Autosomal recessive insulin-dependent diabetes

Coralie Leblicq, M.D.

Department of Human Genetics,
McGill University, Montreal, Quebec, Canada
August 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Sciences

Acknowledgements-Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, le Dr Constantin Polychronakos, pour m'avoir donné l'opportunité de découvrir le monde de la recherche fondamentale. Malgré le peu d'expérience que j'avais dans ce domaine, le Dr Polychronakos m'a ouvert les portes de son laboratoire et m'a soutenue dans ma démarche. Les connaissances exceptionnelles du Dr Polychronakos en génétique, en immunologie, en physiologie, en bioinformatique, en statistiques comme en clinique ont été une réelle source d'inspiration. Dr Polychronakos m'a poussée à dépasser mes connaissances cliniques et à constamment remettre en question mes acquis. Je remercie également chaleureusement le Dr Polychronakos pour son soutien et son humanité au cours de ces deux années.

Mes remerciements vont également au Département de Génétique Humaine de l'Université McGill. Je remercie le département pour m'avoir acceptée dans le programme de maîtrise et les professeurs pour leur enthousiasme à transmettre leur savoir et leur amour des sciences. Un merci tout particulier à Ross MacKay pour sa grande disponibilité et son aide en toutes circonstances.

Je souhaiterais remercier le Dr Quan Li pour son travail d'analyse bioinformatique. Un grand merci également à Tedi Quendro qui a débuté ce travail avec beaucoup d'enthousiasme et au Dr Tuğba Demirci qui a rejoint le laboratoire pendant 6 mois, reprenant le flambeau durant mon congé de maternité.

Un merci particulier à mes collègues Maha Al-Riyami, Meihang Li et le Dr Xiaoyu Du pour leur bonne humeur quotidienne, leur soutien et leurs conseils.

Je tiens également à remercier les Drs Rima Rozen et Laurent Legault, membres de mon comité de suivi de thèse, pour leurs questions pertinentes, leurs conseils et leur soutien. Ils m'ont offert une vision nouvelle et critique de mon projet me permettant de progresser et d'élargir mes connaissances.

Ce travail n'aurait pu être réalisé sans la confiance et le financement du Canadian Pediatric Endocrine Group (CPEG) et de la Fondation Richard et Edith Strauss. Un énorme merci pour avoir cru en notre projet et m'avoir fait confiance.

Un tout grand merci à mes amis et ma famille pour leur soutien et leurs encouragements dans mes choix professionnels. Un merci spécial à Claudia, une amie très chère mais également un grand médecin et une scientifique brillante. Merci pour ta présence, ton soutien, tes pensées positives et nos longues discussions. Merci à mes parents et ma sœur, qui malgré la distance, ont toujours été très présents dans ma vie et ont accepté et soutenu mes choix professionnels. Merci à mes deux

merveilleux enfants, Kenan et Adèle, pour leurs sourires, leurs rires, leur joie de vivre, leurs câlins réconfortants. Vous êtes une source de bonheur quotidienne et une inspiration constante dans ma vie!

Finalement, un merci tout particulier à mon mari adoré, Uğur, sans qui cette aventure n'aurait pu se réaliser. Merci d'être toujours auprès de moi, merci d'être le pilier de notre famille, merci d'être un modèle pour moi et nos enfants, merci de m'avoir soutenue dans ce projet et de me soutenir encore dans mes projets futurs, merci simplement d'être toi et de faire partie de ma vie.

L'enthousiasme de tous ceux qui ont travaillé sur ce projet m'a permis d'y croire et de continuer d'espérer que notre travail mènera à la découverte d'une nouvelle forme de diabète. Saint-Exupéry disait, « Ce qui embellit le désert, c'est qu'il cache un puits quelque part… ». La motivation des chercheurs est leur amour des sciences, leur volonté de comprendre et percer les mystères du corps humain. Face à l'immensité de l'inconnu que représente encore le monde des sciences, ils gardent toujours l'espoir de trouver un jour leur puits.

Pour ma part, je dis « au revoir » à la recherche fondamentale car je m'en retourne à mon premier amour, la pédiatrie. Comme le dirait la chanson populaire, « ce n'est qu'un au revoir… oui nous nous reverrons… ». La recherche restera toujours présente dans ma vie professionnelle car elle a définitivement changé ma vision de la médecine.

Preface and Contribution of Authors

Dr Constantin Polychronakos supervised the project and the writing of the thesis.

Dr Coralie Leblicq evaluated the WES results, did all PCRs (except for Tedi Quendro's and Dr Tuğba Demirci's parts; see below), interpreted all sequencing results, made a review of the literature, and wrote the thesis.

Dr Quan Li performed the bioinformatics analysis (linkage analysis, homozygosity-by-descent analysis).

Tedi Quendro genotyped all families for the SNVs found within the first 8 candidate genes (Paragraph 3.1.2.1).

Dr Tuğba Demirci genotyped some SNVs within new candidate genes and sequenced some exons in families linked to the *SCAMP2* gene (Paragraphs 3.2.2.1.2 and 3.2.2.2).

Research Ethics

We received ethics committee approval for our project in January 2015, which was renewed in January 2016.

Conflict of interest

None of the authors has conflict of interest to disclose.

Abbreviations

- (i)DEND: (intermediate) Developmental delay, Epilepsy and Neonatal Diabetes
- ABCC8: ATP-binding cassette, subfamily C, member 8
- ACIN1: apoptotic chromatin condensation inducer 1
- ACTH: Adrenocorticotropic Hormone
- Ala (A): Alanine
- ANKRD11: ankyrin repeat domain 11
- Arg (R) : Arginine
- Asn (N): Asparagine
- Asp (D): Aspartic acid
- ASXL3: additional sex combs like 3 (Drosophila)
- ATP: adenosine triphosphate
- Blast: Basic Local Alignment Search Tool
- BLK: Tyrosine kinase, B-lymphocyte specific
- bp: bases pairs
- C18orf65: chromosome 18 open reading frame 65
- Ca^{2+}: calcium
- CACNB3: calcium channel, voltage-dependent, beta 3 subunit
- CDA: Canadian Diabetes Association
- CEL: Carboxyl Ester Lipase
- CF: cystic fibrosis
- CFRD: cystic fibrosis-related diabetes
- CFTR: Cystic Fibrosis transmembrane conductance regulator
- CHPF: chondroitin polymerizing factor
- CLASRP: CLK4 (CDC like kinase 4) associating serine/arginine rich protein
- cM: centimorgan
- CNS: central nervous system
- CP: chronic pancreatitis
- CTLA4 : cytotoxic T-lymphocyte associated protein 4
- Cys (C): Cysteine
- ddNTP: dideoxynucleotide
- DM: diabetes mellitus
- DNA: desoxyribonucleic acid
- dNTP: deoxynucleotide
- EIFA2K3: Eukaryotic Translation Initiation Factor 2-Alpha Kinase 3
- FOXP3: Forkhead Box P3
- FRY: furry homolog (Drosophila)
- G-6-P: Glucose-6-phosphate

- GATA4: GATA binding protein 4
- GC: glucocorticoid
- GCK : glucokinase
- GDM: gestational diabetes mellitus
- GLIS3: GLIS family Zinc Finger 3
- Gln (Q): Glutamine
- Glu (E): Glutamic acid
- Gly (G): Glycine
- GWAS: Genome Wide Association Study
- HBD: homozygosity-by-descent
- HGMD: Human Gene Mutation Database
- HH: Hereditary Hemochromatosis
- HH: hyperinsulinemic hypoglycemia
- His (H): Histidine
- HLA : Human Leukocyte Antigen
- HMGXB3: HMG-box containing 3
- HNF-1α: Hepatocyte Nuclear Factor 1 alpha
- HNF-1β: Hepatocyte Nuclear Factor 1 beta
- HNF-4α: Hepatocyte Nuclear Factor 4 alpha
- HSPBAP1: heat-shock protein-associated protein 1
- IBD: identical by descent
- IFIH1 : interferon induced with helicase C domain 1
- IL2RA : interleukin 2 receptor subunit alpha
- Ile (I): Isoleucine
- INS : insulin
- INSR: insulin receptor
- IPF-1: Insulin promoter factor 1
- K^+: potassium
- KATP channels: ATP- sensitive potassium channels
- KCNJ11: potassium inwardly-rectifying channel, subfamily J, member 11
- KIAA2018: KIAA2018 ortholog
- KLF 11: Fruppel-like factor 11
- Leu (L): Leucine
- LOD: logarithm of an odds ratio
- Lys (K): Lysine
- MD: monogenic diabtes
- Met (M): Methionine
- MIDD: maternally inherited diabetes and deafness

- MODY: Maturity Onset Diabetes of the Young
- MT-T: mitochondrially encoded tRNA threonine
- MT-TK: mitochondrially encoded tRNA lysine
- MT-TL1: mitochondrial tRNALeu
- NADH: Nicotinamide Adenine Dinucleotide
- NCKAP5L: NCK (non-catalytic region of tyrosine kinase)- Associated protein 5-like
- NEUROD1: Neurogenic differentiation 1
- NGS: next generation sequencing
- NIDDK: the National Institute of Diabetes and Digestive and Kidney Diseases
- NPHP4: nephronophthisis 4
- Nucleotides: A: adenine; T: thymine; G: guanine; C: cytosine
- PARP4: poly(ADP-ribose) polymerase family member 4
- PAX4 : Paired box gene 4
- PCBD1: pterin-4 alpha-carbinolamine dehydratase 1
- PCOS: polycystic ovary syndrome
- PCR: polymerase chain reaction
- Phe (F): Phenylalanine
- PKP2: plakophilin 2
- PNDM: permanent neonatal diabetes mellitus
- Polyphen: polymorphism phenotyping
- Pro (P): Proline
- PRR12: proline rich 12
- PTF1A: pancreas specific transcription factor, 1a
- PTPN22 : protein tyrosine phosphatase, non-receptor type 22
- RBM23: RNA binding motif protein 23
- RfX6: regulatory factor X6
- RPKM: reads per kilobase per million mapped reads
- rsID: reference SNP cluster identity
- SCAMP2: secretory carrier membrane protein 2
- Ser (S): Serine
- SIFT: sorting intolerant from tolerant
- SNP: single nucleotide polymorphism
- SNV: single nucleotide variant
- T1ADM: type 1A diabetes mellitus
- T1BDM: type 1B diabetes mellitus
- T1DGC: Type 1 Diabetes Consortium Group
- T1DM: type 1 diabetes mellitus
- T2DM: type 2 diabetes mellitus

- T_a: annealing temperature
- Thr (T): Threonine
- TKI: tyrosine kinase inhibitor
- TNDM: transient neonatal diabetes mellitus
- TRAPPC9: trafficking protein particle complex 9
- Trp (W): Tryptophan
- Tyr (Y): Tyrosine
- UTP20: UTP20, small subunit processome component
- Val (V): Valine
- WES: whole exome sequencing
- WFS: Wolfram Syndrome
- WFS1: wolframin
- WHO: World Health Organization

Table of contents

Abstract

Background and hypothesis: Type 1 diabetes (T1DM) is due to the autoimmune destruction of the insulin-producing pancreatic beta-cells. T1DM is a multifactorial disease with a strong genetic (polygenic) component. However, not all cases of insulin-dependent diabetes starting early in life are of autoimmune and polygenic etiology. Indeed, neonatal diabetes and MODYs (Maturity-Onset Diabetes of the Young) are both monogenic diseases that can still be misdiagnosed as T1DM. Our hypothesis is that a new form of monogenic, autosomal recessive, non-autoimmune, insulin-deficient, young-onset diabetes exists but has escaped identification because of lack of strongly suggestive family or clinical history.

Objective: To discover genes whose mutations are responsible for this type of diabetes by exome sequencing of selected multiplex families diagnosed as T1DM.

Materials and Methods: Data from 2,345 sibling pairs affected with T1DM, from the T1DM Genetics Consortium (T1DGC) were analyzed for the presence of autoimmune markers (3 available autoantibody tests (GAD65, ZNT8 and IA2)). From 167 families where both affected siblings were autoantibody-negative, we found 37 families in which both were also negative for any of the known predisposing HLA haplotypes (e.g. HLA-DRB1*04/DQB1*0302 or HLA-DRB1*0301/DQB1*0201). Although some of these 37 families may still have autoimmune T1DM, we posited that most could have diabetes for a reason other than autoimmunity.
DNA from all members of these 37 families was obtained from the NIDDK repository and DNA from 14 affected children was sent to McGill/Génome Québec Innovation Centre for whole-exome sequencing (WES).

Results: Before filtering WES results to select new candidate genes, we identified 3 families with mutations in genes already known to be involved in monogenic diabetes: *KCNJ11* and *WFS1.* Thereafter, we selected 20 new candidate genes according to 3 criteria: 2 or more protein-altering variants, positive LOD (Logarithm of Odds) score in the nuclear family, and significant expression of the gene in the pancreatic islets (RPKM (reads per kb per million) higher than 3).
We genotyped probands and their families for SNVs found in 16 of those candidate genes and we found 3 genes carrying variants that segregated with the disease under a recessive model (*NCKAP5L, HSPBAP1* and *SCAMP2*). We sequenced all families linked to those 3 loci but we did not find any other family with mutations inherited on a recessive Mendelian fashion within one of those 3 genes.
Based on our first findings (*KCNJ11* and *WFS1* diabetes misclassified as T1DM), we posited that some patients with MODY3, the most frequent type of MODY, must also be misdiagnosed as T1DM. Among all samples listed in the NIDKK repository, we selected families where one parent

was diabetic and where the affected children and the affected parent were negative for autoimmune markers. Nine families fulfilled these criteria. All protein-coding exons of *HNF-1α* were sequenced in 2 families that showed linkage at the *HNF-1α* locus and whose DNA was available. We found c.599 G>A missense mutation (p.R200Q) within exon 3 in the affected mother and in the 3 affected children of the first family, confirming MODY3 diagnosis.

Conclusions: Different cases of monogenic diabetes are still misdiagnosed as T1DM. Usually, clinical and/or family history should help us making the right diagnosis. Some monogenic diabetes (e.g. *KCNJ11*) represent life-changing diagnosis as they can be treated with sulfonylureas for perfect glycemic control without risk of hypoglycemia. In addition to missed diagnosis, some mutations appear to alter to a lesser extent the protein function which can influence clinical evolution of the disease and explain misdiagnosis.

In the same way, we have identified three highly likely candidate genes for a new form of autosomal recessive form of diabetes that remain to be confirmed in at least 1 additional unrelated proband.

Résumé

Introduction et hypothèse: Le diabète de type 1 (DT1) est une maladie causée par la destruction auto-immune des cellules β du pancréas, productrices d'insuline. L'origine du DT1 est multifactorielle avec une composante génétique (polygénique) importante. Tous les cas de diabète insulino-dépendant débutant précocement ne sont cependant pas d'origine polygénique. En effet, les diabètes néonataux et les MODYs sont des maladies monogéniques qui sont encore sous-diagnostiquées ou mal catégorisées. Notre hypothèse est qu'une nouvelle forme de diabète monogénique, autosomique récessif, non auto-immun, insulino-dépendant, à début précoce existe et serait non diagnostiquée en raison de l'absence d'histoire familiale ou de tableau clinique spécifiques.

Objectif: Découvrir un gène dont la mutation serait responsable de ce type de diabète, par séquençage d'exons de familles multiplexes diagnostiquées avec un DT1.

Méthodologie: 2,345 paires de frères et sœurs atteints de DT1 et regroupées au sein du T1DM Genetics Consortium (T1DGC) furent testées pour les 3 auto-anticorps disponibles (GAD65, IA2 et ZnT8)). Des 167 familles où les 2 enfants diabétiques étaient négatifs pour les auto-anticorps, 37 étaient également négatifs pour les haplotypes prédisposant du HLA (HLA-DRB1*04/DQB1*0302 or HLA-DRB1*0301/DQB1*0201). Bien que certaines de ces 37 familles doivent malgré tout souffrir d'un diabète auto-immun, nous avons émis l'hypothèse que la majorité devait être atteinte d'une autre forme de diabète. L'ADN des membres des 37 familles a été obtenu du NIDDK et 14 échantillons ont été envoyés à McGill/Génome Québec Innovation Centre pour un séquençage complet d'exons

Résultats: Nous avons identifié 3 familles avec des mutations dans les gènes *KCNJ11* et *WFS1*, impliqués dans des diabètes monogéniques connus. Nous avons ensuite sélectionné 20 gènes candidats sur 3 critères : la présence d'au moins deux variants modifiant la séquence de la protéine, un LOD (Logarithm of Odds) score positif dans la famille du proband et l'expression significative du gène dans le pancréas (RPKM (reads per kb per million) supérieur à 3). Nous avons génotypé les probands et leurs familles pour les variants trouvés dans 16 de ces 20 gènes candidats et nous avons trouvé 3 gènes d'intérêt où les variants étaient transmis selon un mode autosomique récessif (*NCKAP5L, HSPBAP1* et *SCAMP2*). Nous avons séquencé toutes les familles qui avaient un LOD score positif pour ces loci mais n'avons trouvé aucune autre famille qui portait une variation transmise selon un mode autosomique récessif.

Étant donnée la découverte de familles porteuses de diabètes monogéniques connus au sein de notre cohorte initiale, nous avons émis l'hypothèse que certains patients étiquetés de DT1 pouvaient être atteints de MODY3. Parmi les échantillons du T1DGC, nous avons sélectionné les familles où un parent était diabétique et où les enfants et le parent atteints étaient négatifs

pour tout marqueur d'auto-immunité. Neuf familles remplirent ces critères. Nous avons séquencé tous les exons codants d'*HNF-1α* dans 2 d'entre elles qui présentaient un LOD score positif pour le locus de ce gène et pour lesquelles nous disposions d'ADN. Nous avons trouvé la mutation faux-sens c.599 G>A (p.R200Q) chez la mère et les 3 enfants diabétiques de la première famille, confirmant le diagnostic de MODY3.

Conclusions: Différents cas de diabètes monogéniques sont encore mal diagnostiqués et étiquetés de DT1. Le tableau clinique et/ou l'histoire familiale devraient aider le clinicien à poser un diagnostic adéquat permettant dans certains cas de changer le schéma thérapeutique pour une médication orale, plus confortable avec un meilleur contrôle de la glycémie.

Nous avons également identifié 3 gènes candidats très prometteurs pour une nouvelle forme autosomique récessive de diabète qui doivent cependant encore être validés chez un proband non lié aux familles séquencées.

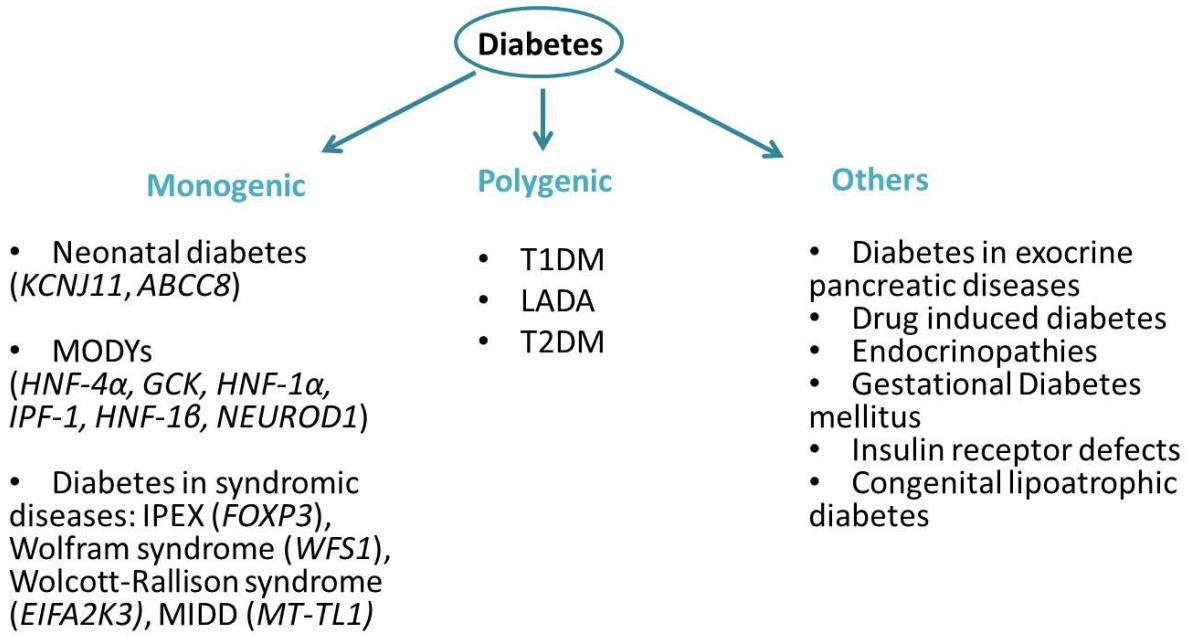
List of Figures

List of Tables

1. Introduction- Literature review

Diabetes mellitus (DM) is a chronic metabolic disease occurring when the pancreas does not produce enough insulin or when insulin is not properly used by cells. This results in an increased blood glucose level (hyperglycemia), responsible for several symptoms such as polyuria, polydipsia, and weight loss. Hyperglycemia is the main cause of the acute complications of diabetes (ketoacidosis, hyperosmolar coma) as well as debilitating chronic complications (nephropathy, retinopathy, peripheral neuropathy, and macro- and microvascular diseases). Diabetes has become a major health problem worldwide. Indeed, in 2014, the World Health Organization (www.who.int) estimated that 9% of adults worldwide had diabetes. Likewise, in 2015, more than 3.4 million Canadians (9.3% of Canadian population) were living with diabetes (www.diabetes.ca). The two main forms of diabetes are type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM) but many other forms of diabetes exist and are summarized in **Figure 1**.

Figure 1: Diabetes classification.
Diabetes can be classified within 3 large groups based on the origin of the disease (polygenic, monogenic or resulting from another condition). T1DM, T2DM and Latent Autoimmune Diabetes in Adults (LADA) are polygenic diseases. Neonatal diabetes, Maturity Onset Diabetes of the Young (MODYs) and some syndromic diseases result from mutations in a single gene. All the other forms of diabetes are the consequence of a previous condition.



1.1. Type 2 diabetes mellitus (T2DM)

T2DM is the most common form of diabetes in adults (85-95% of diabetes in patients >18 years of age) and becomes more prevalent in young, in relation with the explosion of obesity this last decade (Writing Group for the SEARCH for Diabetes in Youth Study Group *et al*., 2007) (IDF diabetes atlas, sixth edition, 2013) (D'Adamo E. and Caprio S., 2011).

T2DM is a complex trait due to both insulin resistance and subsequent failure of the insulin-producing β-cells of the pancreas. Genome wide association studies (GWASs) highlighted more

than 60 loci linked to T2DM even if only 10% of the genetic component of T2DM can be explained by those loci (Hara K *et al*., 2014). T2DM results from the interaction of genetic susceptibility and environment. Obesity and physical inactivity are the two main factors influencing development of T2DM in genetically predisposed patients (positive family history). Other risk factors for T2DM are ethnicity, female gender, and age. Progression from glucose intolerance to T2DM is faster in children underlining the importance of focusing the attention on prevention, early diagnosis and treatment (D'Adamo E. and Caprio S., 2011).

T2DM can be initially treated with life-style modifications (physical activity, diet) and oral drugs even if insulin injections are often required in the course of the disease.

1.2 Type 1 diabetes mellitus (T1DM)

T1DM is the most common form of diabetes in childhood, accounting for more than 70% of new diagnosis of diabetes in patients younger than 19 years of age in the United States (Writing Group for the SEARCH for Diabetes in Youth Study Group *et al*., 2007).

In 2013, about 500,000 children worldwide were living with T1DM and its incidence continues to increase in many countries with an annual augmentation of 3% (IDF diabetes atlas, sixth edition, 2013). Canada is one of the countries that show the highest incidence of T1DM with more than 24 new cases of T1DM (0-14 years) per 100,000 children per year (IDF diabetes atlas, sixth edition, 2013).

T1DM is an autoimmune disease characterized by progressive destruction of the insulin-producing β-cells by infiltration of the pancreatic islets with dendritic cells, macrophages and T-

lymphocytes (insulitis). Patients also develop humoral autoimmunity to islet antigens but it is widely accepted that anti-islet autoantibodies do not by themselves cause beta-cell destruction. Autoantibodies, that are therefore determined mainly for diagnostic purposes and disease prediction, are the following: islet cell autoantibodies (ICA), directed against unspecified islet cell surface and cytoplasmic antigens; anti-insulin antibodies (IAA); anti-glutamic acid decarboxylase (GAD) antibodies directed against the enzyme responsible for GABA synthesis in the β-cells and the neurons (the GAD65 isoform is specific of pancreatic cells); anti-protein tyrosine phosphatase-like antigen (IA-2) in the β-cells, and Zinc transporter protein (ZnT8) autoantibodies (Taplin C.E. and Barker J.M., 2008) (Fabris M. *et al.,* 2015)

T1DM is a multifactorial disease caused by environmental factors in genetically susceptible individuals (Bluestone JA, *et al.* 2010). The first loci linked to T1DM were highlighted by candidate-gene approach studies (HLA, INS, PTPN22, CTLA4, IL2RA, IFIH1). GWASs showed more than 40 additional susceptibility alleles associated with T1DM (Polychronakos C. and Li Q*.,* 2011). The genetic component is very strong, entailing a sibling relative risk of 15, one of the highest in common complex diseases. However, this genetic component is polygenic with a large number of loci, each with small contribution to disease risk. The one exception is the HLA class II *DR-DQ* locus, where haplotypes common in the general population (~20% of Europeans) are found in 80% of T1DM cases and closer to 90% in cases from multiplex families (more than one case per family, indicating heavier genetic load). The highest risk haplotypes are HLA DRB 301-DQB 201 and HLA DRB 401-DQA 301-DQB 302 (Polychronakos C. and Li Q., 2011). In genetically susceptible individuals, environmental factors appear to trigger the autoimmune process leading to destruction of the pancreatic β-cells. Many studies have tried to pinpoint

those environmental agents but proof is lacking for the implication of any specific one in T1DM development and/or progression. The first environmental factor involved in T1DM occurrence is the diet: cow's milk consumption during childhood (Lamb M.M. *et al.,* 2015), vitamin D status (The EURODIAB Substudy 2 Study Group, 1999), age at gluten introduction (Norris J.M. *et al.*, 2003), and omega-3 fatty acids intake (Skyler J.S., 2012). Secondly, viruses (especially enteroviruses) have been described as major actors in development of autoimmunity. The mechanism of action of viruses would be based on molecular mimicry (controversial hypothesis because this phenomenon is quite common in nature), direct pancreatic injury (strong pancreotropism of enteroviruses) or virus-induced exocrine pancreas inflammation with production of pro-inflammatory cytokines and activation of antigen-presenting cells (Bergamin C.S. and Dib S.A., 2015). A third hypothesis is the "Hygiene Hypothesis" suggesting that the improvement in sanitary conditions in developed countries leads to less exposure to infections and therefore to impaired immune response to environmental triggers that may, somehow, promote autoimmunity (Bergamin C.S. and Dib S.A., 2015). More recently, alterations in the gut microbiome have been suggested as risk factors for T1DM development (Alkanani A.K. *et al.* 2015).

Individuals with genetic predisposition in contact with some environmental triggers can develop autoimmunity (T-cell induced insulitis and autoantibodies) and progress to diabetes following a period as long as 10 to 15 years (70% to 85% of individuals with 2 or more autoantibodies, respectively) (Steck A. S. *et al*., 2015). By the time of the appearance of the symptoms, the β-cell mass has been almost completely destroyed.

Clinically, T1DM is characterized by early onset and immediate and lifelong requirement for exogenous insulin treatment. This diagnosis is made practically automatically in any pediatric-age case that presents with hyperglycemia and ketosis. The polygenic and autoimmune nature of the disease can be corroborated by testing for HLA typing and serum autoantibodies but this is rarely done as part of routine clinical care.

This autoimmune diabetes is sometimes referred to as type 1A, to distinguish it from non-autoimmune disease with similar clinical presentation (type 1B diabetes mellitus). T1BDM is characterized by absolute insulin deficiency but with no evidence of autoimmunity and no known cause for β-cell destruction. The underlying mechanism of T1BDM remains to be understood (Sang Y. *et al.,* 2014) (Moritani N. *et al*., 2013).

In T1DM, insulin deficiency is absolute and irreversible. The treatment is then life-saving and lifelong subcutaneous insulin injections. However, alternative treatments are effective in some cases that fall under the T1BDM concept.

1.3 Latent Autoimmune Diabetes in Adults (LADA)

LADA is slowly progressive autoimmune diabetes. Clinically, LADA reminds T2DM (late-onset, slow evolution with no insulin requirement within the first 6 months) but with autoantibodies specific of T1DM (mostly anti-GAD). The physiopathology of this disease is still unclear but markers point to autoimmunity. LADA would represent 2-10% of all diabetes (Yousefzadeh G. et al., 2016).

1.4 Monogenic diabetes (MD)

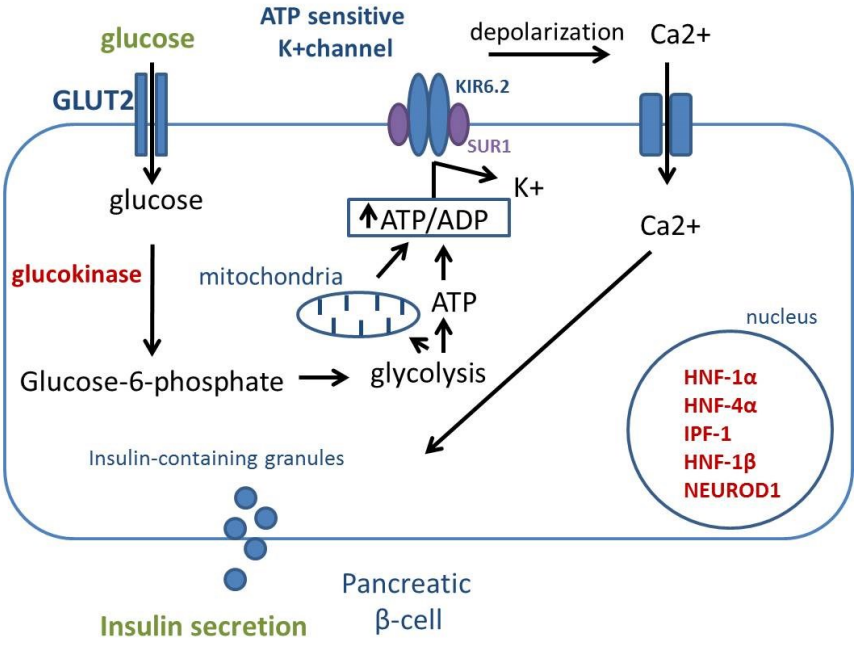
Not all cases of insulin-dependent diabetes starting early in life are of autoimmune and polygenic etiology (Nakhla M. and Polychronakos C., 2005). It is estimated that 1-5% of all cases of diabetes mellitus are caused by mutation in a single gene (Irgens H. U. *et al.,* 2013). The pancreatic β-cells are affected in all such cases.

1.4.1 Physiology of pancreatic β-cells

The β-cells of the pancreas produce and release insulin depending on blood glucose concentrations. They play a major role in maintaining glucose levels in normal ranges.

The metabolic cascade between blood glucose level and insulin secretion is described in **Figure 2**.

Figure 2: Glucose-sensing insulin secretion by pancreatic β-cell. Glucose moves into pancreatic β-cells by facilitated diffusion through glucose transporters (Glut-2). Glucose is phosphorylated to glucose-6-phosphate (G-6-P) by the pancreatic glucokinase enzyme. Once phosphorylated, glucose is metabolized by glycolysis to produce pyruvate, NADH (nicotinamide adenine dinucleotide) and ATP. Pyruvate is in turn metabolized in mitochondria to produce ATP that acts on the β-cell-specific ATP-sensitive K^+ (KATP) channel. This is an octameric structure composed of four pore-forming potassium channel subunits (Kir6.2) and four high-affinity regulatory Sulphonylurea Receptor 1 subunits (SUR1). ATP produced by G-6-P glycolysis and by mitochondria binds to the Kir6.2 subunit. This leads to K^+channel closure and β-cell membrane depolarization which results in opening of the voltage-dependent calcium channels with an influx of calcium. Ultimately, Ca^{2+} influx induces fusion of insulin-containing granules with the β-cell membrane and insulin secretion (Adapted from Nakhla M. and Polychronakos, 2005).



1.4.2 Genes involved in monogenic diabetes

More than 20 genes are highly expressed in the pancreatic β-cell and their mutations have been reported in patients with diabetes (e.g. *GCK* (encoding glucokinase), *ABCC8* (encoding SUR1), *KCNJ11* (encoding KIR6.2) and genes encoding transcription factors involved in pancreas

development and/or β-cells differentiation and function (*HNF-1α, HNF-4α, IPF-1, HNF-1β, NEUROD1*)). Other proteins are common to many different cells and their modifications have been implicated in syndromes including diabetes.

Different etiological mechanisms of dysfunction are involved including impairment of pancreatic beta-cell development and/or gene expression, failure of glucose sensing, disruption of insulin synthesis, disorders of ion channels and increased endoplasmic reticulum stress leading to destruction of the beta-cells (Alkorta-Aranburu G. et al., 2014).

1.4.3 Neonatal diabetes

The term neonatal diabetes, defined as persistent hyperglycemia (lasting more than 2 weeks) occurring within the first 6 months of life and requiring insulin therapy for management, refers to several different monogenic diseases. These are rare metabolic disorders affecting 1 in 100,000 live births (Slingerland A. S. *et al*., 2009). About 50% of neonatal diabetes are transient (TNDM), beginning in the first weeks of life with recovery by 18 months of age. In up to half of patients, diabetes relapses later in life. TNDM is mostly (70% of TNDM) caused by abnormalities within 6q24 imprinted region (e.g. paternal uniparental disomy of chromosome 6 ($UPD_{pat}6$), paternal duplication of 6q24 or point mutation in 6q24 genes (e.g. *ZAC/PLAG1*)) even if mutations in *ABCC8* and less frequently *KCNJ11* have also been reported. Neonatal diabetes can also be permanent (PNDM). The most common mutations in patients with PNDM are found within *KCNJ11* and *ABCC8* even if several other genes have also been implicated in PNDM (cf.

RfX6, IPF-1, Ins2, GCK, GLIS3, or PTF1A) (Edghill E.L, *et al.* 2010)(Nakhla M. and Polychronakos C., 2005).

1.4.3.1 *KCNJ11* (potassium inwardly-rectifying channel, subfamily J, member 11) (chr11p15.1) encodes Kir6.2 protein, the pore-forming subunit of the KATP channels. Heterozygous activating mutations in Kir6.2 subunit increase the number of opened K^+channels (decreasing their sensitivity to ATP) on the cell membrane that remains hyperpolarized. This leads to Ca^{2+} channel closure and absence of insulin release (Nakhla M. and Polychronakos C., 2005). Those mutations are mostly responsible for PNDM even if *KCNJ11* mutations have also been described in TNDM. 80% of *KCNJ11* mutations appear *de novo* but familial cases with dominant pattern of inheritance as well as germline mosaicism have also been reported.

Clinically, patients with PNDM have low birth weight due to lack of insulin secretion in utero and 1/3 of them present with ketoacidosis. In 20 % of PNDMs due to *KCNJ11* mutation neonatal diabetes is associated with developmental delay and epilepsy (DEND syndrome). A milder form is iDEND (intermediate DEND) that combines neonatal diabetes with developmental delay and/or muscle weakness. Association between diabetes and neurological symptoms is explained by *KCNJ11* expression in β-cells as well as in central nervous system (CNS). Indeed, KATP channels regulate electrical activity of the plasma membrane in many tissues and cell types such as pancreatic β-cells and brain but also kidney, skeletal and smooth muscle.

Mutant KATP channels remain sensitive to sulfonylureas. This molecule binds to the sulfonylurea-receptors of the K^+ channels resulting in their closure independently of ATP and in

insulin secretion. Patient with neonatal diabetes due to *KCNJ11* mutations can be treated with sulfonylureas for better glycemic control without risk of hypoglycemia (Edghill E.L, *et al*. 2010).

1.4.3.2 *ABCC8* (ATP-binding cassette, subfamily C, member 8) (chr11p15.1) encodes SUR1 protein, the regulatory Sulphonylurea Receptor 1 subunit of KATP channels. Activating, gain-of-function mutations in this gene are responsible for 10% of all neonatal diabetes, mainly TNDM, but also for monogenic late-onset diabetes (usually mislabeled as type 2 diabetes) (Babenko A.P. *et al*., 2006). DEND syndrome has also been reported in patients with *ABCC8* mutations even if this syndrome is more often associated with *KCNJ11* mutations (Patch A. M. *et al*., 2007). 50% of *ABCC8* mutations arise *de novo* and in contrast to *KCNJ11*, both dominant and recessively inherited *ABCC8* mutations have been described (Edghill E.L, et al. 2010). Mutant KATP channels are overactivated but remain sensitive to sulfonylurea. Neonatal diabetes caused by *ABCC8* mutations can then be treated with oral sulfonylurea, which is more convenient for patients with a better and easier control of blood glucose levels.

Today, more than 100 activating mutations within *KCNJ11* and *ABCC8* have been reported in patients with neonatal diabetes (Edghill E.L, et al. 2010).

Before *ABCC8* and *KCNJ11* activating mutations were reported in neonatal diabetes, inactivating/loss-of-function mutations within those 2 genes were described in babies with hyperinsulinemic hypoglycemia (HH) (Thomas P.M *et al.,* 1995 and 1996). Many genes are involved in HH but mutations in the *ABCC8* and *KCNJ11* genes are by far the most common cause of HH (*ABCC8* mutations account for 45% of HH and *KCNJ11* mutations account for 5% of HH). Most mutations identified in the KATP channel genes are recessively acting, even if a few

reports mentioned dominant mutations. These dominant mutations could be involved in a biphasic phenotype with patients that are at risk of developing young-onset diabetes (Flanagan S.E *et al*., 2011).

1.4.4 Diabetes as part of Syndromes

1.4.4.1: Mutations within *EIFA2K3* (Eukaryotic Translation Initiation Factor 2-Alpha Kinase 3) (encoding the membrane protein PERK located in the endoplasmic reticulum) are associated with **Wolcott-Rallison syndrome** (OMIM # 226980), a rare autosomal recessive disease, characterized by neonatal non-autoimmune insulin-requiring diabetes, bone dysplasia and hepatic dysfunction (Julier C. and Nicolino M., 2010).

1.4.4.2: IPEX (Immunodysregulation Polyendocrinopathy Enteropathy X-linked syndrome) (OMIM # 304790) is a rare neonatal X-linked disorder caused by mutation in *FOXP3* (Forkhead Box P3) leading to dysfunction of the transcription factor Scurfin/FOXP3, a major regulator in immune system. Clinically, IPEX is a heterogeneous disease classically presenting with severe chronic diarrhea (autoimmune enteropathy), diabetes and/or thyroid dysfunction (autoimmune endocrinopathy), eczematous dermatitis, hemolytic anemia and variable immunodeficiency. The prognosis is poor and most of the cases reported have been lethal (Bin Dhuban K. *et al.* 2015).

1.4.4.3: Wolfram syndrome (OMIM # 222300), also called DIDMOAD (Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy, Deafness), is an autosomal recessive disorder, characterized

by early-onset insulin-requiring diabetes and optic atrophy. Both are required to validate the diagnosis. Diabetes insipidus (loss of vasopressin-secreting neurons) appears later in 50 to 82% of individuals as well as progressive sensorineural deafness in 62% of patients. Hydronephrosis (due in part to the high urine flow in diabetes insipidus), neurologic dysfunction (mostly ataxia of the trunk), or psychiatrist illness can complete the syndrome. Wolfram syndrome is caused by loss-of-function mutations in the gene of wolframin (*WFS1*), a transmembrane protein of the endoplasmic reticulum, expressed mainly in pancreas and neurons where it regulates Ca^{2+} homeostasis (Rigoli L. and Di Bella C., 2012).

1.4.4.4: Maternally Inherited Diabetes and Deafness (MIDD) (OMIM # 520000) is a rare form of diabetes (1% of all diabetes) associated with hearing loss, occurring in mid-adulthood. The mutation m.3243A>G in the mitochondrial gene *MT-TL1* (mitochondrial tRNALeu) is responsible for 85% of MIDD. Other mitochondrial gene mutations (*MT-TK*, or *MT-T)* have also been involved in MIDD (Murphy R. *et al*., 2008).

1.4.5 Maturity Onset Diabetes of the Young (MODY)

MODY is a clinically and genetically heterogeneous group of disorders diagnosed in adolescence or early adulthood (<25 years of age). MODY is the most common form of monogenic diabetes, accounting for 1 to 5 percent of diabetes (Eide S.A. *et al.,* 2008) (Shields B.M. *et al.,* 2010). MODY is caused by dominantly inherited loss-of-function mutation (haploinsufficiency) within genes regulating pancreatic β-cell development, function and regulation, glucose sensing and expression of insulin gene itself (Nakhla M. and Polychronakos C., 2005) (Naylor R. and

Philipson L.H., 2011). MODYs are characterized by hyperglycemia associated with possible other clinical features depending on the underlying molecular genetic defect.

Table 1 summarizes the most common forms of MODYs (MODY 1 to 6) and their characteristics.

Table 1 describes clinical features in patients carrying heterozygous inactivating mutations. Homozygous mutations within the *GCK* gene result in a complete deficiency of glucokinase and lead to permanent neonatal diabetes. Likewise, homozygosity for *IPF-1* mutation causes pancreatic agenesis and PNDM requiring pancreatic enzymes replacement in addition to insulin therapy (Nakhla M. and Polychronakos C., 2005) (Naylor R. and Philipson L.H., 2011).

Other very rare types of MODYs have also been described. MODY 7 is caused by mutations in the Fruppel-like factor 11 (*KLF 11*) gene, MODY 8 in the Carboxyl Ester Lipase (*CEL*) gene, MODY 9 in the Paired box gene 4 (*PAX4*), MODY 10 in the Insulin (*INS*) gene, MODY 11 in the Tyrosine kinase, B-lymphocyte specific (*BLK*) gene, MODY 12 in *ABCC8* and MODY 13 in *KCNJ11*. They all present with diabetes but *CEL* mutations also induce pancreas atrophy and exocrine pancreatic dysfunction (www.OMIM.org) (Naylor R. and Philipson L.H., 2011) (Bowman P. *et al*., 2012) (Bonnefond A. *et al,* 2012).

MODYs are often misdiagnosed as type 2 or type 1 diabetes, as they may present with ketosis and /or high blood glucose levels. MODY diagnosis should be considered in “T1DM patients” in the presence of a diabetic parent and absence of islet autoantibodies and if endogenous insulin secretion seems present outside the honeymoon period. Moreover, a diagnosis of T2DM in a non-obese young individual without evidence of insulin resistance should also be questioned (Naylor R. and Philipson L.H., 2011).

MODYs can be diagnosed based on family history which entails an affected parent and, on further careful family history, grandparents and often aunts, uncles and cousins. These extended pedigrees have originally made the monogenic cause obvious and have led to linkage analysis that identified the distinct MODY types.

MODY treatment depends on the metabolic status of the patient. However, MODYs 3 and 1, the most frequent types of MODY, are particularly sensitive to sulfonylureas that are the treatment of choice in those HNF-1α and -4α diabetes (Naylor R. and Philipson L.H., 2011) (Tuomi T. *et al.,* 2006). MODY2 is characterized by asymptomatic mild hyperglycemia with rare complications and generally this glucokinase-related diabetes does not require any therapy.

Table 1: Most common MODY types and their characteristics

	Affected gene (gene symbol)	Function	Clinical features in heterozygous state	Frequency
MODY 1	Hepatocyte Nuclear Factor 4 alpha (***HNF-4α***)	Transcription factor important to pancreatic development and β-cell differentiation and function.	Diabetes-reduction in serum concentration of triglycerides, low HDL, reduction in apolipoproteins AII and CIII and Lp(a) lipoprotein. Microvascular complications are common.	<10%
MODY 2	Glucokinase (***GCK***)	Enzyme (in β-cells and hepatocytes)-glucose sensor in β-cells.	Impaired fasting glucose, impaired glucose tolerance, mild hyperglycemia. Microvascular complications are rare.	15-30%
MODY 3	Hepatocyte Nuclear Factor 1 alpha (***HNF-1α***)	Transcription factor important to pancreatic development and β-cell differentiation and function.	Diabetes-renal glycosuria (low renal glucose threshold), increased HDL. Microvascular complications are common.	50-65%
MODY 4	Insulin promoter factor 1 (***IPF-1***)	Transcription factor that regulates pancreas development.	Mild diabetes	rare
MODY 5	Hepatocyte Nuclear Factor 1 beta (***HNF-1β***)	Transcription factor important to pancreatic development and β-cell differentiation and function.	Diabetes- renal cysts- progressive non-diabetic renal dysfunction-genitourinary abnormalities (uterine abnormalities-hypospadias)-pancreas atrophy- exocrine pancreatic dysfunction- joint laxity-learning difficulties-hyperuricaemia/gout.	rare
MODY 6	Neurogenic differentiation 1 (***NEUROD1***)	Transcription factor that regulates neuron development, pancreatic islets development and insulin gene expression.	Diabetes	rare

(Fajans S.S. *et al.,* 2001) (Nakhla M. and Polychronakos C., 2005) and (Naylor R. and Philipson L.H., 2011)

1.5 Other diabetes

1.5.1 Diseases of exocrine pancreas

1.5.1.1 Cystic fibrosis-related diabetes (CFRD). Cystic fibrosis (CF) is a multisystem disorder caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene. CFTR protein is a chloride/bicarbonate channel whose dysfunction results in an ionic imbalance of secretions from several organs such as lung, liver, gastrointestinal tract and pancreas. CF initially affects the exocrine pancreatic function but, in many patients, the pancreatic islet cells are also progressively affected, with decreasing in insulin secretion. Diabetes is now the most common complication of CF and insulin is the only recommended treatment for CFRD (Barrio R., 2015).

1.5.1.2 Hemochromatosis. Hereditary hemochromatosis (HH) is an autosomal recessive disorder characterized by excessive iron storage. Both impaired insulin secretion (toxic increased iron content of β-cells) and insulin resistance ("second hit" in obese individuals unable to compensate by increasing insulin secretion) have been implicated in the pathophysiology of diabetes in HH. Early diagnosis and therapeutic phlebotomy can improve insulin secretion capacity and stop diabetes progression (Mitchell T.C. and McClain D.A., 2014).

1.5.1.3. Chronic pancreatitis (CP). 41 to 86% of patients with chronic pancreatitis develop diabetes as a result of disease progression or surgical resection. Diabetes is the major late sequel of CP, requiring insulin therapy (Malka D. *et al.*, 2000).

1.5.2 Endocrinopathies

Acromegaly (growth hormone hypersecretion), Cushing's disease (cortisol hypersecretion in response to ACTH hyperproduction), and hyperthyroidism (excess of thyroid hormones) can be associated with glucose intolerance or overt diabetes, requiring insulin therapy (Resmini E. *et al*., 2009).

1.5.3. Drug-induced diabetes

Several drugs such as antipsychotics, cardiovascular drugs, immunosuppressants, hormones (growth hormone, sex hormones), anti-infectives, and glucocorticoids (GCs) have the side effect of inducing glucose intolerance or diabetes (Luna B. and Feinglos M.N., 2001) (Fathallah N. *et al*., 2015). Glucocorticoids are a common cause of clinically significant drug-induced hyperglycemia. Indeed, GCs are used to treat a wide range of diseases through their anti-inflammatory and immunosuppressive effects. However, they are also associated with many side effects including hyperglycemia, occurring in about 20% of patients and requiring insulin for management (Liu X.X. *et al*., 2014).

1.5.4 Gestational diabetes mellitus (GDM)

GDM is defined by The American College of Obstetricians and Gynecologists as hyperglycemia discovered during pregnancy (with onset or first recognition at any time during pregnancy). Several organizations (e.g. WHO, American Diabetes Association) have proposed a new definition of GDM because of the increased incidence of undiagnosed T2DM in women of childbearing age. In order to distinguish GDM from preexisting diabetes (T2DM or MODY), the

definition of GDM has been adapted as hyperglycaemia developing in the second half of pregnancy. GDM occurs in about 5% of pregnancies with an expected increase as the epidemic of obesity expands. In normal pregnancy, peripheral tissues become progressively insensitive to insulin, partly caused by placental hormones and weight-gain. GDM occurs when pregnant woman is not able to produce enough insulin to compensate for this normal insulin resistance. Many risk factors influence development of GDM such as obesity, maternal age, ethnicity, family history of diabetes, polycystic ovary syndrome (PCOS) and personal history of GDM. By definition, GDM is a pregnancy disease and women with GDM mostly return to euglycaemic state shortly after delivery. GDM is responsible for increased morbidity in the mother and the baby. The treatment is diet and insulin if needed (Kampmann U. *et al.,* 2015) (Voormolen D.N. et al. 2016).

1.5.5 Insulin receptor (INSR) defects

Several very rare syndromes are associated with insulin receptor gene mutations leading to severe insulin resistance. Type A severe insulin resistance is the mildest form of inherited insulin resistance that usually starts during adolescence. In addition to diabetes mellitus, patients develop acanthosis nigricans and in affected women, hyperandrogenism manifests as hirsutism, polycystic ovaries, virilization, and oligomenorrhea (Nakhla M. and Polychronakos C., 2005).

Donohue Syndrome (OMIM # 246200) is the most severe form of genetic insulin resistance that presents with intrauterine and postnatal growth retardation associated with dysmorphic

features and biological alterations. Most patients die within the first year of life (Nakhla M. and Polychronakos C., 2005) (de Kerdanet M. *et al.,* 2015).

Rabson-Medenhall syndrome (OMIM # 262190) is milder than Donohue Syndrome, with progressive hyperglycemia and refractory ketoacidosis, growth retardation, dysmorphic features, acanthosis nigricans and abnormal dentition. Patients with Rabson-Medenhall syndrome die within the third decade. Donohue and Rabson-Medenhall patients are homozygous or compound heterozygous for mutations in the *INSR* gene (Ben Abdelaziz R. *et al.,* 2016).

1.5.6 Congenital lipoatrophic diabetes

Lipoatrophic diabetes is a heterogeneous group of disorders that associate complete or partial absence of adipose tissue with insulin resistance and hypertriglyceridemia. The insulin resistance leads to hyperglycemia and hyperinsulinemia with the associated clinical features of acanthosis nigricans and hyperandrogenism in female patients. Both autosomal recessive and dominant mode of inheritance are described in this spectrum of disorders (Nakhla M. and Polychronakos C., 2005) (Oral E.A., 2003).

2. Objective of the project

Our hypothesis is that monogenic, autosomal recessive, non-autoimmune form of insulin-deficient, young-onset diabetes exists. Indeed, in the large group of T1DM, patients with absolute insulin deficiency but with no evidence of autoimmunity (classified as T1BDM) have been described in the literature with attempts in finding mutations within genes involved in β-cell insulin secretion (Moritani N. *et al.,* 2013). However, unlike neonatal diabetes that are characterized by their very early onset, or MODYs that present with a strong family history, an autosomal-recessive form of diabetes would have neither specific clinical characteristics nor family history. Indeed, the only family history would be diabetes in ¼ of the siblings, common in autoimmune T1DM and given the size of contemporary families, most cases will have a completely negative family history.

Today, next generation sequencing (NGS) and large T1DM cohorts make our hypothesis testable.

Therefore, our objective is to test DNA of selected cases diagnosed as T1DM in search of autosomal recessive, non-autoimmune diabetes that clinically presented as T1DM.

3. Materials and Methods

Position of all genes, single nucleotide variants (SNVs) and single nucleotide polymorphisms (SNPs) have been expressed according to GRCh37/hg19 genome reference.

3.1. First part of the project:

3.1.1 Materials

The Type 1 Diabetes Consortium Group (T1DGC) was created by the National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK) and the Juvenile Diabetes Research Foundation to identify genes that contribute to the risk of T1DM. The T1DGC built a resource base of well-characterized families from multiple ethnic groups in different countries to facilitate the localization and characterization of T1DM susceptibility genes (Rich S.S. *et al*., 2009). In the T1DGC database, all information about clinical features, onset of diabetes, autoantibodies, HLA typing, ethnic origin, and country of origin is available and DNAs from all those families are stocked in the NIDDK repository. Family members are labeled based on the following code: 01= father; 02=mother; 03-04-07= affected children; 05-06= unaffected children. This code is added to the six-digit family code

Data from 2,345 sibling pairs affected with T1DM were analyzed. Results for the two available autoantibody tests (GAD65 and IA2) found approximately 30% of all subjects to be negative. However, most of these cases had a positive affected sibling, indicating that the disease was nevertheless autoimmune. Among 167 families where both affected siblings were autoantibody-negative, we found 17 families in which both affected siblings were also negative

for any of the known predisposing HLA genotypes (i.e. negative for the presence of at least one copy of the predisposing haplotypes, HLA-*DRB**301–*DQB**201 and *DRB**401–*DQA**301–*DQB**302 (Polychronakos and Li Q*.,* 2011)) and in which the parents were non-diabetic. Although some of these 17 families may still have autoimmune T1DM, we posited that most could have diabetes for a reason other than autoimmunity. In that case, the occurrence of the phenotype in two or three affected siblings born to non-diabetic parents makes it likely that this is an autosomal recessive trait.

DNAs from the 17 families were obtained from NIDDK repository. Given that the vast majority of mutations causing monogenic disease involve alterations in protein sequence, we decided to perform whole exome sequencing (WES) on selected samples. Indeed, exome is the protein-coding part of the genome (<2%) that contains about 85% of known variants causing monogenic disease (van Dijk E.L*. et al.*, 2014). DNAs from 3 affected children from 3 different families were sent to McGill/Génome Québec Innovation Centre for WES by capture on the SureSelect 50 Mb exome library (reduced genome library specific for exome capture) that covers most of the ~200,000 protein-coding exons in the human genome and was sequenced on the *Illumina* hiSeq (very high throughput sequencing). Next generation sequencing (cf Illumina, SOLiD, Ion Torrent…) refers to massively parallel sequencing, which means that millions of small fragments of DNA can be sequenced at the same time, creating a massive pool of data. Today, NGS allows sequencing the whole exome even the entire human genome much faster and for an affordable price and it has become an essential tool for genetics studies.

We sequenced our 3 samples at a mean depth of 100, a level at which large deletions or duplications can also be detected, based on read counts. Sequencing depth of coverage refers to the average number of times each base is read during a sequencing run. At higher levels of coverage, each base is covered by a greater number of aligned sequence reads, so base calls can be made with a higher degree of confidence. This is important to remove sequencing errors and to highlight low frequency mutations (Sims D. *et al*., 2014).

From WES data of the three affected children, we obtained extensive information about single nucleotide variants such as the genes (and the chromosomes) where the SNVs are located, the exact nucleotide position of the SNVs, the reference and alternate alleles, as well as the rsID (=SNP identity number listed in dbSNP database) (when available).

Those data were filtered in order to select candidate genes for autosomal recessive diabetes using the presence of **2 or more protein-altering variants** within the gene, a positive LOD score for the locus of interest in the proband's nuclear family **(linkage analysis**, using negative LOD scores to eliminate variants not shared by the untested sibling) and a significant **expression of the gene** in the pancreatic islets.

3.1.1.1 Protein-altering variants

As we expect to find a new autosomal recessive form of diabetes, candidate genes must have at least 2 variants in the same base pair (homozygous) or in two different base pairs within the same gene (compound heterozygous). Each genome carries 5-10 such genes with homozygous or compound heterozygous mutants but the probability of the same gene having them in both

siblings, in more than one family, is very small. Presence of diallelic variants on the same gene in two or more families will be considered strong evidence of causal relationship.

The variants must lead to amino-acid changes resulting in modification in the structure or the function of the protein. Synonymous variants were then removed from the candidate gene list.

The different protein-altering variants we expected finding were the following: 1) missense variant (point mutation leading to a change in the amino acid sequence), 2) nonsense variant (point mutation leading to a premature stop codon), 3) frameshift variant (deletion or insertion of a number of nucleotides not divisible by 3 that changes the reading frame and then the translation from the original), 4) splicing variants, or 5) in-frame variant (deletion or insertion of a number of nucleotides divisible by 3 that does not changes the reading frame but can change the amino-acid sequence).

3.1.1.2 Linkage analysis

3.1.1.2.1 Genetic linkage

In 1865, Gregor Mendel's second law of "independent assortment of traits" stated that the alleles of one gene sort in the gametes independently of the alleles of another gene. Progressively, scientists showed that this law could only be applied to loci that are far from each other, e.g. located on different chromosomes. In the beginning of the 20th century, T.H. Morgan was one of the first scientists suggesting that genes could be linked as a consequence of being located close together on a chromosome (Morgan T.H., 1911). One of his students, Alfred Sturtevant developed genetic maps based on Morgan's work (Emerson S., 1971).

Genetic linkage refers to the physical proximity of loci along the chromosome. Two loci are "linked" if they are sufficiently close together such that their alleles tend to cosegregate within families meaning that they are transmitted together from parents to offspring more often than expected under independent inheritance.

Meiosis involves a process known as genetic recombination, during which chromosomes exchange segments with their homologous (crossing-over). As a result, the gametes produced during meiosis are genetically unique. Genetic recombination occurs more frequently between loci that are far apart. The more the distance between 2 loci is short; the lower will be the probability of recombination events. Recombination is then a function of the distance between 2 loci. Recombination frequency (θ) between two loci is defined as the ratio of the number of recombined gametes to the total number of gametes produced. θ ranges between 0 and 0.5 (0 indicating perfect linkage and 0.5 meaning complete independence of the two loci). Recombination frequency can also be expressed in terms of genetic map unit or centimorgan (cM). 1 cM represents 1% (1 product of meiosis in 100) of recombination between two genes. Recombination frequency can then be used to construct a linear gene map that would reflect a gene's position, relative to the other genes, along a chromosome (Pulst S.M., 1999) (Borecki I.B. and Province M.A., 2008).

3.1.1.2.2 Linkage studies

The purpose of linkage analysis is finding the approximate location of a gene for a trait. In linkage studies, we seek to identify trait loci that cosegregate with a specific genomic region, tagged by polymorphic markers, within families.

For this purpose, linkage studies use affected relatives (mostly siblings) to identify regions on chromosomes that are shared more frequently than expected by chance. Samples are genotyped for a panel of markers (e.g. microsatellites), that have been selected to be regularly spaced across the genome, in order to search for alleles that are inherited together. Regions in the genome that show evidence of linkage are then fine-mapped, meaning that additional markers are typed to narrow down regions associated with the disease. Usually, linkage studies are most effective in finding rare alleles with large effect sizes (Zemunik T. and Boraska V., 2011*).* In T1DM, linkage studies showed evidence of linkage between the *HLA* region on chromosome 6 and T1DM (Nakhla M. and Polychronakos C., 2005) but failed to detect any of the other loci known from replicated association studies.

Parametric (model-based) linkage analysis is the analysis of cosegregation of genetic loci in pedigrees when the disease is expected to be inherited on a Mendelian fashion.

Linkage is usually reported as a LOD score, which is the logarithm of the odds that the loci are linked, divided by the odds the loci are independent (Pulst S.M., 1999). Expression of the likelihood as a logarithm allows summing the likelihoods observed in different pedigrees.

Prior to performing parametric linkage analysis, the disease expression model must be specified, including penetrance (reflecting the mode of inheritance), frequency of the disease allele, marker allele frequencies, a full marker map for each chromosome and a classification of individuals as affected or unaffected. As a result, the higher the LOD score is, the greater the evidence for linkage will be. Generally, a score of 3 (more recently, this threshold has been increased to 3.3 to reach p=0.05) or more is considered significant evidence for linkage (odds

1000:1) while a score of -2 or less is accepted as eliminating the possibility of linkage (Pulst S.M., 1999) (Borecki I.B. and Province M.A., 2008).

In our study, genome-wide linkage analysis was performed by Dr Quan Li, the bioinformatics post-doc in Dr Polychronakos' lab, on MERLIN program using genotyping data on 5k SNPs available in the T1DGC database. As we expect to find a new rare, recessive form of diabetes, prior to perform the linkage analysis, we introduced the following parameters: a penetrance of 0.000001, 0.001 and 0.9 for 0, 1 or 2 shared allele(s), respectively, and a very low disease allele frequency set to 0.000001.

Considering the genetic heterogeneity of T1DM, we are not expecting finding very high LOD scores. Therefore, we decided to only consider negative LOD scores within the nuclear family as arguments against linkage (LOD <0 makes it increasingly unlikely that the untested sibling has the mutation) and we kept all regions with positive or weakly negative (>-0.5) LOD scores as potential regions of interest.

3.1.1.3. Expression of the genes in the pancreas

Expression of each gene in the insulin-producing pancreatic beta cells was estimated from a RNA-sequencing database, using RPKM (reads per kilobase per million mapped reads) values (Eizirik D.L. *et al*., 2012). RPKM quantifies gene expression from RNA sequencing data by normalizing for total read length and for number of sequencing reads (i.e. normalized measure of exonic read density) (Mortazavi A. *et al*., 2008). This facilitates transparent comparison of transcript levels both within and between samples. RPKM value of >3 was taken as evidence of meaningful expression.

After filtering the results from WES database using the 3 criteria described above (2 variants, positive LOD score in proband's family and RPKM >3), we selected eight candidate genes (*CLASRP*, *TRAPPC9*, *HMGXB3, NCKAP5L, NPHP4, PARP4, PKP2,* and *UTP20)* (See **Table 2**). Despite a slightly negative LOD score for *HMGXB3* (LOD -0.031)*,* we kept it in the candidate list given the high RPKM values this gene showed in pancreas (RPKM= 14).

3.1.2 Methods

3.1.2.1 Genotyping of family members for SNVs found by WES

This part of the project (3.1.2.1) was performed by Tedi Quendro, an undergrad student, who worked in Dr Polychronakos's lab before I took over the project.

For each candidate gene mentioned above, he sequenced (Sanger sequencing (see below)) the variants found by WES in each member of the probands' families (See **Table 3**).

We observed that only in one gene *(NCKAP5L)*, the variants seemed segregating with the disease under a recessive model. Thereafter, we had to find at least a second family with a protein-altering mutation in the same gene to validate our hypothesis.

3.1.2.2. Sequencing of all protein-coding exons of *NCKAP5L* in linked families

All protein-coding exons (exons 4 to 12) of *NCKAP5L* were sequenced (Sanger Sequencing) in each member of the families that showed a positive LOD score at this locus (5 families of the initial pool of 17 families).

The sequence of the whole gene was found on www.ensemble.org as well as the sequence of the coding exons. We selected the exons of the 2 coding transcripts (ENST00000335999 and ENST00000433948) and we decided not to sequence the 5'UTR nor the 3'UTR regions. Indeed, those regions play a role in gene expression that is usually not involved in Mendelian disease.

3.1.2.2.1 Primer design and polymerase chain reaction (PCR)

We designed primers for each coding exon. First, we selected the sequence we wanted to be amplified starting at least 60 bases pairs before the starting point of the exon and ending 60 bases pairs past the last base of the exon. This is important to make sure that the exon will be entirely sequenced including the splicing regions surrounding exon sequence. We used primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/) to design the most appropriate primers and we checked that no SNP, polymorphisms found in more than 1% of the general population with no clinical implication, was found within the primer sequence using a list of variants available on Ensembl (www.ensembl.org).

When primer 3 suggested the best primers for the sequence of interest, we checked them on Gene Runner to reject those making hairpin loops, dimers, internal loops or bulge loops at any

temperature used in PCRs cycles. Finally, we validated the primers after checking that they were specific for the region we wanted to be amplified using Blat on the UCSC genome browser.

Polymerase chain reaction or PCR is a technic for amplifying DNA segments. The PRC product is a mix of primers (forward and reverse) binding the DNA target, free nucleotides (i.e 10mM dNTP mix = equimolar solution of deoxynucleotides = dATP, dCTP, dGTP and dTTP), Mg^{2+} solution (50mM MgCl2 from Invitrogen), Buffer solution (10X PCR Rxn Buffer from Invitrogen), UltraPure water and Taq Polymerase (Taq DNA polymerase from Invitrogen). Taq polymerase is a DNA polymerase, an enzyme that assembles a new DNA strand with free nucleotides, using single-stranded DNA as a template and DNA primers for initiation of DNA synthesis. Buffer solution provides a suitable chemical environment for optimum activity and stability of the DNA polymerase and Mg^{2+} complexes the single nucleotides that are the substrate for DNA polymerase.

The recipes we used for PCR mixes were: 1 µl of forward and reverse primers (100ng/ µl) + 0.5 µl of 10mM dNTP mix + 0.5-0.75 or 1 µl of 50 mM $MgCl_2$ solution + 2.5 µl of 10 x Taq Buffer + 17.25-17 or 16.75 µl of UltraPure H2O (volume adapted to $MgCl_2$ volume, respectively) + 0.25 µl of Taq polymerase for a total volume of 23 µl which was added 2 µl of DNA at a mean concentration of 10 to 20 ng/ µl. This recipe was multiplied by 2 for a total volume of 50 µl given that the PCR product had to be sent for sequencing which requires 2x20 µl of amplification product to sequence the sense and the antisense strands.

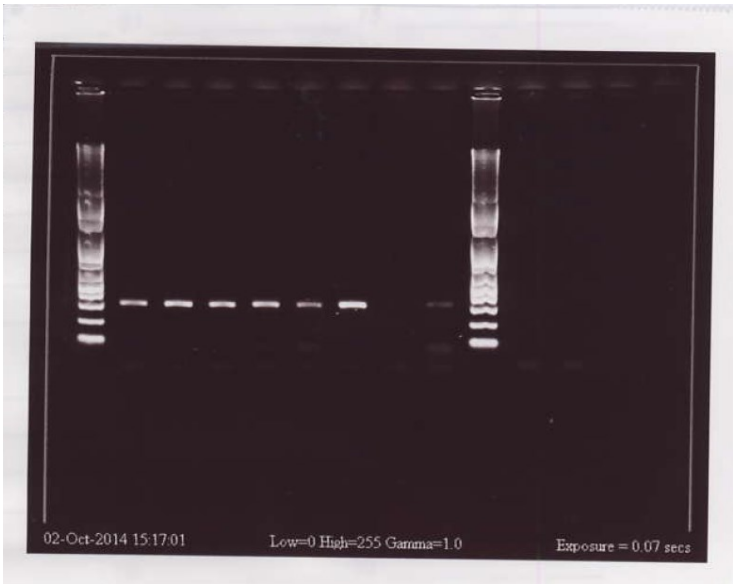
The thermal cycler performs PCR by increasing and decreasing the temperature of the samples following programming steps. Initially, the mixture is heated to denature the double-stranded DNA template into single strands. The mixture is then cooled so that the primers anneal to the DNA template. The annealing temperature is primer specific and can then vary from a reaction to another. At this point, the DNA polymerase begins to synthesize new strands of DNA starting from the primers. Following synthesis and at the end of the first cycle, each double-stranded DNA molecule consists of one new and one old DNA strand. PCR continues with additional cycles that repeat the aforementioned steps. The newly synthesized DNA segments serve as templates in later cycles, which allow the DNA target to be exponentially amplified millions of times. The thermal cycler we used was Gene Amp ® PCR System 9700 and the cycling parameters were: 94˚C for 3 minutes then 35 cycles of 94 ˚C for 30 seconds- variable temperature (=annealing temperature (T_a)) for 30 seconds-72 ˚C for 30 seconds, then 72 ˚C for 10 minutes and 4 ˚C for preservation of the product.

Thereafter, electrophoresis of PCR products was performed to verify that the reaction worked in amplifying the right DNA fragment. PCRs products (10 µl) were mixed with 5 µl of Orange G (30% Glycerol) to be loaded into 1% Agarose Gel (0.8g of agarose, 80 ml of TBE 1x) added with 3.5µl of DNA staining agent (GelRed from Biotium or RedSafe from iNtRON). The DNA fragments were separated according to their weight by using electricity (120 Volts). After running the gel, we could see the DNA bands by exposure of the gel to UV light that revealed the DNA staining agent bound to DNA fragments. Size-marker (1Kb+ ladder) was co-electrophorated with DNA samples for fragment size determination (see **Figure 3**).

For each coding-exon of *NCKAP5L* gene, we designed primers (**Table 4**) and optimized PCRs ($MgCl_2$ and annealing temperature) on DNA controls. Patients DNA were only used when all steps of PCRs were perfectly optimized.

When gel electrophoresis confirmed the presence of the right DNA fragments, PCR products were sent to McGill Genome Centre for Sanger sequencing.

Figure 3: Gel electrophoresis.
Gel electrophoresis of *NCKAP5L* Exon 6 PCR product (290bp). The first 6 wells contain patients' samples; the well 7 contains a negative control (PCR performed without DNA) and the last well contain a positive control (DNA control on which the PCR was designed). We can observe a clear band of DNA for the 6 first wells and for the last well on position 290 of the 1Kb+ ladder confirming the amplification of the right band.



3.1.2.2.2 Sanger sequencing

DNA sequencing with chain-terminating inhibitors ("Sanger Sequencing") was first described by F. Sanger in 1977 (Sanger F. *et al*., 1977). Sanger's method is based on the use of dideoxynucleotids (ddNTP's) in addition to the normal nucleotides (A, T, C and G) found in DNA. Dideoxynucleotides are essentially the same as nucleotides except that they contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH). These modified nucleotides, when incorporated into a growing strand of DNA, prevent the addition of further nucleotides and stop the replication process. This results in varying lengths of short DNA (they differ in length from each other by a single base). These short DNA strands are ordered by size, and by reading the end letters from the shortest to the longest piece, the whole sequence of the original DNA is revealed. Each ddNTP is labeled with a specific fluorescent dye emitting light at different wavelength.

All PCRs products of *NCKAP5L* protein-coding exons from each member of the 5 linked families were sent to McGill Genome Centre for sequencing using Sanger's method. We sent 2 x 20 µl of each PCR product and 10 µl of each primer (5 µM).

The sequencing results were analyzed following several steps:

1) We compared the sequence of our samples with the reference sequence using Blast (Basic Local Alignment Search Tool) on NCBI. This is essential to find homozygous variants that are not locatable on chromatograms.

2) Even if sample sequences were matching perfectly with the reference sequence, we checked carefully each chromatogram looking for heterozygous variants. Indeed, some heterozygous variants could be mislabeled with the reference nucleotide (see **Figures 4 and 5**).

3) Each variant found within the sense sequence had to be confirmed in the antisense to reject sequencing errors. Indeed, sometimes, the background of the chromatogram was not perfect and some fluctuations could be misinterpreted as heterozygous variants. Only variants found from a clear chromatogram on the sense and the antisense sequences were considered as real variants.

Chromaspro was the program we used to read chromatograms.

Figure 4: Chromatogram. This is the picture of a classical chromatogram after Sanger Sequencing. The four nucleotides are differently colored: blue= C; red= T; green= A and black= G. In this sequence, all base pairs are identical on both alleles.

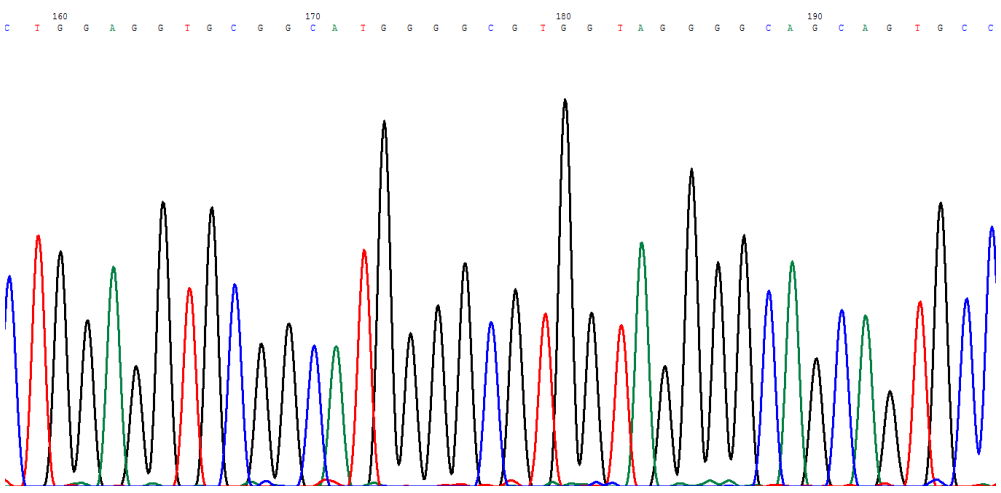
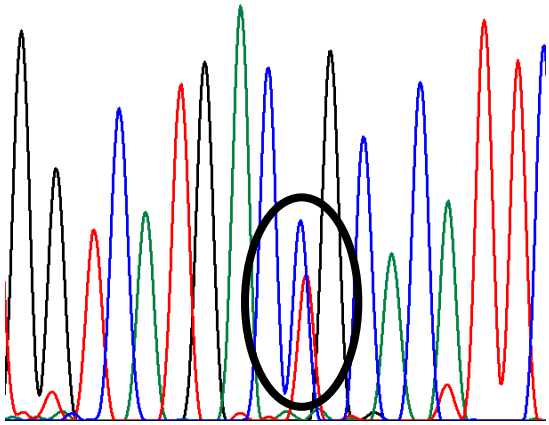


Figure 5: Heterozygous SNV. This chromatogram shows a heterozygous variant (C/T) within the sequence. This variant is shown on the graph by a double peak (blue=C and red=T) representing the nucleotide carried by each allele. This individual is then heterozygous for this variant.



Each variant was identified and labeled with its rsID when available or with its position on the chromosome (using GRCh37/hg19 genome reference). For new variants (i.e. not listed on dbSNP), we used Mutation Taster (www.mutationtaster.com) to predict their disease potential.

3.2 Second part of the project

3.2.1 Materials

3.2.1.1 NIDKK repository

After this work had been performed, an additional twenty families, for a total of thirty-seven families fulfilling the conditions we sought (no autoantibodies (GAD65 and Anti-IA2) and no

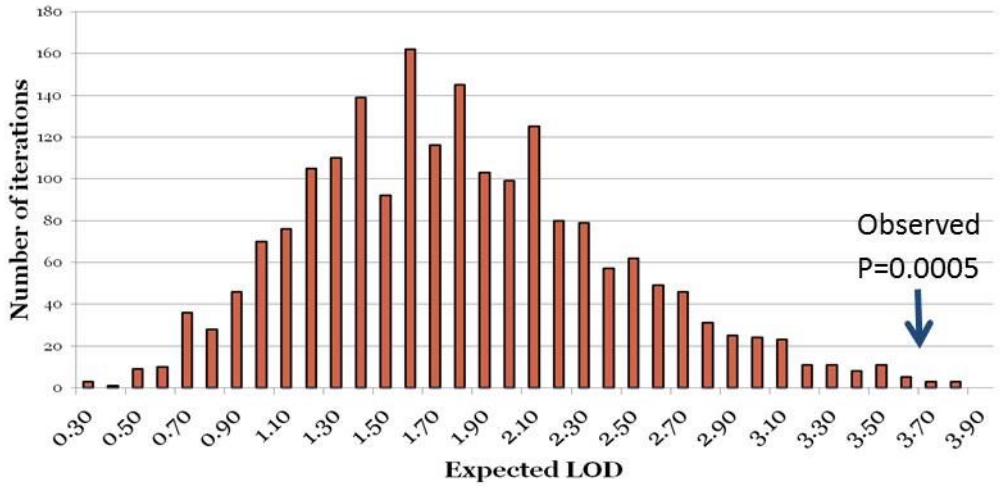
predisposing HLA) were identified from T1DGC repository, as more complete HLA and antibody testing became available.

We decided to send 11 more samples for WES and we tried to select them in order to optimize our results.

A concept that we didn't take into account in the first part of the project was the genetic heterogeneity of the disease. This is a phenomenon in which a single phenotype or genetic disorder may be caused by any one of a multiple number of mutations in one or different genes. If only a small proportion of families are truly linked to a given locus, the associated positive LOD scores can be overwhelmed by negative scores from unlinked families, masking the linkage evidence.

To work around this issue, we performed heterogeneity-insensitive linkage analysis (Grant A.V., *et al*., 2011). At each marker, we calculated the LOD score only for families that show linkage (LOD>0). We identified the locus with the highest LOD score (LOD=3.7) and in order to assess its statistical significance, we compared this result to what was expected under the null hypothesis. The distribution of the expected results was obtained through permutations analysis (moving the position of the markers along each chromosome by a random distance, different for each family), repeated 10,000 times (**Figure 6**). Each time, we calculated the LOD score taking into account only the positive values. We were thus able to compare the expected result with the observed result and we identified significant linkage to two loci on chromosomes 6 and 11 ($p<0.0001$).

Figure 6: Permutation analysis. Frequency distribution, expected under the null hypothesis, of LOD scores obtained by taking into account only the families with LOD>0 at each locus. It was generated by translating the position of markers along each chromosome by a random physical distance different for each family and counting only the maximum LOD. It serves to assess the statistical significance of the heterogeneity-insensitive linkage analysis. The Observed LOD (blue arrow) is plotted against this distribution. Only 5 of the 10k iterations produced an equal or higher LOD.



Twenty two of the 37 families showed linkage for one of those loci. Eleven of those 22 samples were prepared (DNA concentration validated by Picogreen) and sent to McGill/Génome Québec Innovation Centre for WES by capture on Illumina Nextera Rapid Exome (requiring less DNA (50 ng)) and sequencing on Illumina HiSeq.

We asked the bioinformatics of McGill Genome Centre to reanalyze our first 3 samples and to combine them with the 11 samples to homogenize and optimize our results.

WES gave us extensive information about all SNVs found in every sample (location of SNVs (genes and chromosomes associated), reference and alternate alleles, minor allele frequency (from 1000 Genomes), amino-acid change, gene mutation frequency, and SIFT and PolyPhen2

scores (i.e. bioinformatics programs that use sequence conservation to predict whether a nonsynonymous SNV, resulting in an amino acid substitution is likely to be deleterious to protein function (Flanagan S.E, 2010))). SIFT (Sorting Intolerant From Tolerant) score ranges from 0 to 1. The amino-acid substitution is predicted damaging if the score is ≤0.05 and tolerated if the score is >0.05 (www.sift.jcvi.org). PolyPhen2 (Polymorphism Phenotyping) score also ranges from 0 to 1. The variant is predicted to be benign if the score is 0.0 to 0.15, possibly damaging if the score is 0.15 to 1.0 and more confidently damaging if the score is 0.85 to 1.0 (https://ionreporter.thermofisher.com). Those scores are only available for nonsynonymous SNV.

Before filtering the results, we looked for genes involved in monogenic diabetes and we removed 4 of the 14 samples (3 families) because of heterozygous variants found in *KCNJ11* and compound heterozygous variations found in *WFS1.* We also observed heterozygous variants in *ABCC8* in 2 other families.

Thereafter, we filtered the WES results based on the aforementioned criteria: the presence of 2 or more protein-altering variants within the gene, a positive LOD score for the locus of interest in the probands' families and a significant expression of the gene in the pancreatic islets (RPKM >3). We removed genes located on sex chromosomes and we finally selected a list of 12 candidate genes (see **Table 5**). We also made a list of seven hemizygous X-chromosome variants found in male subjects, if the affected sibling was also a male, consistent with X-linked inheritance but, as of now, we have done no further work with them. Even if their LOD scores were slightly negative, we still kept 2 genes (*CACNB3* and *ACIN1*) in our candidates list given their high RPKM values and the number of linked families to these loci. Moreover, we also kept

C18ORF65 in our list because of the positive LOD score even if we did not have RPKM score for that gene (see **Table 5**).

3.2.1.2 Homozygosity-by-descent (HBD)

A cohort of 1,000 diabetic patients from the pediatric diabetes clinics across Canada has been genotyped for 550,000 polymorphic markers along the genome (*Illumina* Hap550 array). In addition to GWAS, these results can be used for determining homozygosity-by-descent regions. This method is usually used to highlight regions of interest in individuals with a recessive disease whose parents are related. Indeed, it tests the hypothesis that a homozygous mutation in a recessive disease gene is ‘‘identical by descent’’ (IBD) by segregating twice to the affected child from a common ancestor through both the maternal line and the paternal line. In addition, homozygosity mapping supposes that a short chromosomal segment surrounding the homozygous mutation has not been recombined by crossing over meaning that all SNP markers in this segment will therefore also be homozygous by descent. Homozygosity mapping has been extended to individuals who are from outbred populations but carry homozygous disease gene mutations by descent from an unknown distant ancestor, even though the parents are unaware of any relation (Hildebrandt F. *et al*., 2009) (Lalitah K *et al.*, 2002)

In our cohort, we identified patients who had 1 Mb segments of chromosome covering a candidate gene where all variants were homozygous. These segments, considered as homozygous-by-descent, were defined by Dr Quan Li using the PLINK software package. Homozygosity mapping can only be applied to individuals that bear homozygous mutations of

recessive disease genes. We then expected finding homozygous mutations and not compound heterozygous mutations in those samples unlike what we looked for in samples from NIDKK.

3.2.2 Methods

3.2.2.1 Further evaluation of variants found in genes known to be mutated in monogenic DM

We genotyped all members of probands' families for the variants we found in the three genes already described in the literature as responsible of monogenic forms of diabetes (*KCNJ11*, *ABCC8*, and *WFS1*) (see **Table 6**). We designed primers surrounding the variations by at least 120 bp upstream and downstream (see **Table 7**). As described previously, we used Primer3 to design primers. For *WFS1*, we were not able to design primers in regions free of SNPs. We then ordered primers with both the reference and the possible alternate nucleotides.

All PCR products that revealed the right length band of DNA on 1% Agarose Gel were sent to McGill Genome Centre for Sanger sequencing. The results were interpreted following the same rigorous steps as previously described.

Variations within those three genes were not all point variants. Indeed, within *WFS1*, one variant was a single base deletion and another variant was a 16-bases deletion. The point deletion was confirmed by Sanger sequencing and the long deletion was revealed by high concentration gel electrophoresis (NuSieve GTG® 3.5% Agarose gel) that ensured fine resolution for small DNA fragments.

3.2.2.2 Candidate genes for autosomal recessive diabetes

We genotyped every member of probands' families for the variants found within *CHPF, FRY, HSPBAP1, CACNB3, ACIN1, C18orf65, RBM23,* and *SCAMP2* genes. Primers were designed to amplify sequences surrounding the variants by at least 120 bp up- and downstream (see **Table 8**). All PCR products from DNA samples were finally sent to McGill genome Centre for sequencing by Sanger's method. This part of the work was done in collaboration with Dr Tuğba Demirci who was working in the lab for 6 months.

HSPBAP1 variant in family 220356 seemed segregating with the disease under a recessive model. Eleven families (including the WES family) were linked to *HSPBAP1* locus but we only had DNA from 7 of them. All members of those 6 families (other than the WES family) were genotyped for the variant using the same primers and PCR conditions as described in the previous paragraph.

SCAMP2 variant in family 223581 also seemed to segregate with the disease under a recessive model. Six families (including the WES family) were linked to *SCAMP2* locus but we only had DNA for 4 of them. We sequenced all protein-coding exons of *SCAMP2* (exons 1 to 9; transcripts 001 (ENST00000268099) and 003 (ENST00000566480)) in 1 affected child from each of the 3 linked families (other than the WES family) and in 6 additional patients from McGill cohort showing homozygosity-by-descent at this locus. To increase our number of patients, we included 5 more families from NIDDK cohort that showed very slightly negative LOD score (-0.068) at this locus. We followed the same steps (primer design, PCR optimization, Sanger

sequencing) as previously described. The primers are listed in **Table 9**. This part of the work was done in collaboration with Dr Tuğba Demirci.

3.3 Third part of the project

Given that we found variants in genes involved in monogenic diabetes within our recessive cohort, we suspected that other known mutations could be found in NIDKK cohort. We then added a supplemental part to the project that is not in relation with our initial hypothesis but could be really life-changing for patients. Indeed, we decided to test relevant families for MODY3, an autosomal dominant form of diabetes, misdiagnosed as T1DM.

3.3.1 Materials

Among all samples listed in the NIDKK repository, we selected families where one parent was diabetic and where the affected children and the affected parent were autoantibody-negative (GAD65, IA2 and ZnT8 when available) besides being negative for any of the known predisposing HLA haplotypes (i.e. negative for the presence of at least one copy of the predisposing haplotypes, HLA-*DRB**301–*DQB**201 and *DRB**401–*DQA**301– *DQB**302 (Polychronakos C. and Li Q., 2011)). Nine families fulfilled these criteria.

Dr Quan Li performed linkage analysis using dominant model parameters (disease allele frequency of 0.000001 and penetrance of 0.000001, 0.95 and 0.95 for 0, 1 and 2 shared allele(s)). These results were used to select families that were linked to *HNF-1α* locus (chr12:

121,416,549-121,440,314 (GRCh37/hg19)). Four families showed positive LOD score at this region but we only had DNA for 2 families (423172 and 492182).

3.3.2 Methods

All protein-coding exons of *HNF-1α* (exons 1 to 10) were sequenced in every member of families 423172 and 492182. For the PCR primers, see **Table 10**.

For one sample (49218204), PCR of exon 7 amplified more than one DNA band. We then extracted the right band from the gel using QIAquick® Gel extraction kit (250) and we sent the purified PCR product to McGill Genome Centre for Sanger sequencing.

3' Materials and Methods: Tables

Table 2: First 8 candidate genes characteristics. RPKM in pancreatic islets (Eizirik D.L *et al*, 2012), LOD score in proband's family, chromosome location and number of the 17 families linked to the locus (i.e. showing a positive LOD score at the locus).

Candidate Gene	RPKM in pancreas	LOD	Chromosome number	Number of families linked to the locus
CLASRP	17.9	0.211	19	8
TRAPPC9	9.3	0.211	8	5
HMGXB3	14	-0.031	5	0
NCKAP5L	6.6	0.247	12	6
NPHP4	4.8	0.134	1	7
PARP4	23.1	0.247	13	8
PKP2	4	0.247	12	4
UTP20	3.5	0.247	12	4

Table 3: Families genotyping for variants found by WES in 8 candidate genes. (Adapted from T. Quendro's report, 2013)

Gene Name (SNVs position*)	**Family ID**	**Father**	**Mother**	**Affected Child 1**	**Affected Child 2**	**Unaffected Child 1**	**Unaffected Child 2**
CLASRP (chr19: 45561063)	460380	Hom (R)	Hom (R)	Hom (R)	Hom (R)	–	–
TRAPPC9 (chr8: 140922400)	460380	Hom (R)	Hom (R)	Hom (R)	Hom (R)	–	–
TRAPPC9 (chr8: 141381168)	460380	Hom (R)	Hom (R)	Hom (R)	Hom (R)	–	–
HMGXB3 (chr5: 149384413)	437721	–	Hom (R)	Hom (R)	Hom (R)	–	–
HMGXB3 (chr5: 149389803)	437721	–	Hom (R)	Hom (R)	Hom (R)	–	–
NCKAP5L (chr12: 50185732)	417769	–	Hetero	Hetero	Hetero	Hom (R)	Hetero
NCKAP5L (chr12: 50187215)	417769	–	Hom (R)	Hetero	Hetero	Hetero	Hom (R)
NPHP4 (chr1: 5965750)	417769	–	Hetero	Hetero	Hetero	Hetero	Hom (R)
NPHP4 (chr1: 6027365)	417769	–	Hetero	Hetero	Hetero	Hetero	Hom (R)
PARP4 (chr13: 25075947)	417769	–	Hom (R)	Hom (R)	Hom (R)	No data	Hom (R)
PKP2 (chr12: 32949047)	417769	–	Hetero	Hetero	Hetero	Hom (R)	Hetero
PKP2 (chr12: 32975502)	417769	–	Hom (R)	Hom (R)	Hom (R)	Hom (R)	Hom (R)
PKP2 (chr12: 33031108)	417769	–	Hetero	Hetero	Hetero	Hom (R)	Hom (R)
UTP20 (chr12: 101693499)	417769	–	Hom (R)	Hetero	Hom (R)	Hetero	Hom (R)
UTP20 (chr12: 101693818))	417769	–	Hom (R)	Hetero	Hom (R)	Hetero	Hom (R)
UTP20 (chr12: 101734308)	417769	–	Hetero	Hetero	Hom (R)	Hom (R)	Hom (R)
UTP20 (chr12: 101750392)	417769	–	Hetero	Hetero	Hom (R)	Hom (R)	Hom (R)

Hom (R) stands for homozygous for the reference allele

Hetero stands for heterozygous

*Human Genome reference: GRCh37/hg19

Table 4: Primers for PCR amplification of *NCKAP5L* coding-exons.

Exon	Forward primer (5'-3')	Reverse primer (5'-3')	Length of DNA fragment (bp)	T_a (˚C)	MgCl2 50mM (µl)
E4	TTCATCTTTGTCCTCCTCCTCCTTC	AAAAACCTGCAGGAAGCTCA	299	54	0.5
E5(1)	CACCCTGTCATCCTCCCTAA	GGGCTATCACAACCTGGAAA	815	54	0.5
E5(2)	GCGGGTCCATTCTGTTCTAC	TCTCTAAAGTTTATCACCCCTGGT	669	57	1
E6	CCTCATAGAAGGAGGTTGCAG	AGTGAAAGGAGCGCAGATG	290	54	0.5
E7	GCTGGGAGTCTGGGAGTC	CGTCTGAGCACTGGGAAA	323	54	1
E8(1)	CTGGGCCTGGAACTGGAT	TCATCAGAAGAGGAGGAGGA	511	54	0.5
E8(2)	GAGGGGAAGAGGACACTGG	TTGAGACCCCTAGCTGGAGA	577	54	0.5
E8(3)	TCCCCTTCCTTAGCATGTTC	AAAGGTGCTCCGAGAAAGGT	494	54	0.5
E8(4)	GTCGGAACAGTGGCTCAGA	CTGCTGCTTCAGGCTGTTT	750	54	0.75
E8(5)	TGGGGAAGCTGAAGACAG	GCTGGCAGCTTGCTCTTCT	847	54	0.5
E8(6)	TAGTACCTGGCCCCACTGAC	CTAGCTCCAGACGCAGGTGT	591	57	0.75
E9	GGTGGATGGCAGGAAGTG	CACTCATGGTCTTGGGGAAG	372	54	0.5
E10	CAGACACATTCCCACCCTCT	GGGGACGGGAAGAAGTAAA	247	54	0.5
E11	CTGTGGAATGAGCCTGTTGG	CCGAGTCCTTTCCCTTTCCT	709	58	1
E12	CCAGCCTGTTTCCTCTGAAG	CCTTTTCACCTTCTTATCCACCT	720	57	0.5

Table 5: Second 12 candidate genes characteristics. RPKM in pancreatic islets (Eizirik D.L *et al.*, 2012), LOD score in proband's family, chromosome location and number of the 37 families linked to the locus (i.e. showing a positive LOD score at the locus).

Candidate Genes	RPKM	LOD	Chromosomes	Number of linked families
CHPF	89.8	0.211	2	6
FRY	7.66	0.231	13	7
HSPBAP1	3.94	0.231	3	11
CACNB3	8.6	-0.031	12	10
ACIN1	27.6	-0.015	14	15
C18orf65	?	0.850	18	9
RBM23	7.8	0.211	14	12
SCAMP2	30.97	0.210	15	6
ANKRD11	16	0.231	16	10
ASXL3	3.8	0.849	18	11
KIAA2018	3.2	0.850	3	9
PRR12	13.40	0.034	19	10

Table 6: Variants found within *WFS1-KCNJ11* and *ABBC8* genes.

Gene	Sample ID	Status	Chromosomes	Variant (cDNA)	Amino-acid change
WFS1	15271203	multiple hetero	4	c.2149G>A	p.E717K
WFS1	15271203	multiple hetero	4	c.2207_2207delG	p.G736fs
WFS1	41990004	multiple hetero	4	c.961A>C	p.T321P
WFS1	41990004	multiple hetero	4	c.1362_1377delCACGCGCAGGGCCCTG	p.Y454fs
KCNJ11	43752104	hetero	11	c.679G>A	p.E227K
KCNJ11	43752103	hetero	11	c.679G>A	p.E227K
ABCC8	45676503	hetero	11	c.742C>T	p.R248Q
ABCC8	26036704	hetero	11	c.1384A>G	p.I462V

c.=cDNA; del= deletion; p.=protein; E= Glutamic acid; K= Lysine; G= Glycine; fs= frame shift; T= Threonine; P= Proline; Y= Tyrosine; R= Arginine; Q= Glutamine; I= Isoleucine; V= Valine

Table 7: Primers for PCR amplification of *KCNJ11-WFS1* and *ABCC8* variants.

Genes	Variant	Forward primer (5'-3')	Reverse primer (5'-3')	Length of DNA fragment (bp)	T_a (˚C)	MgCl2 50mM (µl)
KCNJ11	c.679G>A	GATCCTCATCGTGCAGAACA	TGTAACACCCTGGATGAGCA	865	55	0.75
WFS1	c.2149G>A	TCTCAGGGCTCCGTCATC	GAAGAGATASAGCAGRTAGA	1063	54	1
WFS1	c.2207_2207 delG	TATGCCYRTTTCCTGYTCTY	GAAGAGATASAGCAGRTAGA	242	54	1
WFS1	c.961A>C + c.1362_1377 delCACGCGCA GGGCCCTG	CTCATCCTGGTGTGGCTCA	RATGTGGCAGGSGTGYTT	399	54	0.75
ABCC8	c.742C>T	CCCCATCTGTTAGAGATCAAGTG	GCCTCAGTTTCCCCTCTTGT	475	58	0.75
ABCC8	c.1384A>G	CACTCTAGGGGACTGGGGAA	ATGACAGTGTGGGTGTGTGG	279	58	0.75

S = G or C; R = G or A; Y = C or T

Table 8: Primers for PCR amplification of candidate genes variants.

Genes (SNVs position*)	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Length of DNA fragment (bp)**	**T_a (˚C)**	**MgCl2 50mM (µl)**
CHPF (chr2:220404334 + chr2: 220404335)	CGCCTGACTTCCTGAACC	TCCTGTTCAAAGAGTAGCATGG	448	55	0.75
FRY (chr13:32798380)	GCCCTGCAGTGAGTTTTGAT	GATTTCCTTCTCCCCACACA	382	56	1
FRY (chr13: 32783777)	TGATTTGTGCATTATTAACTACTGTTT	GGGGCAGAAGGGAACTTTTA	493	56	0.75
HSPBAP1 (chr3: 122459292)	GAAGAATTAAATGTGTGCAACCAC	TGAAAGAAGGTGGCAACAGA	557	55	0.75
HSPBAP1 (chr3: 122459595)	GAGAACAGAAGATGAAATTAAAACAGA	GACCAGATCAGGGCCAAA	572	55	0.75
CACNB3 (chr12:49221416)	GAATCTCGCACCTCACCAGAAGA	TGGTACAGGTCCTGGTAGGCAT	245	58	0.75
ACIN1 (chr14: 23548787)	TTCTCGGTCTCATTCACCGC	GCTCTCCAGGGGCTGAAATG	238	58	0.75
C18orf65 (chr18: 74280553)	ATTCCGTCCAAGACGCCCAA	TGACCTTTAACATGCCCGCTC	387	58	0.75
RBM23 (chr14:23371267)	GATGAAGGGTTGGGGCTTAAT	GGCTAGAATGAGCTAGACCCTG	360	58	0.75
SCAMP2 (chr15: 75137840 + chr15: 75137428)	CAGCCCTGTCTACACTGGATAA	AACCACCACCACATAAGGCAC	585	58	0.75

*Human Genome reference: GRCh37/hg19

Table 9: Primers for PCR amplification of all protein-coding exons of *SCAMP2*.

Exon	Forward primer (5'-3')	Reverse primer (5'-3')	Length of DNA fragment (bp)	T_a (˚C)	MgCl2 50mM (µl)
E1	GTGCCGTCATAGGGCCA	GCGTTACGATAGGCCAGACC	369	58	0.75
E2-3	CTGGCCCGGATTTTCTCCAC	GGCCAGAGTTCTTTGTCAGC	705	58	0.75
E4	CTCCTTGATGGCTGTCCCTG	CTCAGGGTGAATGCTGGGAG	324	58	0.75
E5	CATACACTATGCTGGCGTGC	GATGGGGCTCTCAAGTCACA	274	58	0.75
E6	TTCGCAGCATGAACCTCACC	AATCTCTGCTCTGGGTGGGA	427	58	0.75
E7	TCTGTCTCTTGCCCATGCTG	AGAAGGCCCAAAGCAGATCC	377	58	0.75
E8-9	ACAGCAGCCACTTGGAGAAG	GCACAACCACCACCACATAA	688	58	0.75

Table 10: Primers for PCR amplification of all protein-coding exons of *HNF-1α*.

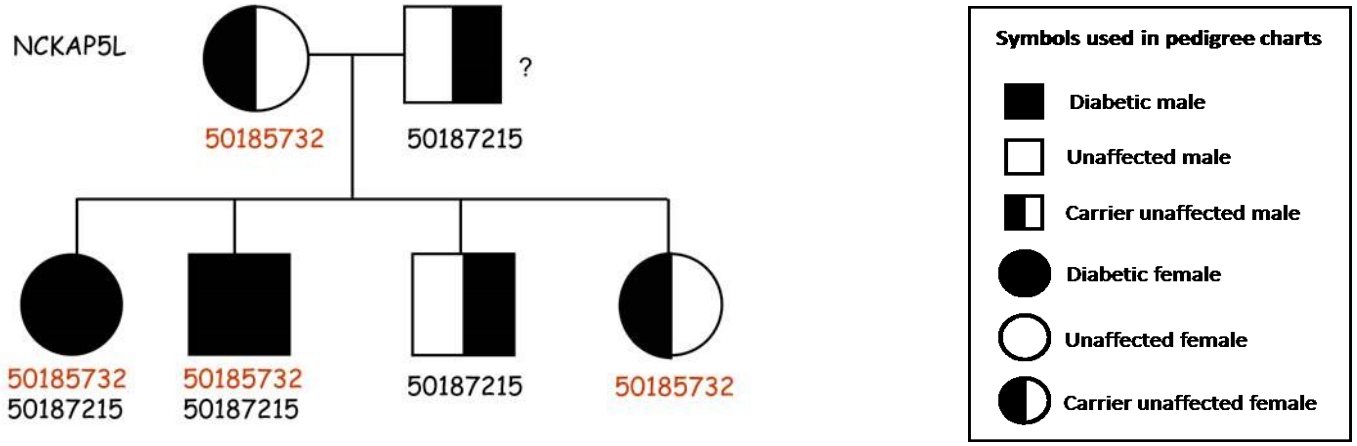
Exon	Forward primer (5'-3')	Reverse primer (5'-3')	Length of DNA fragment (bp)	T_a (˚C)	MgCl2 50mM (µl)
E1	CACAGGGCTTGGCTAGTGG	GGGGACTCAACTCAGAAGG	566	58	0.75
E2	GGGCTCCATAACTGCTTTCA	CCTTTCCATCTACCTGTCTGTG	447	60	0.75
E3	CAAGGTCAGGGGAATGGA	CCCAGACCAAACCAGCAC	638	54	0.5
E4	CAGATCTGCCAGCCTCAAAC	AAGGAGTGGCATGAATGGAA	540	56	0.5
E5-6	TGGAGTTTGAAGTGCTGAGG	GCCCATAGACCCTTCACCTT	960	58	0.75
E7	CCCCCTCCTCCAAACCAC	GTCCCAGAGACACATGCAGA	418	58	0.75
E8-9	CCCTGGATCTCCAACTGCT	CCAGTCCCTGAGATGTTCTTG	840	58	0.75
E10	GTGTGACTTTGGGGTTCCT	GCACAGCTGTCCAGGAAGG	390	58	0.5

4. Results

4.1 First part of the project

All family members of the 3 initial patients were genotyped for the SNVs found in the 8 candidate genes. Only one of those genes (*NCKAP5L*) carried variations that segregated with the disease under a recessive model within family 417769 (see **Figure 7**).

Figure 7: Genotyping of *NCKAP5L* SNVs (chr12: 50185732; T to C and chr12: 50187215; G to A) in family 417769. The mother is heterozygous for chr12:50185732 variant; the 2 affected children are compound heterozygous for both variants; the 2 unaffected children are heterozygous for either the chr12: 50185732 or the chr12:50187215 variants. The father, for whom we did not have DNA, must be heterozygous for chr12: 50187215 variant.
This pedigree was designed according to standard symbols.



The two variations were missense mutations. The first variation (chr12: 50185732; T to C) resulted in replacement of methionine by valine on position 1299 (SIFT= 0.52; PolyPhen2= 0.001). The second variation (chr12: 50187215; G to A) also induced modification in the amino

acid sequence with the substitution of proline by leucine on position 1087 (SIFT= 0.65; PolyPhen2= 0.006).

In order to prove the link between the *NCKAP5L* gene and T1DM, we had to find another family with protein-altering variations in the same gene. All linked families were then sequenced for all protein-coding exons of *NCKAP5L*. Many polymorphisms and variants were identified in several samples but none of these variations seemed to be inherited following an autosomal recessive pattern (see **table 11**). Some variants were synonymous, meaning that they do not change the amino-acid sequence (rs3741555, rs2603104, rs2720298) which is not relevant for our project. Variants rs3741554 and rs3813526 were missense variants, leading to amino acid changes. Variant rs3813526 results in the replacement of serine by proline on position 681 (p.Ser681Pro), and rs3741554 changes a leucine for a methionine on position 326 (p.Leu326Met). One single family (113625) carried those variants but unfortunately the affected children were homozygous for the reference alleles and only the unaffected parents were heterozygous for these variants. Another variant (rs1470907) was located either in an intron or in the 3'UTR sequence depending on the transcript (ENST00000335999 and ENST00000433948 respectively) we analyzed. Its physiological implication could be discussed but only one single family was carrying this variant in the two unaffected parents (heterozygous) and the affected children were homozygous for the reference allele.

The variant found at position chr12:50197819 (C>T) is predicted to be potentially disease causing by Mutation Taster by changing a splice site or by (maybe) modifying protein features although no amino-acid change seemed resulting from that point mutation. Only one family

(460380) carried this variant: the father and 1 affected child were heterozygous for the variant and the mother and the other affected child were homozygous for the reference allele.

The next variant was located on chr12:50189817 (A>T) and is considered by Mutation Taster as a polymorphism (probably harmless) despite a change in the amino-acid sequence (p. Leuc609His). Two families carried this variant: one affected child in family 239152 and the father in family 283703 (both heterozygous for the variant).

The last variant's location was chr12:50186044 (C>A). Mutation Taster predicts that this variant, located within an intron or the 3'UTR sequence depending on the transcript, would be a polymorphism. The 2 parents of only one family (113625) carried this polymorphism in a heterozygous form; the 2 affected children of this family were homozygous for the reference allele.

The two initial variants highlighted in the prime family (chr12:50185732 in exon 12 and chr12:50187215 in exon 10) were not found in any member of the other linked families.

Some DNA samples were not sufficient enough to repeat PCRs till we obtain clear and well interpretable chromatograms.

Table 11: Sequencing of *NCKAP5L* coding-exons in 5 linked families

NCKAP5L Exons	SNVs (ID or position (GRCh37/hg19))	Families																
		113625				239152		283703				460380				480941		
		Mo	AC1	AC2	AC3	AC1	AC2	Fa	Mo	AC1	AC2	Fa	Mo	AC1	AC2	Fa	AC1	AC2
4	chr12:50197819	/	/	/	/	/	/	/	/	/	/	Homo ref G/G	Hetero G/A	Homo ref G/G	Hetero G/A	/	/	/
5(1)-(2)	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
6-7	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
8(1)	/	**?**	**?**	/	/	/	**?**	/	/	/	/	/	/	/	/	/	/	**?**
8(2)	rs3741554	Hetero A/T	Hetero A/T	**?**	Homo ref T/T	/	/	/	/	/	/	/	/	/	/	/	/	/
	rs3741555	Hetero A/G	Hetero A/G	**?**	Homo ref G/G	/	/	/	/	/	/	/	/	/	/	/	/	/
8(3)	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
8(4)	rs2603104	/	/	/	/	/	**?**	/	/	/	/	**?**	**?**	Hetero A/C	Hetero A/C	/	Hetero A/C	**?**
	rs2720298	/	/	/	/	Hetero G/T	/	Hetero G/T	Homo ref G/G	Homo ref G/G	Homo ref G/G	Hetero G/T	Hetero G/T	Homo ref G/G	Hetero G/T	Hetero G/T	Hetero G/T	/
	rs3813526	Hetero T/C	Hetero T/C	Homo ref T/T	Homo ref T/T	/	/	/	/	?	/	/	/	/	/	/	/	**?**
	chr12:50189817	/	/	/	/	Hetero A/T	/	Hetero A/T	/	/	/	/	/	/	/	**?**	**?**	**?**
8(5)-(6)	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
9-10-11	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
12	rs1470907	Hetero C/A	Hetero C/A	Homo ref C/C	Homo ref C/C	/	/	/	/	/	/	/	/	/	/	/	/	/
	chr12:50186044	Hetero A/G	Hetero A/G	Homo ref G/G	Homo ref G/G	/	/	/	/	/	/	/	/	/	/	/	/	/

Fa = father; Mo =mother; Ac = affected child; Hetero =heterozygous; Homo = homozygous; ref = reference allele; / = no variant; ? = no DNA Synonymous variants, missense variants, intron/3'UTR variant

4.2 Second part of the project

4.2.1 *KCNJ11-WFS1-ABCC8*

WES results were analyzed before being filtered to target new candidate genes.

4.2.1.1 We found 2 samples (2 affected children from the same family (437521)) that carried the same variant in the *KCNJ11* gene. The variant is located on chr11:17408960 (c.679G>A) and it results in the replacement of the 227th amino acid namely glutamic acid to lysine (p.E227K). SIFT and PolyPhen2 scores are 1 and 0.998, respectively. Mutation taster confirms PolyPhen2 score predicting that the variant is disease causing (HGMD (Human Gene Mutation Database) CM071808). Moreover, this variant was already described in a family where some cases had non-neonatal onset, consistent with the missed diagnosis in our cases (Abbasi F. *et al*., 2012). We confirmed WES results in sequencing the region surrounding the variation in the 2 affected children and in one unaffected sibling. No DNA from the parents was available. We found that all the three siblings were carrying the same variant in a heterozygous state.

The affected children were diagnosed with diabetes at 9 years and 11 years of age, respectively. The last child, unaffected by the time the DNA was taken, was only 10 years-old.

4.2.1.2 We found 2 samples from 2 different families (152712 and 419900) that carried 2 variants (compound heterozygous) within the *WFS1* gene. The first affected child from family 152712 carried a point variant located on chr4:6303671 (c.2149G>A) leading to the replacement of glutamic acid by lysine on position 717 of the protein (p.E717K) (SIFT = 0.89 and PolyPhen2= 0.705). This mutation is known to be disease causing (CM032681). This child also

carried a frameshift deletion on chr4: 6303727 (C.2207_2207delG) (p.G736fs), also susceptible to cause disease based on Mutation Taster. We sequenced the region surrounding the 2 variants in the affected sibling and we found that he was also heterozygous for both. The DNA of the parents was degraded so we were not able to genotype them for these variants. Moreover, we had no more DNA from the proband to confirm the WES results.

Clinically, the proband was diagnosed with diabetes at 17 years of age and the affected sibling at 20 years of age. No other manifestations of Wolfram syndrome were registered at the time of recruiting (2005).

The other patient from family 419900 was carrying a missense variant located on chr4:6302483 (c.961A>C) resulting in amino acid change (p.T321P) and a frameshift deletion on chr4:6302883 (c.1362-1377delCACGCGCAGGGCCCTG) (p.Y454fs). The missense variant is predicted to be disease causing by Mutation Taster and PolyPhen2 (PolyPhen2= 0.995) while SIFT score predicts that this would probably be benign (SIFT= 1). The deletion region is known to carry many protein-altering point mutations. We genotyped the 2 parents and the affected child and we confirmed that the child was heterozygous for both variations and that the mother was carrying the point mutation and the father the deletion. This patient was diagnosed with diabetes at 10 years of age and recruited in 2007, at which time no syndromic manifestations were recorded.

4.2.1.3 Heterozygous variants within *ABCC8* were found in 2 distinct families (260367 and 456765). The first variant (rs376005080; missense mutation; CM994413) was located on chr11:17483209 (c.742 C>T) and changed the amino-acid sequence (R248Q). We genotyped the proband (45676503) and we confirmed that he was carrying the variation on a heterozygous

state. We also genotyped the affected sibling but we found that he was homozygous for the reference allele.

The second variant (rs117874766; missense mutation; CM092537) was located on chr11:17464808 (c.1384 A>G) and resulted in modification in the amino-acid sequence (I462V). We found that the father and the patient were heterozygous for the variant and that the mother and the affected sibling were homozygous for the reference allele.

4.2.2 Candidate genes

Eight of 12 candidate genes variants were sequenced in the probands to confirm WES results and in their families to evaluate the transmission pattern (see **Table 12**). The first candidate gene we selected was ***CHPF*** (RPKM 89.9; LOD 0.211). The family 462477 was genotyped for the 2 variants located on chr2:220404334 (T>A) and chr2:220404335 (C>T) and we found that the mother, the patient and the affected sibling were heterozygous for both variants while the father was homozygous for both reference alleles. The second gene was ***FRY*** (RPKM 7.66; LOD 0.231). The family 220356 was genotyped for the 2 variants located on chr13:32798380 (A>G) and chr13:32783777 (A>C): the father was homozygous for the reference alleles and the mother, the patient and the affected sibling were heterozygous for both variants, meaning that they were both on the same chromosome. The next candidate gene was ***HSPBAP1*** (RPKM 3.94; LOD 0.231). The family 220356 was genotyped for one variant (nonframeshift insertion) located on chr3:122459292 (TGAG>TGAGAAG) (p.T455delinsTP) showing that the patient and the affected sibling were carrying homozygous insertion while both parents were carrying the

insertion on a heterozygous state. Another family (260367) was genotyped for another variant in the same gene (chr3:122459595 T>G) (p.K355T): the father and the patient were heterozygous for the variant while the mother and the affected sibling were homozygous for the reference allele.

WES results showed possible homozygous variants in 4 genes (***CACNB3, ACIN1, C18orf65*** and ***RBM23***) but Sanger Sequencing did not confirm those results. Indeed, all members of 3 families were homozygous for the reference allele (***ACIN1, C18orf65*** and ***RBM23***) and in the last family, the father was homozygous for the reference allele and the mother, the patient and the affected sibling were heterozygous for the variant (***CACNB3***).

The last candidate gene we tested was ***SCAMP2*** (RPKM 30.97; LOD 0.210). The family 223581 was genotyped for the 2 variants located on chr13:75137840 (C>T) (p.V277M) and chr15:75137428 (T>C) (p.N329S): the father was heterozygous for the first variant, the mother was heterozygous for the second variant and the patient and the affected sibling were compound heterozygous.

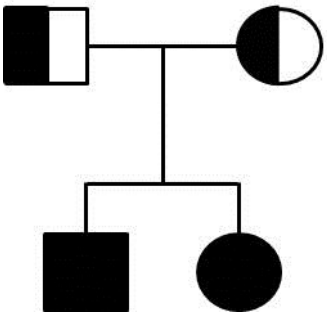
Table 12: Families genotyping for variants found by WES in 8/12 candidate genes.

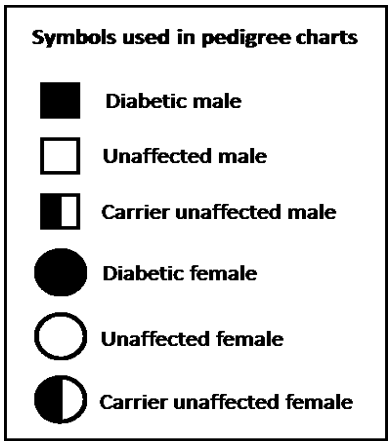
Gene Name (variants position)	**Family ID**	**Father**	**Mother**	**Affected Child 1**	**Affected Child 2**
CHPF (chr2:220404334)	462477	Hom (R)	hetero	hetero	Hetero
CHPF (chr2:220404335)	462477	Hom (R)	hetero	hetero	hetero
FRY (chr13:32798380)	220356	Hom (R)	hetero	hetero	Hetero
FRY (chr13:32783777)	220356	Hom (R)	hetero	hetero	hetero
HSPBAP1 (chr3:122459292)	220356	hetero	hetero	Hom (A)	Hom (A)
HSPBAP1 (chr3:122459595)	260367	hetero	Hom (R)	Hetero	Hom (R)
CACNB3 (chr12:49221416)	223011	Hom (R)	hetero	hetero	hetero
ACIN1 (chr14:23548787)	456765	/	/	Hom (R)	Hom (R)
C18orf65 (CHR18: 74208553)	462477	/	/	Hom (R)	Hom (R)
RBM23 (chr14: 23371267)	462477	Hom (R)	Hom (R)	Hom (R)	Hom (R)
SCAMP2 (chr15: 75137840)	223581	hetero	Hom (R)	hetero	Hetero
SCAMP2 (chr15:75137428)	223581	Hom (R)	hetero	hetero	hetero

Hom = homozygous; R = reference allele; A: alternate allele; hetero = heterozygous; / = no DNA available

Two candidate genes stood out from the crowd: ***HSPBAP1*** and ***SCAMP2***. Indeed, they carried variations that segregated with the disease under a recessive model (homozygous for the mutated/alternate allele in *HSPBAP1* (**Fig.8**) and compound heterozygous for *SCAMP2* (**Fig.9**)).

Figure 8: Genotyping of *HSPBAP1* insertion (chr3: 122459292; TGAG>TGAGAAG) in family 220356. The 2 parents are heterozygous for the insertion and the 2 affected children are homozygous for the insertion. This pedigree was designed according to standard symbols.

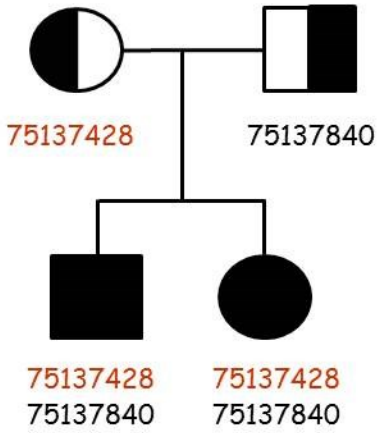


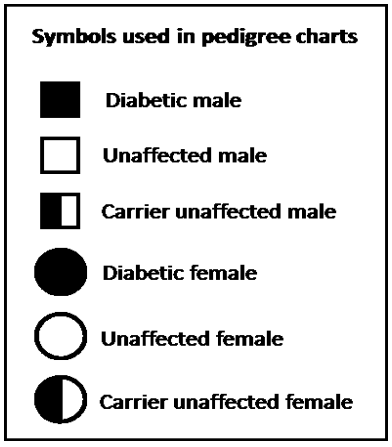


The variant found in *HSPBAP1* (p.T455delinsTP) is an in-frame insertion that changes the amino-

acid sequence while being predicted as a polymorphism by Mutation Taster.

Figure 9: Genotyping of *SCAMP2* SNVs (chr15: 75137840 (C>T) and chr15:75137428 (T>C) in family 223581. The father is heterozygous for the first variant, the mother is heterozygous for the second variant and 2 affected children are compound heterozygous. This pedigree was designed according to standard symbols.





The 2 variants found in *SCAMP2* are missense mutations (p.V277M and p.N329S) that are both predicted to be disease causing in affecting protein features and/or changing splice sites by Mutation Taster while only the second mutation is predicted to be damaging by PolyPhen2 (p. V277M: SIFT= 0.89 and PolyPhen2= 0.023; p.N329S: SIFT= 0.98 and PolyPhen2= 0.953).

In order to validate *HSPBAP1* or *SCAMP2* as potentially linked to a new form of recessive diabetes, we had to find another family with a homozygous or compound heterozygous mutation within one of those genes.

We genotyped linked families for *HSPBAP1* insertion and then, we sequenced all protein-coding exons of *SCAMP2* in affected children from families linked to *SCAMP2* locus.

4.2.2.1 The *HSPBAP1* gene

Eleven families were linked to *HSPBAP1* locus but we only had DNA for 7 of them (including the index family). We genotyped the 6 other families for the insertion described in the prime family (see **Table 13**) but we did not find any results compatible with a recessive mode of inheritance. Indeed, 2 affected children were heterozygous for the variant while all the others were homozygous for the reference allele.

Table 13: Genotyping of 6 families linked to *HSPBAP1* locus

Families ID	Father	Mother	Affected Child 1	Affected child 2	Affected child 3	Unaffected child
417769	/	/	Homo (R)	/	n/a	n/a
460380	Hetero	/	hetero	/	n/a	n/a
437521	/	/	Hom (R)	Hom (R)	n/a	n/a
239152	/	/	Hom (R)	/	n/a	n/a
491015	Hom (R)	hetero	Hom (R)	Hom (R)	n/a	Hom (R)
275392	hetero	Hom (R)	Hom (R)	Hom (R)	hetero	n/a

Hom = homozygous; R = reference allele; A: alternate allele; hetero = heterozygous; / = no DNA available

4.2.2.2 The *SCAMP2* gene

15 affected children (9 from NIDDK repository and 6 from pediatric diabetes clinics across Canada showing HBD) were sequenced for all protein-coding exons of *SCAMP2*. The two variants initially found in family 223581 (within exons 8 and 9) were not found in those 15 children. Moreover, no other protein-altering variant was found in the other exons except for 1 variant in the sequencing product of exon 7. Four patients from NIDKK and 5 patients from diabetes clinics' cohort carried this variant in a heterozygous and homozygous state, respectively. This intron variant is described by Mutation Taster as a polymorphism (rs3765066). It is located 88 bp after the end of the coding part of exon 7.

4.3 Third part of the project

Among the 46 families totally negative for autoimmune markers, the nine "dominant" families represent a 19.5% frequency of affected parents, twice that observed in autoimmune diabetes (10.3%, p=0.00013). This, despite the absence of high-risk HLA, the main driver of familial clustering in autoimmune T1DM. This is compelling evidence of many cases of misdiagnosed MODY.

HNF1-α protein-coding exons were sequenced in families 423172 and 492182.

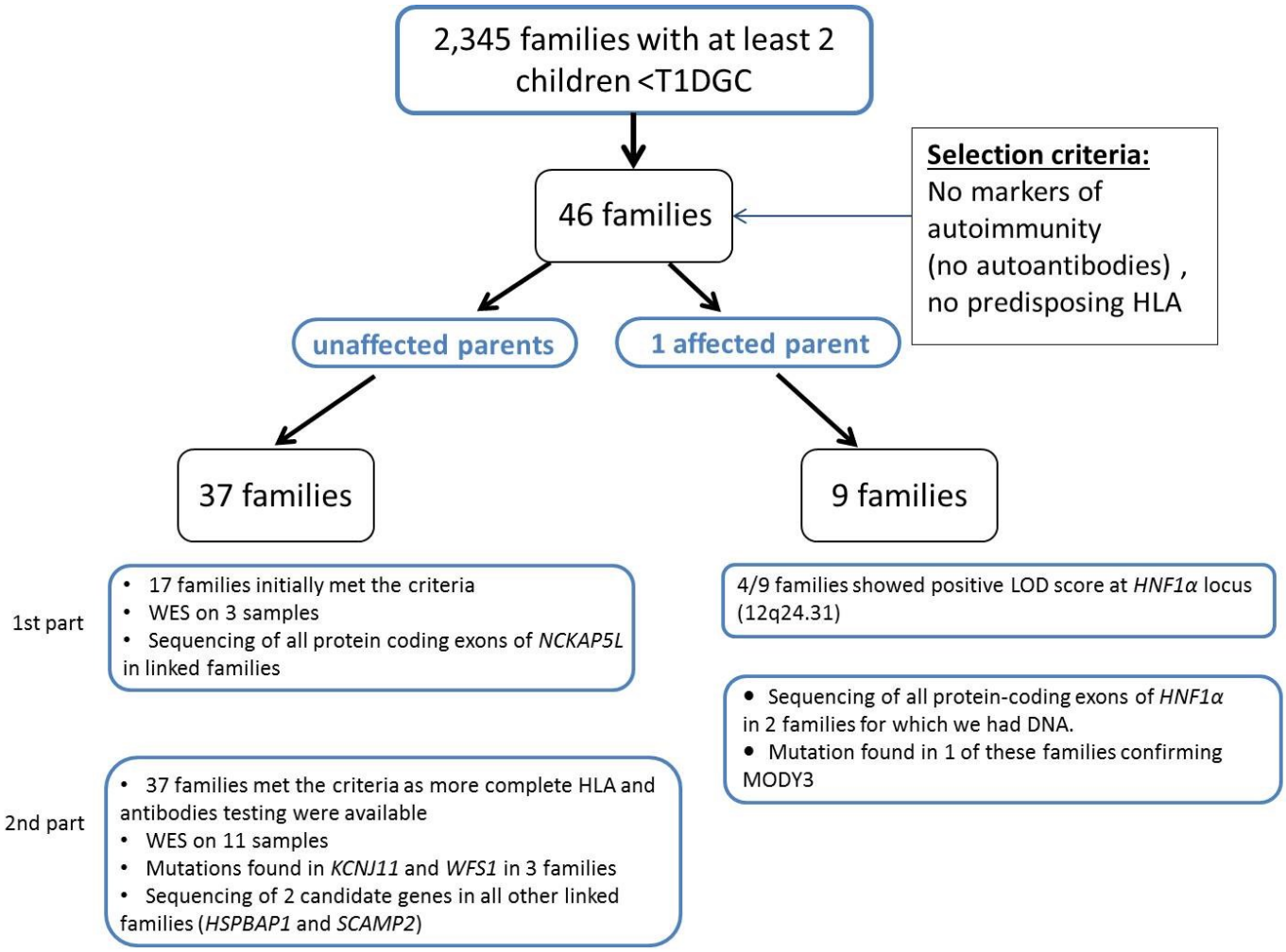
In family 423172, we found c.599 G>A missense mutation (p.R200Q) within exon 3 in the affected mother and in the 3 affected children who were diagnosed with diabetes at 26, 8, 20 and 15 years of age, respectively. This is a known disease-causing mutation for MODY3, a disease treatable with oral agents rather than insulin.

No protein-altering variant was found in the second family.

The different parts of the project (Methods and Results) are summarized in **Figure 10**

Figure 10: Overview of the different parts of the project.

From 2,345 families with at least 2 diabetic children, we selected those without neither markers of autoimmunity nor predisposing HLA. Forty six families fulfilled these criteria. The main project was to detect genes involved in a new type of autosomal recessive diabetes. Therefore, we only selected families in which the parents were unaffected. Initially 17 families met those criteria and WES was performed on 3 samples. *NCKAP5L* was selected because of segregating with the disease under a recessive model in the prime family and all *NCKAP5L* protein-coding exons were sequenced in linked families. Subsequently, more complete HLA and antibodies testing were available and 37 families in total met the selection criteria. WES was performed on 11 samples. Mutations in *KCNJ11* and *WFS1* were found in 3 families. Two candidate genes were selected and sequenced in all linked families. The last part of the project involved families with 1 affected parent looking for mutations in *HNF1α* (MODY3).



5. Discussion

Monogenic diabetes would represent 1 to 5% of all cases of diabetes mellitus (Irgens H.U. *et al*., 2013) even if many of them remain probably undiagnosed or misclassified as T1DM or T2DM (Alkorta-Aranburu G. *et al*. , 2014). Variable awareness of other forms of diabetes and limited access to genetic testing are potential reasons that many MDs go unrecognized (Alkorta-Aranburu G. et al. , 2014). Moreover, all diabetes can successfully be treated with insulin injections which could limit the interest of looking further for a rare diagnosis. Nevertheless, the most frequent MDs (e.g. MODY3 and *KCNJ11* mutations) respond well to oral sulfonylurea (Tuomi T. *et al.,* 2006) which can be life-changing for patients through the comfort of oral medication and avoidance of long-term complications with better glycemic control and minimal risk of hypoglycemia.

The T1DGC database was developed to create a large cohort of T1DM families with all clinical and biological information in order to identify genes that contribute to the risk of T1DM (Rich S.S. *et al.*, 2009). Despite rigorous clinical selection, monogenic diabetes cases were still found in this "T1DM" cohort.

Gain-of-function mutations in *KCNJ11* which encodes Kir6.2 protein of the β-cell KATP channels have been shown to cause neonatal diabetes since 2004 (Gloyn A.L. *et al.,* 2004) while homozygous mutations were first described in babies with hyperinsulinemic hypoglycemia in 1996 (Thomas P.M. *et al*., 1996) (Nestorowicz A. *et al*., 1997). Usually, activating mutations in *KCNJ11* cause permanent neonatal diabetes or more rarely transient neonatal diabetes with relapse later in life (Edghill E.L *et al.,* 2010) (Gloyn A.L. et al., 2004). Kir6.2 mutations associated

with PNDM or TNDM are gain-of-function mutations acting by reducing the ability of ATP to inhibit the KATP channels that stay open. This leads to permanent membrane hyperpolarization which prevents Ca^{2+} influx. A lack of insulin secretion will then result from that overactivation of KATP channels (Girard C.A.J. *et al.,* 2006). As those activating mutations have been known for only 12 years, many patients, presently adolescents or young adults might have been misdiagnosed as T1DM. Some publications describe *KCNJ11* heterozygous mutations in young adults who were misclassified as "T1DM" very early in life, with successful replacement of insulin injections by oral sulfonylureas (Thewjitcharoen Y. *et al.,* 2014). In addition to missed neonatal diabetes, we can also observe that some specific mutations within *KCNJ11* seem to be associated with variable phenotypes. Indeed, E227K mutation, which is the mutation found in family 437521, has been described in patients with TNDM as well as in patients with highly variable age at diabetes onset (Abbasi F. *et al.,* 2012) (Girard C.A.J. *et al.,* 2006). Abbasi *et al.* reported a diabetic family in which the father was diagnosed with "T1DM" at 15 years of age, the first child at 2 years of age and the second child at 40 days of life. They all were treated with insulin injections for many years until a diagnosis of monogenic diabetes (mutation E227K within *KCNJ11*) was made with successful response after they were transferred to sulfonylureas treatment (Abbasi F. *et al*., 2012). Girard *et al*. studied the role of specific Kir6.2 mutations on K^+channel activity. E227K mutation lies distant from the ATP-binding site at the interface between adjacent subunits and it probably alters the channel ATP sensitivity indirectly by increasing the intrinsic open probability of the channel (Girard C.A.J. *et al*., 2006). Some genotype-phenotype correlation exists since mutations induce more severe or precocious disease, with possibly neurological symptoms (DEND syndrome) while others result in TNDM or

late-onset disease (Hattersley A.T *et al.,* 2005). However, the same mutation can also result in neonatal onset, suggesting a possible genetic-background effect, or an environmental effect (Girard C.A.J. *et al.,* 2006). The E227K mutation of Kir6.2 seems to induce diabetes at variable ages, which could explain missed diagnosis in family 437521 in which the first 2 children developed diabetes at 9 and 11 years of age. Moreover, it is not surprising to find the same mutation in the unaffected sibling who was 10 years-old at the time of recruitment. Indeed, given that the second child developed diabetes at 11 years of age and that diabetes-onset can be as late as 15 years of age as mentioned by Abbasi et al., this child probably developed diabetes later in adolescence (Abbasi F. *et al.,* 2012).

We could also note that the 3 children carry the same mutation while usually 80% of *KCNJ11* mutations appear *de novo*. We did not have DNA from the parents but they are listed as unaffected, implying that they could not be carrier of E227K in their somatic cells. Therefore, the mutation was certainly transmitted through germline mosaicism (Gloyn A.L. *et al*., 2004). Germline mosaicism implies that some germ cells carry the mutation while others not. Because the germline mutation is present in the egg or sperm cell, it will be present in all cells of the child developing from that germ cell. If it is an autosomal dominant mutation, the child will then be affected with the disorder. Germline mosaicism can be observed with any inheritance pattern, but it is most commonly seen with autosomal dominant and X-linked disorders that severely compromise survival or reproduction and thus manifest mostly through new mutations. The recurrence risk for a parent with germline mosaicism for an autosomal dominant disorder is hard to predict since it depends on the proportion of germline cells with the mutation. Based on family studies, Edghill *et al*. evaluated the risk for another affected child

as low (around 1%), moderate (as in 6%), or high (30%), depending on both the proportion of mutated germ cells and the disorder (Edghill E.L. *et al.,* 2007).

The right diagnosis is crucial for family 437521 in terms of treatment and genetic counseling.

The second gene involved in monogenic disease that we found in our cohort was *WFS1*. To date, *WFS1* mutations have been implicated in a large spectrum of phenotypes (e.g. Wolfram syndrome, low-frequency non-syndromic hearing loss, or psychiatric diseases), and common variants have also been associated with T2DM (Cryns K. *et al.,* 2003) (Zalloua P.A. *et al.*, 2008) (Chaussenot A. *et al.* 2015). This may explain why some studies found *WFS1* mutations in up to 0.7 to 6% of T1DM patients (Boutzios G. *et al*., 2011) (Blanco-Aguirre M.E. *et al.*, 2015). Some genotype-phenotype correlation also exists between *WFS1* mutations and clinical presentation, with both inter- and intra-familial clinical heterogeneity (Blanco-Aguirre M.E. *et al.,* 2015) (Cryns K. *et al*., 2003). Indeed, Cryns *et al*. described a relationship between mutation types (missense vs nonsense, deletions or insertions) and severity of clinical features (Cryns K. *et al*., 2003). In the same way, Zalloua P. *et al.* showed that some *WFS1* mutations (missense mutations and frameshift deletions) may result in non-syndromic juvenile onset diabetes, especially in some population subgroups with high level of consanguinity (Zalloua P.A. *et al*., 2008). Moreover, Chaussenot *et al.* classified patients with *WFS1* mutations into early-onset Wolfram syndrome and late-onset Wolfram syndrome depending on the time they developed diabetes and optic atrophy, before or after 15 years of age, respectively (Chaussenot A. *et al.* 2015).

Our WFS1 patients developed diabetes without any other symptoms (based on clinical information from T1DGC database), at 17, 20 and 10 years of age. It seemed late and unlikely

for Wolfram syndrome but considering data found in the literature, those patients can have monogenic diabetes resulting from *WFS1* mutations without other symptoms or late-onset Wolfram syndrome. The right diagnosis could also be beneficial for those families for close medical follow-up (screening of other associated symptoms) and for genetic counselling. However, no therapeutic change can be suggested to those patients.

Heterozygous variants within *ABCC8* were found in 2 distinct families (260367 and 456765). In the first family, the proband was heterozygous for c.742 C>T (R248Q) mutation but the affected sibling was homozygous for the reference allele. Tarasov *et al.* (2008) suggested that the R248Q mutation could be associated with adult-onset diabetes. Indeed, they found heterozygous R248Q in late-onset diabetic patients but they struggled to prove its implication in the abnormal insulin secretion. The concentration-inhibition curves for KATP channels carrying SUR1-R248Q were practically identical to the wild type, suggesting either that this mutation affected other properties of the channel or was not responsible for diabetes (Tarasov A.I *et al*., 2008). This mutation could be considered as disease causing in family 260367 if both affected children carried the same mutation. Given that the second affected child did not carry the mutation, the probability that this mutation was responsible for this familial diabetes is null.

In the second family, we found the mutation c.1384 A>G (I462V) that was described by Sandal *et al*. in 2009 in patients with HH (Sandal T. *et al*., 2009). We found that the father (unaffected) and the patient were heterozygous for the variation and that the mother and the affected sibling were homozygous for the reference allele. Given that the unaffected father was carrying the mutation while the affected sibling did not, this mutation cannot be considered as potentially responsible for diabetes in this family.

We then kept these 2 families in our list for candidate gene selection.

After we removed families with *KCNJ11* and *WFS1* diabetes, we were able to apply our selection criteria to find candidate genes for new recessive diabetes.

Some recent publications found new genes involved in early-onset non-autoimmune diabetes. In the vast group of T1DM, T1BDM, which is not from autoimmune origin, still intrigues the scientific community by lack of understanding of its etiology. Simaite *et al.* reported recessive mutations in *PCBD1* as a cause of new type of early-onset diabetes. They studied a consanguineous family with non-autoimmune diabetes and no mutations in the *HNF1-* or *4α*, *ABCC8, KCNJ11* or *INS* genes. They found homozygous mutations in *PBCD1* that is known to play a role in supporting HNF1B action in pancreatic development but was mainly reported in hyperphenylalaninemia. In Simaite *et al*.'s cohort, some patients had transient hyperphenylalaninemia during the neonatal period with or without diabetes. They were able to successfully transfer the patients to sulfonylurea (Simaite D. *et al.*, 2014).

GATA4 mutations can also be a cause of neonatal and childhood-onset diabetes (Shaw-Smith C. *et al*., 2014). Five patients with *GATA4* mutations and variable phenotype of transient or permanent diabetes diagnosed in neonates or during childhood were reported. The exocrine pancreatic phenotype ranged from complete agenesis to hypoplasia with subclinical exocrine insufficiency or normal exocrine function. Additional features included neurocognitive defects and congenital heart malformations. Indeed, *GATA4* is mainly known to be involved in heart development (Tomita-Mitchell A. *et al*., 2007). Because of the exocrine deficiency, this type of diabetes is less likely to be misdiagnosed as T1DM.

In both Simaite *et al.* and Shaw-Smith *et al.*, diabetes was part of other clinical features which highlighted the role of *PBCD1* and *GATA4* in variable cells types.

In our study, we tried to find a new gene only involved in early-onset non-autoimmune diabetes, without additional features that would distinguish it from T1DM. T1DGC inclusion criteria were strict, excluding individuals with other clinical features or medical history. In the entire WES results table, no variation was found in *PBCD1* and a simple heterozygous variation was found in *GATA4* which is not compatible with an autosomal recessive disease.

After having applied our selection criteria to target new candidate genes and genotyped probands and their families for the variations, we found 3/16 genes that were the only genes with variations segregating with the disease under a recessive model in the corresponding family.

NCKAP5L (*NCK (non-catalytic region of tyrosine kinase)- Associated protein 5-like*) is a gene whose function remains unclear. It was found to be differentially expressed in various tissues in obesity and T2DM (Chen J. *et al*., 2013) and upregulated during neuronal depolarization, suggesting that *NCKAP5L* could be involved in neuronal function and/or development and that mutations could lead to autism spectrum disorders (Chahrour M.H. *et al*., 2012). In our cohort, family 417769 carried chr12:50185732 and chr12: 50187215 missense variants. The 2 affected children were compound heterozygous while the unaffected siblings and the parents were heterozygous for one mutation. This pedigree was highly suggestive for an autosomal recessive transmission pattern but in order to confirm the potential role of *NCKAP5L* in diabetes, it was crucial to find another family with mutations in the same gene. All

families with a positive LOD at this locus were sequenced. Many polymorphisms and variants were discovered but none of them segregated with the disease under a recessive model.

Tyrosine kinases are enzymes playing an important role in the modulation of growth factor signaling. They represent a target for anti-cancer therapy (tyrosine kinase inhibitor; TKI). Side effects of TKI include hypoglycemia or hyperglycemia but the molecular mechanism by which TKI controls glucose homeostasis remains unknown (Vergès B. *et al.,* 2014). *NCKAP5L* encodes a tyrosine kinase associated protein that could therefore potentially play an indirect role in glucose metabolism. Given the potential role of *NCKAP5L* in glucose metabolism and the interesting pedigree of family 417769, we must keep this gene in our candidates list even if we were not able to confirm our results in a second family.

HSPBAP1 (= *heat-shock protein-associated protein 1*) encodes a protein that binds to one of the small heat shock proteins (hsp27). The common functions of small heat shock proteins are chaperone activity, thermo-tolerance, inhibition of apoptosis, regulation of cell development, and cell differentiation. HSPBAP1 is widely expressed and was found to be abnormally expressed in brain tissue (surgical resection of epileptogenic zone in the anterior temporal neocortex) of patients with intractable epilepsy, although how brain function is affected remains unknown (Xi Z-Q *et al.,* 2007). The exact function of HSPBAP1 remains presently unclear. Heat shock proteins have already been described in T1DM as marker of β-cell stress that contributes to T1DM pathogenesis and progression (Watkins R.A. *et al.,* 2016).

In our cohort, 2 affected siblings were homozygous for the in-frame insertion p.T455delinsTP while both parents were carrying this insertion on a heterozygous state. This pedigree was

suggestive of an autosomal mode of inheritance but we did not confirm a causal role for *HSPBAP1,* as we did not find another family carrying any mutation in the same gene.

SCAMP2 (=*secretory carrier membrane protein 2*) encodes a protein which functions as carrier to the cell surface in post-Golgi recycling pathways and that is associated with the pool of Glut 4 transporters in the cell. Park *et al.* studied the role of *SCAMP2* in insulinoma cells, but they did not find any changes of knockdown of Scamp2 on insulin secretion (basal or stimulated) (Park M.J. *et al*., 2009).

This gene was an interesting candidate gene for our study due to its high RPKM in pancreas (RPKM= 30.97). In one family, missense variants (p.V277M and p.N329S) segregate with the disease under a recessive model. Fifteen affected children linked to this locus were then sequenced for all protein-coding exons of *SCAMP2* but no potential disease-causing variations were found. Indeed, the only variant found in that group of 15 affected children was an intron variant, described as a polymorphism, unlikely to be functional.

Even if we did not find any other family that carried variants segregating under a recessive model in one of those three genes, we cannot reject *NCKAP5L*, *HSPBAP1* and *SCAMP2*. Indeed, we are looking for a very rare disease and although we selected families based on strict criteria (no autoantibodies, no predisposing HLA), some of them may still have autoimmune diabetes.

Our method appeared really promising as we worked on a large group of non-autoimmune diabetic families for which another diagnosis than T1DM was very likely.

When we sent our second batch of samples for WES, we tried to apply some selection criteria to keep a dozen of samples. We hoped to optimize our results in selecting more specifically the

samples. Therefore, we performed heterogeneity-insensitive linkage analysis and we selected 11 samples among those showing positive linkage to two loci on chromosomes 6 and 11. Ultimately, the candidate genes we selected from WES results were not preferentially located on those two chromosomes. This method did not actually prove of any interest in finding better candidate genes but did not nevertheless prevent finding interesting results. Indeed, we found *KCNJ11* (located on chromosome 11) and *WFS1* monogenic diabetes in this cohort.

We did not design the study to find known monogenic diabetes in our cohort with unaffected parents, but these unplanned positive controls provide strong support for the validity of our screening approach for the discovery of unknown ones.

SIFT and PolyPhen2 score are used to predict the impact of a nonsynonymous SNP, which results in an amino acid substitution, on the activity of proteins. Mutation Taster also provides information about the disease causing potential of variations. SIFT and PolyPhen2 scores often provided conflicting results, underscoring their limitations. Johnson M. *et al.* studied SIFT and PolyPhen score concordance in genes involved in steroid hormone metabolism and response (Johnson M.M. *et al*., 2005). They explained that those scores may differ because PolyPhen, in addition to using sequence information, incorporates structural information of the protein. Mutation Taster mostly gave similar results as PolyPhen2 so we used their prediction to classify our variants as potentially disease causing or not. However, these computational predictions should be used only for prioritizing, not for excluding or confirming a causal role.

Given that, despite strict selection criteria to create “T1DM” cohort by T1DGC, we found 3 families with known monogenic diabetes in our recessive cohort, we were convinced to find *HNF-1α* mutations (MODY3) in families with an affected parent. Indeed, 19.6% (9/46) of non-autoimmune families from T1DGC cohort have an affected parent. This is almost twice the percentage of affected parents that we observe in the remaining cohort (10.3%, p=0.00013) despite the absence of high-risk HLA, the main driver of familial clustering in autoimmune T1DM. This strongly suggests misclassified MODYs.

4/9 families that showed positive linkage to *HNF-1α* locus were prioritized but all those 9 families must be sequenced, as we know that linkage analysis is less powerful in dominant model as only one allele is affected.

We found *HNF-1α* R200Q mutation in all affected members of family 423172. This mutation has been known to be involved in MODY3 for almost 20 years (Hattersley A.T., 1998). In family 423172, diabetes-onset matched perfectly with MODY3 (26-20-15 years of age) except for one child who developed diabetes at 8 years of age which is precocious for MODY diseases (Fajans S.S. *et al.* 2001). However, Tatsi *et al.* described this R200Q mutation in a child diagnosed at 4 years of life (mother’s diabetes-onset = 17 years of age and grand-mother’s =31 years of age) (Tatsi C. *et al*., 2013). He was diagnosed based on the family history rather than on clinical presentation. By definition, MODYs start before the age of 25 years, in adolescents and young adults but facing strong family history, we should question diagnosis, even in a young child.

In conclusion, different cases of monogenic diabetes are still misdiagnosed as T1DM. Usually, clinical and/or family history should help us making the right diagnosis. However, some mutations appear to alter to a lesser extent the protein function, which can influence clinical evolution of the disease and explain misclassification.

In the same way, we are convinced that a new form of autosomal recessive diabetes exists and is also misdiagnosed as T1DM. The interesting cohort of families with no marker of autoimmunity nor predisposing HLA, have allowed us to find highly likely candidate genes that still remain to be confirmed in at least 1 additional family.

Our results should convince clinicians to always question T1DM or T2DM diagnosis if clinical or family history (i.e. dominant pedigree) suggests an alternative diagnosis. Monogenic diabetes can often be treated with oral sulfonylurea and a right diagnosis could lead to an appropriate genetic counseling and follow-up.

6. Future directions

We ordered new DNA from all 46 non-autoimmune families (37 recessive and 9 dominant families) in the NIDDK repository.

Some missing results can be confirmed such as the parents' genotyping in the *KCNJ11* family and the genotyping of the parents and the proband in the first *WFS1* family. Even if these results are not expecting to change the clinical history of these families, they are important to validate and conclude our findings. Moreover, some candidate genes (*ANKRD11, ASXL3, KIAA2018* and *PRR12*) selected after the second WES were not genotyped because of lack of DNA. Given that those genes are as interesting (RPKM and LOD scores) as the ones we already genotyped and in which we found some interesting candidates, we will prioritize those genes when we will get the new DNAs. We also plan to send 10 more samples for WES in order to increase the size of our cohort and select new candidate genes.

We really hope finding candidate genes with variants segregating with the disease under a recessive model in more than 1 family. If causal mutations are identified, we will sequence the gene involved in HLA-neg and Antibody-neg patients from our 1,000 Canadian T1DM case collection, starting with cases that show homozygosity-by-descent at the candidate locus. If homozygotes are identified we will test the effect of oral antidiabetics on their C-peptide production. To understand the functional importance of these genes in the beta-cell, with possible relevance in the search for novel therapeutics for T2DM, we will also collaborate with Dr. Guy Rutter, Imperial College, London, who has expressed a strong interest in developing *in vitro* and animal models to this effect.

Finally, all the 9 dominant families will also be sequenced to exclude MODY3 as we are convinced that more than 1 family should be affected with *HNF-1α* diabetes in that cohort. Indeed, we found almost twice the percentage of affected parents in the autoimmune negative cohort that we observed in the remaining cohort, which strongly suggests misclassified MODYs.

7. References

- Abbasi F., Saba S., Ebrahim-Habibi A. *et al.* "Detection of KCNJ11 Gene Mutations in a Family with Neonatal Diabetes Mellitus. Implications for Therapeutic Management of Family Members with Long-Standing Disease". *Mol Diagn Ther* 2012; 16 (2): 109-114

- Alkanani A.K., Hara N. , Gottlieb P.A. *et al.* « Alterations in Intestinal Microbiota Correlate With Susceptibility to Type 1 Diabetes » *Diabetes* 2015; 64: 3510–3520.

- Alkorta-Aranburu G., Carmody D., Cheng Y.W. *et al.* « Phenotypic heterogeneity in monogenic diabetes: The clinical and diagnostic utility of a gene panel-based next-generation sequencing approach » *Mol Genet Metab*. 2014; 113(4): 315–320.

- Babenko A.P., Polak M., Cavé H. *et al.* "Activating Mutations in the ABCC8 Gene in Neonatal Diabetes Mellitus" *N Engl J Med.* 2006;355(5):456-66

- Barrio R. "Cystic fibrosis-related diabetes: novel pathogenic insights opening new therapeutic avenues" *European Journal of Endocrinology*, 2015; 172, R131–R141

- Ben Abdelaziz R., Ben Chehida A., Azzouz H., *et al.* "A novel homozygous missense mutation in the insulin receptor gene results in an atypical presentation of Rabson-Mendenhall syndrome" *European Journal of Medical Genetics,* 2016; 59: 16e19

- Bergamin C.S. and Dib S.A. "Enterovirus and type 1 diabetes: What is the matter?" *World J Diabetes* 2015; 6(6): 828-839.

- Bin Dhuban K. and Piccirillo C.A. "The immunological and genetic basis of immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome" *Curr Opin Allergy Clin Immunol* 2015, 15:525–532

- Blanco-Aguirre M.E., Rivera-De la Parra D., Tapia-Garcia H. *et al.* "Identification of unsuspected Wolfram syndrome cases through clinical assessment and WFS1 gene screening in type 1 diabetes mellitus patients" *Gene* 2015; 566: 63–67

- Bluestone J.A., Herold K. and Eisenbarth G. "Genetics, pathogenesis and clinical interventions in type 1 diabetes." *Nature.* 2010; 464(7293):1293-300.

- Bonnefond A., Philippe J., Durand E. *et al.* "Whole-Exome Sequencing and High Throughput Genotyping Identified KCNJ11 as the Thirteenth MODY Gene" *PLoS One.* 2012;7(6):e37423.

- Borecki I.B., and Province M.A. "Linkage and association: basic concepts." *Adv Genet*. 2008;60:51-74.

- Boutzios G., Livadas S., Marinakis E., *et al.* "Endocrine and metabolic aspects of the Wolfram

syndrome" *Endocrine* 2011; 40:10–13

- Bowman P., Flanagan S. E., Edghill E. L. *et al.* "Heterozygous ABCC8 mutations are a cause of MODY" *Diabetologia,* 2012; 55:123–127

- Chahrour M.H., Yu T.W., Lim E.T. *et al.* "Whole-Exome Sequencing and Homozygosity Analysis Implicate Depolarization-Regulated Neuronal Genes in Autism" *PLoS Genet.* 2012;8(4):e1002635.

- Chaussenot A., Rouzier C., Quere M., *et al.* "Mutation update and uncommon phenotypes in a French cohort of 96 patients with WFS1-related disorders" *Clin Genet* 2015; 87: 430–439

- Chen J., Meng Y., Zhou J. *et al.* "Identifying Candidate Genes for Type 2 Diabetes Mellitus and Obesity through Gene Expression Profiling in Multiple Tissues or Cells" *J Diabetes Res.* 2013;2013:970435.

- Cryns K., Sivakumaran T.A., Van den Ouweland J. M.W.*et al*. "Mutational Spectrum of the WFS1 Gene in Wolfram Syndrome, Nonsyndromic Hearing Impairment, Diabetes Mellitus, and Psychiatric Disease" *Hum Mutat* 2003;22:275–287

- D'Adamo E. and Caprio S. "Type 2 diabetes in youth: epidemiology and pathophysiology." Diabetes Care, 2011;34 Suppl 2:S161-5.

- de Kerdanet M., Caron-Debarle M., Nivot S. *et al.* "Ten-year improvement of insulin resistance and growth with recombinant human insulin-like growth factor 1 in a patient with insulin receptor mutations resulting in leprechaunism." *Diabetes & Metabolism*, 2015; 41 331–337

- Edghill E.L., Gloyn A.L., Goriely A*. et al.* "Origin of de Novo KCNJ11 Mutations and Risk of Neonatal Diabetes for Subsequent Siblings" *J Clin Endocrinol Metab* 2007;92: 1773–1777

- Edghill E.L., Flanagan S.E. and Ellard S. "Permanent neonatal diabetes due to activating mutations in ABCC8 and KCNJ11" *Rev Endocr Metab Disord*, 2010; 11:193–198

- Eide S.A., Ræder H., Johansson S. *et al.* "Prevalence of HNF1A (MODY3) mutations in a Norwegian population (the HUNT2 Study)" *Diabet. Med.* 25, 775–781 (2008)

- Eizirik D.L, Sammeth M., Bouckenooghe T. "The Human Pancreatic Islet Transcriptome: Expression of Candidate Genes for Type 1 Diabetes and the Impact of Pro-Inflammatory Cytokines" *PLoS Genet.* 2012;8(3):e1002552.

- Emerson S."ALFRED HENRY STURTEVANT (November 21, 1891-April6, 1970)" *Annu Rev Genet.* 1971; 5:21-4

- Fabris M., Zago S., Liguori M. et al. "Anti-zinc transporter protein 8 autoantibodies significantly improve the diagnostic approach to type 1 diabetes: an Italian multicentre study on paediatric patients" Autoimmun Highlights, 2015; 6:17–22.

- Fajans S.S., Bell G.I., Polonski K.S.. "Molecular mechanisms and clinical pathophysiology of Maturity-Onset Diabetes of the Young*" N Engl J Med* 2001; Vol. 345, No. 13:971-80

- Fathallah N., Slim R., Larif S*. et al.* "Drug-Induced Hyperglycaemia and Diabetes" *Drug Saf* 2015; 38:1153–1168

- Flanagan S.E., Patch A.M., Ellard S. "Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations." *Genet Test Mol Biomarkers.* 2010;14(4):533-7.

- Flanagan S.E., Kapoor R.R., Hussain K. "Genetics of congenital hyperinsulinemic hypoglycemia" *Seminars in Pediatric Surgery,* 2011; 20: 13-17

- Girard C. A. J., Shimomura K., Proks P. *et al.* " Functional analysis of six Kir6.2 (KCNJ11) mutations causing neonatal diabetes" *Pflugers Arch - Eur J Physiol*, 2006; 453:323–332

- Gloyn A.L., Pearson E.R., Antcliff J.F. *et al.* "Activating Mutations in the Gene Encoding the ATP-Sensitive Potassium-Channel Subunit Kir6.2 and Permanent Neonatal Diabetes" *N Engl J Med.*2004; 350, 18: 1838-49

- Grant A.V., Boisson-Dupuis S., Herquelot E. *et al.* "Accounting for genetic heterogeneity in homozygosity mapping: Application to Mendelian susceptibility to mycobacterial disease" *J Med Genet.* 2011August ; 48(8): 567–571

- Hara K., Fujita H., Johnson T.A. et al. "Genome-wide association study identifies three novel loci for type 2 diabetes" Human Molecular Genetics, 2014; 23: 239-46.

- Hattersley A.T. "Maturity-onset Diabetes of the Young: Clinical Heterogeneity Explained by Genetic Heterogeneity" *Diabet. Med*. 1998; 15: 15–24

- Hattersley A.T. and Ashcroft F.M. "Activating Mutations in Kir6.2 and Neonatal Diabetes New Clinical Syndromes, New Scientific Insights, and New Therapy" *Diabetes* 2005; 54:2503–2513

- Hildebrandt F., Heeringa S.F., Franz Ruschendorf F. *et al.* "A Systematic Approach to Mapping Recessive Disease Genes in Individuals from Outbred Populations" *PLoS Genet*. 2009;5(1):e1000353.

- IDF atlas-Sixth Edition, 2013: http://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf

- Irgens H. U., Molnes J., Johansson B. B. *et al.* « Prevalence of monogenic diabetes in the population-based Norwegian Childhood Diabetes Registry » *Diabetologia,* 2013 ; 56:1512–1519

- Johnson M.M., Houck J. and Chen C. “Screening for Deleterious Nonsynonymous Single-Nucleotide Polymorphisms in Genes Involved in Steroid Hormone Metabolism and Response” *Cancer Epidemiol Biomarkers Prev* 2005;14(5):1326–9

- Julier C. and Nicolino M. “Wolcott-Rallison syndrome” *Orphanet Journal of Rare Diseases* 2010, 5:29

- Kampmann U., Ring Madsen L., Oeskov Skajaa G., *et al.* “Gestational diabetes: A clinical update” *World J Diabetes* 2015; 6(8): 1065-1072

- Lalitha K. , Jalali S. , Kadakia T. et al. “Screening for homozygosity by descent in families with autosomal recessive retinitis pigmentosa” *Journal of Genetics,* 2002; Vol. 81, No. 2: 59-63

- Lamb M.M., Miller M., Seifert J.A., *et al.* “The effect of childhood cow’s milk intake and HLA-DR genotype on risk of islet autoimmunity and type 1 diabetes: The Diabetes Autoimmunity Study in the Young”. *Pediatric Diabetes.* 2015; 16: 31–38.

- Liu X.X., Zhu X.M., Miao Q*., et al.* “Hyperglycemia induced by glucocorticoids in nondiabetic patients: a meta-analysis.” *Ann Nutr Metab.* 2014; 65(4):324-32

- Luna B. and Feinglos M.N. “Drug-induced hyperglycemia” *JAMA*, 2001; 286: 1945-48

- Malka D., Hammel P., Sauvanet A. *et al.* “Risk Factors for Diabetes Mellitus in Chronic Pancreatitis” *Gastroenterology*, 2000; 119: 1324–1332

- Mitchell T.C. and McClain D.A. “Diabetes and Hemochromatosis” *Curr Diab Rep,* 2014 14:488

- Morgan T. H. “Random Segregation Versus Coupling in Mendelian Inheritance” *Science*, New Series, 1911; Vol. 34, No. 873, p. 384

- Moritani M., Yokota I., Tsubouchi K., *et al.* « Identification of INS and KCNJ11 gene mutations in type 1B diabetes in Japanese children with onset of diabetes before 5 yr of age » *Pediatric Diabetes* 2013; 14: 112–120

- Mortazavi A., Williams B.A., McCue K. *et al*. “Mapping and quantifying mammalian transcriptomes by RNA-Seq.” *Nat Methods.* 2008;5(7):621-8.

- Murphy R., Turnbull D. M., Walker M. *et al.* “Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A>G mitochondrial point mutation” *Diabet. Med.,* 2008; 25: 383–399

- Nakhla M. and Polychronakos C. «Monogenic and Other Unusual Causes of Diabetes Mellitus.» *Pediatr Clin N Am*, 2005 ;52 : 1637–1650.

- Naylor R. and Philipson L. H. “Who should have genetic testing for maturity-onset diabetes of the young?” *Clinical Endocrinology*, 2011; 75: 422–426

- Nestorowicz A., Inagaki N., Gonoi T. *et al*. "A Nonsense Mutation in the Inward Rectifier Potassium Channel Gene, Kir6.2, Is Associated With Familial Hyperinsulinism" *Diabetes* 1997; 46:1743-1748

- Norris J.M., Barriga K., Klingensmith G., *et al.* "Timing of Initial Cereal Exposure in Infancy and Risk of Islet Autoimmunity" *JAMA,* 2003; 290:1713-20.

- Oral E.A. "Lipoatrophic Diabetes and Other Related Syndromes" *Reviews in Endocrine & Metabolic* Disorders 2003; 4: 61–77

- Park M.J., Farrell J., Lemmon K. *et al*. "Enterostatin alters protein trafficking to inhibit insulin secretion in Beta-TC6 cells" *Peptides* 2009; 30(10): 1866–1873.

- Patch A. M., Flanagan S. E., Boustred C. *et al.* "Mutations in the ABCC8 gene encoding the SUR1 subunit of the KATP channel cause transient neonatal diabetes, permanent neonatal diabetes or permanent diabetes diagnosed outside the neonatal period" *Diabetes, Obesity and Metabolism*, 2007; 9 (Suppl. 2), 28–39

- Polychronakos C. and Li Q. "Understanding type 1 diabetes through genetics: advances and prospects." *Nat Rev Genet.* 2011;12(11):781-92

- Pulst S.M. "Genetic linkage analysis." *Arch Neurol*. 1999 ;56(6):667-72.

- Resmini E., Minuto F., Colao A *et al.* "Secondary diabetes associated with principal endocrinopathies: the impact of new treatment modalities" *Acta Diabetol*, 2009; 46:85–95

- Rich S.S., Akolkar B., Concannon P. *et al*. « Overview of the Type I Diabetes Genetics Consortium" *Genes Immun.* 2009 ; 10(Suppl 1): S1–S4.

- Rigoli L and Di Bella C. "Wolfram syndrome 1 and Wolfram syndrome 2" *Curr Opin Pediatr* 2012, 24:512–517

- Sandal T., Laborie L.B., Brusgaard K., *et al.* "The spectrum of ABCC8 mutations in Norwegian patients with congenital hyperinsulinism of infancy" *Clin Genet* 2009: 75: 440–448

- Sang Y., Yang W. Yan J, *et al.* « KCNJ11 gene mutation analysis on nine Chinese patients with type 1B diabetes diagnosed before 3 years of age. » *J Pediatr Endocr Met* 2014; 27(5-6): 519–523.

- Sanger F., Nicklen S., Coulson AR. "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci USA*. 1977 Dec;74(12):5463-7

- Shaw-Smith C., De Franco E., Lango Allen H. *et al*."GATA4 Mutations Are a Cause of Neonatal and Childhood-Onset Diabetes" *Diabetes* 2014;63:2888–2894

- Shields B. M., Hicks S., Shepherd M. H. *et al.* "Maturity-onset diabetes of the young (MODY):

how many cases are we missing?" *Diabetologia*, 2010; 53:2504–2508

- Simaite D., Kofent J., Gong M. *et al.* "Recessive Mutations in PCBD1 Cause a New Type of Early-Onset Diabetes" *Diabetes* 2014;63:3557–3564

- Sims D., Sudbery I., Ilott N.E., *et al.* "Sequencing depth and coverage: key considerations in genomic analyses" *Nat Rev Genet*. 2014;15:121-32

- Skyler J.S. "Primary and secondary prevention of Type 1 diabetes" *Diabet. Med.* 2013; 30: 161–169.

- Slingerland A. S., Shields B. M., Flanagan S. E. *et al.* "Referral rates for diagnostic testing support an incidence of permanent neonatal diabetes in three European countries of at least 1 in 260,000 live births" *Diabetologia*, 2009; 52:1683–1685

- Steck A. S., Vehik K., Bonifacio E. *et al*., « Predictors of Progression From the Appearance of Islet Autoantibodies to Early Childhood Diabetes: The Environmental Determinants of Diabetes in the Young (TEDDY). » *Diabetes Care* 2015;38:808–813.

- Tarasov A.I., Nicolson T.J., Riveline J-P., *et al.* "A Rare Mutation in ABCC8/SUR1 Leading to Altered ATP-Sensitive K Channel Activity and -Cell Glucose Sensing Is Associated With Type 2 Diabetes in Adults" *Diabetes* 2008; 57:1595–1604

- Taplin C.E. and Barker J.M . "Autoantibodies in type 1 diabetes" Autoimmunity, 2008; 41:1: 11-18.

- Tatsi C., Kanaka-Gantenbein C., Vazeou-Gerassimidi A. *et al.* "The spectrum of HNF1A gene mutations in Greek patients with MODY3: relative frequency and identification of seven novel germline mutations" *Pediatric Diabetes* 2013: 14: 526–534

- Thewjitcharoen Y., Wanothayaroj E, Himathongkam T. *et al.* "Permanent neonatal diabetes misdiagnosed as type 1 diabetes in a 28-year-old female: A life-changing diagnosis" *Diabetes Res Clin Pract*. 2014;106(2):e22-4

- The EURODIAB Substudy 2 Study Group "Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus." *Diabetologia.* 1999;42(1):51-4.

-Thomas P.M., Cote G.J., Wohilk N. et al. "Mutations in the Sulfonylurea Receptor Gene in Familial Persistent Hyperinsulinemic Hypoglycemia of Infancy" *Science* 1995; 268: 426-9

- Thomas P.M., Ye Y. and Lightner E. "Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy." *Human Molecular Genetics*, 1996; Vol. 5, No. 11: 1809–1812

- Tomita-Mitchell A., Maslen C.L., Morris C. D., *et al.* "GATA4 sequence variants in patients with congenital heart disease" *Diabetes* 2014;63:2888–2894

- Tuomi T., Honkanen E.H., Isomaa B., *et al.* "Improved prandial glucose control with lower risk of hypoglycemia with nateglinide than with glibenclamide in patients with maturity-onset diabetes of the young type 3." *Diabetes Care.* 2006; 29:189-94.

- van Dijk E.L., Auger H., Jaszczyszyn Y. *et al*. "Ten years of next-generation sequencing technology" *Trends in Genetics*, 2014; Vol. 30, No. 9: 418-26

- Vergès B., Walter T. and Cariou B. "Effects of anti-cancer targeted therapies on lipid and glucose metabolism" *Eur J Endocrinol.* 2014; 170: R43–R55

- Voormolen D.N., Abell S.K., James R., *et al.* "Diagnostic Criteria and Treatment for Gestational Diabetes Mellitus" *Semin Reprod Med.* 2016;34(2):102-9

- Watkins R.A, Evans-Molina C., Terrell J.K. *et al.* "Proinsulin and heat shock protein 90 as biomarkers of beta-cell stress in the early period after onset of type 1 diabetes" *Translational Research* 2016;168:96–106

- Writing Group for the SEARCH for Diabetes in Youth Study Group, Dabelea D., Bell R.A. et al. "Incidence of Diabetes in Youth in the United States." JAMA, 2007; 297(24):2716-24.

- Xi Z-Q., Sun J-J.,Wang X-F., et al. "HSPBAP1 is Found Extensively in the Anterior Temporal Neocortex of Patients With Intractable Epilepsy" *SYNAPSE* 2007; 61:741–747

- Yousefzadeh G., Gozashti M., Najafipour H. *et al.* « Common autoimmune biomarkers, thyroid hormonal abnormalities, and beta cells dysfunction in patients with latent autoimmune diabetes in adults with type II diabetes mellitus. » Diab Met Syndr: Clin Res Rev (2016), http://dx.doi.org/10.1016/j.dsx.2015.09.013 (in press)

- Zalloua P.A., Azar S.T., Delépine M. *et al.* "WFS1 mutations are frequent monogenic causes of juvenile-onset diabetes mellitus in Lebanon" *Human Molecular Genetics*, 2008; Vol. 17, No. 24: 4012–4021

- Zemunik T. and Boraska V. *"*Type 1 Diabetes - Pathogenesis, Genetics and Immunotherapy"*, book edited by David Wagner*, ISBN 978-953-307-362-0, Published: November 25, 2011 "Genetics of type 1 diabetes"