stored BAM and VCF data to a partner of our choice⁴⁹.
17. **Option 2.** The University of Nijmegen offers to take over the repatriation of the FASTQs and the generation of BAMs and VCFs at no extra cost. This option should be given careful consideration.

It should be noted that:

- 1) this option could potentially threaten the integrity of data to be transmitted - as the volume would be almost doubled (BAM files being larger than FASTQ and VCF);
- 2) this option would place the F101G in a situation of dependence on a single partner for access to sequenced data, which could be a problem in the event that sequenced data from other projects were to be repatriated and;
- 3) that this option could create possible small interoperability difficulties.

⁴⁴ Which guarantees that it can be linked to larger genomic projects using Illumina technology such as the 100,000 Genomes UK Project.

⁴⁵ In particular, we will remember the surprising claims of the new entrant Roswell Biotechnologies, which announces a WGS 30x sequencing for \$ 100 in 1 hour: <https://www.roswellbiotech.com/>

⁴⁶ **Bluebee** - Laan van Zuid Hoorn 57, 2289 DC Rijswijk, The Netherlands; Tel: +31 (0)64 160 3158. Website: www.bluebee.com. Contact: Pierre Bourbon - Sales Director Benelux & Nordics. Email: pierre.bourbon@bluebee.com; Tel: +31 (0)88 214 0200.

⁴⁷ For “Variant Call Format”. On the website of the “Global Alliance for Genomics and Health” (GA4GH) there is the following definition of VCF File : “VCF is a text file format (most likely stored in a compressed manner). It contains meta-information lines, a header line, and then data lines each containing information about a position in the genome. The format also has the ability to contain genotype information on samples for each position”. The full description of this internationally used bioinformatics standard is available at: <https://samtools.github.io/hts-specs/VCFv4.2.pdf> (Annex 9).

⁴⁸ See Annex 6: Bluebee white paper explaining the measures put in place to ensure end-to-end security of the data it retains.

⁴⁹ It provides a maximum cost of 0.10 euros per gigabyte for downloading files out of the platform and outside its network (=Off-net).

iii) Data crossing

18. **Medisapiens.** This partner for the crossing could be the Finnish company Medisapiens⁵⁰ recommended by Bluebee.

This company has developed a platform for cross-checking genomic and phenotypic data. This platform should not require any adaptation costs to accommodate the tool and could be used as is according to its designers.

Medisapiens could handle the integration of phenotypic data for an amount ranging from 5 to maximum 30 euros per patient (regardless of the language in which these data were collected).

Once the phenotypic data have been integrated, they could then be cross-referenced with genomic data on the secured access platform⁵¹ developed by Medisapiens.

An unlimited number of authorised persons could obtain secure access to the platform's anonymised data in accordance with procedures to be defined by the Foundation within the framework of the project.

Once set up by Medisapiens, the platform will be able to host up to 1,000 genomes (which is much higher than the 101 genomes of the initial cohort).

At the current stage, Medisapiens charges a license fee for the use of the platform, which allows cross-checking genomic and phenotypic data collected of up to 39,000 euros for the first year and thereafter: either 19,000 euros per year as long as the cohort does not exceed 101 patients or 28,000 euros per year when the cohort exceeds 101 patients.

⁵⁰ **Medisapiens Ltd** - Mikonkatu 17 C, 00100 Helsinki, Finland; Tel: +358 45 847 8878. Website: <https://www.medisapiens.com/>, Email: contact@medisapiens.com via Pierre Bourbon (Email: pierre.bourbon@bluebee.com; Tel: +31882140200).

⁵¹ See Annex 7 for a description of the security measures taken by Medisapiens.

2.3 DRAFT BUDGET

19. **Budget.** The tariff information collected from the potential partners identified at this stage enables the preparation of a first draft budget retaining for the moment the services offered at the highest price (Option 1 for sequencing and processing)⁵².

On the basis of this information, it therefore appears that the overall budget for making available to researchers a cross-genomic/phenotypic platform for a cohort of 101 patients over a ten-year period would be in the order of 880,000 euros:

101 Genomes Marfans - Ebauche de budget					TVA	121%
	Prix	unité	cohorte initiale	Total		
<u>Base de données génomiques</u>					€ 232.565,63	27%
1 Assurance responsabilité sans faute ETHIAS			en cours de négociation			
2 Collecte et conservation d'ADN	€ 181,50	1	101	€ 18.331,50		
3 MACROGEN séquençage du génome entier (WGS) -> FastQ	€ 1.452,00	1	101	€ 146.652,00		
4 BLUEBEE Génération des BAM et VCF à partir des FastQ et stockage 1 an	€ 211,75	1	101	€ 21.386,75		
5 BLUEBEE Stockage actif des BAM et VCF durant 9 années supplémentaires	€ 34,85	9	101	€ 31.676,83		
6 BLUEBEE Stockage des FastQ durant 1 an	€ 34,85	1	101	€ 3.519,65		
7 BLUEBEE Stockage des FastQ durant 9 années supplémentaires	€ 12,10	9	101	€ 10.998,90		
<u>Base de données phénotypiques et plateforme croisée</u>					€ 395.966,30	45%
1 MEDISAPIENS Plateforme existante pas de coût	€ -	1	1	€ -		
2 MEDISAPIENS intégration des données phénotypiques pour 101 patients	€ 36,30	1	101	€ 3.666,30		
3 MEDISAPIENS Mise en place et licence d'utilisation 1ère année	€ 47.190,00	1	1	€ 47.190,00		
4 MEDISAPIENS Licence d'utilisation annuelle jusqu'à 101 génomes*	€ 22.990,00	9	1	€ 206.910,00		
5 CYTOMINE Développement de l'algorithme	€ 20.000,00	1	1	€ 20.000,00		
6 CYTOMINE accompagnement	€ 7.800,00	9	1	€ 70.200,00		
7 CYTOMINE Serveurs	€ 4.800,00	10	1	€ 48.000,00		
*Accès illimité mais contenance de la plateforme limitée à 1000 génomes						
<u>Comité scientifique</u>					€ 72.600,00	8%
Forfait annuel pour la composition et la gestion annuelle de la cohorte	€ 7.260,00	10	1	€ 72.600,00		
<u>Administration</u>					€ 175.346,37	20%
1 Administration première année	€ 57.016,62	1	1	€ 57.016,62		
2 Administration annuelle années suivantes	€ 8.468,00	9	1	€ 76.212,03		
3 Frais annuels de la FRB première année	€ 18.453,88	1	1	€ 18.453,88		
4 Frais annuels de la FRB années suivantes	€ 2.629,31	9	1	€ 23.663,83		
					€ 876.478,30	100%

This would be the breakdown of costs for an initial cohort of 101 over a 10 year period :

- (1) setting up the **genomic** database is of around 235,000 euros (representing 27% of the total budget);
- (2) setting up the **phenotypic** database and licence to use the platform provided by Medisapiens around 395,000 euros (representing 45% of the total budget).
- (3) composition and management of the cohort by the **scientific committee** around 72,600 euros (representing 8% of the total budget).
- (4) **administrative costs** would be around 175,000 euros (representing 20% of the total budget).

⁵² See draft budget in Annex 2.

To cover this budget for 10 years, an amount of 350,000 euros would need to be collected for the actual setting up of the platform in the first year followed by an amount of 60,000 euros each year for the following 9 years :

Première année	€ 348.276,70	1	€ 348.276,70
9 années suivantes	€ 58.689,07	9	€ 528.201,59
			€ 876.478,30

As a precautionary measure⁵³, we retain an alternative budget of 500,000 euros during the first year and 75,000 euros during each of the following 9 years. Representing a total budget of 1,175,000 euros over 10 years:

Première année	€ 500.000,00	1	€ 500.000,00
9 années suivantes	€ 75.000,00	9	€ 675.000,00
			€ 1.175.000,00

It should be noted that this ideal budget may be adapted according to the sums collected. For example, the medical collection option can be considered at a later stage and the BGI sequencing option allows an overall saving of around 150,000 euros.

⁵³ It should be noted that these budgets are made without considering any potential VAT exemption or the possibility of recovering it.

2.4 FROM 101 TO 1000 GENOMES

20. **Incremental cost per additional genome.** The Medisapiens platform can accommodate up to 1000 genomes for an adjusted annual license cost of 28,000 euros . In this context, it seems justified to calculate an incremental cost per genome added to the initial cohort.

Subject to further examination and additional checks, at the current stage, this incremental cost would be more or less 2,550 euros (TVA Incl.) in variable costs (without contribution to licensing, management and administration costs).

It is possible to retain an incremental cost in only variable costs of around 705 euros (TVA Incl.) for the integration on the platform of a genome already available in FASTQ format.

a) *Extension of the original cohort*

21. **Two approaches.** Determining these cost levels allows us to consider two complementary approaches to expanding the original cohort:

- 1) partnerships with other projects;
- 2) call for new participants already genetically diagnosed.

These two approaches are discussed successively in the following paragraphs.

22. **Partnerships with other projects.** The first cohort enlargement approach is part of the second strategy described above⁵⁴ for the identification of possible modifier gene(s) and involves exchanges with other large-scale genomic information collection projects such as: GnomAD/Exac⁵⁵, Decode Genetics⁵⁶, “*The 100.000 genomes project*”⁵⁷, “*Genomics and health*”⁵⁸, SequOIA⁵⁹, etc.

The idea being to check whether among the large sequenced populations that make up these different projects, individuals affected by harmful mutations of FBN1 already previously identified as confirming a diagnosis of Marfan syndrome could not be found⁶⁰.

⁵⁴ See *supra* paragraph 7.

⁵⁵ USA: <http://gnomad.broadinstitute.org/>.

⁵⁶ Iceland: <https://www.decode.com/>.

⁵⁷ GB: <https://www.genomicsengland.co.uk/the-100000-genomes-project/>.

⁵⁸ Global projet: <http://genomicsandhealth.org/>.

⁵⁹ France: <http://www.aphp.fr/contenu/selection-du-projet-segoia-porte-par-lassistance-publique-hopitaux-de-paris-linstitut-curie>.

⁶⁰ Public genetic databases related to Marfan syndrome (and especially the UMD FBN1 database <http://www.umd.be/FBN1/>) or generalist genetic databases (such as “*The Human Gene Mutation Database*” - HGMD (<http://www.hgmd.cf.ac.uk/ac/index.php>)) and the sharing of private genetic databases of the various centers of expertise that study Marfan syndrome could be particularly useful in carrying out this comparison.

The identification of such mutations in these databases could enable, if patient identification or feedback to patients is permitted⁶¹ :

- either to reclassify as benign a previously identified deleterious mutation (and thus allow to try and identify another mutation as being the actual deleterious one in this patient) ;
- or to identify a Marfan patient (and possibly, if necessary, to warn a participant in one of these projects who does not know that he or she is affected by the syndrome and advise of the precautions he or she should consider) ;
- or to identify an individual who should be affected by Marfan symptoms but who does not present any (strong) symptoms and who could possibly be naturally protected from the damages of a deleterious mutation of FBN1 by the action of modifier genes⁶².

In this last hypothesis, the objective would be to try to obtain access to the FASTQ file⁶³ of this patient whose genome has already been sequenced in order to be able to add it to the cohort at a lower cost (in the ideal hypothesis it would be possible to recontact the individual concerned to invite and obtain their consent to participation in the study).

23. **Call for new participants already genetically diagnosed.** The second approach to cohort expansion is to extend, if necessary (i.e., if the original cohort would not have identified any modifier/protector genes), the cohort beyond 101 individuals by identifying other prospective genetically diagnosed participants with characteristics that would potentially identify the protective gene action in them.

⁶¹ Or possible. The current version of Article 11 of the Belgian law of 19 December 2008 on the procurement and use of human body material for human medical applications or for scientific research purposes (*Moniteur Belge* of 30 december 2008, page 68.774) foresees:

“If, during an operation carried out on human body material or during the use of human body material, analyses generate information with significant consequences for the donor's state of health, the donor is entitled to this information. (...)”

Physicians who take cognizance of the information referred to during an operation or use, managers of human body material, and the chief physician of the hospital from which the sample was taken, are each, within the framework of their functions and skills, responsible for the application of paragraphs 1 and 2” (Free translate).

And that in future law, it is already established that this provision of Belgian law will be retained and extended to analyses carried out on “*blood and blood derivatives of human origin*”.

This Belgian legal provision therefore implies not only that: (1) the person who has consented to have his or her blood tested has the right to be informed of the significant consequences for his or her health which the tests would have revealed, but also (2) that all doctors involved in the process are liable if this information has not been transmitted to him or her.

⁶² It is this approach which, for Marfan syndrome alone, would be similar *mutatis mutandis* to the approach adopted by the promoters of the above-mentioned project “*The Resilience project*” (<http://resilienceproject.com/>) (see footnote 20).

⁶³ Bluebee recommends that data be repatriated to the FASTQ format to ensure consistency of VCF data conversion and that all data on the platform has been converted to this format in a harmonised manner.

b) Reuse and/or extension of the platform

24. **Reuse and/or extension.** As explained above, once implemented for 101 genomes, the platform developed by Medisapiens can contain up to 1000 genomes.

In this context, account should be taken of the possibility of reusing or expanding the use of the platform to make it available for other projects involving other rare diseases, such as Loeys-Dietz, Beals and even Elher-Danlos syndromes, which also affect connective tissue.

The cohort model of 101 Genomes could therefore be replicated for other rare diseases that could benefit from the bioinformatics platform and the partnerships set up.

2.5 FINANCIAL PARTNERS TO FINANCE THE PLATFORM

a) *Partners & Funding sources*

25. **Two partners.** At this stage, two main partners have been identified for the financing of the platform's implementation. These are the King Baudouin Foundation (hereafter "KBF") and Delen Private Bank.

26. **Three sources of funding for the platform.** At this stage, three funding sources have been identified for the implementation of the platform.

The platform's funding will mainly come from **donations** and fundraising events.

The F101G will also actively explore the possibilities of obtaining **subsidies** to finance the establishment of the platform⁶⁴.

An additional source of funding could be found on an ad hoc basis by charging companies a **fee for secured access** to the platform, while guaranteeing royalty-free access to academic and non-profit structures⁶⁵.

27. **Tax deductibility of donations.** The collection of donations is currently the platform's main funding channel. Belgian law offers the possibility of benefiting from a tax reduction of 45% for donations when certain conditions are met⁶⁶. It is therefore essential to ensure that donors have the possibility of claiming tax deductions for their donations.

However, this possibility is only offered by Belgian Law to Belgian organisations specifically designated by the law or to organisations which have previously obtained specific approval⁶⁷. At this stage, the F101G has not obtained this approval and the donations it would receive would therefore not be tax deductible.

⁶⁴ By examining, for example, the opportunities offered by:

- the Belgian Health Care Knowledge Center (KCE): <http://kce.fgov.be/en/content/funding-opportunities-2017>;
- the *European Research Area Network on Cardiovascular Diseases* (ERA-CVD) : <http://www.era-cvd.eu/> and;
- the *European Joint Programme on Rare Disease* (EJP-RD) <https://www.era-learn.eu/public-to-public-partnerships/european-joint-programme-cofund-ejp-cofund>.

Possibly by asking for the occasional help of Monique Marrec-Fairley of BioWin for this purpose: Monique.Marrec-Fairley@biowin.org +32(0)485/46.46.27 <http://www.biowin.org/biowin/fr/5573-lequipe.html>).

⁶⁵ This approach is comparable to that adopted by *Genecards* (www.genecards.org):

"Other companies allow access to their databases for a fee, such as GeneCards (organised and run primarily by the Weizmann Institute of Science and LifeMap Sciences) which is freely available for academic, non-profit use but requires a commercial licence for any other organisation wishing to access the data. This system therefore allows research to progress without hindrance whilst also making a profit from commercial enterprises wishing to use their data" GENOMIC DATA 101, Edition 2017, *An introduction to Genomic data*, page 9.

⁶⁶ https://finances.belgium.be/fr/particuliers/avantages_fiscaux/dons.

⁶⁷ https://finances.belgium.be/fr/particuliers/avantages_fiscaux/dons and certain international organisations.

After a series of meetings, initiated by Bank Delen, with representatives of the King Baudouin Foundation, it became clear that the King Baudouin Foundation⁶⁸ would be an ideal partner with the added advantage of ensuring donors might benefit from a tax deductibility of 45% of the amount of any donations in excess of 40 euros.

28. **Clarification on the deductibility of donations.** Concerning the tax deductibility of donations, it should be noted :

Firstly: that the amount of the tax reduction is **capped**.

- Thus, **for an individual**, the tax reduction corresponds to 45% of the amount donated, duly authenticated by certificate, however this reduction cannot exceed:
 1. either 10% of total net revenues;
 2. or 376,350 euros (tax year 2017 - income 2016).
- And **for a company**, the amount of tax reduction cannot exceed:
 1. either 5% of taxable income;
 2. or an amount of 500,000 euros.

Secondly: that the excess amount of any tax reduction above the threshold, **cannot be spread over several fiscal years**. The tax reduction that is not absorbed in one year is lost. In order to avoid losing the benefit of the tax reduction above the threshold, two approaches are suggested:

- **either by allowing another person to take advantage of the surplus tax deduction** *by remitting the excess amount to that person and entrusting them to proceed with donation to the Foundation*. This intermediary person will therefore be the ad hoc recipient of the tax certificate.
- **or by granting a loan** for the total amount intended to be transferred to the Foundation, of which the lender would renounce annually an amount corresponding to the maximum annual amount according to the thresholds in force.

Thirdly: that the tax reduction is charged on the overall taxable income (according to the progressive scale in instalments). These are mainly **professional and property income** (but not withholding tax), as well as equity income (if their taxation according to the progressive scale results in a lower rate than withholding tax - i.e. in principle 30%).

⁶⁸ Mrs Brigitte Duvieusart (Philanthropy Advisor, duvieuart.b@kbs-frb.be) and Mrs Annemie T'Seyen (Scientific secretariat, Fund for Scientific Medical Research, tseyen.a@kbs-frb.be).

b) King Baudouin Foundation

29. **The King Baudouin Foundation.** The King Baudouin Foundation (hereafter KBF) is a public benefit foundation created in 1976, on the occasion of King Baudouin's 25 years of reign. The KBF's mission is:

“[T]o contribute to a better society. The Foundation is an actor for change and innovation, serving the public interest and increasing social cohesion in Belgium and Europe. We seek to maximize our impact by strengthening the capacity of organizations and individuals. We also stimulate effective philanthropy by individuals and corporations. The Foundation's key values are integrity, transparency, pluralism, independence, respect for diversity, and promoting solidarity. The Foundation's current areas of activity are poverty and social justice, philanthropy, health, civic engagement, developing talents, democracy, European integration, heritage and development cooperation”⁶⁹.

The KBF has been chaired since 2015 by Her Majesty Queen Mathilde of Belgium.

According to the figures provided by the KBF, it managed more than 747 million assets in 2016 and during the same year 2016, the KBF and the funds it manages allocated 41.9 million euros in support to 1,911 organisations and 300 individuals for financing projects in the fields of action it covers⁷⁰.

The establishment of a partnership with the KBF within the framework of the P101GM presents the undeniable advantages of benefiting from the vast experience of the KBF and grants real visibility and legitimacy to the project.

This partnership could also enable the possible development of specific synergies with other projects of which the funds are also hosted by the King Baudouin Foundation.

30. **Agreement with the KBF.** An Agreement between the KBF and representatives of the F101G has been drawn for the joint opening of a 101 Genomes Fund, the “Fonds 101 Genomes”. This Agreement, together with the “*Règles générales de fonctionnement pour les Fonds gérés par la Fondation Roi Baudouin*” (General operating rules for the funds managed by the KBF), is set out in the Annex 4.

The Agreement provides that this “Fonds 101 Genomes”⁷¹ will be hosted by the KBF and its asset management will be entrusted to the Delen Bank⁷². All the sums collected for the financing of the P101GM will therefore be paid into the bank account of the “Fonds 101 Genomes”. The funds collected will be released from the account of the “Fonds 101 Genomes” to finance the activities of the F101G and, primarily, the implementation of its genomic/phenotypic platform created within the framework of the P101GM.

Three articles of this Agreement with the KBF are mentioned briefly below.

⁶⁹ <https://www.kbs-frb.be/en/About-us/What-we-do/Mission-vision-values>.

⁷⁰ <http://en.calameo.com/read/00177429575c0c52f76e2?authid=jQaRtXnxhXtw>.

⁷¹ Annex 4: Convention, Article 1 – Dénomination.

⁷² Annex 4: Convention, Article 4 – Dotation: “Funds that are not immediately used will be invested in defensive funds, with guaranteed capital and a high level of liquidity so that the amounts invested are always available very quickly” (Free translation).

31. **Synergies.** Article 2 of the Agreement provides that the object of the “Fonds 101 Genomes” is:

“[T]o support the research and application of therapies for people suffering from rare diseases, particularly connective tissue abnormalities and specifically to work , as a first step, to discover, adopt and implement innovative therapies that will benefit children and adults with Marfan Syndrome. The Fund will give priority to projects developed at the initiative of the 101 Genomes Foundation.

With a view to seeking the most effective impact, the Fund may combine its action with that of other funds created within the Foundation which share the same objective towards rare diseases”(Free translation of Annex 4)⁷³.

The reference to the possibility of joining the “Fonds 101 Genomes” action with that of other funds created within the KBF is particularly interesting because one of the current privileged fields of action of the KBF is that of health.

The KBF currently groups 68 funds supporting scientific medical research⁷⁴.

These funds include numerous funds dedicated to research on specific rare diseases⁷⁵ and, in particular, the “*Rare Diseases and Orphan Drugs Fund*” specifically dedicated to rare diseases in general, which “*brings together various partners involved in health care with the aim of stimulating knowledge on rare diseases, as well as the development and availability of orphan drugs*”⁷⁶.

The pilot project of the F101G could be of interest to some of those responsible for these various Funds and some synergies could be envisaged⁷⁷.

It is worth noting that Professor Anne De Paepe (Rector of the University of Ghent, geneticist, specialist in connective tissue diseases) sits on the Advisory Board of the KBF.

⁷³ Annex 4 : Convention, Article 2 - Objet.

⁷⁴ <https://www.kbs-frb.be/>

⁷⁵ Such as for example ; the “[Fonds Cremers-Opdebeeck](#)” (Support for medical research against Duchenne-Becker muscular dystrophy); the “[Fonds Eye Hope](#)” (Support for research into ophthalmological problems in people with Wolfram syndrome) ; the “[Fonds Marie-José Martin](#)” (supporting cardiac surgery through the “Fonds Live” of “*Bruxelles-Animation*”, the Smiles association at “*CHU Saint-Pierre*” in Brussels and research on orphan diseases); the “[Fonds Walter Pyleman](#)” (Three-year financial support for scientific research on Steinert’s disease and related rare diseases of the neuromuscular or musculoskeletal system. Research into the causes, clinical symptoms or treatment of disease) and the “[Fonds Candle in the Dark](#)” (Support for research on genetic mutations that cause Leber’s congenital amaurosis, a rare genetic disorder).

⁷⁶ For a longer description of the Fund: <http://www.bonnescauses.be/fund/fund?id=J5811350>.

⁷⁷ Mrs. Annemie T’Seyen, from the scientific secretariat of the King Baudouin Foundation, alluded to the possibility of organising a round table with the heads of these various Funds during which the F101G and the P101GM model could be presented. This round table could eventually allow other projects to emerge by joining the platform.

32. **Administration.** With regard to the operation of the “Fonds 101 Genomes”, Articles 6 and 7 of the Agreement with the F101G refer essentially to the “General Operating Rules for Funds managed by the KBF”⁷⁸.

When we read these two documents together, we see:

1. that the KBF undertakes to contribute all its experience and expertise to the “Fonds 101 Genomes” and to ensure the day-to-day management of the Fund⁷⁹;
2. that the KBF is responsible for collecting the payments, providing a receipt to donors and sending them a certificate of tax exemption for donations⁸⁰;
3. that the accounts of the “Fonds 101 Genomes” are kept by the accounting department of the KBF. A financial report on the management of the Fund shall be submitted annually to the Fund’s Management Committee⁸¹;
4. that the KBF is responsible for the administrative management of the “Fonds 101 Genomes” and places at its disposal secretaries who will ensure its general coordination and day-to-day management⁸². The designated secretary for the Fund:
 - *“studies and proposes strategic developments appropriate for the chosen objective;*
 - *convenes the Management Committee;*
 - *draw up minutes for each meeting of the Committee, containing the decisions taken by the Committee;*
 - *implements the decisions provided that they comply with the King Baudouin Foundation’s statutes and the available resources;*
 - *co-ordinates communication with the public and the press”*⁸³.

The partnership with the KBF relieves the F101G of numerous administrative procedures, ensures rigorous accounting and allows it to focus its action on the P101GM.

33. **Fees.** Article 10 of the Agreement provides that in order to “cover its secretarial expenses and the general promotion costs of the “Fonds 101 Genomes”, the KBF is authorised to make an annual withdrawal of 5% of the donations made during the year”⁸⁴.

The percentage of costs retained by the KBF was taken into account in the preparation of the budget. It is stressed here that this percentage refers only to donations made in the course of the year and not to all the sums capitalized since the opening of the “Fonds 101 Genomes”.

⁷⁸ Annex 4: *Convention, Articles 6 et 7 – Rôle et Fonctionnement.*

⁷⁹ Annex 4: *Convention, Article 6 – Rôle de la Fondation à l’égard du Fonds.*

⁸⁰ Annex 4: *Règles générales de fonctionnement (...), p. 20.*

⁸¹ Annex 4: *Règles générales de fonctionnement (...), p. 17.*

⁸² Annex 4: *Règles générales de fonctionnement (...), p. 13.*

⁸³ Annex 4 : *Règles générales de fonctionnement (...), p.13.* Free Translation.

⁸⁴ Annex 4: *Convention, Article 10 – Frais de secrétariat et de gestion financière du Fonds.* Free Translation

34. **Bonnescauses.be.** Since the signature of the agreement, the “Fonds 101 Genomes” is included among the Funds of the KBF.

The Fund is listed on the KBF website (<https://www.kbs-frb.be>) and on <http://www.bonnescauses.be/>. This will give additional visibility to the P101GM and will enable those interested to contribute to the financing of the project in just a few clicks.

35. **Donorinfo.be.** On the recommendation of a partner, the F101G will also make use of the donorinfo.be⁸⁵ platform to contribute to the financing by the “Fonds 101 Genomes” of the P101GM.

⁸⁵ <http://donorinfo.be/en>: “Donorinfo is a public interest foundation created in 2005 on the initiative of Erik van Baren with the following missions: * to promote the financial transparency of Belgian philanthropic organisations entitled to collect private donations (recipients) * to encourage the generosity of donors (private donors, companies, public services) informed on the basis of objective and controlled data. All the philanthropic organizations listed by Donorinfo are dedicated to helping people in need and/or supporting ecological and social transition. Donorinfo is an independent, self-funded centre of expertise. The donorinfo.be platform is a research tool for donors who seek information and philanthropic organisations that offer a guarantee of control and transparency. Thanks to donorinfo.be, we strengthen the links between vigilant donors and more than 240 non-profit organisations, NGOs and transparent foundations that act in a sustainable manner for the development of a more just, humane and supportive society” (Free Translation).

2.6 CONSENT

a) *Protection of personal data*

36. **Personal data.** In order to comply with the Belgian legislation in force since 1992 on personal data protection and with the “General Data Protection Regulation” (hereafter GDPR) SINCE 25 May 2018, participants in the project will be asked to sign a consent form. The draft of this document, which will be submitted to an ethics committee, is attached in Annex 3.
37. **GDPR.** The Consent Project in Annex 3 is currently under review to ensure compliance with the RGDP.
38. **Controller.** A Processing Activities Controller will be designated.
39. **Data Protection Officer.** A “*Data Protection Officer*” (hereafter DPO) will be appointed. If appointed internally, he will have to devote at least 20% of his working time within the F101G to data protection.
40. **Privacy Commission.** Before being effectively launched, the P101GM will be submitted to the new version of the former Privacy Commission for approval⁸⁶.

b) *Non-medicinal interventional clinical study*

41. **Clinical Study.** Assuming that the P101GM could be qualified as a non-medicinal interventional clinical study, the consent project in Annex 3 is designed to meet the requirements of the Belgian law of 7 May 2004 on experiments on the human being.

The obligations arising from this Law include, inter alia, the following obligations:

- to take out a no-fault insurance (cost of which could be added to the draft budget, if necessary) and;
- to submit to an ethics committee for an opinion the scientific protocol defined to set up the tool .

⁸⁶ <https://www.privacycommission.be/en/sc-h/competences-and-organisation>

2.7 ETHICS COMMITTEE AND SCIENTIFIC COMMITTEE

42. **Ethics Committee.** The F101G plans to submit its action to the ethical control of an already existing Ethics Committee to oversee its action within the framework of the P101GM.
43. **Scientific Committee.** A specific scientific committee will oversee the work of the F101G as part of the P101GM.

This committee will immediately be asked to set up a protocol for:

- 1) the identification of the main phenotypic data to be retained;
- 2) the identification of the parameters of the genomic/phenotypic cross-data platform;
- 3) the determination together with the academic partners of the criteria that will enable the 101 best candidates to be identified for sequencing, the genome of which would be sequenced as a priority

A specific “*Code of conduct*” lay down the most flexible working methods for this Committee: video-conference meetings, exchanges of agreement by e-mail, alternating chairmanship, annual report, remuneration, etc.

44. **Ideal composition of the Scientific Committee.** Ideally, the scientific committee of the project would be formed by:
 - Professor Anne De Paepe⁸⁷, Rector of Ghent University and specialist in Marfan syndrome;
 - Professor Julie De Backer⁸⁸ (Julie.DeBacker@UGent.be), Chair of HTAD WG of the VASCERN. [Center for Medical Genetics Ghent](#), [Ghent University Hospital](#), Ghent, Belgium;
 - Professor Bart Loeys⁸⁹ (bart.loeys@uantwerpen.be), member of HTAD WG of the VASCERN. [Center of Medical Genetics](#), [University Hospital of Antwerp](#), University of Antwerp, Belgium;
 - Professor Thierry Sluysmans (thierry.sluysmans@uclouvain.be). Responsible for Marfan children’s consultations at UCL, Université Catholique de Louvain, Brussels, Belgium;
 - Dr Guillaume Smits⁹⁰ (guillaume.smits@huderf.be), geneticist of the HUDERF, the Queen Fabiola Children’s University Hospital, Brussels, Belgium. Bioinformatics Specialist⁹¹.

⁸⁷ <https://www.ncbi.nlm.nih.gov/pubmed/?term=De+paepe+A>.

⁸⁸ <https://www.ncbi.nlm.nih.gov/pubmed/?term=De+Backer+J>.

⁸⁹ <https://www.ncbi.nlm.nih.gov/pubmed/?term=Loeys+B>.

⁹⁰ <https://www.ncbi.nlm.nih.gov/pubmed/?term=Smits+G>.

⁹¹ <http://ibsquare.be/>.

Ideally, the Scientific Committee should include a European representative of the “HTAD WG” working group in the VASCERN,

- Professor Guillaume Jondeau⁹² (guillaume.jondeau@aphp.fr), co-Chair of HTAD WG of the VASCERN. Cardiology Department [AP-HP](#), [Hôpital Bichat-Claude Bernard](#), Paris, France.

Outside Belgium but still within the European Union, the project should involve:

- Professor Catherine Boileau⁹³ (catherine.boileau@aphp.fr), Head of the Department of Genetics, AP-HP, Bichat-Claude Bernard Hospital, Paris, France.

Outside Belgium and the European Union, other experts could be involved if the P101GM were to take on a global scale, such as:

- Professor Hal Dietz⁹⁴, internationally recognized American specialist in Marfan syndrome;
- Professor Daniel MacArthur⁹⁵, creator of the ExAC and GnomAD databases.

45. **First reactions.** The reactions of the scientists to whom the P101GM was submitted were positive. Professors De Backer, Jondeau, Boileau, Smits and Loeys showed interest in joining the Scientific Committee.

Professors De Backer and Boileau immediately made constructive observations which enabled the project to evolve⁹⁶.

Professor Loeys and Dr. Aline Verstraeten welcomed us to their genetics center in Antwerp to discuss the scientific, budgetary and geographical aspects of the project. During this discussion, the options for partnerships with the BGI and the University of Nijmegen were discussed.

Bank Delen asked us to invite the main Belgian researchers involved in our project to a philanthropic evening organized on Thursday 23rd November with the King Baudouin Foundation (KBF) in the presence of a member of the Belgian Royal Family. Professors Anne De Paepe, Julie De Backer, Bart Loeys and Dr. Aline Verstraeten and Dr. Guillaume Smits accepted our invitation to attend the event.

⁹² <https://www.ncbi.nlm.nih.gov/pubmed/?term=Jondeau+G>.

⁹³ <https://www.ncbi.nlm.nih.gov/pubmed/?term=Boileau+C>.

⁹⁴ <https://www.ncbi.nlm.nih.gov/pubmed/?term=Dietz+H>.

⁹⁵ <https://www.ncbi.nlm.nih.gov/pubmed/?term=Macarthur+D>.

⁹⁶ See Annex 11: Correspondence with Professors Jondeau, De Backer and Boileau.



From left to right: Mrs. Ludivine Verboogen, Professor Anne De Paepe, Dr. Michel Verboogen, Mr. Romain Alderweireldt, Mr. René Havaux, Mrs. Cécile Jacquet, Professor Bart Loeys, Dr. Aline Verstraeten, Professor Julie De Backer and Dr. Guillaume Smits

The F101G was put on the spotlight during this evening and Professor Anne De Paepe had the opportunity to talk at length with Prince Lorenz of Belgium about Marfan syndrome and related diseases. Professor Anne De Paepe confirmed her willingness to get personally involved in the F101G and the P101GM and explained that this is what needs to be put in place to advance research. To formalise its involvement, it is planned to meet at the KBF with Professor Anne De Paepe before the end of the year to set up the management committee (according to Article 5 of the Agreement of the 17th of November 2017).

That same evening Professors Julie De Backer, Bart Loeys, Aline Verstraeten and Guillaume Smits began discussing the scientific dimension of the P101GM and the different issues to be addressed. They agreed to meet on Friday 19 January 2018 from 9am to 1pm for the first proper scientific meeting of the P101GM, which will take place in Brussels at the bank's premises.

It is proposed to discuss at this meeting the main issues relating to:

- criteria for collecting phenotypic data⁹⁷;
- criteria for composition of the cohort whose genomes will be sequenced as a priority⁹⁸;

⁹⁷ The “Supplemental Table A” that accompanies Franken’s article (FRANKEN R., “*Relationship between fibrillin-1 genotype and severity of cardiovascular involvement in Marfan syndrome*”, *Heart Online First*, published on May 3, 2017 as 10.1136/heartjnl-2016-310631), in which reference is made to age, gender, aortic diameter, ectopia lentis, complications and follow-up of participants, could serve as a basis for discussion on the essential phenotypic data to be collected.

⁹⁸ This implies choosing between the four approaches considered at this stage:

- 1) *The “family” approach*, which consists in giving priority to families with a broad spectrum of symptoms;
- 2) *The “sibling” approach*, which consists in prioritizing children from the same parents who have different symptoms;
- 3) *The “extremes” approach*, which consists in retaining patients with severe symptoms and at the other end of the spectrum of patients with mild symptoms;
- 4) *The “recurrence” approach*, which studies the variability of symptoms among patients affected by recurrent FBN1 pathogenic mutations. With regard to this approach, it appears from UMD-FBN1 (and other databases) that mutations c.7754T>C (exon 62) and c.8176C>T (exon 64) are relatively frequent and would affect about 50 individuals from different families in different countries.

- the method of validation of pathogenic mutations⁹⁹.

Professors Jondeau and Boileau were invited to join them on this occasion. Professor Boileau confirmed that she will be present in Brussels on the 19th of January to represent French researchers. Professor Bart Loeys offered to prepare a draft document that could serve as a basis for discussion on questions of cohort composition and determination of the phenotypic criteria to be addressed during this first scientific meeting.

2.8 PATIENTS ASSOCIATIONS

46. **Patients associations.** Support from patients' associations is essential for the implementation of the project.

At this stage, the Belgian Association of Marfan Syndrome ("*Association Belge du Syndrome de Marfan*" hereafter ABSM: <http://www.marfan.be/>) appeared receptive and its President expressed her readiness to support the project. An article published in the association's magazine introducing the project received a positive response¹⁰⁰.

The President of the French Marfan Syndrome Association, Mr. Patrice Touboulie, appeared particularly enthusiastic about the F101G and he is evaluating the best way to collaborate with it.

If the P101GM is to be implemented as we wish, we will solicit the support of other national patients associations, in particular through the European network "Marfan Europe Network" (<http://www.marfan.eu/>) and the American "Marfan Foundation" (<https://www.marfan.org/>).

⁹⁹ This implies, according to our current understanding, of at least considering the establishment of a verification mechanism that could query all relevant public and private databases for the identification of pathogenic genetic mutations (see footnote 60).

¹⁰⁰ See Annex 8.

3. FINAL REMARK

47. **Listen to nature.** The P101GM is born from the reading of an interview with Professor Hal Dietz, who mentioned that the intersection of genomic and phenotypic data could help us understand “*how natural genetic variants can protect some people from the consequences of a fibrillin-1 mutation*” and on this basis possibly be able to “*identify drugs that can mimic nature’s successful strategy*”¹⁰¹.

In the meantime, Professor Hal Dietz gave a presentation to the American Marfan Foundation on 6th August 2017 entitled: “*Listening to Nature’s Cues*”¹⁰².

During this presentation, Professor Dietz reported on research conducted within 5 “exceptional” Marfan families that, although presenting a FBN1 mutation previously confirmed deleterious, appeared to be protected by the action of a modifier gene (MAP3K4)¹⁰³. Tests conducted on mice seem to confirm the interest in pursuing this track¹⁰⁴ and led Professor Dietz and his team to identify 6 potential modifier genes in mice that could play a protective role¹⁰⁵.

For the time being Research is at this stage.

Had the P101GM already succeeded in making the phenotypic / genomic cross-platform available, Professor Dietz and his team could have immediate access and verify within a few hours whether their observations with respect to mice could also be observed in humans.

This would truly pave the way for the development of drugs that could change the lives of Marfan children.

¹⁰¹ WEISMAN R., “*Meet Your Gene: An Introduction to the Marfan Gene and Current Research*”, 10 january 2017. Available at: <http://blog.marfan.org/meet-your-gene-an-introduction-to-the-marfan-gene-and-current-research>.

¹⁰² DIETZ H., “*Listening to Nature’s Cues*”, 6 august 2017. Available in Annexe 11 and at: <http://www.marfan.org/sites/default/files/Research%20Update%202017%20Annual%20Conference.pdf>.

¹⁰³ DIETZ H., *op. cit.*, page 13.

¹⁰⁴ DIETZ H., *op. cit.*, pages 15 et 16 (which, by the way, influences p38 protein).

¹⁰⁵ DIETZ H., *op. cit.*, page 23.

LIST OF ANNEXES

- Annex 1: Statutes of the private foundation “*Fondation 101 Genomes*” (10/11/2017)
- Annex 2: Draft of budget
- Annex 3: Draft Information and Consent Document to be provided to participants in the “101 Genomes Marfan” Project
- Annex 4: Agreement of the 17th November 2017 between the King Baudouin Foundation and the “Foundation 101 Genomes” for the opening of a “*Fonds pour les 101 Genomes*” followed by the « *Règles générales de fonctionnement pour les Fonds gérés par la Fondation Roi Baudouin* ».
- Annex 5: RIORDAN J.D., NADEAU J. H., “*From Peas to Disease: Modifier Genes, Network Resilience and the Genetics of Health*” in *The American Journal of Human Genetics*, 101, 177–191, 3rd August 2017, <http://dx.doi.org/10.1016/j.ajhg.2017.06.004>.
- Annex 6: Bluebee Genomics Platform - Data Security & Compliance – White Paper.
- Annex 7: Document describing the security measures of Medisapiens: “*General security measures and implementations*”.
- Annex 8: Article published in “*Le Cœur & la Main*”, the ABSM (“*Association Belge du Syndrome de Marfan*”) magazine: “*Actualités du syndrome de Marfan : Pistes pour de nouvelles thérapies, Registre marfan et Fonds européens (Inhibition des protéines p38 et nos2, Identification des variants génétiques naturels)*”.
- Annex 9: The Variant Call Format (VCF) Version 4.2 Specification, 24th May 2017
- Annex 10: IRDiRC, *The European Perspective on Data Sharing –omics Data for Personalised Medicine in Rare Disease*, third IRDiRC conference on 8th February 2017, Jussieu, Paris.
- Annex 11: Correspondence with Professors Jondeau, De Backer and Boileau.
- Annex 12: DIETZ H., “*Listening to Nature’s Cues*”, 6th August 2017.
- Annexe 13 : Organigram of the « Fondation 101 Génomes ».