University Campus Bio-Medico, Rome

International PhD in Endocrinology and Metabolic Disease

A multilevel approach to the complexity of etiopathogenesis of Type 1 Diabetes: from genetic risk to the loss of beta-cell communication

Rosa Alba Valentina Portuesi

Supervisor

Prof. Paolo Pozzilli

Co-Supervisor

Prof. Simonetta Filippi

01 February 2013

TABLE OF CONTENTS

ABSTRACT 8

Acknowledgements 10

I PART. INTRODUCTION TO TYPE 1 DIABETES 13

CHAPTER 1. DIABETES MELLITUS 13

1.1 Definition of Diabetes 13

1.2 Classification 14

1.2.1 Type 1, type 2 and gestational diabetes mellitus 14

1.3 Diagnostic Criteria 16

1.4 Epidemiology of diabetes mellitus 17

1.5 Incidence and risk factors for Type 1 Diabetes 21

1.5.1 Susceptibility genes. 22

1.5.2 Environmental risk factors 26

1.6 Pathogenesis of Type 1 Diabetes 28

1.7 Morbidity and mortality 30

II PART. COMPUTATIONAL BIOMEDICINE: NEW MULTILEVEL APPROACH TO THE MULTIFACTORIALITY AND COMPLEXITY OF TYPE 1 DIABETES 41

CHAPTER 2. "A NETWORK-BASED APPROACH" 45

2.1 The reasons for using a network-based approach 45

2.2 An interactive multilevel network organization for a complex disease: a new approach for the "Etiology question" 47

2.3 Bayes' Theorem and Bayesian Networks 51

2.3.1 Bayes' Theorem 52

2.3.2 From Bayes' Theorem to Bayesian Networks 56

2.4 "How should we approach complexity and multifactoriality? ": the role of Conditional Independence 63

2.5 Updating our beliefs by introducing new knowledge 64

2.6 "How can we quantify the uncertainty? ": the relevance of information status 67

2.7 Strengths and limitations of Bayesian Networks. 69

REFERENCES 72

CHAPTER 3. BAYESIAN NETWORK IN GENETIC ANALYSIS FOR RISK ASSESSMENT IN T1D 77

3.1 Introduction 77

3.2 Data sets used for the study 78

3.3 First aim: Assessment of the risk conferred by HLA-DR, INS-VNTR and PTPN22 genes on the onset of T1D and evaluation of the joint risk

conferred by the three susceptibility loci using the Bayesian network approach.....81

Rationale.....81

3.3.1 Methods.....83

3.3.2 Results: Case-control study (Data set: A).....88

3.3.3 Conclusions.....107

3.3.4 Results: Family study (Data sets D).....110

3.3.5 Conclusion.....130

REFERENCES.....132

3.4 Second aim. Do susceptibility genes have any influence on age of T1D onset?.....137

Rationale.....137

3.4.1 Materials and Methods.....137

3.4.2 Results.....140

Case-control study.....161

Family study.....162

3.4.3 Conclusion.....165

REFERENCES.....168

3.5 Third aim. Do HLA-DR, INS-VNTR and PTPN22 interact with each other?.....170

Rationale.....170

3.5.1 Materials and methods.....171

3.5.2 Results 172

3.5.3 Conclusion 175

REFERENCES.....178

CHAPTER 4. INTRODUCTION TO SYSTEM BIOLOGY.....181

4.1 From Bayesian Network to the Systems Biology**Error! Bookmark not defined.**

4.2 The Systems Biology: an answer to the complexity of biological systems 181

4.2.1 Structure of Pancreas and the Islet of Langerhans.....185

4.2.2 Islet structure in NOD mouse 187

4.4 Role of beta-cells coupling.....193

4.5 From experimental data to mathematical models to explain the pancreatic beta-cell electrical behavior 198

4.6 From a Deterministic to a Stochastic model 199

REFERENCES.....205

CHAPTER 5. A STOCHASTIC MATHEMATICAL MODEL TO STUDY THE AUTOIMMUNE PROGRESSION TOWARDS TYPE 1 DIABETES.....212

5.1 Introduction 212

5.2 Material and Methods 216

5.3 Results 220

5.4 Conclusion 224

REFERENCES 228

Appendix 239

CHAPTER 6. FINAL REMARKS 242

CHAPTER 7. METHODOLOGICAL APPENDIX 247

7.1 Predictive methods for multifactorial diseases: the different statistical approaches 247

7.2 Updating 248

7.3 Tolerance to missing data and supplementary clinical information 248

7.4 Models Description 249

7.4.1 Decision Trees 249

7.4.2 The k-nearest neighbour model 251

7.4.3 Logistic Regression 252

7.4.4 Artificial Neural Networks 253

7.4.5 Bayesian Networks 254

7.5 Logistic Regression Versus Bayesian Networks 257

7.6 Used Software Packages 259

REFERECES 260

Abstracts arising from this work 262

Publication and paper submitted arising from this work 265

ABSTRACT

Various factors come into play in the onset and progression of Type 1 Diabetes (T1D) and for this reason it is defined as a multifactorial disease. This feature is evident since the early disease stages, during which the intervention of various factors (both genetic and environmental) is necessary for the onset of the disease. Moreover, during the progression of the disease, a number of inter and intra cellular phenomena interact, resulting in the complete loss of pancreatic beta-cell capacity to produce and release insulin.

One can easily realize that a multilevel approach, taking into account the network of interacting factors, is necessary for better understanding of T1D.

This thesis aims at approaching the multifactorial nature of T1D in two different aspects:

1) **Etiology**: interactions between T1D susceptibility genes were examined separately and the joint risk of their interaction was estimated, using the *Bayesian probabilistic approach*.

2) **Progression of the beta-cell damage**: this was implemented in a virtual portion of diabetic pancreas, by evaluating the effects of apoptotic events induced by autoimmune responses on communication between beta-cells, and hence on the ability of the pancreas to produce insulin. The damage was studied by implementing a stochastic mathematical model that simulates the complex relationships that exist within a virtual cluster made up of pancreatic beta-cell.

As concerns the first aim, our results confirmed that HLA-DR is the most relevant susceptibility gene compared to INS and PTPN22 and proved that the INS and PTPN22 genotypes marginally influence T1D risk in all HLA genotype

risk categories. The absolute risk conferred by genes simultaneously carrying high, moderate or low risk HLA and risk genotypes at the other two loci, compared with non-risk-associated genotypes at all three loci, was 19.8%, 6.6% and 2.2%, in the family cohort and 11.5%, 1.7% and 0.1% in the case-control sample, respectively. The present work represents, to the best of our knowledge, the first study based on both case-control and familiar data sets, showing the joint effect of HLA, INS and PTPN22 in T1D in a Caucasian population with a heterogeneous age of T1D onset, generalizing previous findings regarding data sets consisting of patients and controls < 15 years published in literature. Finally, this study shows that a feasible and accurate risk assessment can be performed by applying the BN method.

Whit regard to the second aim, at 31% (normoglycaemia) and 69% (hyperglycaemia) of dead beta-cells the system appeared to be sufficiently robust biologically to maintain regular Ca2+ ions oscillations guaranteeing an effective insulin release. Simulations at 84%, 94% and 98% grades (severe hyperglycemia) showed that intracellular Calcium oscillations were absent. In such conditions insulin pulsatility is not expected to occur. Our results suggest that the islet tissue is biophysically robust enough to compensate high rates of beta-cell loss. These predictions can be experimentally tested 'in vitro' quantifying space and time electrophysiological dynamics of animal islets kept at different glucose gradients. The model indicates the necessity of maintaining glycaemia within physiological levels as soon as possible after diabetes onset in order to avoid a drastic interruption of Ca2+ pulsatility with the subsequent reduction of insulin release.

ACKNOWLEDGEMENTS

This section is to acknowledge the contribution of all those who in various ways allowed me to confidently achieve goals fulfilling my very own expectations.

I would like to gratefully and sincerely thank my supervisor Prof. Paolo Pozzilli for his guidance and understanding since I was a medical student and during my PhD studies. He encouraged me to not only grow as a medical doctor and a researcher but also as an independent thinker. He gave me the opportunity to go to the University of Ulm to develop part of my PhD project. He also taught me the basics of scientific reasoning, giving me both knowledge and method.

I could not have completed my work without the help of my co-Supervisor Prof. Simonetta Filippi. I want to thank her for encouraging me to have an optimistic and interdisciplinary vision of research. The enthusiasm she has for research was contagious and motivational for me, even during tough times in the pursuit of the Ph.D. I also thank all the those I worked with in the Non Linear Physics and Mathematical Modeling Laboratory, at the University Campus Bio-Medico of Rome. They provided me with their intellectual and technical support during this project. In particular, I would like to mention Drs. Christian Cherubini and Alessio Gizzi.

I would like to express my deep and sincere gratitude to Professor Bernhard Boehm, Head of the Division of Endocrinology and Diabetes, University of Ulm. His wide knowledge and expertise in clinical and experimental diabetes immunology have been of great value for me. He offered me the opportunity to work on big data sets, really relevant for my project. He also patiently train

me on bench work and his understanding, encouraging and personal guidance have provided a good basis for the present thesis.

I wish to thank Dr. Hans Kestler, member of the Institute of Neural Information Processing, University of Ulm, for his guidance in statistical analysis and especially for introducing me in the Bayesian probabilistic reasoning. Moreover, I would like also thank his collaborators, Ludwig Lausser, Johann Kraus, Martin Hopfensitz, Alexander Groß, Christian Wawra and Markus Maucher, for their help and technical support during my stay in Ulm.

A special thank goes to Professor Giulio D'agostini, Dr Serena Cenatiempo and Dr Francesco Sanfilippo for their help and expert support in the probabilistic approach.

Also I would like to express my deep gratitude to all the members of Endocrinology and Metabolic Disease Department, in particular Dr Silvia Manfrini, Dr Chiara Guglielmi, Dr Luciana Valente, Dr Riccardo Murari, Dr Giusi Beretta Anguissola, Dr Elvira Fioriti, Dr Manon Khazrai, Dr Angelo Lauria, Dr Annarita Maurizi, Dr Daria Maggi, Dr Andrea Palermo, Dr Sara Fallucca, Dr Shady Kyanvash, Dr Rocky Strollo, Dr Ernesto Maddaloni, Dr Dario Tuccinardi, Dr Giuseppe De Feudis, Dr Simona Miglietta, Dr Alessandra Lanzara and Dr Nicola Napoli. The group has been a source of friendships as well as good advice and collaboration.

I would also like to express my sincere gratitude to Prof. Roberto Angioli, Head of Gynecology and Obstetrics Department at University Campus Bio-Medico of Rome, for welcoming me in his group and for giving me the opportunity to complete this thesis.

I thank Dr Chiara Moretti, a collegue and a friend. The German adventure in Ulm started with her.

Finally I would like to thank Jane, for her friendship, time, interest, and helpful comments regarding this dissertation.

Special thanks to Porta Nevia's friends, for their availability, smiling and welcoming me in any situation.

I wish to thank all my friend, who followed me no matter how far away I have been. In particular Marija and Ivana who shared with me the innumerable years and never made me feel distant.

I would like to give special thanks to my parents for tirelessly supporting me through all the life, for letting me do it 'my way' and for encouraging to reach for my dreams. Thanks to my brother Antonio, for all the inspiring and motivating conversations and moreover, for his unending and contagious enthusiasm for life.

I PART. INTRODUCTION TO TYPE 1 DIABETES

CHAPTER 1. DIABETES MELLITUS

1.1 Definition of Diabetes

The term diabetes mellitus represents a set of disease conditions sharing certain characteristics. Foremost among these is the presence of elevated plasma glucose levels. The presence of hyperglycemia in a patient is used both to diagnose diabetes and to guide management therapeutic decisions, which are largely directed toward avoiding hyperglycemia. An important characteristic of the various disease states that are labeled as diabetes is the development of "end-organ damage" in vital organs of the body. The so-called *microvascular complications* include damage of the retina, the renal glomerulus and peripheral nerves. The damage results, at least in part, from the chronic effects of hyperglycemia and is mediated through glycation of tissue proteins, increased activity of the polyol pathway, or other, still unknown, mechanisms. Moreover, people with diabetes have a considerably greater risk of developing atherosclerotic disease affecting the coronary, cerebrovascular, peripheral arterial, or other parts of the circulatory system. A cause-and-effect relationship between chronic hyperglycemia and these so-called *macrovascular complications* of diabetes has not been clearly established, although evidence linking the two is accumulating (Stern M, 1996).

1.2 Classification

As reported in the last *Standards of Medical Care in Diabetes—2012*, the classification of diabetes includes four clinical classes (Standards ADA, 2012):

1) *type 1 diabetes* (results from beta-cell destruction, usually leading to absolute insulin deficiency)

2) *type 2 diabetes* (results from a progressive insulin secretory defect on the background of insulin resistance)

3) *gestational diabetes mellitus* (GDM) (diabetes diagnosed during pregnancy)

4) *other specific types of diabetes* due to other causes, e.g., genetic defects in beta-cell function, genetic defects in insulin action, diseases of the exocrine pancreas (such as cystic fibrosis), and drug or chemical-induced diabetes (such as in the treatment of AIDS or after an organ transplant).

1.2.1 Type 1, type 2 and gestational diabetes mellitus

Type 1 and type 2 diabetes are the most common clinical forms of the disease.

Type 1 diabetes (T1D) (10% of diabetes cases): It is a multifactorial and polygenic disease characterized by the development of a state of complete insulin deficiency due to the autoimmune destruction of pancreatic insulin producing beta-cell. In its fully developed form, patients will, if deprived of insulin, develop ketoacidosis, coma, and death. Biochemical testing reveals the absence of circulating C peptide (a marker of insulin secretion) despite

hyperglycemia. Although the peak incidence occurs in childhood and early adolescence, this form of diabetes can occur at any age. Recent epidemiological and immunological research has led to recognition of two major forms of T1D based on the presence or absence of certain immunological markers.

Type 2 diabetes: represents the most common form of diabetes seen in most parts of the developed world. The condition is characterized by hyperglycemia that results from a combination of defects in insulin secretion and insulin action. In any given individual, the degree to which these defects contribute to the hyperglycemia may vary. The disease usually has its onset after the age of 40, although increasingly, type 2 diabetes is being seen among the youth (Mokdad AH, 2001). While progressive beta-cell failure is believed by many to be an important part of the natural history of this form of diabetes (Rudensky, 1988), the beta-cell destruction is not autoimmune-mediated and does not progress to a point at which the patient becomes dependent on insulin for survival (Leahy J, 1990). Ketoacidosis is unusual in this form of diabetes, and when it occurs, it is usually in the setting of a major intercurrent illness such as myocardial infraction, stroke, or treatment with glucocorticoid. Individuals with type 2 diabetes are not at increased risk for autoimmune diseases but have a higher prevalence of metabolic abnormalities, including obesity, hypertension, and a typical dyslipidemia that is characterized by hypertriglyceridemia and low levels of HDL (Cowie C, 1995). The causes of type 2 diabetes remain to be determined. Evidence to support a genetic component of the disease comes from the strong concordance for the disease that is seen among monozygotic twins (Barnett A, 1981). On the other hand, the dramatic increase in incidence and prevalence of this form of diabetes that

accompanies the change to a so-called westernized lifestyle strongly supports an environmental component as well (Knowler W, 1993).

Gestational diabetes: is defined as diabetes whose onset or first recognition occurs during pregnancy. The prevalence of gestational diabetes increases concurrently with the prevalence of type 2 diabetes in a population. Risk factors include age (it is more common among older women), ethnicity (higher rates are seen among women from ethnic groups with a high incidence of type 2 diabetes), pre-pregnancy body mass index (the risk increases with degree of obesity), parity (the risk increases with the number of previous pregnancies), and family history of diabetes. The diagnosis of gestational diabetes is important, since, if left untreated, adverse fetal outcomes can occur. The main adverse fetal or maternal outcomes are macrosomia and neonatal hypoglycemia. Maternal complications include a higher rate of dystocia and caesarian section as well as a greater risk of future type 2 diabetes.

1.3 Diagnostic Criteria

In their 1997 report, the Expert Committee of the ADA, recommended changes in the previous diagnosis criteria for the disease published in 1979 report. The main differences between the new and the old criteria were a reduction in the fasting plasma glucose cut point used to diagnose the disease and an emphasis on the use of fasting plasma glucose as opposed to the oral glucose tolerance test (OGTT) to screen for and diagnose the disease. The fasting plasma glucose level required for a diagnosis of diabetes was reduced from 140 mg/dL to 126 mg/dL. The diagnostic threshold for the 2-hours post-glucose challenge plasma glucose level was left unchanged at 200 mg/dL. Two

other important features of the diagnostic criteria that were retained include the ability to use casual plasma glucose levels in patients with hyperglycemic symptoms and the requirement that a firm diagnosis be based on testing carried out on more than one occasion in asymptomatic patients. ADA has not previously recommended the use of A1C for diagnosing diabetes, in part due to lack of standardization of the assay. However, in the last recommendations A1C ≥ 6.5% appears among the others diagnostic criteria as shown in table 1 below.

A diagnostic test of diabetes should be repeated to rule out laboratory error, unless the diagnosis is clear on clinical grounds, such as a patient with classic symptoms of hyperglycemia or hyperglycemic crisis.

1.4 Epidemiology of diabetes mellitus

Diabetes mellitus is one of the most common and widespread diseases throughout the world and for that reason it is one of the most challenging health problems in the 21st century.

It is the fifth leading cause of death in most high-income countries and there is substantial evidence that it is widespread in many economically developing and newly industrialized nations.

The number of studies describing the epidemiology of diabetes over the last 20 years has been enormous. From 1997 ADA report recommending new diagnostic criteria for diabetes, many studies have assessed the impact of the change on the prevalence of diabetes in different populations. It has been a consistent finding of population-based diabetes studies that a substantial

proportion of all people found to have diabetes had not been previously diagnosed.

Table 1. Criteria for the diagnosis of diabetes (Standard of Medical Care in Diabetes-2012). *In the absence of unequivocal hyperglycemia, results should be confirmed by repeat testing.

A1C ≥6.5%. The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay.*
OR
FPG ≥126 mg/dL (7.0 mmol/L). Fasting is defined as no caloric intake for at least 8 h.*
OR
2-h plasma glucose ≥200 mg/dL (11.1 mmol/L) during an OGTT. The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.*
OR
In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥200 mg/dL (11.1 mmol/L)

Significant data on this topic come from the last edition of the *IDF Diabetes Atlas*: the prevalence of diabetes mellitus has been estimated for different countries for the years 2010 and 2030 (Fig 1-4).

It is estimated that approximately 285 million people worldwide, or 6.6%, in the age group 20-79, will have diabetes in 2010, some 70% of whom live in low and middle-income countries. This number is expected to increase by more than 50% in the next 20 years. By 2030, 7.8% of the adult population is projected to have diabetes. The largest increase will take place in the regions dominated by developing economies (see Fig 1). The 40-59 age group currently has the greatest number of people with diabetes with some 132

million in 2010, more than 75% of whom live in low and middle-income countries (see Fig 2).

Figure 1. Prevalence (%) estimates of diabetes by region, 2010 and 2030. The data presented for adults are for both types 1 and 2 diabetes combined in the age range 20-79. Africa (AFR), Europe (EUR), Middle East and North Africa (MENA), North America and Caribbean (NAC), South and Central America (SACA), South-East Asia (SEA), and the Western Pacific (WP).

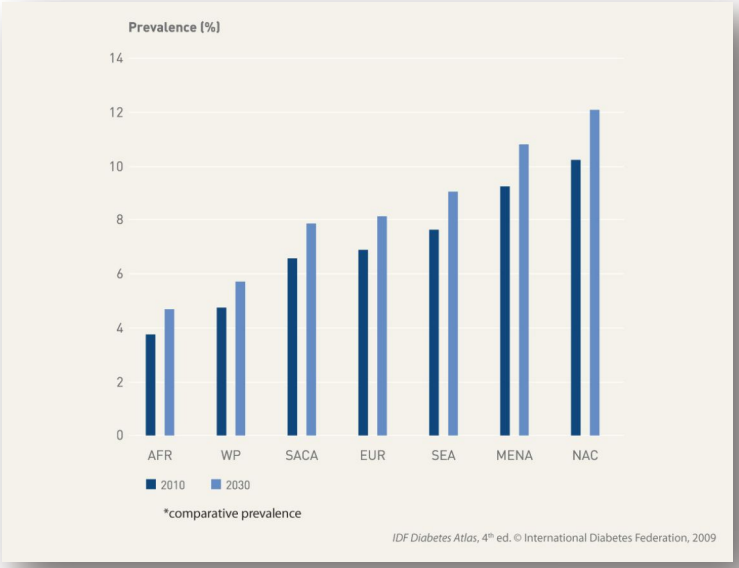


Figure 2. Number of people with diabetes by age group, 2010 and 2030. The figure shows the absolute number, ranging by age, of people affected by diabetes mellitus worldwide. The data presented for adults are for both types 1 and 2 diabetes combined in the age range 20-79.

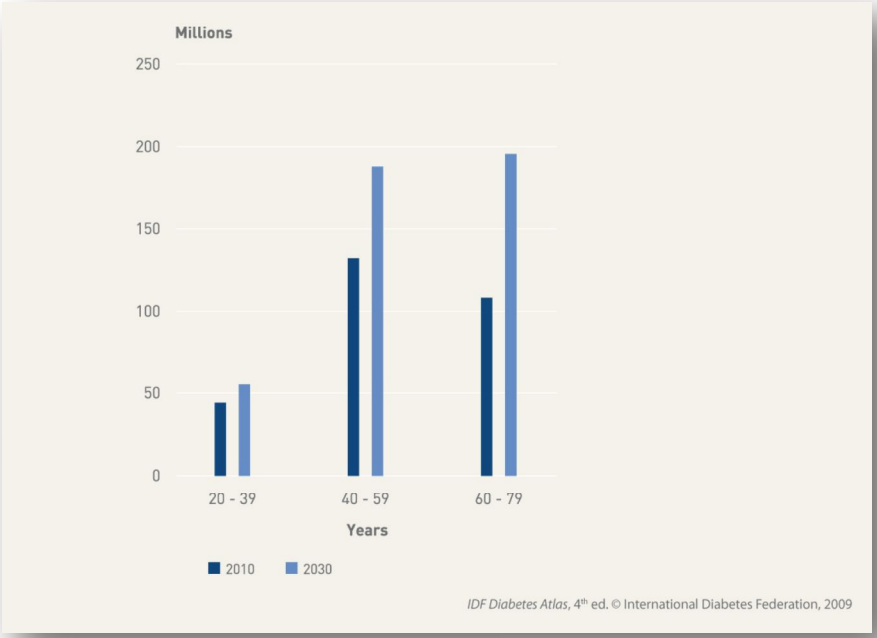


Figure 3. Prevalence (%) estimates of diabetes (20-79 years), 2010.

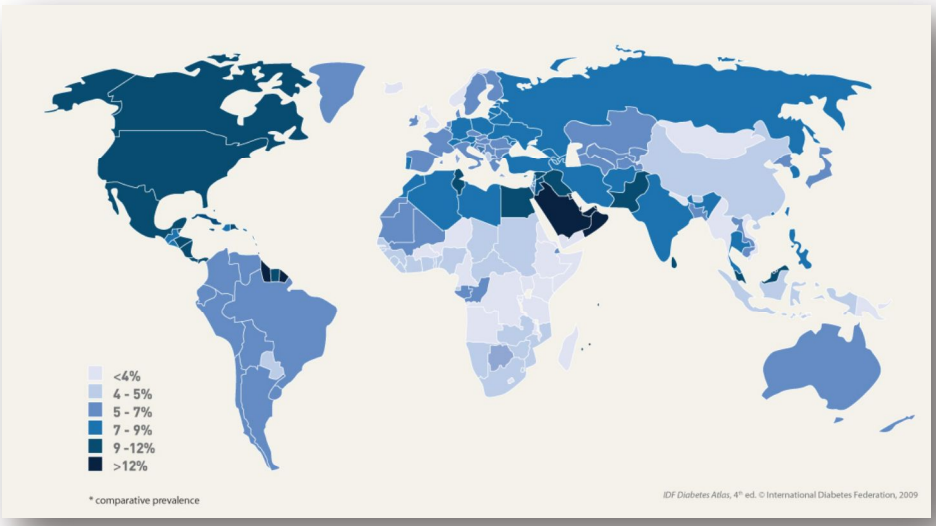
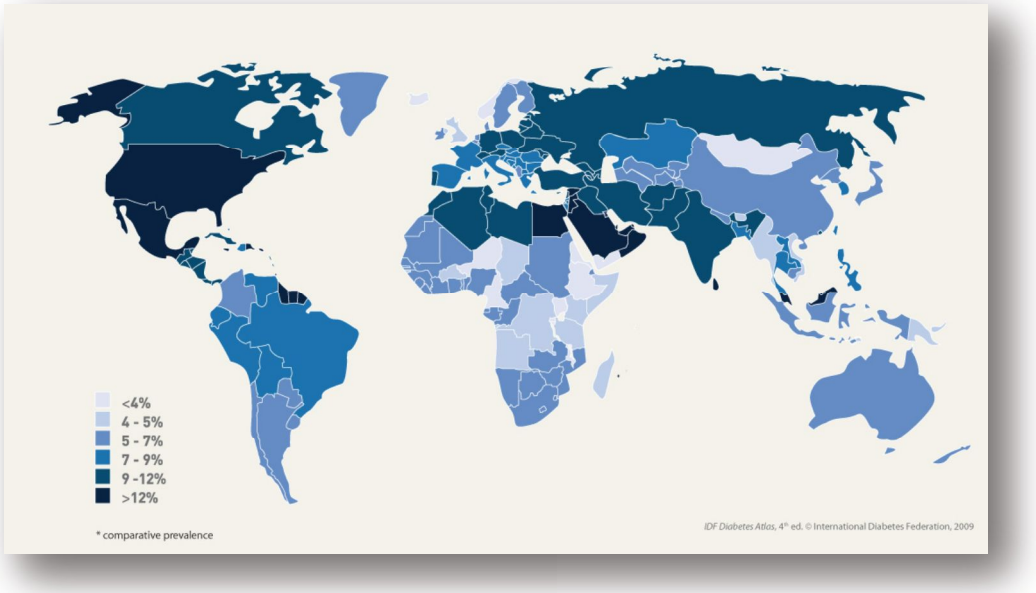


Figure 4. Pevalence (%) estimates of diabetes (20-79 years), 2030.

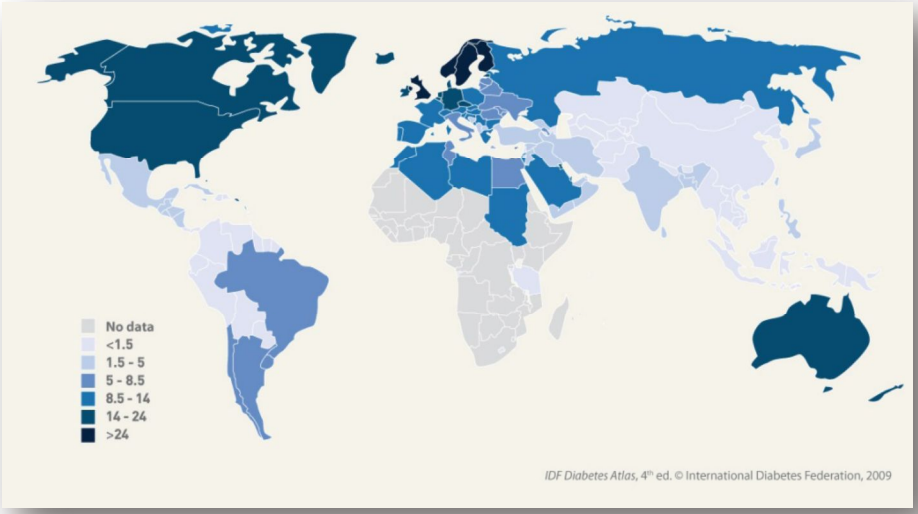


1.5 Incidence and risk factors for Type 1 Diabetes

There is a significant geographical variation in the incidence of T1D. It is more common in Europe and less common in Asia. The incidence of T1D is highest in Scandinavia (30–50/100,000), intermediate in the United States (15–25/100,000) and somewhat lower in Central and Eastern Europe (5–15/100,000). Thus, the absolute risk changes according to the incidence, so that in Finland a child is 40 times more likely to develop T1D than a Japanese child and almost 100 times more likely to get the disease than a child in of China. The incidence of T1D is increasing worldwide by 3% every year (EURODIAB ACE Study Group, 2000; Onkamo P, 1999). Several studies confirmed that these values are increasing worldwide, as shown in fig 5 (Gale EA, 2002; The DIAMOND Project Group, 2006).

Much of the aetiology of T1D is accounted for by genetic predisposition as demonstrated in several studies, by comparing concordance rates in monozygotic and dizygotic twins (Redondo MJ, 2001; Hyttinen V, 2003). The lifetime risk for a member of the general population is often quoted as 0.4%. This increases to 1% if the mother has diabetes, to 3% if the father has T1D (Warram JH, 1988) and 6% in first degree relatives of T1D subjects (15 times greater than in a member of the general population) (Risch N, 1987).

Fig 5. New cases of T1D in children, 0-14 years (cases per 100,000 aged 0-14 years per year), 2010.



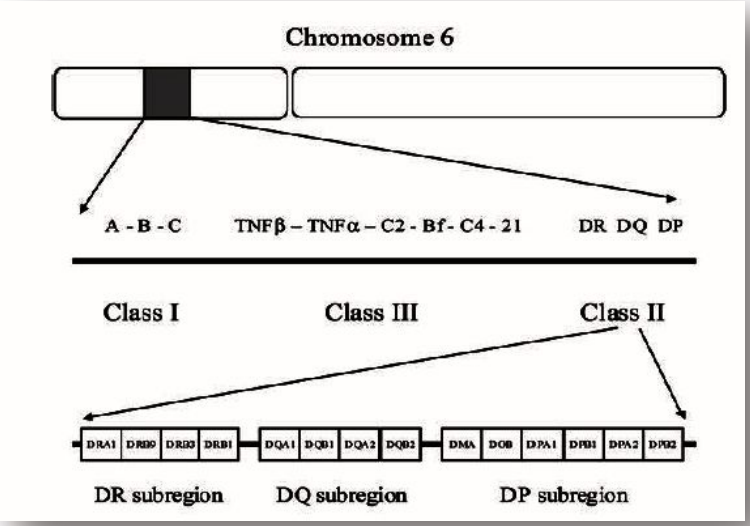
1.5.1 Susceptibility genes.

The study of candidate genes identified the two major susceptibility genes for T1D: IDDM1 (Insulin-Dependent Diabetes Mellitus-1) encoded in the HLA region of the Major Histocompatibility Complex (MHC) on chromosome 6p21 and IDDM2 (Insulin-Dependent Diabetes Mellitus-2) encoded by the insulin

gene region on chromosome 11p15. This approach also revealed the association of other two loci, all with smaller contributions to T1D susceptibility: the CTLA4 (Cytotoxic T Lymphocyte Associated Antigen 4) gene on chromosome 2q33, and the PTPN22 (Lymphoid Tyrosine Phosphatase 22) gene.

Human Leucocyte Antigen (HLA). The HLA are a family of homologous proteins that present antigens to T cells. They are all encoded within the Major Histocompatibility Complex (MHC) region on chromosome 6p21 All HLA proteins are highly polymorphic as they have hundreds of protein-coding alleles (Hughes A, 2008) (fig 6). The class II loci HLA-DRB1 and HLA-DQB1 are known to be strongly associated with T1D risk, explaining ≈ 50% of the genetic susceptibility to T1D, based on functional, structural and genetic evidence (She J-X, 1996; Ounissi-Benkalha H, 2008). It is generally considered that HLA class II DQB1*0302 on the DR4 haplotype and DQB1*0201 on the DR3 haplotype are the principal susceptibility markers for T1D. We know that 90–95% of the young children with T1D carry either or both susceptibility haplotypes, whereas the protective DR2-DQB1*0602 is present in 0.1%. HLA-DPB1, another HLA class II gene, has been shown to modulate HLA class II-mediated risk independent of HLA DRB1 and DQB1 (Cruz TD et al, 2004). The T1D susceptibility or protection associated with HLA may be related to their ability to present peptides of relevance to diabetogenic T cells. Individuals with HLA molecules that are not able to effectively present specific peptides to naïve T cells in the thymus might fail to generate tolerance. Alternatively, specific HLA alleles may selectively present an islet peptide to mature T lymphocytes that have escaped negative selection. These two mechanisms are both possible and they could coexist.

Fig 6. MHC genes on chromosome 6 confer almost 50% of genetic susceptibility to T1D

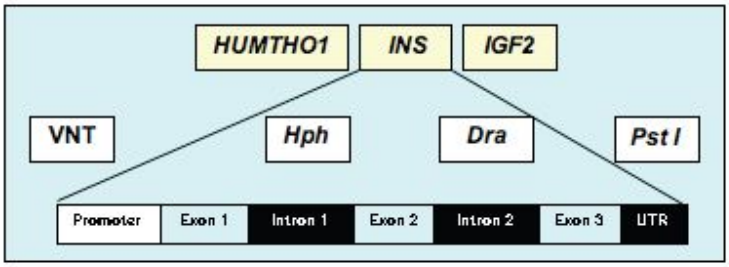


INS (Insulin gene). Its role in disease susceptibility was demonstrated by association studies and was replicated by linkage analysis. Susceptibility in the INS region, in the IDDM2 locus, has been primarily mapped to a variable number of tandem repeats (VNTR) located ~0.5 kb upstream of INS. The VNTR region, located in the promoter of the INS gene, plays an important role for regulating how much INS is produced. Alleles in this region are divided into three classes distinguished by the number of DNA base pair repeats (Fig 7). Class I alleles have a mean of 570 base pairs, class II alleles 1200 base pairs and class III alleles have 2200 base pairs (Bennett ST, 1995). The VNTR may not explain all of the susceptibility in this region and at least two other polymorphisms (-23HphI and +1140A/C) may contribute to the etiological effect (189). Class I alleles are associated to T1D, whereas protection from T1D is conferred by the alleles of class III.

In studies of INS gene expression, it has been shown that class I alleles are associated with higher INS expression in the pancreas when compared with

class III alleles, but the opposite is true in the thymus where class I alleles are expressed at 2–3-fold lower levels. This is likely to alter the selection of T cells in the thymus and may therefore influence tolerance to INS (Vafiadis P, 1997; Pugliese A, 1997).

Figure 7. Insulin VNTR variants and classification



PTPN22, a gene found on chromosome 1p13 that encodes lymphocyte-specific tyrosine phosphatase (negative regulator of T cell kinase signaling that is crucial to the balance between host defense and self-tolerance), was found to be associated with susceptibility to T1D in a case–control study. An SNP in the PTPN22 gene potentially contributes to susceptibility to T1D, as well as many other autoimmune diseases, because of increased negative regulation of T cell activation (Bottini N, 2004; Criswell L, 2005).

CTLA4, is a surface molecule found on activated T cells that produces a negative signal for T cell activation. It is thought that inherited changes in the CTLA-4 gene expression can increase T cell self-reactivity and therefore play an important role in autoimmune diseases such as T1D. The role of this gene in T cell development and in the pathogenesis of autoimmunity is so far incompletely understood.

1.5.2 Environmental risk factors

Despite the clear contribution of genetic factors, it should be noted that the maximum concordance rate for T1D in identical twins does not exceed 50%. Studies in monozygotic twins indicate that only 13–33% are pairwise concordant for T1D (Barnett AH, 1981; Kaprio J, 1992). Moreover, only 1 out of 15 people in the general population with the highest risk HLA genotypes develops T1D. The conspicuous geographic variation in the incidence of T1D in children even among Caucasians and the substantial increase of the incidence worldwide, cannot be the consequence only of genetic factors but must mostly be due to changes in environment and lifestyle. This implies that additional factors are needed to trigger and drive beta-cell destruction in genetically predisposed subjects. Environmental factors have been implicated in the pathogenesis of T1D both as triggers of beta-cell destruction (Dahlquist G , 1995; Åkerblom HK, 1998; Åkerblom HK, 2002).

Among the environmental risk factors, the most prominent example is exposure to rubella during pregnancy. About 20% of children born with congenital rubella develop T1D (Menser MA, 1978; Ginsberg-Fellner F, 1985).

More recent studies have shown an increased risk for childhood T1D if the mother has had an infection with enteroviruses during pregnancy (Dahlquist G, 1995; Hyoty H, 1995).

Other events during pregnancy or at delivery such as pre-eclampsia also confer T1D risk (Dahlquist G, 1989; Jones M, 1998; Dahlquist G, 1999). High birth weight and children born large for gestational age have a higher risk of T1D than controls (Dahlquist G, 1996; Stene Lc, 2001).

Different studies have been designed to investigate the role of environmental risk factors for T1D. Among them:

- *BABYDIAB* study, the first prospective cohort study that examines the natural history of islet autoimmunity in children of patients with T1D. The aim of this study was to determine when islet autoantibodies first appear, which genetic and environmental factors influence their development, and what islet auto-antibody characteristics were most associated with progression in T1D (Ziegler AG, 1993);

- *TEDDY* study (the Environmental Determinants of Diabetes in Youth), a large, international cohort project, was explicitly designed to explore environmental factors with respect to their impact on the development of islet cell autoimmunity and the incidence of T1D among high-risk newborns with specific human leukocyte antigen (HLA) genotypes (TEDDY Study Group, 2008);

- *TRIGR* study, an international, randomized, double-blinded trial. The hypothesis to be tested was whether hydrolyzed infant formula compared to cow's milk-based formula decreases the risk of developing T1D in children with increased genetic susceptibility.

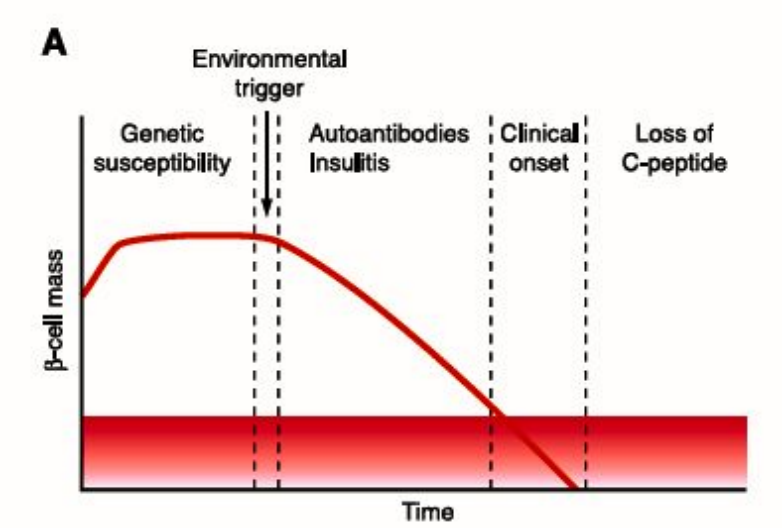
To date environmental causes and mechanism underlying T1D pathogenesis are unknown and results from different cohort studies are not definitive. New studies are therefore needed to identify which risk factors will lead to a better understanding of disease pathogenesis and result in new strategies to prevent, delay or reverse T1D.

1.6 Pathogenesis of Type 1 Diabetes

In T1D, a genetically susceptible host develops autoimmunity against the own beta-cells. What trigger this autoimmune response remain unclear at this time. This autoimmune process results in progressive destruction of beta-cells until a critical mass of beta-cells is lost and insulin deficiency develops. Insulin deficiency in turn leads to the onset of clinical sign and symptoms of T1D. It develops slowly and progressive abnormalities in beta-cell function herald what appears to be a sudden development of hyperglycemia. Rising HbA1c in the normal range, impaired fasting or glucose tolerance, as well as loss of first-phase insulin secretion usually precede overt diabetes. At the time of diagnosis, some viable beta-cells are still present and these may produce enough insulin to leads a partial remission of the disease (honeymoon period) but over time, almost all beta-cells are destroyed and the patients becomes totally dependent on exogenous insulin for survival. The exact beta-cell mass remaining at diagnosis is poorly defined, and there are almost no studies of insulitis before diabetes onset. For patients with long-term T1D, there is evidence of some beta-cell function remaining (C-peptide secretion), although beta-cell mass is usually decreased to less than 1% of normal (Meier JJ, 2005). A large body of evidence indicates that the development of T1D is determined by a balance between pathogenic and regulatory T lymphocytes (Mordes JP, 2004; Chatenoud L, 2005). Several timeline models have been put forward to depict the outcome of the interplay between the genetic and environmental factors. The linear beta-cell decline hypothesis postulated by Eisenbarth in 1986 remains the most widely referenced benchmark model for T1D (Eisenbarth GS, 1986). According to this model (Fig. 8), genetically susceptible individuals at some point encounter certain environmental agents that initiate islet autoimmunity leading to a linear decay in beta-cell mass, development of

autoantibodies, hyperglycemia, and eventually complete loss of C-peptide. There are two characteristics of T1D present in most patients, namely at least one susceptible HLA class II haplotype and islet autoantibodies. Thus, regardless of the etiological factor(s) that may favor the development of islet autoimmunity in a child, it remains highly probable that the initiation of the disease process is the effector immune response to islet beta-cell antigens and that the appearance of islet autoantibodies is the first detectable sign of this process. Islet autoantibodies rarely appear prior to approximately age 6 months, and among children with a family history of T1D there is a peak incidence at approximately 1 to 2 years of age (Naserke, 1999; Hummel, 2004; Bonifacio, 2008). This is a relatively important observation because, in terms of the disease process, it suggests that the events leading to islet autoimmunity are encountered after 6 months of age and potentially at increased frequency in early infancy and/or that neonatal immune mechanisms are relatively protective from disease during the first six months of life.

Figure 8. Timelines of T1D. Model for linear beta-cell mass decay, as originally proposed by Eisenbarth (Eisenbarth GS, 1986). In the context of genetic predisposition, an environmental trigger induces islet autoimmunity and beta-cell death leading to a sequence of "pre-diabetic" stages and eventually clinical onset.



1.7 Morbidity and mortality

Complications are very common, with at least one complication present in a large proportion of people at the time of diagnosis. Diabetes can affect many major organs in the body, including the heart, blood vessels, nerves, eyes and kidneys. Cardiovascular disease, resulting from damage to large blood vessels, causes the death of 50% or more of people with diabetes depending on the population. Long-term complications of T1D develop gradually, over the years.

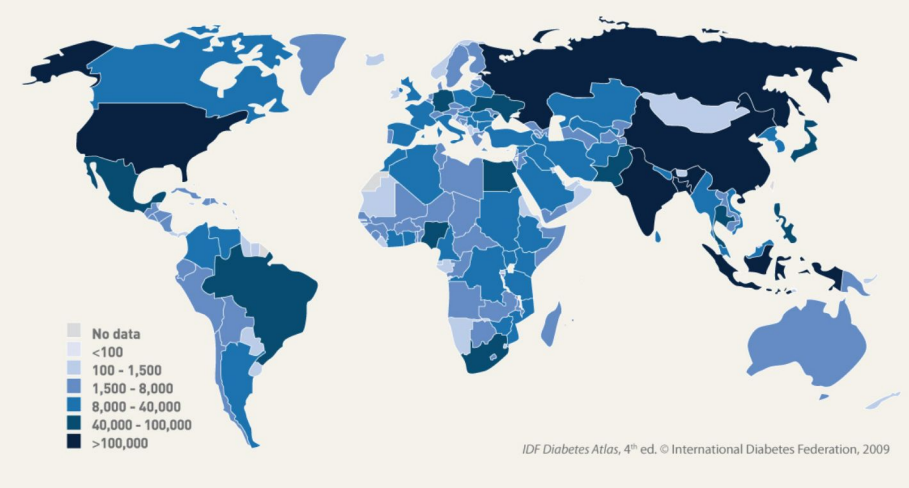
Accurate estimates of *mortality* attributable to diabetes are difficult to obtain because more than a third of countries of the world do not have data on mortality caused by diabetes (Roglic, 2005). For that reason an estimation of

diabetes-related mortality was modeled to estimate the number of deaths in the year 2010, based on the world trend.

The number of deaths attributable to diabetes in 2010 shows a 5.5% increase over the estimates for the year 2007 (IDF, 2006). This increase is largely due to a 29% increase in the number of deaths due to diabetes in the North America and Caribbean Region, a 12% increase in the South-East Asia Region and an 11% increase in the Western Pacific Region. This increase can be explained by a rise in diabetes prevalence in some highly populated countries in each region, particularly in women.

It is well known that a substantial proportion of premature deaths are potentially preventable through public health action directed at primary prevention of diabetes in the population and improvement of care for all people with diabetes (WHO, 2005). However, to date no such decline in the morbidity and mortality has been reported for diabetes. Although some high-income countries have documented an improved survival of persons with diabetes, the increased prevalence is most likely due to a rise in incidence rather than improved survival (Colagiuri S, 2005).

Figure 9. Number of deaths attributable to diabetes (20-79 years), 2010



References

Åkerblom HK, Knip M (1998) Putative environmental factors in type 1 diabetes. Diabetes Metab Rev 14:31– 67

Åkerblom HK, Vaarala O, Hyo¨ ty H, Ilonen J, Knip M (2002) Environmental factors in the etiology of type 1 diabetes. Am J Med Genet 115:18 –29

Barnett A, Eff C, Leslie R, Pyke D (1981) Diabetes in identical twins. A study of 200 pairs. Diabetologia 20:87-93

Barnett AH, Eff C, Leslie RDG, Pyke DA (1981) Diabetes in identical twins: a study of 200 pairs. Diabetologia 20:87–93

Bennett Johnson S, Baughcum AE, Carmichael SK, She JX, Schatz DA (2004) Maternal anxiety associated with newborn genetic screening for type 1 diabetes. Diabetes Care 27:392–397

Bennett ST, Lucassen AM, Gough SC et al. (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9:284–292

Bonifacio E, Pflu¨ger M, Marienfeld S, Winkler C, Hummel M, Ziegler AG (2008) Maternal type 1 diabetes reduces the risk of islet autoantibodies: Relationships with birthweight and maternal HbA(1c). Diabetologia 51:1245–1252

Bottini N, Musumeci L, Alonso A et al. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet

36:337–338

Chatenoud L, Bach JF 2005 Regulatory T cells in the control of autoimmune diabetes: the case of the NOD mouse. Int Rev Immunol 24:247–267

Colagiuri S, Borch-Johnsen K, Glümer C, et al. (2005) There really is an epidemic of type 2 diabetes. Diabetologia 48: 1459-1463

Cowie C, Harris M (1995) Physical and metabolic characteristics of person with diabetes. In National Diabetes Data Group (ed). Diabetes in America. Bethesda, MD, National Institutes of Health 117-164

Criswell LA et al. (2005) Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620 W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 76:561-71

Cruz TD, Valdes AM, Santiago A et al (2004) DPB1 alleles are associated with type 1 diabetes susceptibility in multiple ethnic groups. Diabetes 53:2158-2163

Dahlquist G (1995) Environmental risk factors in human type 1 diabetes: an epidemiological perspective. Diabetes Metab Rev 11:37– 46

Dahlquist G, Bennich S, Kallen B (1996) Intrauterine growth pattern and risk of childhood onset insulin dependent (type I) diabetes: population based case-control study. BMJ 313: 1174–1177

Dahlquist G, Blom L, Tuvemo T, Nystro¨ ML, Sandstro¨ MA, Wall S (1989) The Swedish Childhood Diabetes Study – results from a nine year case register and one year case-referent study indicating that type 1 (insulin-dependent) diabetes mellitus is associated with both type 2 (non-insulin-

dependent) diabetes mellitus and autoimmune disorders. Diabetologia: 32:2–6

Dahlquist G, Ivarsson S, Lindberg B, Forsgren M (1995) Maternal enteroviral infection during pregnancy as a risk factor for childhood IDDM. Diabetes 44:408–413

Dahlquist GG, Patterson C, Soltesz G (1999) Perinatal Risk Factors For Childhood Type 1 Diabetes In Europe. The Eurodiab Substudy 2 Study Group. Diabetes Care 22:1698–1702

DIAMOND Project Group (2006) Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999 Diabet Med 23:857-66

Eisenbarth GS (1986) Type I diabetes mellitus. A chronic autoimmune disease. N Engl J Med 314:1360–1368

EURODIAB ACE Study Group (2000) Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. Lancet. 355:873–876

Gale EA (2006) The rise of childhood type 1 diabetes in the 20th century. Diabetes 51: 3353–3361, 2002. (The DIAMOND Project Group: Incidence and trends of childhood type 1 diabetes worldwide 1990–1999. Diabet Med 23:857–866

Gillespie KM, Bain SC, Barnett AH, Bingley PJ, Christie MR, Gill GV, Gale EA (2004) The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. Lancet 364:1645–1647

Ginsberg-Fellner F, Witt ME, Fedun B et al. (1985) Diabetes mellitus and autoimmunity in patients with congenital rubella syndrome. Rev Infect Dis 7:

170–176

Hughes A L. Major Histocompatibility Complex (MHC) genes: evolution. Encyclopedia of Life Science, 2008

Hummel M, Bonifacio E, Schmid S, Walter M, Knopff A, Ziegler AG (2004) Brief communication: Early appearance of islet autoantibodies predicts childhood type 1 diabetes in offspring of diabetic parents. Ann. Intern. Med. 140, 882–886

Hyoty H, Hiltunen M, Knip M et al. (1995) A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study. Diabetes 44:652–657

Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J (2003) Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes 52:1052–1055

International Diabetes Federation. (2006) The Diabetes Atlas. Third Edition. Brussels: International Diabetes Federation

Jones M, Swerdlow A, Gill L, Goldacre M. (1998) Pre-Natal And Early Life Risk Factors For Childhood Onset Diabetes Mellitus: A Record Linkage Study. Int J Epidemiol 27:444–449

Jönson B. (2002) Revealing the costs of Type II diabetes in Europe, Diabetologia 45:5–12

Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengård J, Kesa¨niemi YA (1992) Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based

cohort of twins in Finland. Diabetologia 35:1060–1067

Knowler W, Saad M, Pettitt D, et al. (1993) Determinants of diabetes mellitus in the Pima Indians. Diabetes Care 16:216-227

Krishnamurthy B, Dudek NL, McKenzie MD Purcell AW, Brooks AG, Gellert S, Colman PG, Harrison LC, Lew AM, Thomas HE, Kay TW (2006) Responses against islet antigens in NOD mice are prevented by tolerance to proinsulin but not IGRP. J Clin Invest 116:3258–3265

Leahy J. (1990) Natural history of beta cell dysfunction in NIDDM. Diabetes Care 13:992-1010

Meier JJ, Bhushan A, Butler AE, Rizza RA, Butler PC (2005) Sustained beta-cell apoptosis in patients with long-standing type 1 diabetes: indirect evidence for islet regeneration? Diabetologia 48:2221–2228

Menser MA, Forrest JM Bransby RD (1978) Rubella infection and diabetes mellitus. Lancet 1:57–60

Mokdad AH, Bowman BA, Ford ES, et al (2001) The continuing epidemics of obesity and diabetes in the United States. JAMA 286:1195-1200

Mordes JP, Bortell R, Blankenhorn EP, Rossini AA, Greiner DL (2004) Rat models of type 1 diabetes: genetics, environment, and autoimmunity. ILAR J 45:278–291

Naserke HE, Bonifacio E, Ziegler AG (1999) Immunoglobulin G insulin autoantibodies in BABYDIAB offspring appear postnatally: Sensitive early detection using a protein A/G-based radiobinding assay. J. Clin. Endocrinol.

Metab. 84:1239–1243

Onkamo P, Vaananen S, Karvonen M, Tuomilehto J. (1999) Worldwide increase in incidence of Type I diabetes-the analysis of the data on published incidence trends. Diabetologia 42:1395–1403

Ounissi-Benkalha H, Polychronakos C (2008) The molecular genetics of type 1 diabetes: new genes and emerging mechanisms. Trends Mol Med 14:268–275

Pugliese A, Zeller M, Fernandez A et al. (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 15:293–297

Redondo MJ, Fain PR, Eisenbarth GS (2001) Genetics of type 1A diabetes. Recent Prog Horm Res 56:69–89

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1–14

Roglic G, Unwin N, Bennett PH, et al. (2005) The burden of mortality attributable to diabetes: realistic estimates for the year 2000. Diabetes Care 28: 2130-2135

Rudensky A, Hadden D, Atkinson A, et al. (1988) Natural history of pancreatic islet beta cell function of type 2 diabetes meelitus studied over six years by homeostasis model assessment. Diabet Med 5:36-41

She J-X. (1996) Susceptibility to type I diabetes: HLA-DQ and DR revisited. Immunol Today 17:323–329

Standards of Medical Care in Diabetes (2012) Diabetes Care. Volume 33,

Suppl 1

Stene Lc, Magnus P, Lie Rt, Sovik O, Joner G. (2001) Birth weight and childhood onset type 1 diabetes: population based cohort study. BMJ 322: 889–892

Stern M. (1996) Do non-insulin-dependent diabetes mellitus and cardiovascular disease share common antecedents? Ann Intern Med 124:110-116

TEDDY Study Group (2008) The Environmental Determinants of Diabetes in the Young (TEDDY) Study. Ann N Y Acad Sci 1150:1-13

Vafiadis P, Bennett ST, Todd JA et al. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet, 15, 289–292

Warram JH, Martin BC, Soeldner JS, Krolewski AS (1988) Study of glucose removal rate and first phase insulin secretion in the offspring of two parents with non-insulin-dependent diabetes. Adv Exp Med Biol. 246:175-9

Williams R et al. (2002) Assessing the impact of complications on the costs of Type II diabetes, Diabetologia 45:13–17

World Health Organization. Preventing chronic diseases: a vital investment. Geneva: World Health Organization; 2005.http://www.who.int/chp/chronic_disease_report/contents/en/index.html)

Ziegler AG, Hillebrand B, Rabl W, Mayrhofer M, Hummel M, Mollenhauer U, Vordemann J, Lenz A, Standl E (1993) On the appearance of islet associated autoimmunity in offspring of diabetic mothers: a prospective study from

birth. Diabetologia 36:402–408

II PART. COMPUTATIONAL BIOMEDICINE: NEW MULTILEVEL APPROACH TO THE MULTIFACTORIALITY AND COMPLEXITY OF TYPE 1 DIABETES

"Networks pervade all aspects of human health"

(Albert-László Barabási, N Engl J Med 2007; 357:404-407)

Introduction to the experimental section

A multilevel approach to the complexity of type 1 diabetes:

the dual purpose of this research project

The thesis has a clear bipartite structure: a first part regarding the use of a Bayesian Network approach to the estimation of the burden of genetic causative component in the T1D and a second part in which is demonstrated that a cluster of beta-cells, modeled by a Network of elementary excitable and interacting units, exhibits an emergent collective behavior resembling biologically relevant properties of islets of Langerhans.

The *trait d'union* between these two parts is clearly the "Network paradigm" that we can summarize with the sentence "*The knowledge of the relations between the parts of a system is much more important than the knowledge of the fine structure of the constituting elements*". The use of the term "network" is more and more widespread in all fields of Biology and Medicine. It evokes a systemic approach to biological and clinical problems able to overcome the evident limitations of the strict reductionism of the past twenty years.

Major consequences follow when not only the single elements, but also the intermingled links connecting different parts of biological entities, are taken into consideration. The network paradigm is nowadays the prevailing methodological approach in Biology. We can read about gene networks, protein networks, metabolic as well as signaling networks. The network paradigm is a horizontal construct, essentially different from the classical top-down perspective of Molecular Biology, dominant until a few years ago. This

approach comprised of a privileged flux of information from DNA down to RNA and proteins.

This kind of strategy is set to enable scientific thought to overcome the current impasse and is a promising solution to the sterile debate occurring due to the 'bottom-up' and 'top-down' approaches.

In general, many diseases are associated with the breakdown of functional modules that are best described as "sub-networks" of a more complex network connecting many components. Therefore, an understanding of the functionally relevant genetic, environmental, metabolic and protein–protein interactions in a disease network will play an important role in increasing of the understanding the pathophysiology of human diseases.

As described earlier, unlike monogenic diseases, there are *various factors that come into play* in the onset and progression of T1D. For this reason it is defined as a multifactorial disease. This feature can be highlighted since the early disease stages, during which the intervention of various factors (both genetic and environmental) is necessary for the onset of the disease. Moreover, during the progression of the disease, *multiple inter and intra cellular phenomena interact*, resulting in the complete loss of pancreatic beta-cell capacity to produce and release insulin.

One can easily see that a multilevel approach, taking into account the network of interacting factors, is necessary for a better understanding of T1D.

Finally, a network-based approach to human disease has multiple potential applications from biological and clinical points of view. A better understanding of the etiopathogenesis and disease progression in T1D may offer more focused targets for preventive strategies and development of

drugs. These advances may also lead to the use of more accurate biomarkers to monitor the functional integrity of pancreatic islets that are perturbed by the autoimmune process, as well as the development of specific disease guidelines for classification and therapy.

This thesis aims at approaching the multifactorial nature of T1D in two different aspects:

1) **Etiology**: interactions between T1D susceptibility genes and environmental risk factors were examined separately and the joint risk of their interaction was estimated, using the *Bayesian probabilistic approach*.

2) **Progression of the beta-cell damage**: this was implemented in a virtual portion of diabetic pancreas, by evaluating the effects of apoptotic events induced by autoimmune responses on communication between beta-cells, and hence on the ability of the pancreas to produce insulin. The damage was studied by implementing a stochastic mathematical model that simulates the complex relationships that exist within a virtual cluster made up of pancreatic beta-cell.

These two different approaches will be introduced in the next chapters, in the context of the multilevel analysis and the new challenge of modern technology.

CHAPTER 2. "A NETWORK-BASED APPROACH"

2.1 The reasons for using a network-based approach

The Human Genome Project has revolutionized gene hunting, leading to an explosion in the number of detected associations between genes and disease phenotypes. However, many of the new disease-associated genetic mutations account for only a small fraction of disease occurrences.

Genetic susceptibility is crucial in the development of T1D. The lifetime risk for a member of the general population is often quoted as 0.4%. This increases to >1% if the mother has diabetes and, intriguingly, to >3% if the father has T1D (Warram JH, 1988). The sibling risk is 6% (15 times greater than in a member of the general population) (Risch N, 1987).

The classic indicator of the role of genes in any disease is found by comparing concordance rates in monozygotic (MZ) and dizygotic (DZ) twins. In a Finnish study authors showed that early age of onset in one twin increased the risk in the co-twin (Hyttinen V, 2003). Concordance rates of 27% in MZ and 3.8% in DZ, respectively, were reported. In particular, the evaluation of the role of genetic and environmental factors in this study showed that 88% of the phenotypic variance was due to genetic factors, and the remaining variance was due to unshared environmental factors. In the prospective study performed by Redondo et al, the analysis of pairs of monozygotic twins from the UK and from USA, discordant for diabetes, underlined the importance of age of onset in one twin for risk of future diabetes in the co-twin: twins of individuals with T1D onset diagnosed at ≤ 24 years of age had a 38%

probability of diabetes against 6% in those diagnosed after the age of 24 years (Redondo MJ, 2001).

These results emphasize the importance of genetics in T1D, but also clearly demonstrate again that having certain combinations of genes is not sufficient to cause T1D. The hypothesis is that other hidden causes in more genes or in more environmental risk factors can be involved.

Considering the recent rapid increase in T1D incidence, and that the gene pool cannot change quickly enough to account for that, environmental factors have been implicated. Three independent studies (Hermann R, 2003; Gillespie KM, 2004; Fourlanos S, 2008), compared genetic susceptibility in individuals developing T1D currently with those developing diabetes at least one generation before and showed that more people with less genetic susceptibility are now developing autoimmune diabetes. It might therefore be argued that genetic susceptibility is now less important than in previous generations. Nevertheless the susceptibility genes have not altered, rather the environment is more 'permissive' for the clinical onset than before. Identification of these environmental triggers is proving more complex than the susceptibility genes. Several studies are ongoing to try to identify the environmental determinants of T1D, e.g. the study performed by an international consortium, the Environmental Determinants of Diabetes in the Young (TEDDY). It was established in order to study genetic–environmental interactions, including gestational events, childhood infections or other environmental factors after birth, in relation to the development of pre-diabetic autoimmunity and T1D (The TEDDY Study Group, 2007).

In some studies, authors have evaluated the relationship between a *single* environmental risk factor and T1D by comparing patients and controls (Chris

R Cardwell, 2010; Cardwell C R 2010, Cardwell C. R. 2010). Moreover, several studies investigated the relationship between environmental triggers individually considered and the genetic susceptibility of subjects (in terms of HLA, INS, PTPN22 and CTLA-4) in case-control groups (Stene L. C., 2006, Soltesz G, 2007).

All these published findings represent a valuable source of information thanks to which we can have a better orientation when we ask "what causes diabetes?", but to date the approaches used have not led to an agreement between different findings. Why is that? Here we will try to find a likely answer to this question.

2.2 An interactive multilevel network organization for a complex disease: a new approach for the "Etiology question"

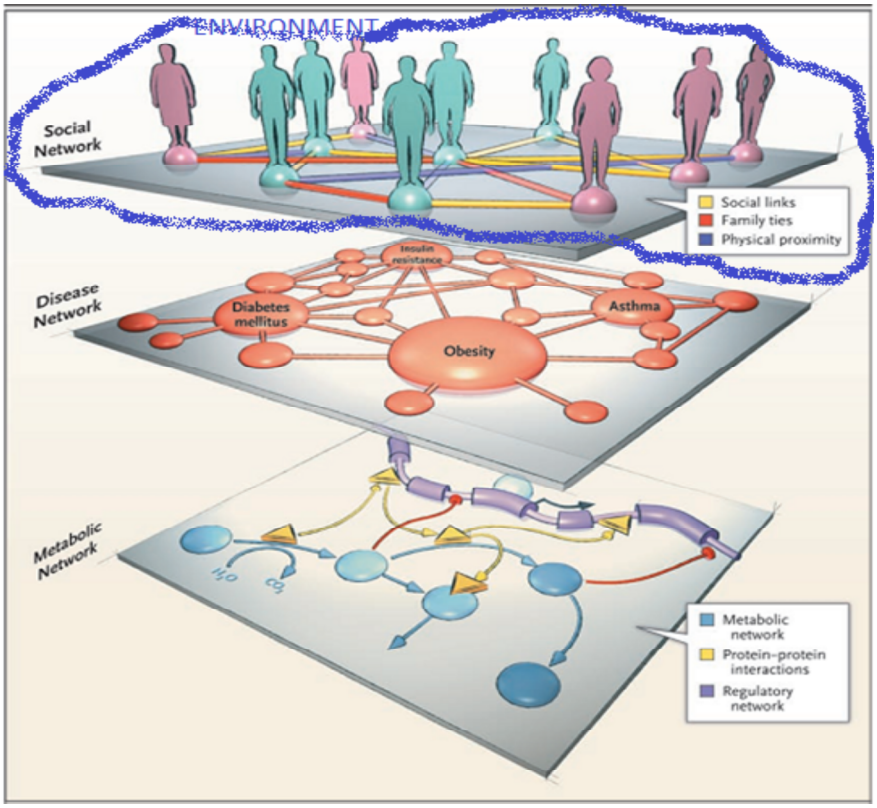
In the last 10 years the idea that an integrated approach is required to understand what comes into play in the development of a multifactorial disease has been steadily consolidating. For this reason, Barabasi et al have introduced a very promising approach that consists of identifying the disease as a complex network, composed of several interacting factors that communicate at multiple levels (Barabási A-L, 2003).

This new concept is illustrated in the figure below (Fig. 3.1). Most of the realities that surround us, (technological, social and biological systems) are characterized by an interactive multilevel network organization and are governed by quantifiable organizing principles.

The growing interest in interconnectedness has brought into focus an recently ignored issue: *"networks pervade all aspects of human health"* (Barabási A-L, 2004).

The existence of intricate molecular links between sub-cellular components and disease genes raises the probability that risk factors may not be as independent of each other as medical practitioners currently consider them to be. The third network level, shown in Fig 1, represents the social network, which encompasses all human-to-human interactions (environmental aspects e.g. family, friendship and proximity-based contacts) that play a role in the spread of pathogens (top layer). To this effect, networks may affect all aspects of medical research and practice, from disease mechanisms to drug discovery (Barabási A-L, 2003).

Fig. 1 Interactions between the cellular, disease, and social levels as part of network disease.



So far, progress towards a reliable network-based approach to disease has been limited by the incompleteness of the experimental data and the limitations of the existing tools to explore the role of network interactions in disease. For example, researchers were forced to apply traditional statistical tools to network data, assuming that the quantities applicable had a normal distribution and often, they did not take into account the fact that measurements and proxies are strongly related to each other. Another assumption of current statistical tools is that quantities characterizing various activity patterns are independent variables. But far a long time we have known since a long time, that most activity patterns at all levels of biological organization, are correlated. Thus, there is a real need to develop statistical tools that are reliable in the context of the interconnected risk factors.

Furthermore, we know that a certain degree of uncertainty considering the complex network of interaction between all the different causes of diabetes known to date, is also involved. Some questions hereby arise:

1) How should we approach the complexity and the multifactoriality of the disease?

2) How should we study the multilevel interaction between different factors involved in the same disease?

3) Which method should be used for an approach that at the same time takes into account all the variables that we want to study as well as their interaction?

4) How do we introduce the information from literature into our evaluation and besides, how can we update our thinking on T1D etiology with new experimental observations or data collected?

5) How can we quantify the uncertainty?

The approach based on "Network paradigm" seeks to offer the answers to these questions. For the first aim of this thesis (assessment of risk of T1D onset based on susceptibility genes) a special kind of Bayesian Network was chosen; whereas for the second aim (mathematical model of progression of beta-cell damage) a network of elementary excitable and interacting units modeling a beta-cell cluster was implemented.

2.3 Bayes' Theorem and Bayesian Networks

Thomas Bayes (1702 – 1761) was an English mathematician and Presbyterian minister, known for having formulated a specific case of the theorem that bears his name: Bayes' theorem, which was published after his death. Since that time the theorem has had a great impact on statistical inference because it enables us to calculate the probability of a cause when its effect is observed. From the philosophy of Thomas Bayes emerges Bayesian statistics and the simple but powerful Bayes theorem. Although it was established two and a half centuries ago, the Bayesian view of the scientific method and its role in prediction has been recognized only recently as a great unifying framework that elegantly ties all the foreseeable components together. In the 1980s, the method was extended to model the probabilistic relationships among many causally related variables. The graphical structures that describe these relationships have come to be known as Bayesian networks.

2.3.1 Bayes' Theorem

Bayesian probability is the name given to several related interpretations of probability, which have in common the notion of probability as something like a partial belief, rather than a frequency. Bayes defines probability as follows:

"The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening."

In other words, following the Bayes' Theorem, we can compute conditional probabilities of events of interest from known probabilities as follows.

Definition of Bayes' Theorem: Given two events E and F such that P(E)≠0 and P(F)≠ 0, we have

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}.$$

In Bayes' theorem, each probability has a conventional name:

* P(E) is the *prior probability* (or "unconditional" or "marginal" probability) of E. It is "prior" in the sense that it does not take into account any information about F;

* P(E|F) is the conditional probability of E, given F. It is also called the *posterior probability* because it is derived from or depends upon the specified value of F.

* P(F|E) is the conditional probability of F given E. It is also called the *likelihood*.

* P(F) is the prior or marginal probability of F, and acts as a normalizing constant.

Furthermore, given mutually exclusive and exhaustive events E1, E2,…E*n* such that P (Ei) ≠ 0 for all *i*, we have for 1≤*i*≤*n*,

$$P(\mathrm{E}_i|\mathrm{F}) = \frac{P(\mathrm{F}|\mathrm{E}_i)P(\mathrm{E}_i)}{P(\mathrm{F}|\mathrm{E}_1)P(\mathrm{E}_1) + P(\mathrm{F}|\mathrm{E}_2)P(\mathrm{E}_2) + \cdots P(\mathrm{F}|\mathrm{E}_n)P(\mathrm{E}_n)}.$$

The first formula enables us to compute (E|F) if we know (F|E), (E), and (F); the second one enables us to compute (Ei|F) if we know (F|Ej) and (Ej) for 1≤*j*≤*n*.

The following example illustrates the use of Bayes' Theorem.

Example 1:

Suppose Jack must take the "Enzyme-Linked Immuno-Sorbent Assay" (ELISA) blood test which tests for the presence of human immunodeficiency virus (HIV). Jack takes the test and it comes back positive for HIV. How likely is it that Jack is infected with HIV? Without knowing the accuracy of the test, Jack really has no way of knowing how probable it is that he is infected with HIV.

The data we ordinarily have when carrying out tests is the true positive rate (sensitivity) and the true negative rate (specificity). The true positive rate is

the number of people who both have the infection and test positive divided by the total number of people who have the infection. For example, to obtain this number for ELISA, 10,000 people who were known to be infected with HIV were identified. This was done using the Western Blot, which is the gold standard test for HIV. These people were then tested with ELISA, and 9990 tested positive. Therefore, the true positive rate is 0.999. The true negative rate is the number of people who both do not have the infection and test negative divided by the total number of people who do not have the infection. To obtain this number for ELISA 10,000 nuns who denied risk factors for HIV infection were tested. Of these, 9980 tested negative using the ELISA test. Furthermore, the 20 positive-testing nuns tested negative using the Western Blot test. So, the true negative rate is 0.998, which means that the false positive rate is 0.002. We therefore formulate the following random variables and subjective probabilities:

$$P(\text{ELISA} = \text{positive} \mid \text{HIV} = \text{present}) = 0.999 \qquad (1)$$

$$P(\text{ELISA} = \text{positive} \mid \text{HIV} = \text{absent}) = 0.002 \qquad (2)$$

One might wonder why we called these subjective probabilities when we obtained them from data. Recall that the frequentist approach says that we can never know the actual relative frequencies (objective probabilities); we can only estimate them from data. However, within the subjective approach, we can make our beliefs (subjective probabilities) equal to the fractions obtained from the data.

It might seem that Jack almost certainly is infected with HIV, since the test is so accurate.

However notice that neither the probability in Equality (1) nor the one in Equality (2) is the probability of Jack being infected with HIV. Since we know that Jack tested positive on ELISA, that probability is

$$P\,(\text{HIV} = \text{present} \mid \text{ELISA} = \text{positive})$$

We can compute this probability using Bayes' Theorem if we know P(HIV = present).

Recall that Jack took the blood test simply because the State required it. He did not take it because he thought for any reason he was infected with HIV. So, the only other information we have about Jack is that he is male and the state in which he resides. Therefore, if 1 in 100,000 men in State is infected with HIV, we assign the following subjective probability:

$$P(\text{HIV} = \text{present}) = 0{,}00001$$

We now apply Bayes' Theorem to compute the

$$
\begin{aligned}
& P(present|positive) \\
= & \frac{P(positive|present)P(present)}{P(positive|present)P(present) + P(positive|absent)P(absent)} \\
= & \frac{(.999)(.00001)}{(.999)(.00001) + (.002)(.99999)} \\
= & .00497.
\end{aligned}
$$

Surprisingly, we are fairly confident that Jack is not infected with HIV.

A probability such as (HIV=present) is called a ***prior probability*** because, in a particular model, it is the probability of some event prior to updating the probability of that event, within the framework of that model, using new information. A probability such as P (HIV = present| ELISA = positive) is called ***posterior probability*** because it is the probability of an event after its prior probability has been updated, within the framework of some model, based on new information.

In the previous example we obtained our beliefs (***subjective probabilities***) directly from the observed fractions in the data. Although this is often done, it is not necessary. In general, we obtain our beliefs from our information about the past, which means that these beliefs are a composite of all our experience rather than merely observed relative frequencies.

2.3.2 From Bayes' Theorem to Bayesian Networks

Since Bayes' time the theorem has had a great impact on statistical inference because it enables us to calculate the probability of a cause when its effect is observed. In the 1980s, the method was extended to model the probabilistic

relationships among many related variables. The graphical structures that describe these relationships have come to be known as Bayesian networks.

Bayesian networks (BNs), also known as belief networks, belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science and statistics.

BNs correspond to another graphical model structure known as a directed acyclic graph (DAG) that is popular in the statistics, the machine learning, and the artificial intelligence societies. BNs are both mathematically rigorous and intuitively understandable. They enable an effective representation and computation of the joint probability distribution (JPD) over a set of random variables (Pearl J. 1988).

The structure of a DAG is defined by two sets: the set of nodes (vertices) and the set of directed edges. The nodes represent random variables and are drawn as circles labeled by the variable names. The edges represent direct statistical dependence among the variables and are represented by an arrow between nodes. In particular, an edge from node X_i to node X_j represents a statistical dependence between the corresponding variables. Thus, the arrow indicates that a value taken by variable X_j depends on the value taken by variable X_I. Node X_I is then referred to as a *parent* of X_j and, similarly, X_j is referred to as the *child* of X_i. An extension of these genealogical terms is often

used to define the sets of "descendants" — the set of nodes that can be reached on a direct path from the node, or ancestor nodes — the set of nodes from which the node can be reached on a direct path (Grifths TL, 2006). The structure of the acyclic graph guarantees that there is no node that can be its own ancestor or its own descendent. Such a condition is crucial for the factorization of the joint probability of a collection of nodes as seen below. Note that although the arrows represent direct relationship connection between the variables, the reasoning process can operate on BNs by propagating information in any direction (Pearl J, 2001). This process is called *inference*. Inference in a Bayesian network means computing the conditional probability for some variables given information (evidence) on other variables.

This is easy when all available evidence is on variables that are ancestors of the variable(s) of interest. But when evidence is available on a descendant of the variable(s) of interest, we have to perform inference opposite the direction of the edges. For this purpose, we employ Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Inference is essentially a clever application of Bayes' Theorem; details can be found in the paper by Jensen 1990 (F. Jensen, 1990)

A BN rejects a simple conditional independence statement. Namely that each variable is independent of its non-descendent in the graph, given the state of its parents. This property is used to reduce, sometimes significantly, the number of parameters that are required to characterize the JPD of the variables. *This reduction provides an efficient way to compute the posterior probabilities, given the evidence.*(Pearl, J. 1988; Friedman N et al1997).

In addition to the DAG structure, which is often considered as the qualitative part of the model, one needs to specify the quantitative parameters of the model. The parameters are described in a manner which is consistent with a Markovian property, where the conditional probability distribution (CPD) at each node depends only on its parents. For discrete random variables, this conditional probability is often represented by a table, listing the local probability that a child node takes on each of the feasible values — for each combination of values of its parents. The joint distribution of a collection of variables can be determined uniquely by these local conditional probability tables (CPTs).

Following the above discussion, a more formal definition of a BN can be given. A Bayesian network is an annotated acyclic graph that represents a JPD over a set of random variables V. The network is defined by a pair $B = \langle G, \Theta \rangle$,where G is the DAG whose nodes $X_1, X_2, \ldots X_n$ represents random variables, and whose edges represent the direct dependencies between these variables. The graph G encodes independence assumptions, by which each variable X_i is independent of its nondescendents given its parents in G. The second component Θ denotes the set of parameters of the network. This set contains the parameter $\theta_{x_i|\pi_i} = P_B(x_i|\pi_i)$ for each realization x_i of X_i conditioned on π_i, the set of parents of X_i in G. Accordingly, P(B) is equal to:

$$P_B(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i|\pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i} \quad (1)$$

If X_i has no parents, its local probability distribution is said to be *unconditional*, otherwise it is *conditional*. If the variable represented by a node is observed,

then the node is said to be an evidence node, otherwise the node is said to be hidden or latent.

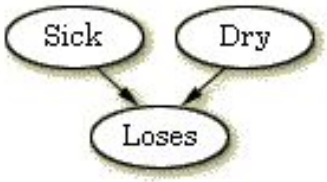
The following is a simple example of BN structure.

Example 3. The Apple Tree Example:

The problem domain of this example is a small orchard belonging to Jack Fletcher (let's call him Apple Jack). One day Apple Jack discovers that his finest apple tree is losing its leaves. Now, he wants to know why this is happening. He knows that if the tree is dry (caused by a drought) there is no mystery - it is very common for trees to lose their leaves during a drought. On the other hand, the loss of leaves can be an indication of disease.

The situation can be modeled by the BN in Fig 2. The BN consists of three nodes: Sick, Dry, and Loses which can all be in one of two states: Sick can be either "sick" or "not" - Dry can be either "dry" or "not" - and Loses can be either "yes" or "no". The node Sick tells us that the apple tree is sick by being in "sick" state. Otherwise, it will be in state "not" . The nodes Dry and Loses tell us in the same way if the tree is dry and if the tree is losing its leaves, respectively.

Figure 2. BN representing the domain of the Apple Jack problem.



The BN in Fig. 2 models one of the possible kind of interaction between nodes: the *causal dependence* from Sick to Loses and from Dry to Loses. This is represented by the two links.

When there is a causal dependence from a node A to another node B, we expect that when A is in a certain state this has an impact on the state of B. One should be careful when modeling the causal dependencies in a BN. Sometimes it is not quite obvious in which direction a link should point. In our example, for instance, we say that there is a causal link from Sick to Loses because when a tree is sick this might cause the tree to lose its leaves. But couldn't one say that when the tree loses its leaves it might be sick and then turn the link in the other direction? No, we cannot! It is the sickness that causes the lost leaves and not the lost leaves that cause the sickness. In Figure 1, we have the graphical representation of the BN. However, this is only what we call the *qualitative* representation of the BN. Before we can call it a BN, we need to specify the *quantitative representation*. The quantitative representation

of a BN is the set of CPTs of the nodes. Tables 1, 2 and 3 show the CPTs of the three nodes in the BN of Figure 3.2

Sick="sick"	**Sick**="not"
0.1	0.9

Table 1: P(Sick).

Dry="dry"	**Dry**="not"
0.1	0.9

Table 2: P(Dry).

	Dry="dry"		**Dry**="not"	
	Sick="sick"	**Sick**="not"	**Sick**="sick"	**Sick**="not"
Loses="yes"	0.95	0.85	0.90	0.02
Loses="no"	0.05	0.15	0.10	0.98

Table 3: P(Loses | Sick, Dry).

Note that all three tables show the probability of a node being in a specific state depending on the state of its parent nodes but since Sick and Dry do not

have any parent nodes, the distributions in Tables 1 and 2 are not conditioned by anything.

2.4 "How should we approach complexity and multifactoriality? ": the role of Conditional Independence

One objection to the use of probability theory is that the complete specification of a probability distribution requires absurdly numerous numbers. For example, if there are ***n*** binary random variables, the complete distribution is specified by 2^{n-1} joint probabilities. Thus, the complete distribution for a BN with 3 nodes would require 7 values, yet we only specified 6. This savings might not seem great, but if we increase the size of the network to 8 nodes (as we have shown in the next fig 4), 2^{n-1} would be 128, but we would only need to give 16 (as shown on the left side column by the green windows in the fig 4). Where does this come from?

The answer is that Bayesian networks have built-in independence assumptions.

Definition of independence assumptions. Two sets of variables, A and B, are said to be (conditionally) independent, given a third set C of variables. If the values of the variables C are known, knowledge about the values of the variables B provides no further information about the values of the variables A:

$$P(A|B,C) = P(A|C)$$

Therefore if two variables are independent given the state of a third variable, then they are said to be conditionally independent. This does not mean that B

and C are independent in any possible occasion, it only means that the knowledge of B is useless when C is already known.

Conditional independence can be read directly from the graph as follows: Let A, B, and C be disjoint sets of variables, then

- identify the smallest sub-graph that contains $A \cup B \cup C$ and their ancestors;
- add undirected edges between nodes having a common child;
- drop directions on all directed edges.

Now, if every path from a variable in A to a variable in B contains a variable in C, then A is conditionally independent of B given C (**Lauritzen S. L, 1990)**

Due to the assumption of independence, the Bayesian network algorithm can estimate in "essentially" the probabilistic relationships that govern the variables of interest thereby making the contemporary analysis of many variables easier and faster.

2.5 Updating our beliefs by introducing new knowledge

"How can information from literature be introduced in our data analysis? How can we update our beliefs on T1D etiology with new experimental observations or collected new data?" Fundamentally, the Bayesian theorem provides us with the answers. The Bayesian formalism allows us to update that confidence (probability) when new information becomes available. Instead of a list of parameters BN begin with the prior joint probability density function (pdf) describing what we know about the parameters before

we begin to build the model. These parameter distributions are gleaned from literature, previous experiments or archival databases. Calibration via Bayesian inference (see paragraph 2.3.2) is used to update the distributions to make the theory (model) consistent with particular observations (data).

To illustrate this concept, let us consider the following hypothetical piece of medical reasoning:

> "Shortness-of-breath (dyspnea) [d] may be due to tuberculosis [t], lung cancer [l] or bronchitis [b], or none of them, or more than one of them. A recent visit to Asia [a] increases the risk of tuberculosis, while smoking [s] is known to be a risk factor for both lung cancer and bronchitis. The result of a single chest X-ray [x] does not discriminate between lung cancer and tuberculosis, neither does the presence or absence of dyspnea". (Lauritzen S.L, 1988).

On the left side probabilities of different events-variables present in the data set imported are shown (in green), whereas on the right side all variables are codified in nodes and are linked by edge of relationship.

The Following are some examples to explain the concept.

If, during the patients interview, we learn that a patient is a smoker, we will adjust our beliefs (increased risks) regarding lung cancer and bronchitis. However, our beliefs regarding tuberculosis are unchanged (i.e., t is conditionally independent of s given the empty set of variables).

Now, suppose we get a positive X-ray result for the patient. This will affect our beliefs regarding tuberculosis and lung cancer, but not our beliefs regarding bronchitis (i.e., b is conditionally independent of x given s). However, had we also known that the patient suffers from shortness-of-

breath, the X-ray result would also have affected our beliefs regarding bronchitis (i.e., b is not conditionally independent of x given s and d).

Figure 3: Graph representing structural aspects of medical knowledge concerning lung disease.

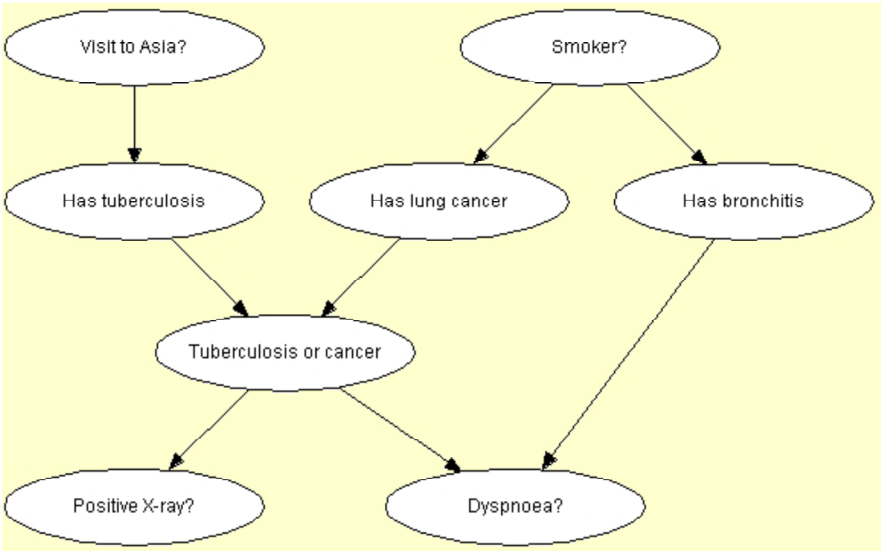


Figure 4: Bayesian Network of lung diseases and conditional probability tables.

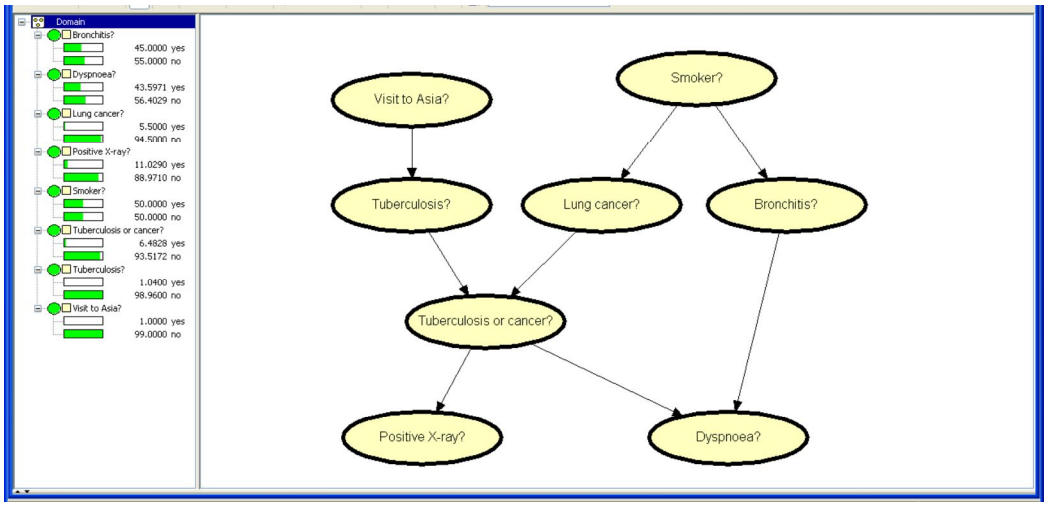
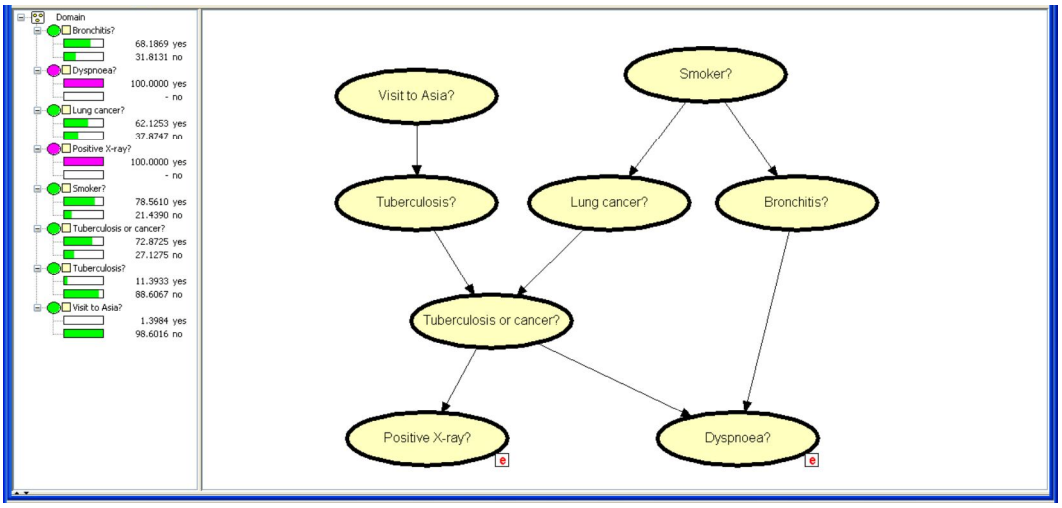


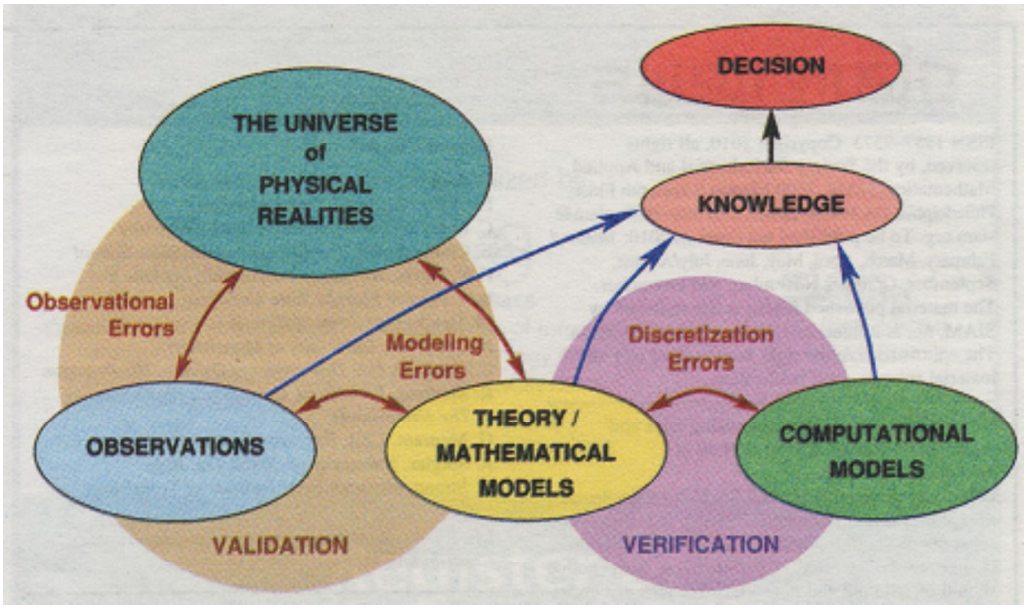
Figure 5: The update of the information status regarding the two variables "Positive X-ray" and "Dyspnoea" (in pink), determines a resulting change in the probability of "Bronchitis" and "Tuberculosis or cancer" than the probabilities shown in the previous figure, when we did not know anything about signs and symptoms of the patient.



2.6 "How can we quantify the uncertainty? ": the relevance of information status

Our goal is to use scientifically based predictions of physical reality to make informed decisions. Probabilistic Bayesian approach gives us the opportunity to make predictions about diseases. As we are talking about complex disease, we need to take into account that the prediction includes concrete, quantifiable measures of uncertainty, that is we must know how good the predictions are. Furthermore, we are aware that from the theories to the experimental data and from instruments used for measuring to computer used for modeling prediction, an abundance of errors and uncertainties affect every aspect of scientifically based predictions.

Figure 6. Imperfect computational modelling: imperfection in the mathematical models, incomplete observational data, observations delivered by imperfect instruments, and corruption of the model itself in the discretization needed for computation all lead to imperfect paths to knowledge. Reproduced from JT Oden, "A brief view of V & V & UQ" a presentation to the Board on Mathematical Sciences and their applications, National Research Council, October 2009."



How can we deal with these imperfections? It is here that old ideas re-emerge to help us with Probability theory of Thomas Bayes. Theories and their consistency with experiments must be judged in terms of probabilities in light of the evidence. In treating uncertainty we should decide how to represent it mathematically. The Bayesian philosophy suggests that ***uncertainty can be represented probabilistically***. In this direction, probability represents our confidence in some proposition, given all currently available information.

If the state of information changes, the evaluation of the probability also has to be modified. More in general, in Schrodinger's words:

"Since the knowledge may be different with different person or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event… Thus whenever we speak loosely of the ***probability of an event****, it is always to be understood: probability with regard to a certain given state of knowledge."*

2.7 Strengths and limitations of Bayesian Networks.

The Bayesian approach provides a framework for making inferences about the parameters of interest, taking into account uncertainty in the nuisance parameters. Owing to computational constraints, Bayesian analyses were not widely used until about 15 years ago, when they became more prevalent in many areas of science and field of research as genetics, oncology, biology, artificial intelligence, management (Bodén M 2010; Yu J 2005; Burnside ES 2004; Chakraborty S 2005; Chen XW 2005), etc. This advance is now extending to genetic association studies and also medicine as recent papers have shown (Ungvári I 2012; Kang J, 2011; Schlosberg CE 2011; Hageman RS 2011).There are some obvious advantages of working with BNs. BNs can facilitate learning about causal relationships between variables (Uusitalo,2007) and can easily be converted into decision support. The graphical nature of a BN clearly displays the links between different system components. This can facilitate discussion of the system structure with people from a wide variety of backgrounds and can encourage interdisciplinary discussion and stakeholder participation. The use of Bayesian inference means that a BN can be readily updated, when new knowledge becomes available. There is often a lack of information about one or more processes involved in natural systems. Models that rely on data alone (e.g. traditional deterministic or process

models) are not suitable to assess uncertain processes in the system. BNs provide a way to overcome data limitations by incorporating input data from different sources. BNs are therefore useful tools for addressing uncertainty in data and combining observations, model simulation and expert knowledge (Uusitalo, 2007). A convenient feature of BNs is the ability to learn about the structure and parameters of a system based on observed data. Knowledge of the structure of a system can reveal the dependence and independence of variables and suggest a direction of causation. It evaluates the 'optimal' BN structure, based on the highest probability score for possible candidate structures, given the data provided and perhaps penalized for the level of complexity. Different score metrics can be used to evaluate the BN structure, varying from entropy methods to genetic algorithms.

Parameter learning entails estimating the CPT at each node, given the link structures and

the data. Parameter learning is based on Bayesian learning algorithms that aim to find the maximum likelihood for the CPTs in a given BN. Of course, the fact that sufficient observations are needed to enable an estimation of conditional probabilities and the availability of 'enough' observed data is precisely a limitation in many natural resource management issues. If there are lots of missing observations, BNs can use complex learning algorithms to learn the tables. The distribution of the unavailable data needs to be defined and may be dependent on the states of other variables or they can be randomly distributed. Kontkanen et al (Kontkanen P, 1998) demonstrate that BNs can yield good prediction accuracy using learning algorithms, even if sample sizes are small.

There are also some clear limitations to BN models. While Bayesian models are a useful way to model expert knowledge, it may be difficult to get experts to agree on the structure of the model and the nodes that ought to be included. Furthermore, experts may be hard put to express their knowledge in the form of probability distributions (Uusitalo, 2007). Elicitation of expert knowledge requires an iterative process, to ensure that experts are comfortable with the nodes, their states and interrelationship in the BN, before they can make statements about distributions and confidence intervals of variables (Pollino, 2008). Furthermore, some BN software packages may have limited ability to deal with continuous data. Such data generally needs to be 'discretized' (broken up into discrete states). The states need to comprise interval values that define the total range of values the continuous variables can assume.

Although discretizing is a convenient way to control the size of the network, discrete states may not capture the original distribution of the variable completely and can lead to lower precision of variable values. Barton et al (Barton M, 2008) show how discretization assumptions can significantly affect the outcome estimates. Another limitation that has been defined in the literature stems from the acyclic nature of BNs. The acyclic property is required to carry out probability calculus, but implies that feedback effects cannot be included in the network (Barton et al, 2008). There is also a limit to the spatial and temporal scales that can be modeled within one BN. The usual approach to account for different scales is to develop a network for each geographical site or time period, and running these separately, thus inevitably increasing the size of the model.

REFERENCES

Grifths TL & Yuille A (2006) A primer on probabilistic inference, Trends in Cognitive Sciences Supplement to special issue on Probabilistic Models of Cognition 10:1—11

Pearl J (1988) Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco

Pearl J. & Russel S (2001) Bayesian networks. Report (R-277), November 2000, in Handbook of Brain Theory and Neural Networks, M.Arbib, ed MITPress, Cambridge, pp. 157—160

Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:10115

Barabási AL. Linked. New York: Plume, 2003

Barton DN, Saloranta T, Moe SJ, Eggestad HO & Kuikka S (2008) Bayesian belief networks as a meta-modelling tool in integrated river basin management. Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. Ecological Economics, 66, 91-104

Bodén M (2010) A Bayesian network model of proteins' association with promyelocytic leukemia (PML) nuclear bodies.J Comput Biol.17:617-30

Burnside ES (2004) Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. Stud Health Technol Inform;107(Pt 1):13-17

Cardwell CR (2010) Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. Diabetologia 51:726–735

Cardwell CR et al. (2010) Birth weight and the risk of childhood-onset type 1 diabetes: a meta-analysis of observational studies using individual patient data. Diabetologia, 53:641-51

Chakraborty S, Ghosh M, Maiti T, Tewari A. Bayesian neural networks for bivariate binary data: an application to prostate cancer study. Stat Med. 2005 Dec 15;24(23):3645-3662

Chris R. Cardwell(2010) Maternal Age at Birth and Childhood Type 1 Diabetes: A Pooled Analysis of 30 Observational Studies. Diabetes, 59;486-494

Fourlanos S, Varney MD, Tait BD et al. (2008) The rising incidence of type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. Diabetes Care, 31, 1546–1549

Friedman N, Geiger, D & Goldszmidt M (1997) Bayesian network classiers, Machine Learning 29, 131—163

Gillespie KM, Bain SC, Barnett AH et al. (2004) The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. Lancet, 364, 1699–1700

Hageman RS, Leduc MS, Korstanje R, Paigen B, Churchill GA (2011) A Bayesian framework for inference of the genotype-phenotype map for segregating populations. Genetics. 187:1163-70

Hermann R, Knip M, Veijola R et al. (2003) Temporal changes in the

frequencies of HLA genotypes in patients with Type 1 diabetes—indication of an increased environmental pressure? Diabetologia, 46, 420–425

Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J (2003) Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes 52:1052–1055

Jensen F & Andersen SK (1990) Approximations in Bayesian belief universes for knowledge-based systems. In Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, 162-169, Cambridge, Massachusetts, July 27-29

Kang J, Zheng W, Li L, Lee JS, Yan X, Zhao H (2011) Use of Bayesian networks to dissect the complexity of genetic disease: application to the Genetic Analysis Workshop 17 simulated data. BMC Proc. ;5 Suppl 9:S37

Kontkanen P, Myllymäki P, Silander T, Tirri H , Grünwald P (1998) Bayesian and Information-Theoretic Priors for Bayesian Network Parameters. X Conference on Machine Learning (ECML-98)

Lauritzen SL, Dawid AP, Larsen BN & Leimer HG (1990) Independence properties of directed Markov fields. *Networks*, 20(5):491-505. Special Issue on Influence Diagrams

Lauritzen SL & Spiegelhalter DJ (1988) Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 50:157-224

Nyberg JB, Marcot BG & Sulyma R (2006) Using Bayesian belief networks in adaptive management1. Canadian Journal of Forest Research, 36, 3104

Pearl, J (1988) Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco

Pollino CA, Woodberry O, Nicholson A, Korb K & Hart BT (2007) Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. Environmental Modelling & Software, 22, 1140-1152

Redondo MJ, Fain PR, Eisenbarth GS (2001) Genetics of type 1A diabetes. Recent Prog Horm Res, 56, 69–89

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet, 40, 1–14

Schlosberg CE, Schwantes-An TH, Duan W, Saccone NL (2011) Application of Bayesian network structure learning to identify causal variant SNPs from resequencing data. BMC Proc. 5 Suppl 9:S109

Soltesz G (2007) Worldwide childhood type 1 diabetes incidence--what can we learn from epidemiology? Pediatr Diabetes 8 Suppl 6:6-14.

Spirtes P, Glymour C & Schienes R (1993) Causation Prediction and Search, Springer-Verlag, New York.

Stene LC (2006) The relation between size at birth and risk of type 1 diabetes is not influenced by adjustment for the insulin gene (-23HphI) polymorphism or HLA-DQ genotype. Diabetologia 49: 2068–2073

The TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. Pediatric Diabetes 2007: 8: 286–298

Ungvári I, Hullám G, Antal P, Kiszel PS, Gézsi A, Hadadi E, Virág V, Hajós G,

Millinghoffer A, Nagy A, Kiss A, Semsei AF, Temesi G, Melegh B, Kisfali P, Széll M, Bikov A, Gálffy G, Tamási L, Falus A, Szalai C (2012) Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. PLoS One.7:e335-373

Uusitalo L (2007) Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling, 203, 312-318

Warram JH, Krolewski AS, Kahn CR (1988). Determinants of IDDM and perinatal mortality in children of diabetic mothers. Diabetes, 37, 1328–1334

Yu J, Chen XW (2005) Bayesian neural network approaches to ovarian cancer identification from high-resolution mass spectrometry data. Bioinformatics 21:i487-494

CHAPTER 3. BAYESIAN NETWORK IN GENETIC ANALYSIS FOR RISK ASSESSMENT IN T1D

3.1 Introduction

Type 1 Diabetes is one of the most common multifactorial and polygenic diseases in which genetic and environmental risk factors interact with each other to determine the onset of disease. It represents a challenge for researchers that have so far tried answer to many questions raised regarding the different causes underlying the etiopathogenesis of the disease. However, there are still many open questions: what is the risk of T1D onset conferred by the three main susceptibility genes HLA, INS and PTPN22? How are these genes associated in order to confer a global genetic risk? Do they have some role in influencing the age at disease onset?

The coexistence of multiple causes underlying the etiopatogenesis of the disease, the complex way in which they can interact with each other, the variability with which the genetic and environmental risk factors exert their protective or predisposing function to the disease and lastly, but not least, the difficulty of managing and evaluating all the available data, make the challenge even bigger.

In the following chapter we will explain the experimental work carried out, using the probabilistic method introduced in the previous chapter, to evaluate the risk for T1D onset conferred by an individual susceptibility gene and the joint action of different association between risk classes of HLA-DR, INS-VNTR and PTPN22 using Bayesian networks and the probabilistic rule theorized by Reverend T. Bayes. The interaction between genes and the role of

the single gene in influencing the age at disease onset were also studied by applying the same probabilistic approach.

3.2 Data sets used for the study

Four independent data sets collecting genetic data (HLA-DR, INS-VNTR and PTPN22) were analysed. All samples were from Caucasian populations from three different countries (see table 1).

The case-control *French* cohort (data set A) consisted of 868 T1D patients (M/F 0.84, mean age at T1D onset 19.63 ±14.40 yearsyears) and 73 French control subjects (0.63 M/F ratio). It was a subset of a cohort consisting of French Caucasian T1D patients and controls, recruited in three hospitals in Paris and Lille (Ghandil P, 2005).

The case *Italian* sample (data set B) collected in Lazio region consisting of 391 T1D patients (M/F 1.06 , mean age at T1D onset **14.31 ± 8.84**); T1D patients were diagnosed by participating centres of the IMDIAB group in continental Italy and the diagnosis of T1D was based on the ADA classification criteria.

The *U.S.* and *Canada* data set (data set C) consisted of 705 T1D patients (M/F 0.84 , mean age at T1D onset 12.38 ± 7.03 years). The American data set consisted of only information in HLA and INS genes from T1D patients recruited by the Genetics of Kidneys in Diabetes (GoKind) study Group. Diabetes was diagnosed before age 31 and all patients were characterized by several other eligibility criteria specified in related references (Mueller PW, 2006).

French nuclear families data set (data set D) with 1694 patients (M/F 1.2, mean age of T1D onset **14.5±*10.3***) and 2340 controls (M/F 0.88). Only two phenotypes, with or without the disease, and no intermediate phenotypes such as antibody positives with normal glucose were considered. Data sets used in our analysis consisted of less data than the originals because we excluded individuals with some missing genotypes.

Gene typing

All individuals taking part in these studies gave their informed consent for genetic studies. DNA was extracted from blood using standard techniques and genotyping of genes was performed on specific methods for each data set (Buzzetti R, 2004; Ghandil P 2005; Muller PW 2006).

Gene risk classification

All genetic information was classified based on the T1D susceptibility or not susceptibility as follows:

HLA. Subjects were grouped from highest (DR3/DR4) to intermediate (DR3/DR3 or DR4/DR4; DR3/DRx or DR4/DRx) to the lowest (DRx/DRx, where x is other than DR3 or DR4 allele) HLA genotypes for T1D risk (Fourlanos S, 2008).

"HLA-DR" node represented in the BN graph consisted of all these three risk classes.

INS gene. A/T and T/T were considered as the non-susceptibility alleles and

AA as the susceptibility allele (Bennett ST, 1995)."INS" node represented in the BN graph, consisted of both risk classes.

PTPN22 gene. C/C was considered as the non-susceptibility allele T/T and C/T as the susceptibility alleles (Bottini N 2004; Lee YH 2007).

"PTPN22" node represented in BN graph, consisted of both risk classes.

For a clearer exposition of the experimental work carried out, below are three different paragraphs (first, second and third aim) each consisting of focused rationale, methods, results and conclusion sections.

3.3 First aim: Assessment of the risk conferred by HLA-DR, INS-VNTR and PTPN22 genes on the onset of T1D and evaluation of the joint risk conferred by the three susceptibility loci using the Bayesian network approach.

Rationale

Determining extreme genetic risk is a fundamental prerequisite for the implementation of primary prevention trials.

About T1D disease, it is well known that the major T1D susceptibility locus maps to the HLA class II genes at 6p21 and accounts for up to 30%–50% of genetic T1D risk (Noble JA, 1996). Other non-HLA T1D loci as the insulin gene (INS) on chromosome 11p15 (Bell GI, 1984), the polymorphic, cytotoxic T-lymphocyte associated protein 4 (CTLA4) gene on chromosome 2q33 (Nistico L, 1996), the protein tyrosine phosphatase, non-receptor type 22 (lymphoid) (PTPN22) gene on chromosome 1p13 (Bottini N , 2004) and other recently discovered loci from genome wide association (GWA) studies, that, compared to HLA alleles, have smaller effects on disease risk. (Todd JA, 2007) (Fig.1).

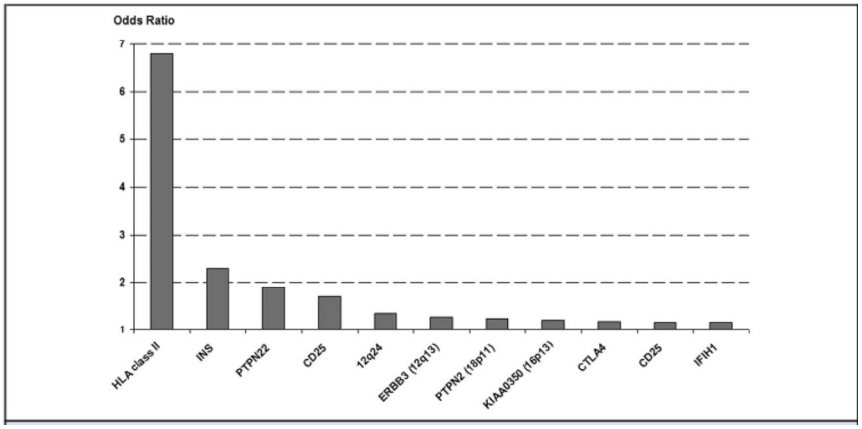


Fig. 1. ORs for confirmed "genes/genetic loci" associated with T1D.
Adapted by permission from Macmillan Publishers, © 2007; Todd et al. (5) and reprinted with permission from Elsevier, © 2008;

The risk for T1D in siblings of patients is 15-fold higher than the risk for T1D in the general population, which suggests that genetic factors play an important role in disease susceptibility. The pattern of inheritance is complex, and the development of disease is thought to be determined by an interaction between genetic predisposition and environmental triggers. There is high familial clustering with a mean prevalence of 6% in siblings compared to 0.4% in Caucasian populations, although more than 85% of patients with T1D lack a positive family history for the disease.

The two primary approaches used to identify risk loci for T1D have been linkage studies and association studies. Linkage studies, typically using affected sibling pairs, can identify regions of the genome that are shared more frequently among affected relatives.

In contrast to linkage studies, association studies can detect alleles with much more modest effects on risk as long as those alleles are relatively common. All of the four well-established risk loci, including HLA, INS, CTLA4, and PTPN22, were identified in candidate gene association studies.

The risk can be further stratified by recruitment of subjects with susceptible genotypes in a case-control study and by selection of children with a multiple family history of diabetes by collecting genetic data of a family cluster.

The aim of this work was to assess the risk conferred by HLA-DR, INS-VNTR and PTPN22 single genes in the onset of T1D and the joint risk conferred by all these three susceptibility loci, using the Bayesian network approach in both population-based case-control and family clustering data sets.

3.3.1 Methods

Data sets A (case-control sample) and D (patients and controls from family clusters) were analysed. Only the data sets consisting of both patients and controls having full genetic information for all three genes of interest, HLA-DR, INS-VNTR and PTPN22 were considered (namely, "A" case-control data sets and "D" family data set). We excluded from this analysis "B" and "C" data sets because they consist of patients data only. We analysed both samples separately using the BNs approach (Pearl 1988).

***BN model*:** Bayesian Networks provide a powerful and flexible tool for reasoning under uncertain conditions (Shafer G, 1997 - Cowell R.G et al, 1999 - Jensen F.V, 2001 - Pearl J, 1998). As explained in more detail in chapter two BN is a graphical model that encodes probabilistic relationships among variables of interest. Each variable is represented graphically by a node and the links (edges) between nodes correspond to the probabilistic dependence between variables. Furthermore, each node has a conditional probability table, quantifying the relationship between connected variables. Users can set the values of any combination of nodes in the network that they have observed. This evidence propagates through the network, producing a new probability distribution over all the variables in the network. (for more details see also chapter two).

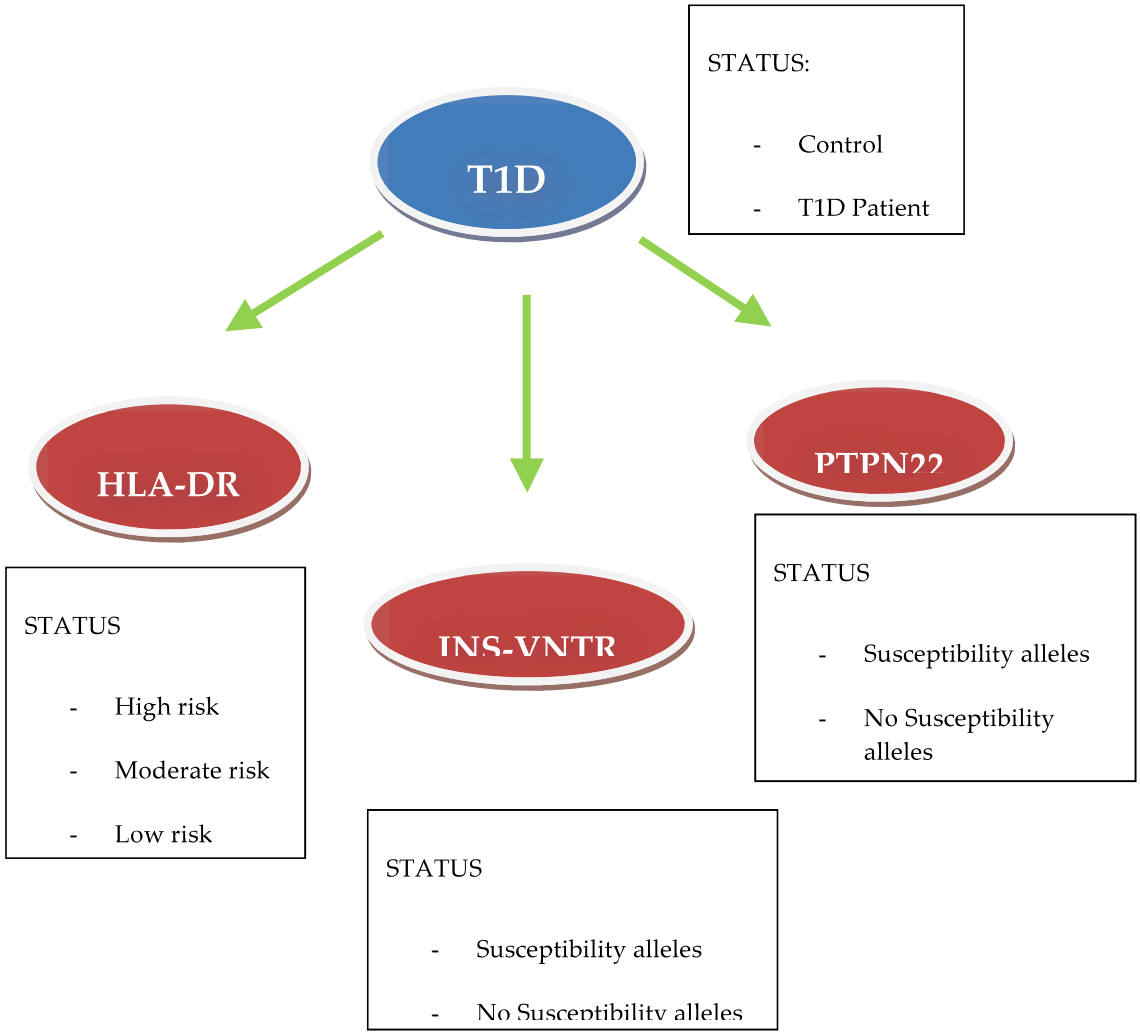
An important reason for choosing BN as a valid representation of such problem solving is that the system is designed to support qualitative reasoning. A physician engaged in medical diagnosis proceeds in a highly selective manner (Elstein AL, 1978). This progressive selectivity can correspond to the kind of links representing independencies/dependencies

between variables, modelled in BN. More generally, there is empirical evidence that qualitative reasoning by relationship between variables, as supported by BN, corresponds closely to human reasoning patterns (Henrion M, 1987 - Heckerman D, 1991 - Pearl J, 1993 - Waldmann MR, 1992).

Two different modes of information flow can be observed in the BN: By entering causes into the network the effects can be observed. This would correspond to a causal flow of information. On the other hand a diagnostic approach would be to enter symptoms into the network and assess which diseases could be responsible for them. In our case, we can enter genetic data and ask BN what is the risk of having the disease.

Variables as nodes: In our graph, HLA-DR, INS-VNTR and PTPN22 genes variables are denoted as nodes. Each node can have a set of states corresponding to risk classes (high, moderate and low risk for HLA, susceptibility/non-susceptibility for INS-VNTR and PTPN22). T1D node consisted of 2 states: control and diabetic subjects (see Fig. 2).

Figure 2. Bayesian Network implemented to assess risk to have T1D.



***The meaning of the direction of the arrows*:**

Associated with the arrow that links two nodes, is a conditional probability table that estimates the value of the likelihood of the state of the second gene given the state of the first gene. It is possible to guide BN learning process by introducing prior knowledge algorithm in the structure.

In the first phase of the study, our primary information was about the healthy and the diabetic status of the subjects recruited and this information was represented in the BN model by the node "T1D". Based on this information ,

the genotyping analysis was carried out for each of the patients and controls. The genetic data was thus obtained based on the "status" of the subject. For this reason the direction of the arrow directed from the node "T1D (healthy / sick)" to the genes nodes has been imposed. By choosing edge direction we can guide the learning process of the BN, following collected data direction, from healthy/diabetic status to the genotype.

Data analysis

Prior knowledge and learning process

Bayesian approach makes it possible to systematically integrate experimental data with multiple sources of "prior" knowledge (called "prior" value), as the existing large body of published literature. In our case the prior was the prevalence value of T1D in French and Italian populations, that is 0.4% (Steck A, 2011). When the family data was analysed a prevalence value of 6% was considered as reported in literature for risk for T1D in siblings (Steck A, 2011) . Prevalence value changes between different populations around the world and with it also the risk conferred by risk factors based on that. For this reason introducing the right prior information is crucial in the evaluation of the risk factors.

After introducing prevalence value, BN algorithm updated automatically population composition and with it also the percentage of three gene risk classes were adjusted accordingly (e.g. percentage of High, Moderate and Low risk classes for HLA were 32%, 55%, 12% before entering the prevalence value and 2.8%, 34.3% and 62.7% after that respectively).

BN was ready to be questioned about the genetic absolute risk values.

After the process of learning from data, implemented with the R program, BN was questioned on the T1D genetic risk conferred by all single susceptibility genes, each one separately, in both allelic risk status.

Moreover, for all possible status combinations of the three genes, a risk value was calculated and the best network fitting data was selected using AIC score and P-value.

This work was done for A and D familiar clustering data set.

Significance parameters: P-Value and AIC score

The measure of the quality of a BN can be computed using several scores. The goal of the score is to figure out how accurately models will predict new data when fitted to the old. Here AIC score (Akaike Information Criterion) (Akaike 1974) was chosen. AIC score, as an estimator of predictive accuracy, figures out how accurately BN models define the relationship between nodes. According to this criterion the network with the highest AIC score was selected as the best network.

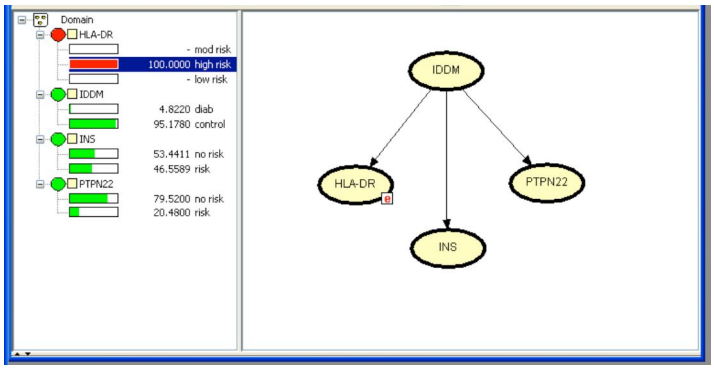
The validity of the edges can be measured by testing the mutual information between a parent node and the corresponding child. The mutual information can then be compared to a chi-square distribution. The corresponding p-value can be seen as the strength of the edge and so of the relationship between the nodes.

3.3.2 Results: Case-control study (Data set: A)

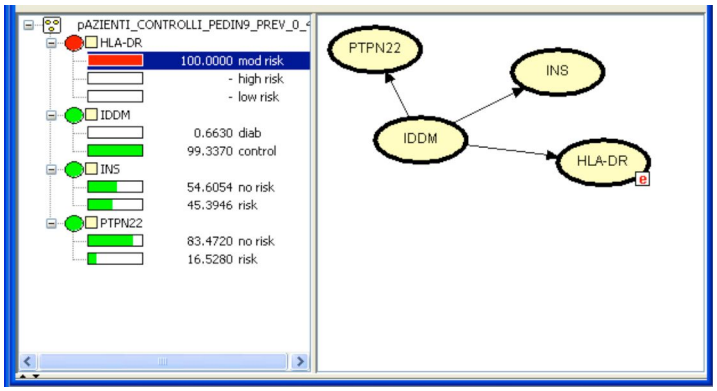
Genes – T1D correlation. When the relationship between HLA, INS and PTPN alleles was evaluated separately, a significant association between genes and T1D was found (p=0.003 for HLA, p=0.9 $*10^{-3}$ for INS gene and p= $1.5*10^{-5}$ for PTPN22).

HLA-DR alleles. When HLA-DR alleles were considered, the risk values of having T1D were 4.8%, 0.6% and 0.05% for high, moderate and low risk HLA alleles, respectively. The following figures show the conditional probability (P) to develop T1D if an individual carries high, moderate and low risk HLA alleles. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

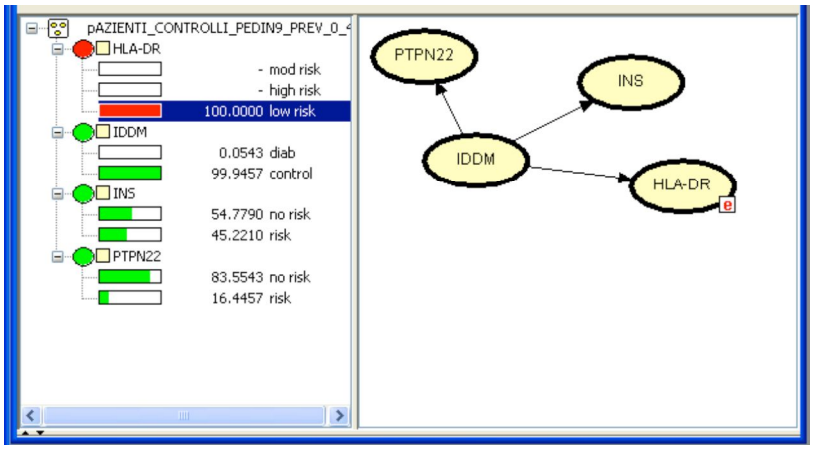
P(D|HLA high risk)



P(D|HLA moderate risk)



P(D|HLA low risk)

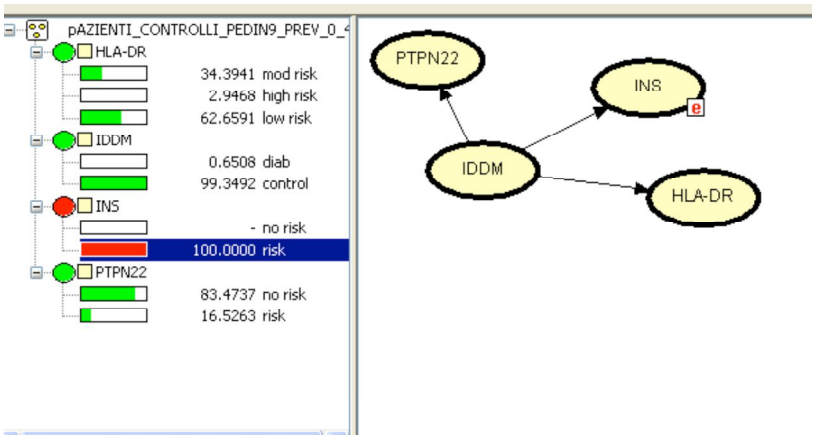


INS gene. When INS alleles were considered, the risk values of having T1D were 0.6% and 0.19% for INS risk and no risk alleles, respectively.

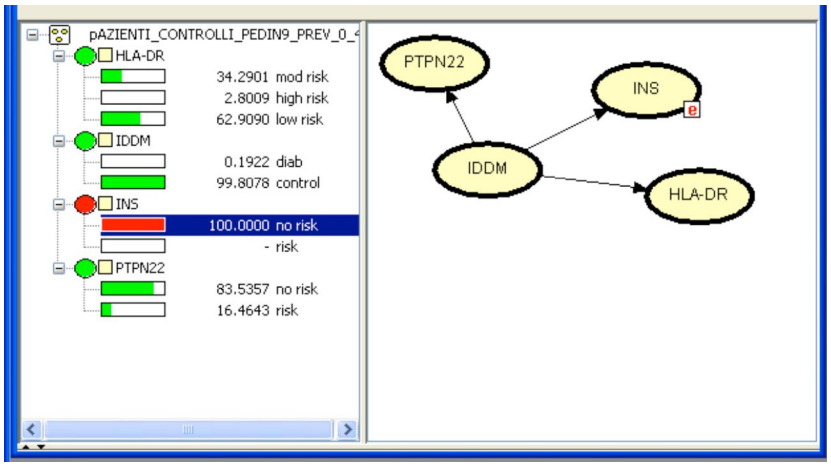
The following figures show the conditional probability (P) to develop T1D if an individual carries susceptibility or non-susceptibility INS genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

Risk conferred by INS alleles :

P(D|INSrisk)



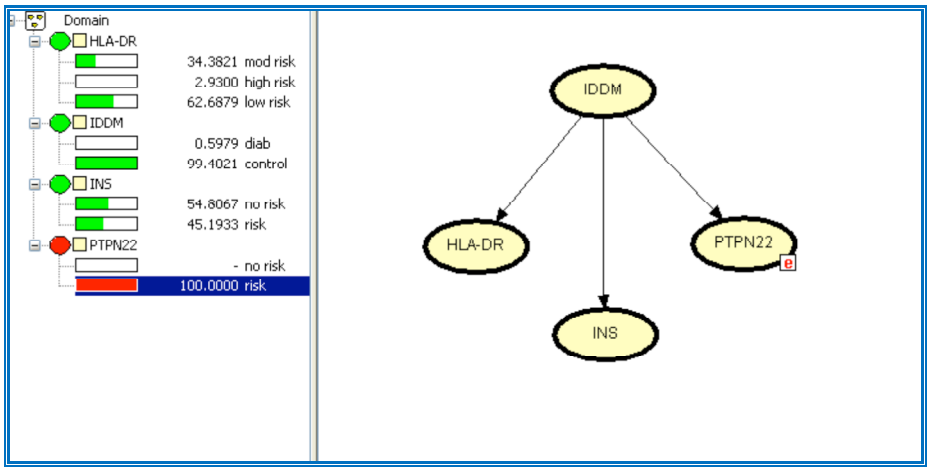
P(D|INS no risk)



PTPN22 alleles. When PTPN22 alleles were considered, the risk values of having T1D were 0.6% and 0.35% for PTPN22 risk and non-risk alleles, respectively.

The following figures show the conditional probability (P) of developing T1D if an individual carrying susceptibility or non-susceptibility PTPN22 genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

P(D|PTPN risk)



P(D | PTPN no risk)

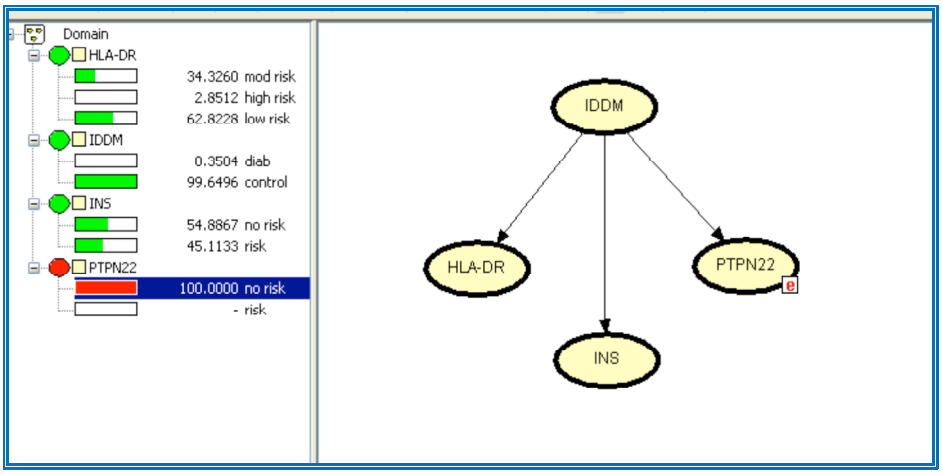


Table 1. Risk analysis using OR parameter and BN algorithm to evaluate single-locus main effects.

Genes	Cases (N)	Controls (N)	OR	95% CI	BN risk (%)
HLA					
High risk	300	2	18.7	4.5-76.9	**4.8**
Moderate risk	494	25	2.51	1.5-4.1	**0.6**
Low risk	74	46	0.05	0.03-0.09	**0.05**
Total	868	73			
INS					
Risk	640	33	3.4	2.09-5.52	**0.65**
No risk	228	40	1		**0.19**
Total	868	73			
PTPN22					
Risk	260	12	2.1	1.1-4.1	**0.6**
No risk	608	61	1		**0.35**
Total	868	73			

Joint effect of HLA and INS: We found significant heterogeneity in the distribution of the INS genotypes (susceptibility/non-susceptibility) in the three HLA risk classes when the patients' group was considered (Tab. 2), in agreement with previous literature results (Motzo C, Diabetes 2004).

Table 2. Distribution of the INS genotype in the three HLA risk classes in case subjects.

	INS susceptibility	INS non-susceptibility
HLA high risk	68.3%	31.4%
HLA moderate risk	77.3%	22.7%
HLA low risk	71.6%	28.4%

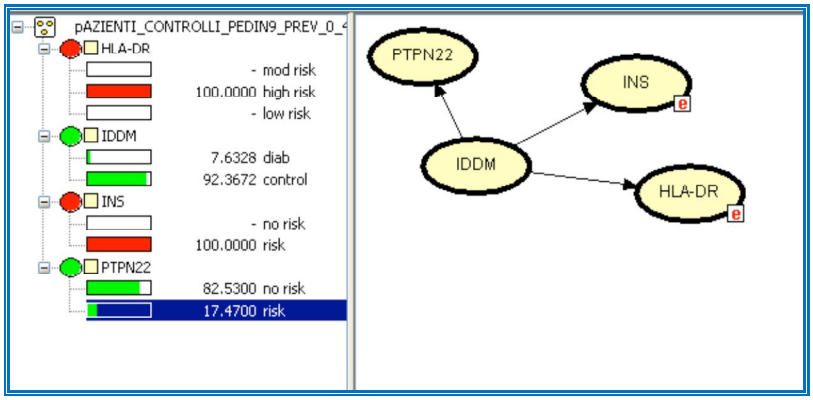
Analysis of this heterogeneity showed that the INS predisposing genotype was more common in both moderate and low-risk HLA genotype patients (77.3%and 71.6%, respectively), than in patients carrying HLA high risk genotypes (68.3%) (chi-square = 7.98, degrees of freedom = 2, p = 0.01).

Moreover, no evidence of heterogeneity was observed in the distribution of the INS genotypes in the control subject group in which the INS genotype was similarly distributed in control subjects within different HLA risk classes (p= 0.2).

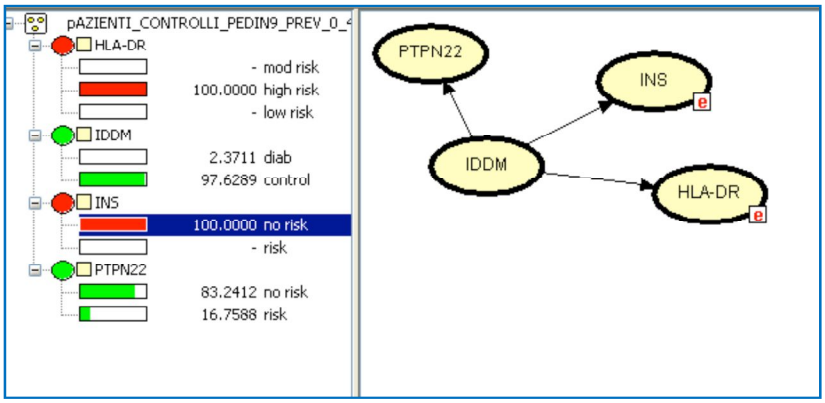
The joint risk of having T1D in a subject with high or moderate or low risk alleles at HLA-DR gene, in association with risk or non-risk INS alleles, was calculated using the BN algorithm. Our results showed that the relative impact of variation at INS within each genotype category at the HLA-DRB genotype were detectable in all of the HLA genotype categories. The absolute risk was

of 7.6% and 2.3% if INS risk or non-risk respectively were present in individuals with high-risk HLA, 1.07% and 0.3% in individuals with intermediate-risk HLA, and 0.08% and 0.02% in individuals with low-risk HLA genotypes (Table 3). The following figures show the conditional probability (P) to develop T1D if an individual carrying either high, moderate or low risk HLA and susceptibility or non-susceptibility INS genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

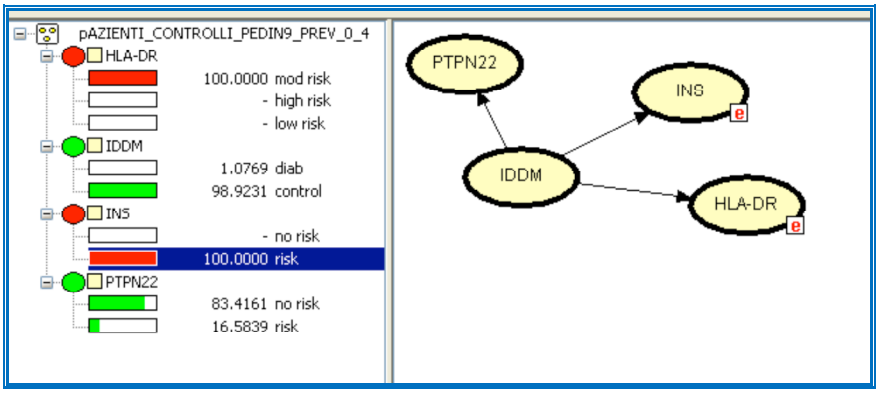
P(D|HLA high, INS risk)



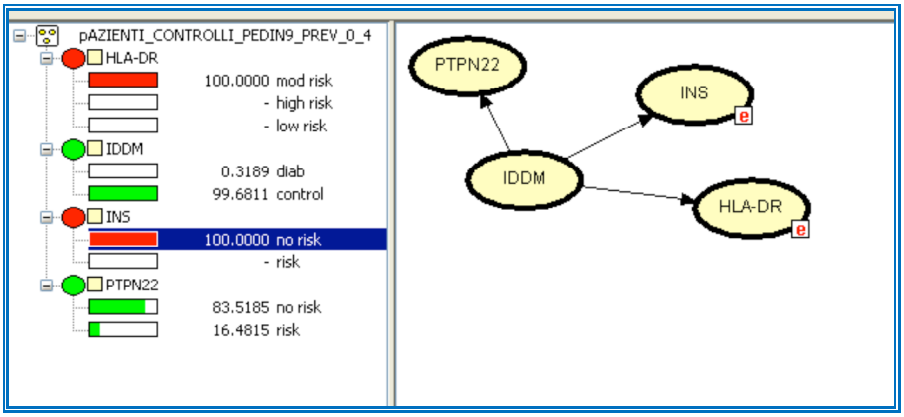
P(D|HLA high, INS no risk)



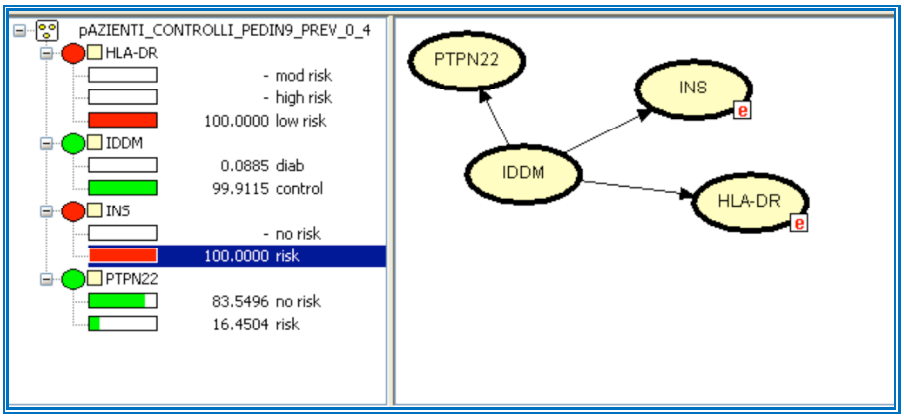
P(D | HLA moderate, INS risk)



P(D | HLA moderate, INS no risk)



P(D | HLA low, INS risk)



P(D | HLA low, INS no risk)

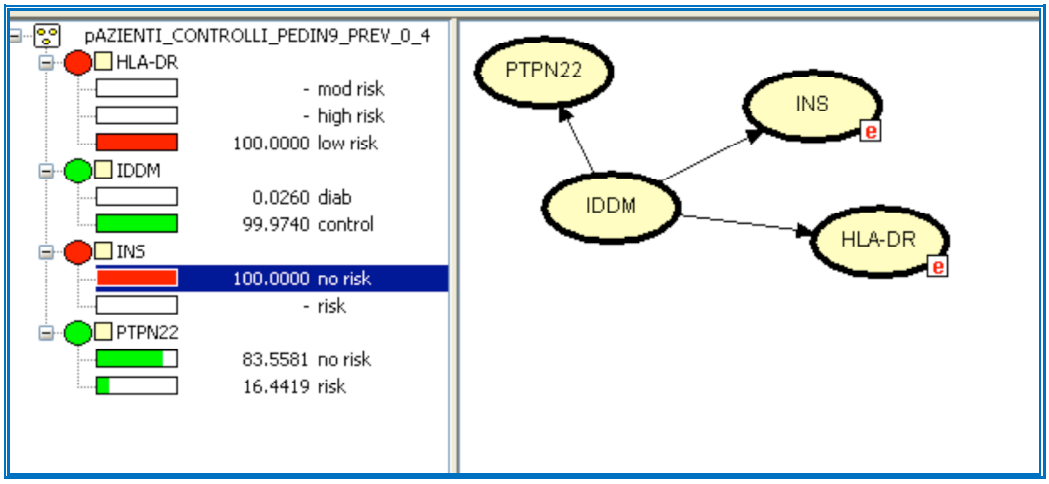


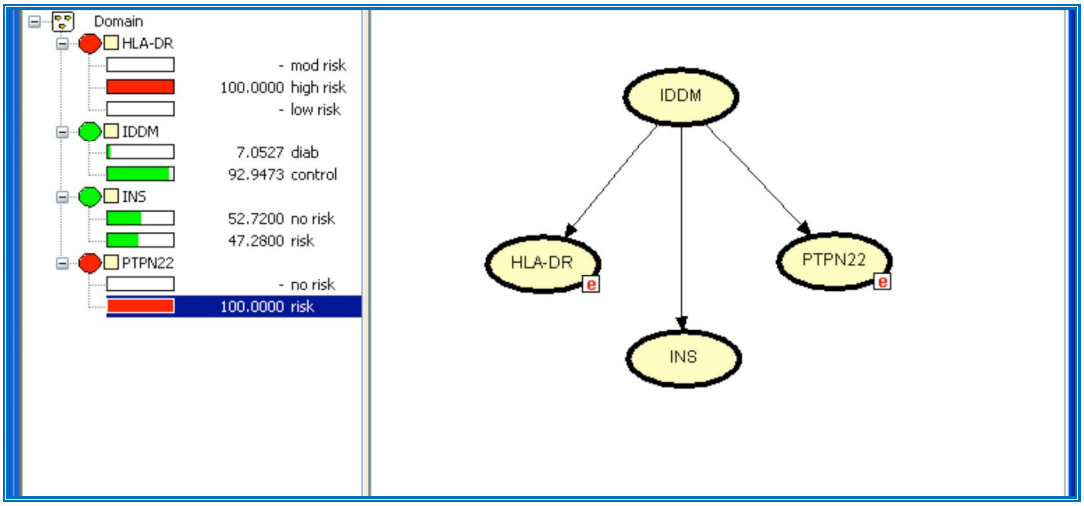
Table 3. Interaction between HLA-INS in the case-control data set using OR and BN algorithm

HLA category	INS alleles	Case N (%)	Controls (%) N (%)	OR	95% CI	BN risk %
High risk	**Risk**	205 (23.6)	2 (2.7)	0.43	0.02-9.05	**7.6**
	No risk	95 (11)	0 (0.0)			**2.3**
Moderate risk	**Risk**	382 (44)	11 (15)	4.29	1.92-9.56	**1.07**
	No risk	112 (13)	14 (19,2)			**0.31**
Low risk	**Risk**	53 (6.1)	20 (27.4)	3.22	0.66-3.79	**0.08**
	No risk	21 (2.3)	26 (35.7)			**0.02**
		868	73			

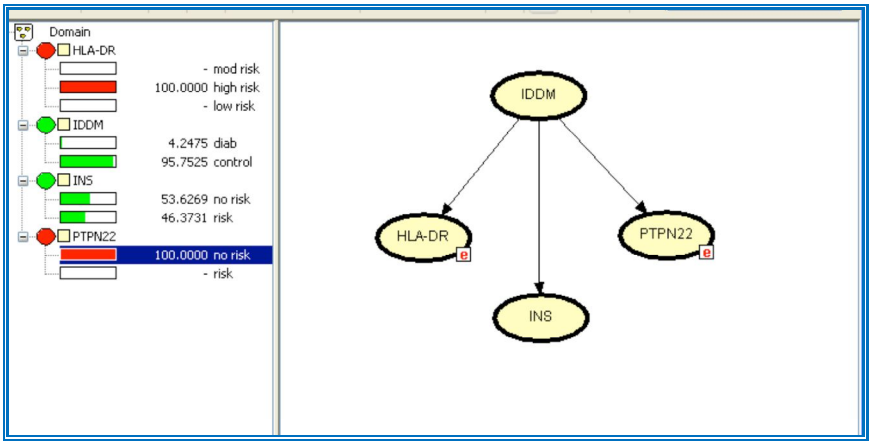
Joint effect of HLA and PTPN22. The joint effect of PTPN22 and HLA varied across the HLA risk categories (as shown in Table 4). The absolute risk was 7% and 4.2% if PTPN22 risk or non-risk respectively were present in individuals with high-risk HLA, 1% and 0.7% in individuals with intermediate-risk HLA and 0.08% and 0.05% in individuals with low-risk HLA genotypes (Table 4).

The following figures show the conditional probability (P) of developing T1D if an individual carrying both high, moderate or low risk HLA and susceptibility or non-susceptibility PTPN22 genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

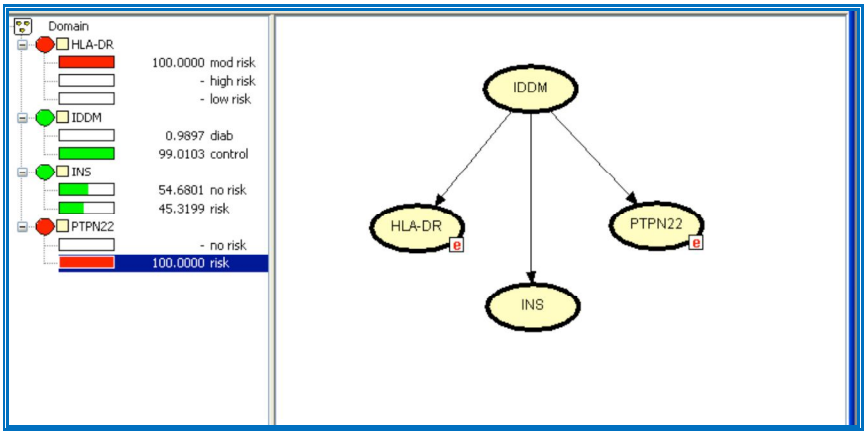
P(D|HLA high, PTPN22 risk)



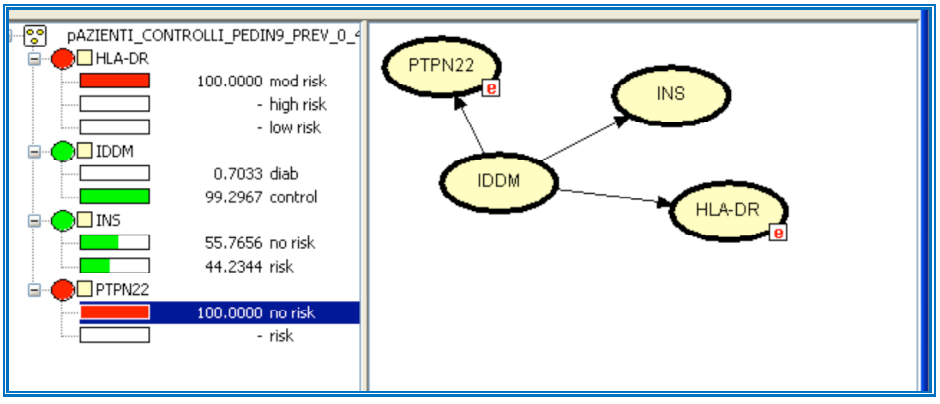
P(D|HLA high, PTPN22 no risk)



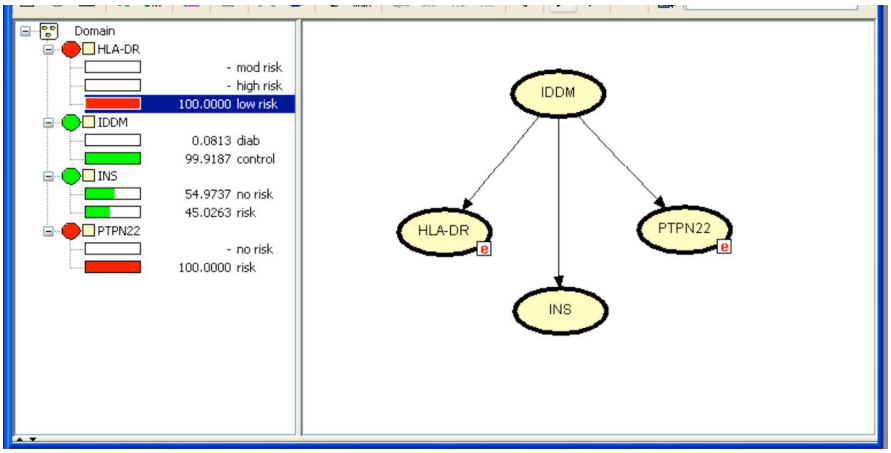
P(D|HLA moderate, PTPN22 risk)



P(D|HLA moderate, PTPN22 no risk)



P(D|HLA low, PTPN22 risk)



P(D|HLA low, PTPN22 no risk)

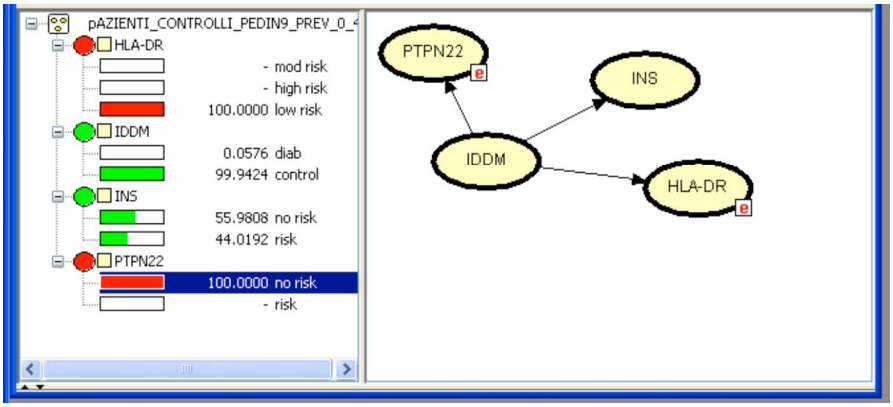
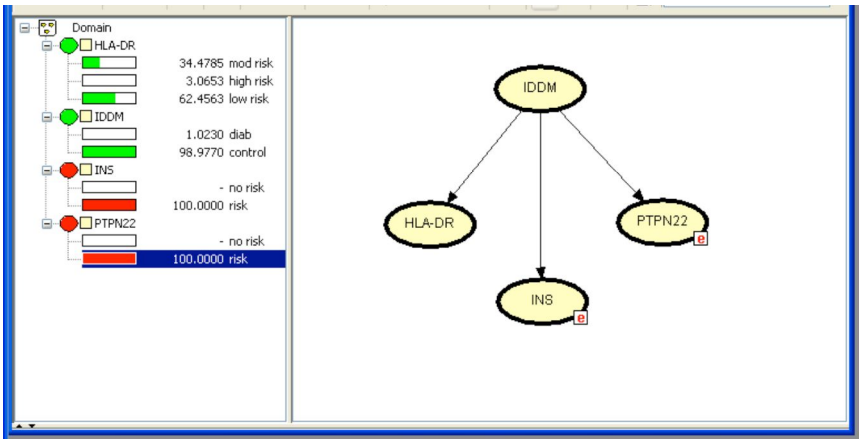


Table 4. Interaction between HLA-PTPN22 in the case-control data set using OR and BN algorithm

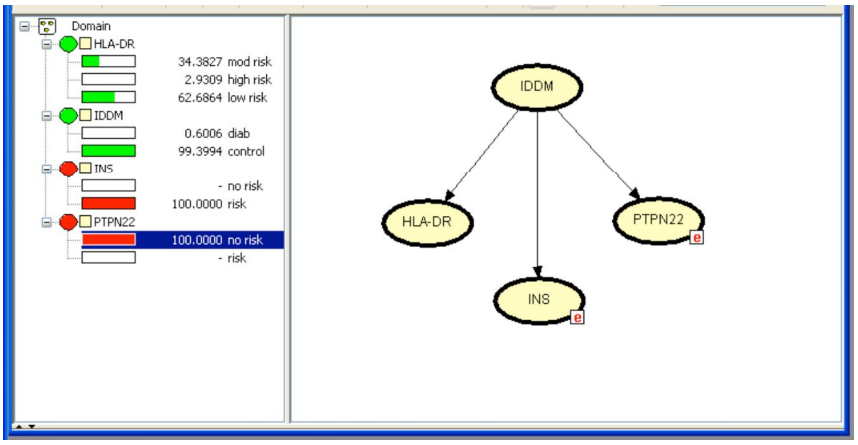
HLA category	PTPN22 alleles	Case N (%)	Controls (%) N (%)	OR	95% CI	BN risk %
High risk	**Risk**	83 (9.5)	0 (0)	**1.91**	0.09-40.4	**7**
	No risk	217 (25)	2 (2,8)	**1**		**4.2**
Moderate risk	**Risk**	156 (18)	3 (4.2)	**3.38**	0.99-11.4	**0.98**
	No risk	338 (39)	22 (30)	**1**		**0.7**
Low risk	**Risk**	21 (2.4)	9 (12)	**1.58**	-0.67-3.95	**0.08**
	No risk	53 (6.1)	37 (50)	**1**		**0.05**
		868	73			

Joint effect of INS and PTPN22. When INS and PTPN22 risk alleles were evaluated in order to establish the risk of T1D onset, the risk values were equal to 1% for subjects with both INS & PTPN22 risk alleles. The risk decreased to 0.15% when both these genes were present with non-risk alleles. The following figures show the conditional probability (P) to develop T1D in an individual carrying both susceptibility and non-susceptibility INS and PTPN22 genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

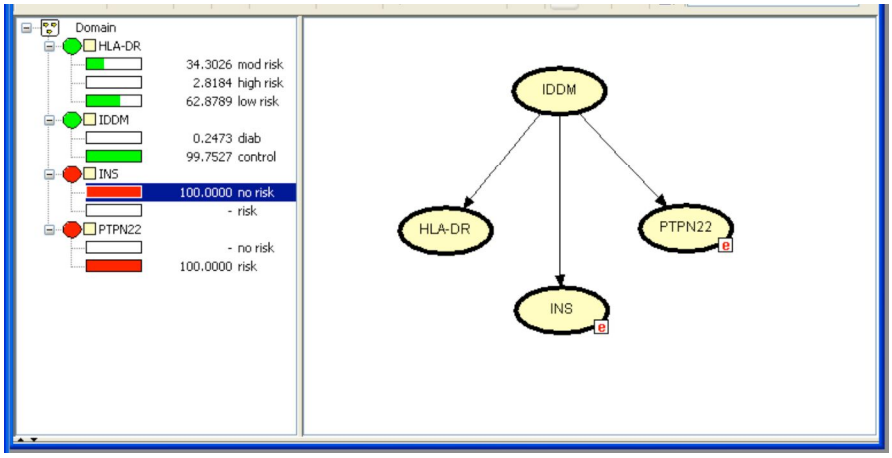
P(D|INS Risk, PTPN22 risk)



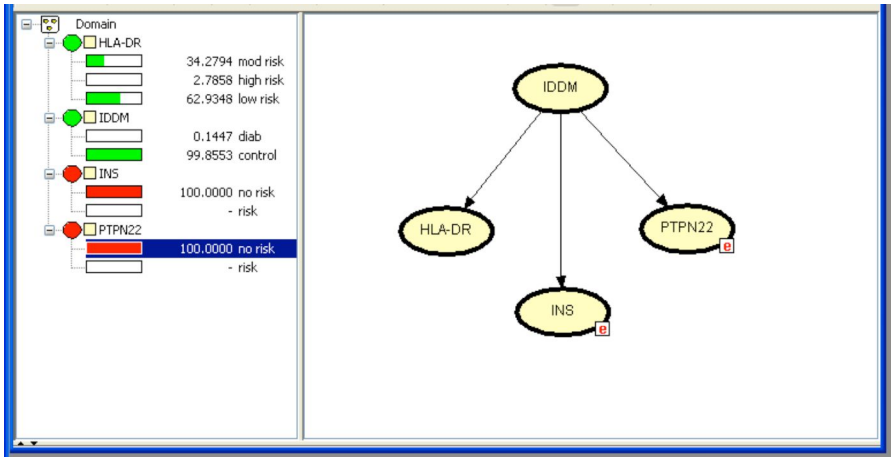
P(D|INS Risk, PTPN22 no risk)



P(D|INS no risk, PTPN22 risk)



P(D|INS no risk, PTPN22 no risk)



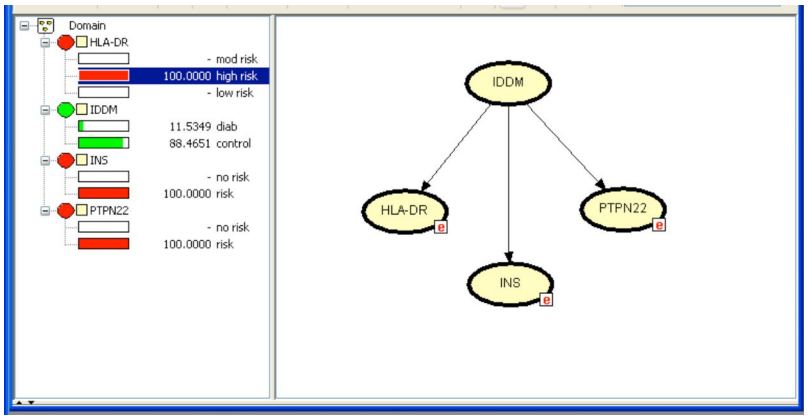
Joint effect of HLA, INS and PTPN22

We also tested models with all three-way interactions involving the three susceptibility loci. The results showed that the more risk loci an individual carries, the higher the absolute risk, but the presence or absence of HLA risk loci influences the absolute risk much more than the other loci, as expected. For instance, carrying risk genotypes as the two non-HLA loci but not at HLA (as shown on page 29, P(D|HLA low, INS risk, PTPN22 risk)) is associated with a much lower risk than HLA risk genotypes together with low-risk genotypes at the other two. The absolute risk (BN risk shown in table 5)

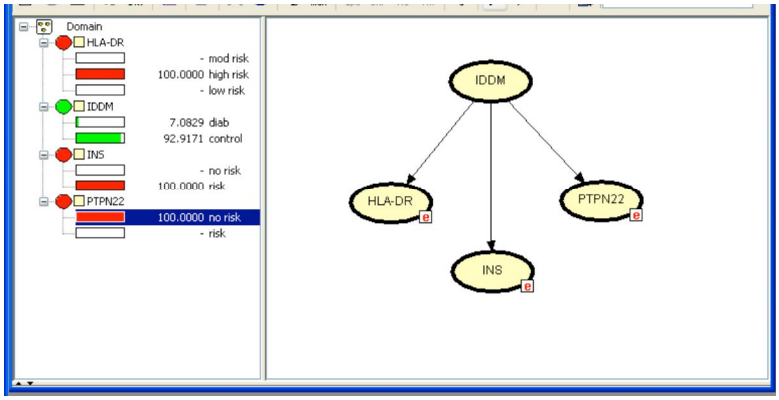
conferred by simultaneously carrying high, moderate or low-risk HLA and risk genotypes to all the other two loci, compared with non-risk-associated genotypes at all three loci, was 11.5%, 1.7% and 0.1%, respectively.

The following figures show the conditional probability (P) to develop T1D in an individual carrying risk or protective genotypes of HLA, INS and PTPN22 genes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

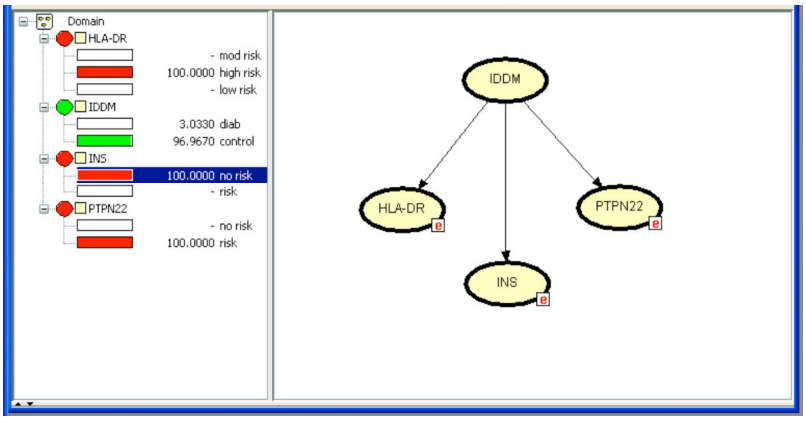
P(D|High Risk, INS Risk, PTPN22 risk)



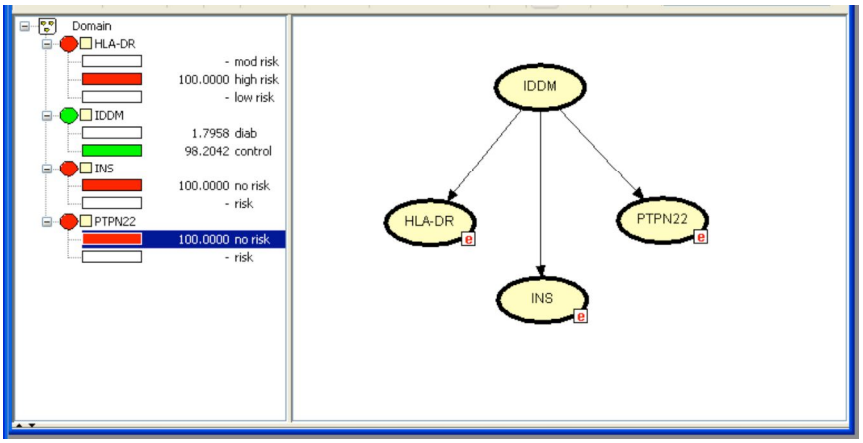
P(D|High Risk, INS Risk, PTPN22 no risk)



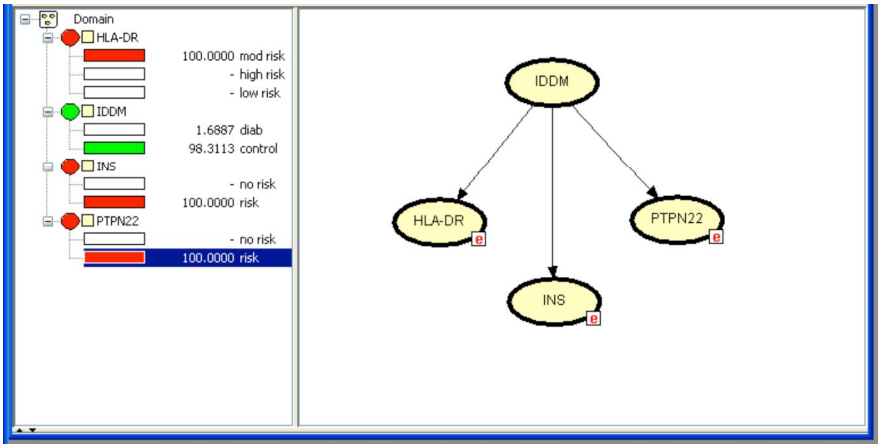
P(D | High Risk, INS no Risk, PTPN22 risk)



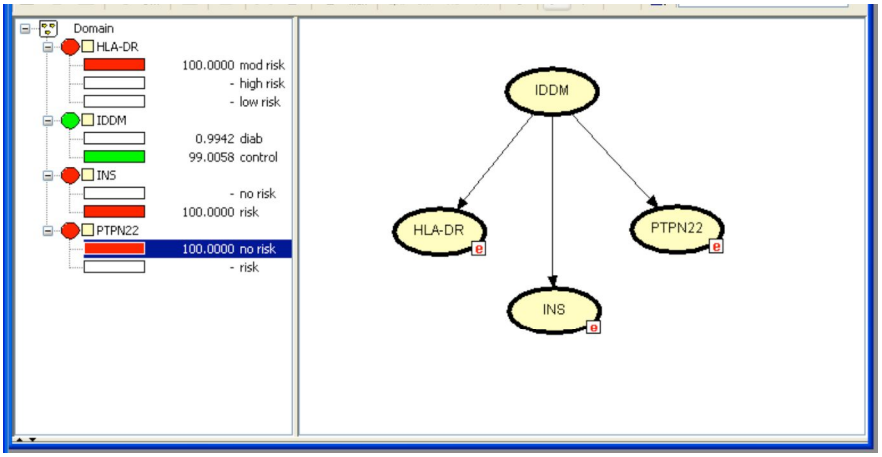
P(D | High Risk, INS no Risk, PTPN22 no risk)



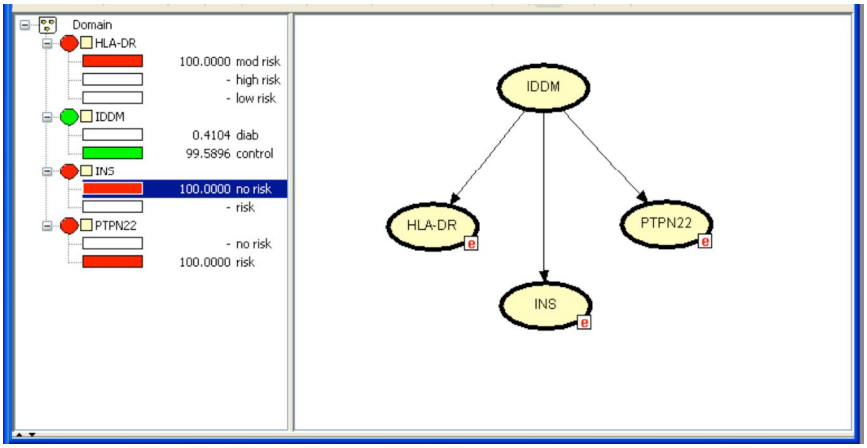
P(D | Moderate Risk, INS Risk, PTPN22 risk)



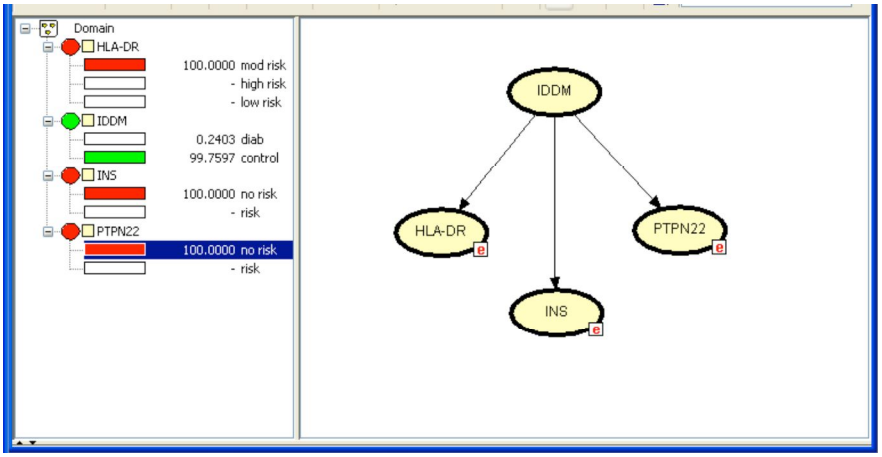
P(D | Moderate Risk, INS Risk, PTPN22 no risk)



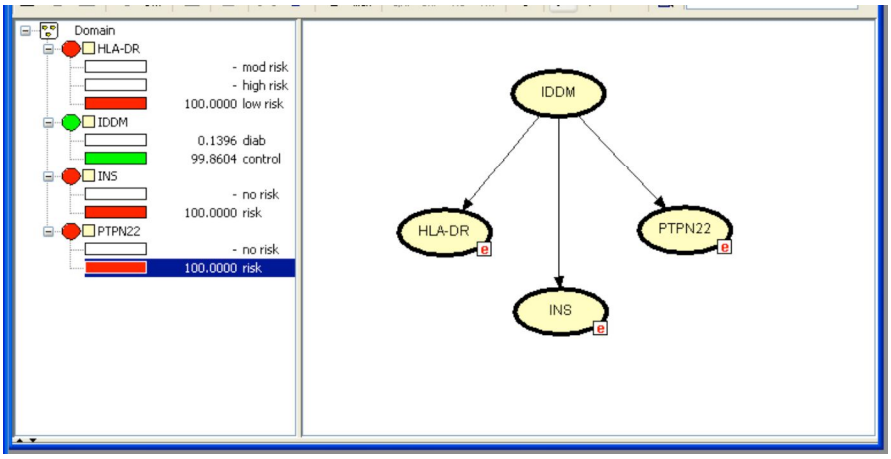
P(D | Moderate Risk, INS no Risk, PTPN22 risk)



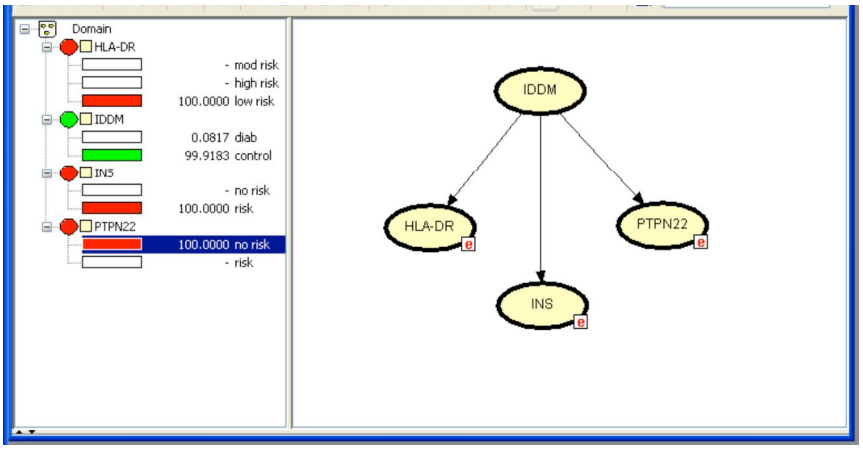
P(D | Moderate Risk, INS no Risk, PTPN22 no risk)



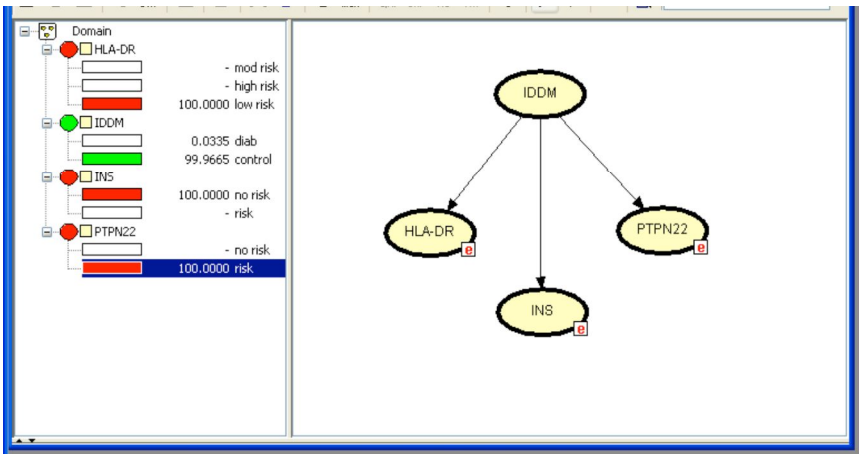
P(D | Low Risk, INS Risk, PTPN22 risk)



P(D | Low Risk, INS Risk, PTPN22 no risk)



P(D | Low Risk, INS no Risk, PTPN22 risk)



P(D | Low Risk, INS no Risk, PTPN22 no risk)

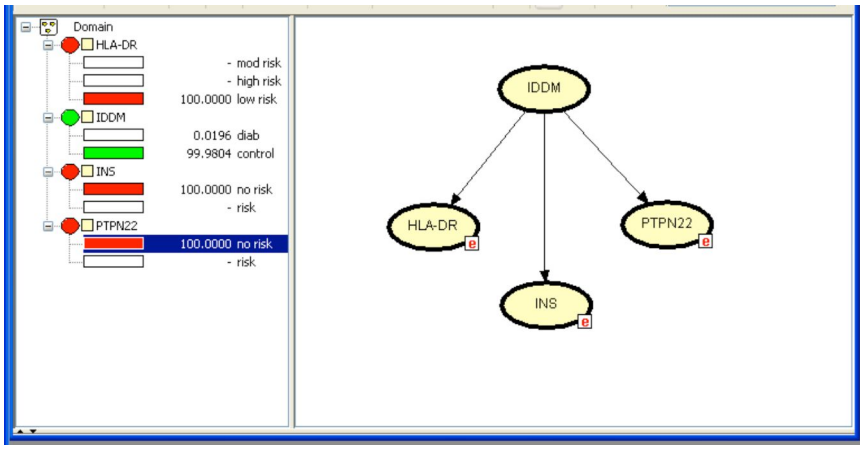


Table 5. Distribution of risk genotypes at all three susceptibility loci among cases and controls

HLA classes	INS1	PTPN1	Cases	Controls	OR (95%CI)[2]	BN risk
High	R	R	51	0	108(6.2-189)	11.5
High	R	NR	154	2	81(17-377)	7
High	NR	R	32	0	68(4-1201)	3
High	NR	NR	319	0	673(39-11597)	1.8
Moderate	R	R	118	1	124(15-988)	1.7
Moderate	R	NR	264	10	27(11.3-68)	1
Moderate	NR	R	38	2	20(4.2-95)	0.4
Moderate	NR	NR	74	12	6.5(2.6-15)	0.24
Low	R	R	18	2	9.5(1.9-46)	0.1
Low	R	NR	35	18	2.05(0.8-4.8)	0.08
Low	NR	R	3	7	0.4(0.1-2)	0.03
Low (reference)	NR	NR	18	19	1	0.02

,[1] "R"genotype at given locus associated with risk; "NR", genotype at given locus not associated with risk; [2] Odds ratio vs single reference group without risk genotype at any of the four loci.

3.3.3 Conclusions

The present study is a comprehensive evaluation of the joint effects of the three most well established T1D susceptibility genes in a case-control data set, from Caucasian French population, to assess the joint genetic risk to develop T1D, based on the genotype variation at these loci.

When the joint risk conferred by susceptibility loci was compared to the joint risk conferred by non-susceptibility genotypes for all three genes, the absolute risk values were 11.5% and 0.02%, respectively.

Our results, in agreement with some of the previous studies, confirmed that HLA-DR is the most relevant susceptibility gene compared to INS and PTPN22 and proved that the INS and PTPN22 genotypes marginally influence T1D risk in all HLA genotype risk categories (**Table 5)** (Bain SC, 1992; Walter M, 2003; Motzo C, Diabetes 2004).

Confirming earlier observations about the heterogeneity in the relative effects of INS (van der Auwera B, 1995; Metcalfe KA, 2004), we also found evidence that the INS predisposing genotype is significantly less frequent in high-risk HLA genotype–positive patients than in those with HLA intermediate and low-risk categories. As regards the composition of the French data set, where we do not have control subjects characterized by the following genotypes HLA high risk + INS risk + PTPN risk, HLA high risk + INS no risk + PTPN risk, HLA high risk + INS no risk + PTPN no risk, the statistically significant relation between HLA and INS genes, needs to be verified against further data sets. For this reason, such a relation has not been taken into account in the construction of the Bayesian Network because of the estimation of the risk of T1D onset.

Moreover, also PTPN22 susceptibility alleles conferred, albeit less than INS gene, a higher risk for T1D, both when compared with absolute risk in the general population and when associated with HLA gene. The relative risk conferred by PTPN22 was stronger in the lower-risk HLA categories than in the high risk HLA category. On the other hand, the protective effect of non-susceptibility genotypes was stronger if INS rather than PTPN22 gene was considered, for all of the HLA risk classes. The joint risks assessed in this study were consistent with findings in literature (Motzo C, 2004; Bjornvold M, 2008). Motzo et al studied the joint effect on T1D onset of HLA and INS genes, in a case-control Sardinian cohort, whereas Bjornvold et al analyzed a sample of case-control subjects under the age of 15 years, with the aim of assessing the joint effect of the four main T1D susceptibility genes. Both studies used a T1D prevalence value of 0.4% and classified HLA and INS alleles in risk categories as we did in our study.

This study showed that a feasible and accurate risk assessment can be performed by applying the BN method. Here the effects of only three genes were evaluated and compared, but the BN method is able to analyze a huge amount of variables with different risk categories for each variable, at the same time. This feature could be crucial in the study of multifactorial diseases, in which the triggers involved in the complex mechanisms underlying disease pathophysiology are multiple. Studies in different populations and ethnic groups have indicated some heterogeneity in HLA-associated risk of T1D and it is also possible that gene–gene interactions may vary across populations. Therefore, genes could play a different role depending on the population in which data are collected (Thomson G, 2007). In fact a geographical stratification of T1D risk is essential because of potentially different mechanisms of gene-environment and gene-gene interaction in triggering

disease in the different countries. Furthermore, increasing the number of susceptibility loci considered simultaneously, generally increases the predictive value for disease. The downside is that the proportion of the population simultaneously carrying multiple risk alleles becomes minute and that even with relatively large data sets, as in our study, the absolute risk estimate becomes imprecise. The high-risk HLA genotype is carried by 2-3% of population controls and confers a very high risk of disease. Moreover, in our data set, as confirmed in literature (Bjørnvold M, 2008), only a small proportion of the population (and cases with T1D) simultaneously carry HLA and multiple non-HLA susceptibility genotypes.

By introducing "prior" knowledge from literature, we can also analyze *small* data sets while maintaining accuracy. In this study, thanks to the BN approach, a small sample consisting of 73 control subjects was analyzed and the training results matched with literature findings. Prevalence value was used here as prior and, based on that, the network was able to learn the correct rate of genotype combinations characterizing both the general population and patients group and to elaborate data, giving coherent results discussed before. Despite to the odds ratio parameter, shown in Table 5, BN analysis was not affected by lack of data about control subjects subgroups with specific genetic combinations (high risk HLA, INS and PTPN22 risk alleles; high risk HLA, INS and PTPN22 non-susceptibility alleles and high risk HLA, INS non-susceptibility alleles and PTPN22 susceptibility genotypes). In conclusion, the present work represents, to the best of our knowledge, the first study based on both case-control and familiar data sets, showing the joint effect of HLA, INS and PTPN22 in T1D in a Caucasian population with a heterogeneous age of T1D onset, generalizing previous findings regarding data sets consisting of patients and controls < 15 years by Bjørnvold M. et al. (Bjørnvold M, 2008) .

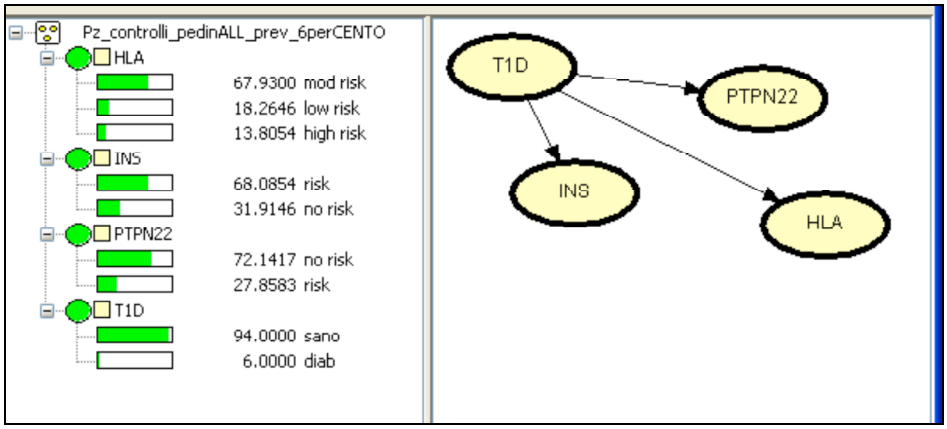
Our results showed that BN represents an alternative way to assess the joint risk to develop T1D, by considering different disease markers at one time. This method allows us to estimate the risk for each individual carrying a specific genes combination, even if a small data set is available. By collecting data regarding more genes and environmental risk factors, the BN approach gives us the opportunity to elaborate data sets with many variables, to create a population-based risk network and to also establish the relationship between risk markers, with the aim of assessing the global risk for such diseases and to define a personalized risk. On this basis, new predictive trials on the general population could be assessed. Although no preventive intervention is available for T1D today, prediction of disease is an important part of prevention strategies, both for recruitment of participants for research studies and for identification of target populations for future preventive interventions. Understanding the joint effect of the established T1D susceptibility genes will enhance this possibility.

3.3.4 Results: **Family study (Data sets_D)**

In this second part of this first aim, data set D consisting of 1964 patients and 2340 controls was analyzed. For each of the subjects HLA, INS and PTPN22 genotypes was considered.

Genes – T1D correlation. When the relationship between HLA, INS and PTPN alleles was evaluated separately, a significant association between genes and T1D was found ($p=0.01$ for HLA, $p= 1*10^{-3}$ for INS gene and $p= 3*10^{-4}$ for PTPN22).

The figure below shows the conditional probability (P) to develop T1D in the general population without any specific gene combination. This network was calibrated against prevalence value of 6% (as a "prior"), as specified in Methods section (Steck A, 2011).

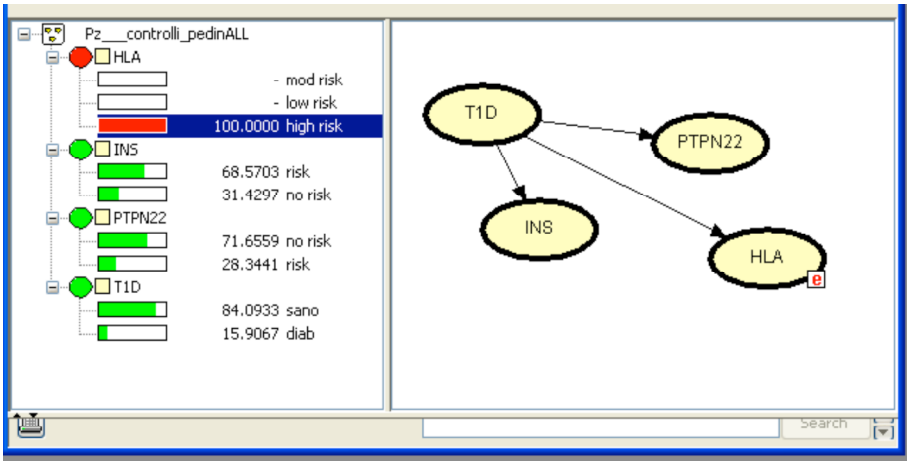


Single-locus analysis

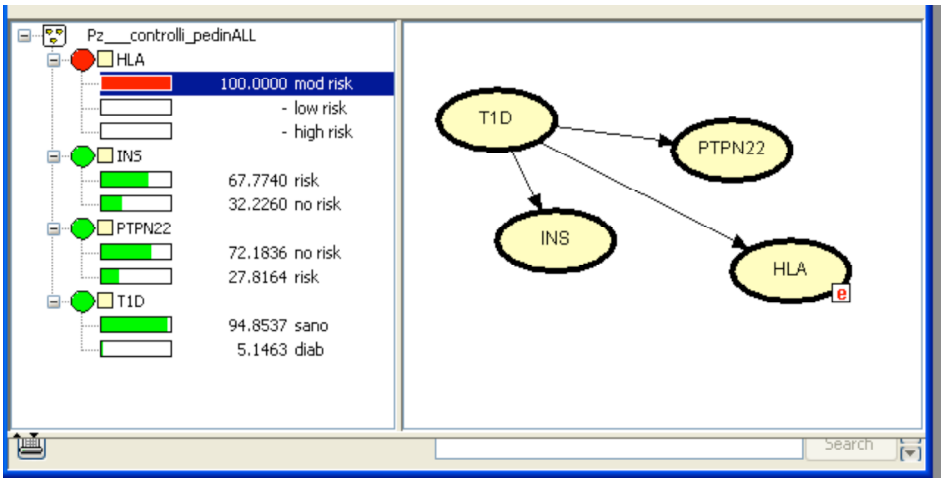
HLA-DR alleles. Compared with the T1D absolute risk of 6% in Caucasian siblings, the HLA high risk alleles conferred a risk of 15.9%. For moderate and low risk categories the risk was 5.1% and 1.68% of having T1D, respectively.

The following figures show the conditional probability (P) to develop T1D if an individual carries selectively high, moderate or low risk HLA. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

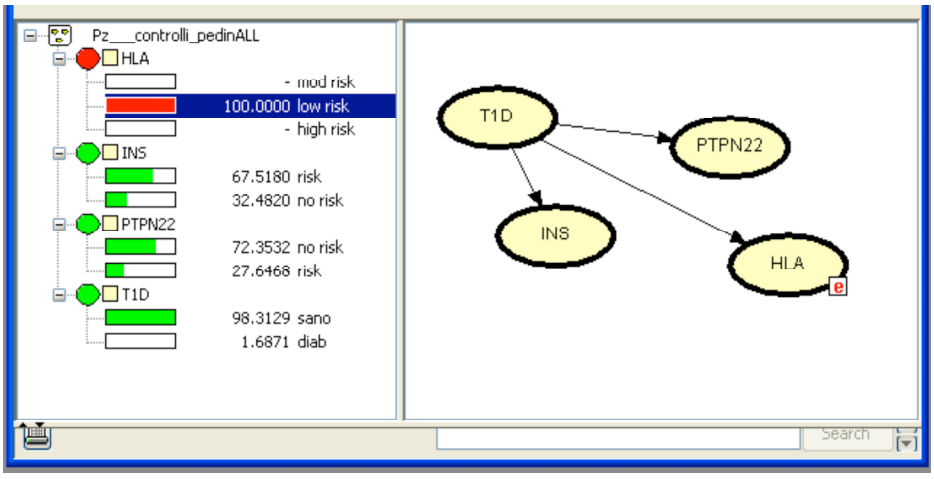
P(D | HLA high risk)



P(D | HLA Moderate risk)



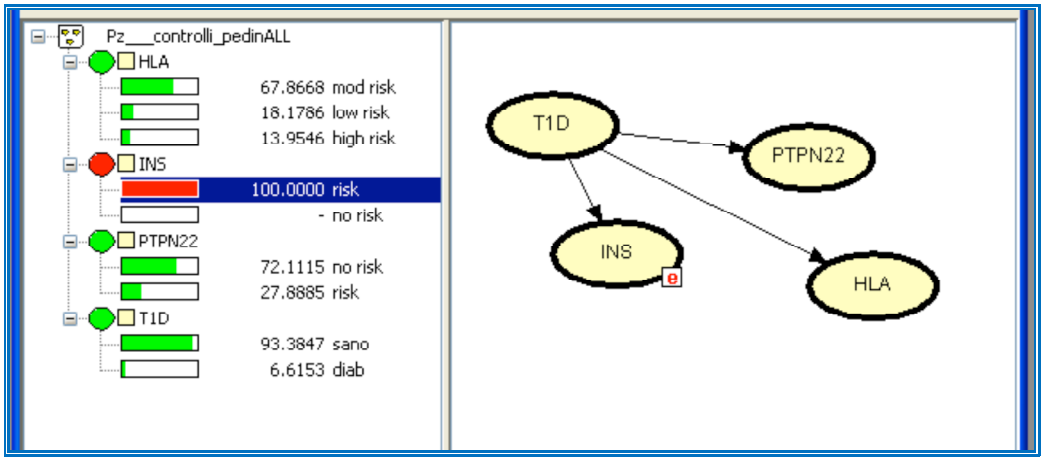
P(D | HLA low risk)



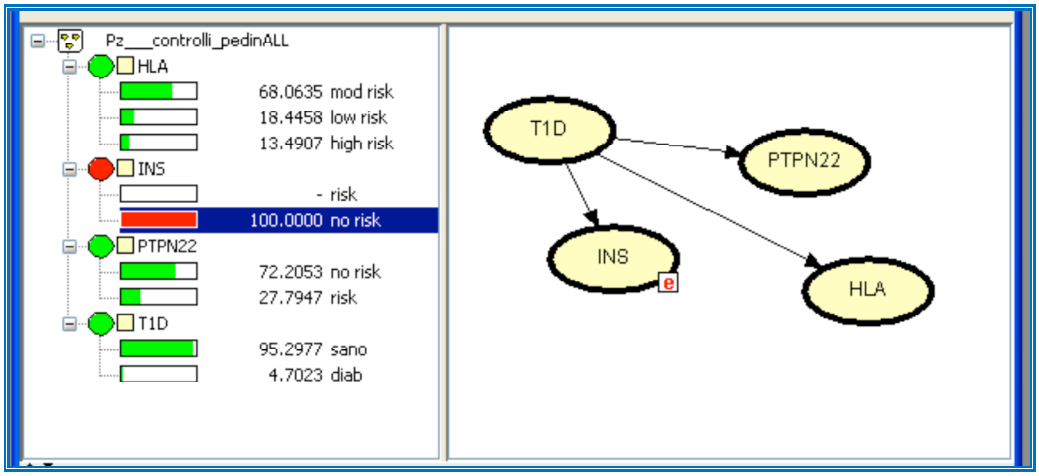
INS gene. When INS alleles were considered, the risk values of having T1D were 6.6% and 4.7% for INS risk with no risk alleles, respectively.

The following figures show the conditional probability (P) to develop T1D if an individual carries susceptibility or non-susceptibility INS genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

:P(D|INS risk)



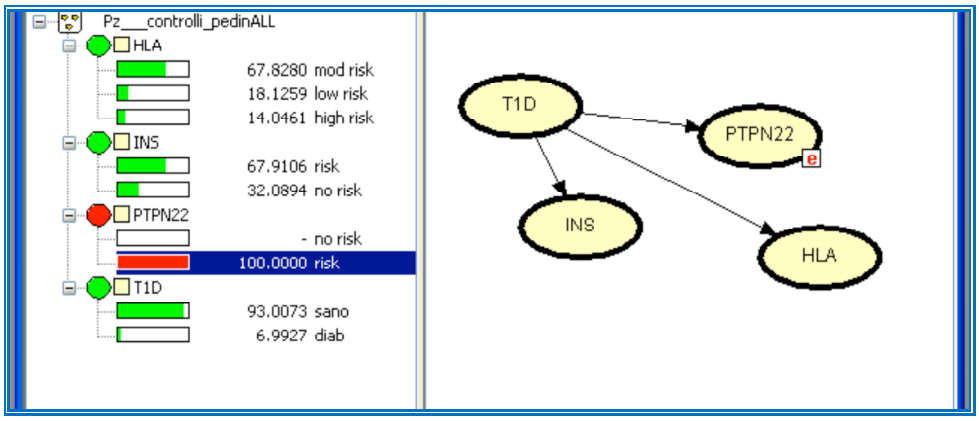
P(D|INS no risk)



PTPN22 alleles. When PTPN22 alleles were considered, the risk values of having T1D were 6.9% and 5.6% for PTPN22 risk and no risk alleles, respectively.

The following figures show the conditional probability (P) to develop T1D if an individual carries susceptibility or non-susceptibility PTPN22 genotypes. In red is shown the genetic condition which was used to calculate the absolute risk (T1D).

P(D|PTPN risk)



P(D|PTPN no risk)

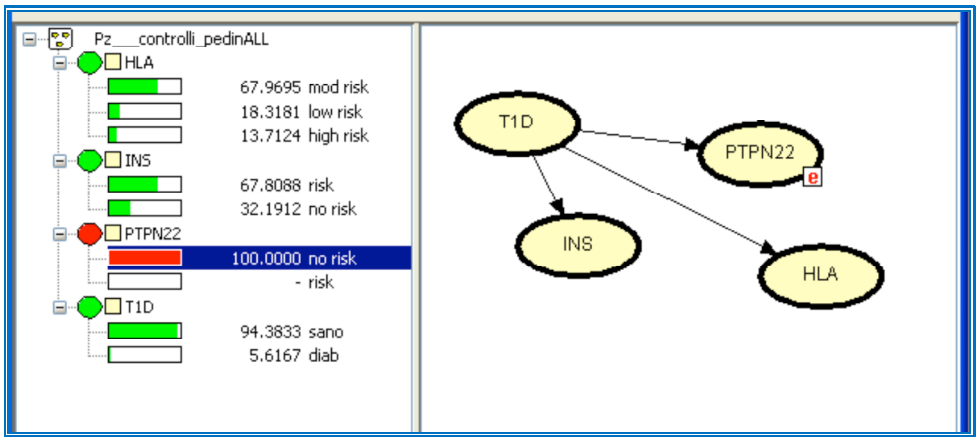


Table 6. Risk analysis using OR parameter and BN algorithm to evaluate single-locus main effects.

Genes	Cases (N)	Controls (N)	OR	95% CI	BN risk (%)
HLA					
High risk	620	289	4.1	3.4-4.7	**15.9**
Moderate risk	987	1604	0.6	0.5-0.7	**5.1**
Low risk	87	447	0.04	0.03-0.05	**1.68**
Total	1694	2340			
INS					
Risk	1267	1577	1.4	1.2-1.6	**6.6**
No risk	427	763	1		**4.7**
Total	1694	2340			
PTPN22					
Risk	550	645	1.2	1.1-1.4	**6.9**
No risk	1144	1695	1		**5.6**
Total	1694	2340			

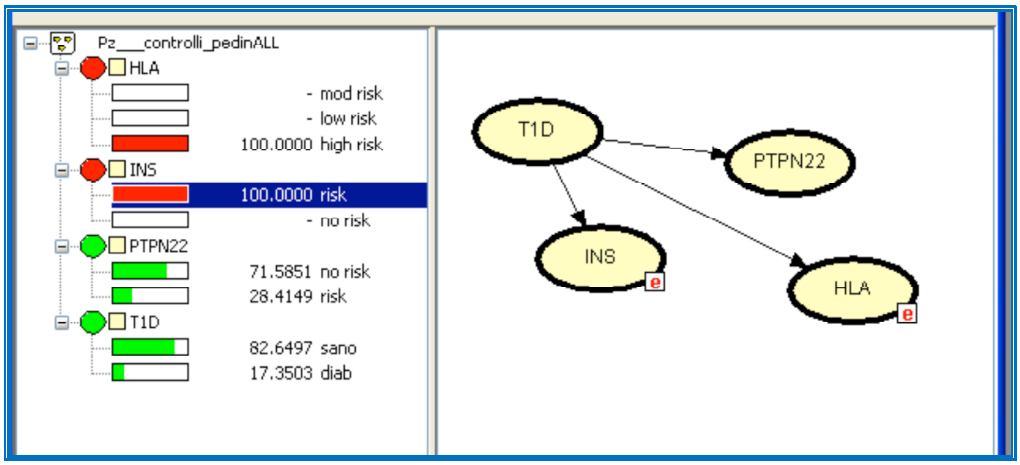
Joint effect of HLA and INS: We did not find a significant heterogeneity in the distribution of the INS genotype (susceptibility/non-susceptibility) in HLA risk classes when patients and controls were considered (Patients group: chi-square= 3.80, degrees of freedom = 2, p = 0.149; Controls group: chi-square= 0.9, degrees of freedom = 2, p = 0.64, NS).

The joint risk values to have T1D in a subject with HLA high or moderate or low risk categories, in association respectively with INS risk or INS no risk

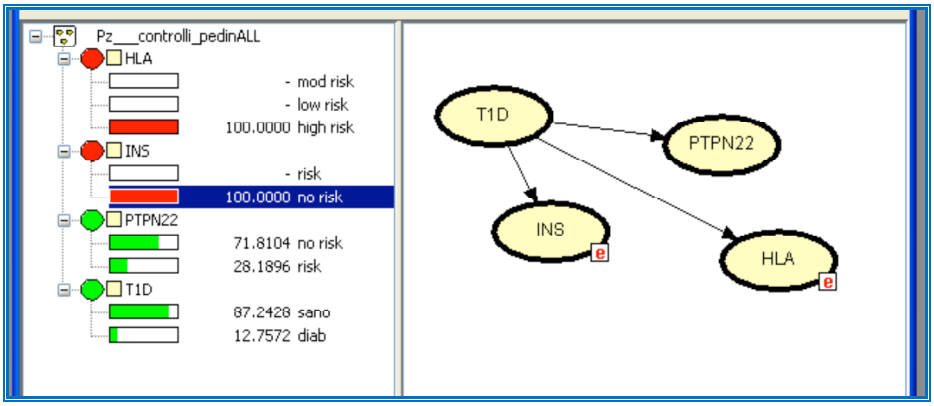
alleles, were calculated using BN algorithm. Our results showed that the relative impact of variation at INS within each genotype category at the HLA-DRB genotype were detectable in all of the HLA genotype categories. The absolute risk was of 17.3% and 12.7% if INS risk or non-risk alleles were respectively present in individuals with high-risk HLA, 5.7% and 4% in individuals with intermediate-risk HLA, and 1.8% and 1.3% in individuals with low-risk HLA genotypes (Table 7).

The following figures show the conditional probability (P) to develop T1D if an individual carries either high, moderate or low risk HLA and susceptibility or non-susceptibility INS genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

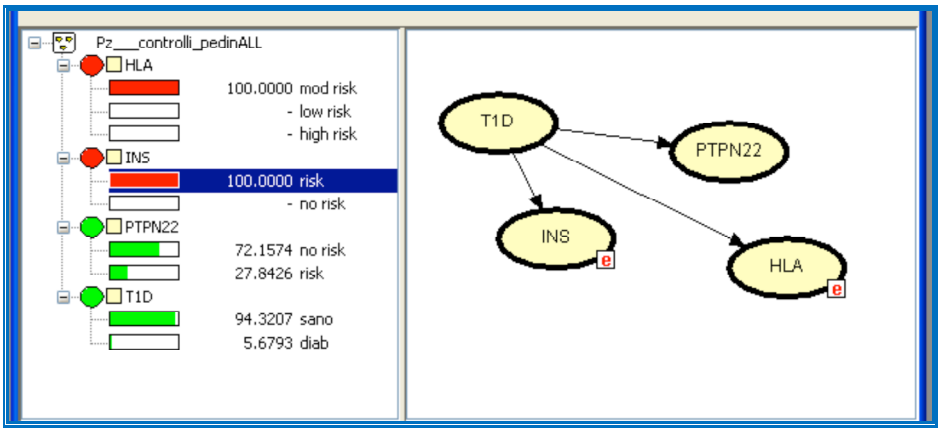
P(D|HLA high, INS risk)



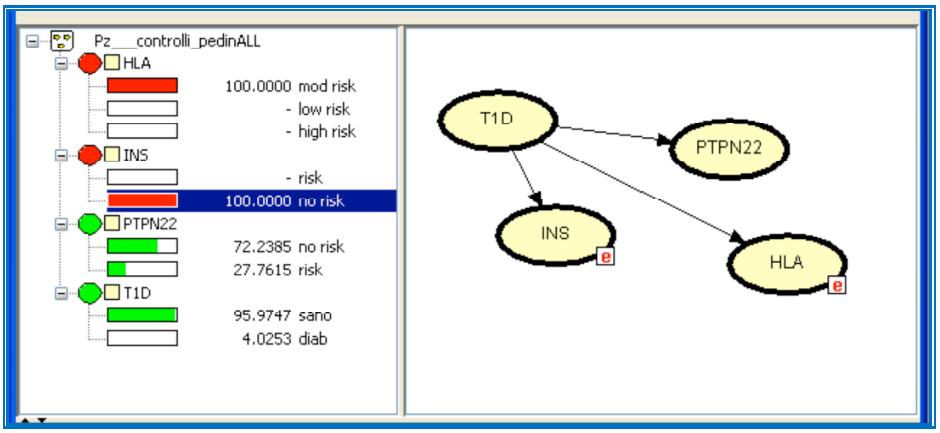
P(D | HLA high, INS no risk)



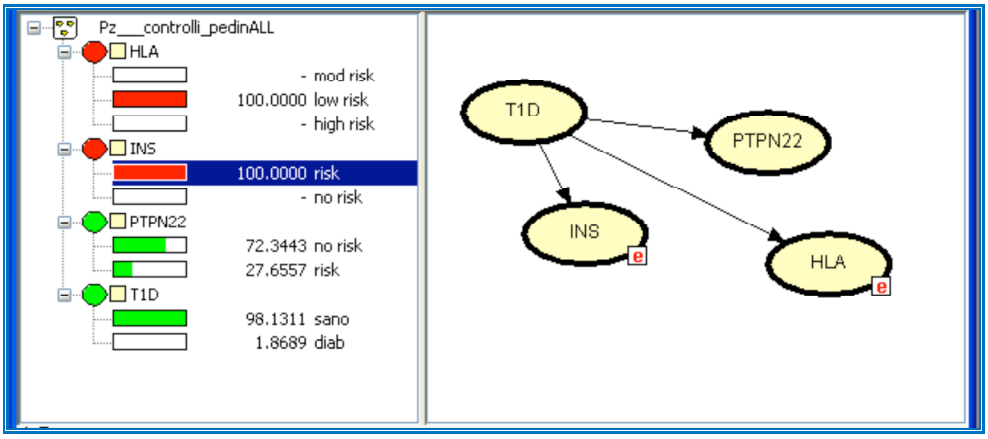
P(D | HLA Moderate, INS risk)



P(D | HLA Moderate, INS no risk)



P(D | HLA low, INS risk)



P(D | HLA low, INS no risk)

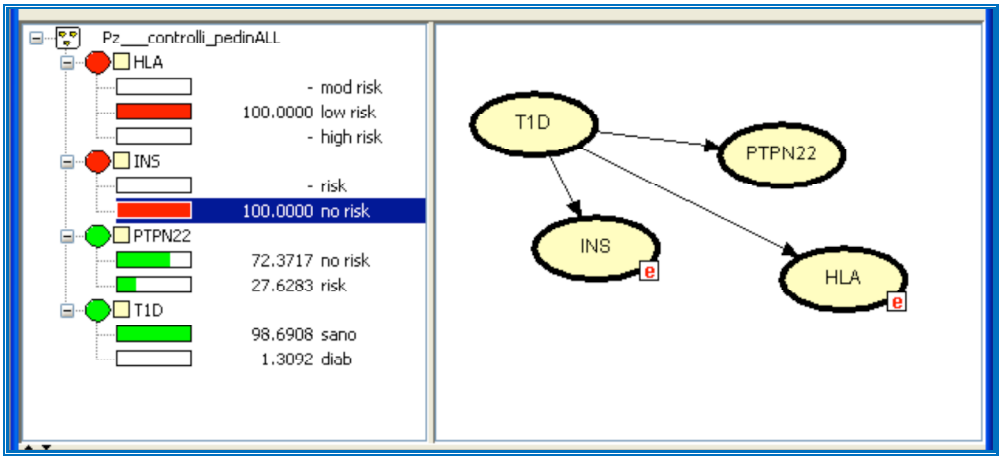


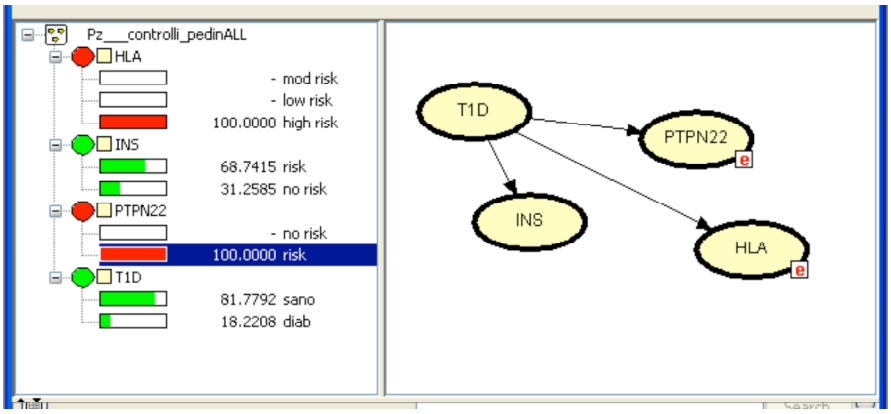
Table 7. Interaction between HLA-INS in the case-control data set using OR and BN algorithm

HLA category	INS alleles	Case N (%)	Controls (%) N (%	OR	95% CI	BN risk %
High risk	**Risk**	448 (26.5)	188 (8)	1.3	1.03-1.8	17.3
	No risk	172 (10.2)	101(4.5)	1		12.7
Moderate risk	**Risk**	750 (44.3)	1084 (46.5)	1.5	1.2-1.8	5.7
	No risk	237 (14)	520 (22)	1		4
Low risk	**Risk**	69 (4)	305 (13)	1.8	1.02-3.1	1.9
	No risk	18 (1)	142 (6)	1		1.3
Total		1694	2340			

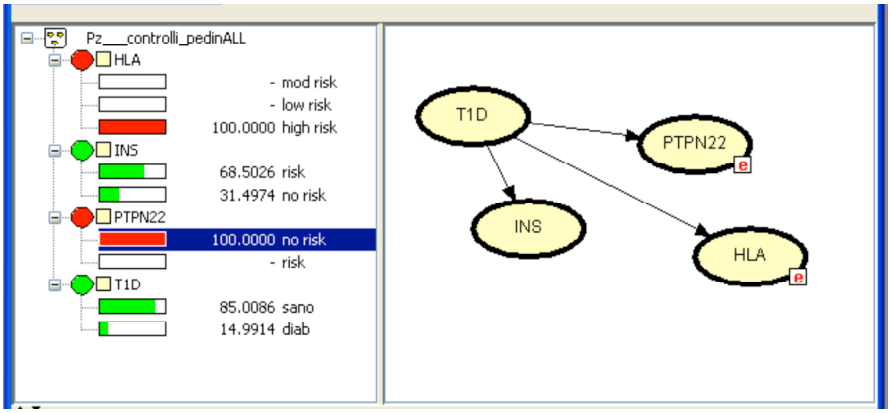
Joint effect of HLA and PTPN22.

Our results showed that the relative impact of variation at PTPN22 within each genotype category at the HLA-DRB genotype were detectable in all of the HLA genotype categories. The absolute risk of 18.2% and 14.9% if PTPN22 risk or non-risk respectively were present in individuals with high-risk HLA, 6% and 4.8% in individuals with intermediate-risk HLA and 1.9% and 1.5% in individuals with low-risk HLA genotypes (Table 8). The following figures show the conditional probability (P) to develop T1D if an individual carrying either high, moderate or low risk HLA and susceptibility or non-susceptibility PTPN22 genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated

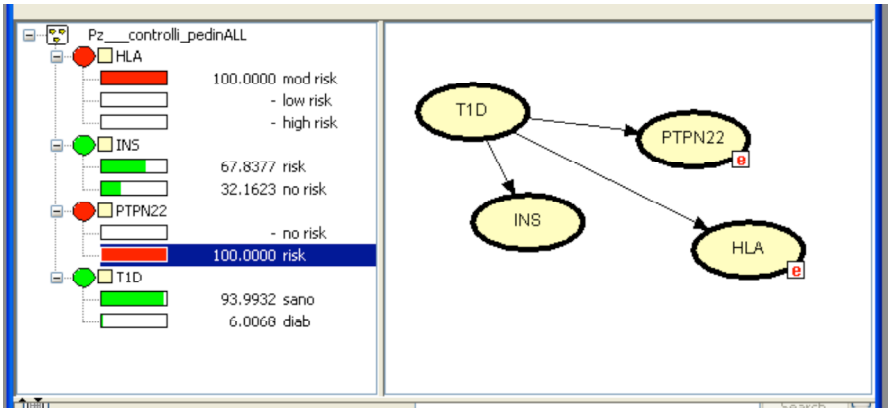
P(HLA High Risk, PTPN22 Risk)



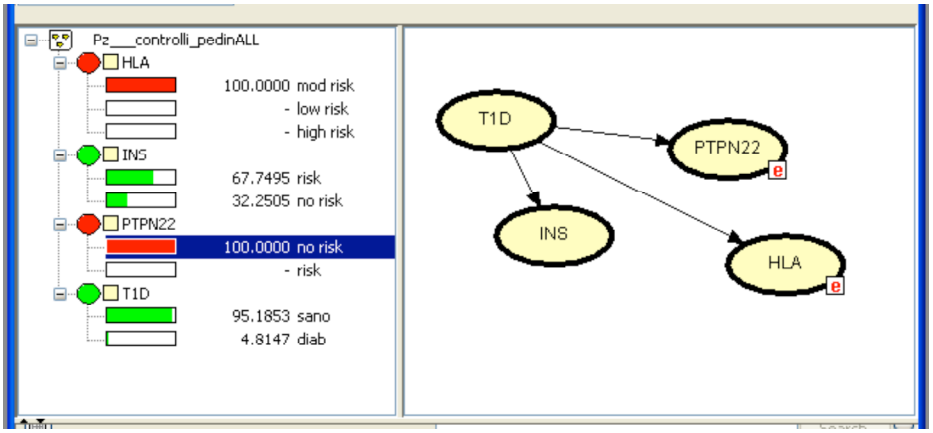
P(HLA High Risk, PTPN22 no risk)



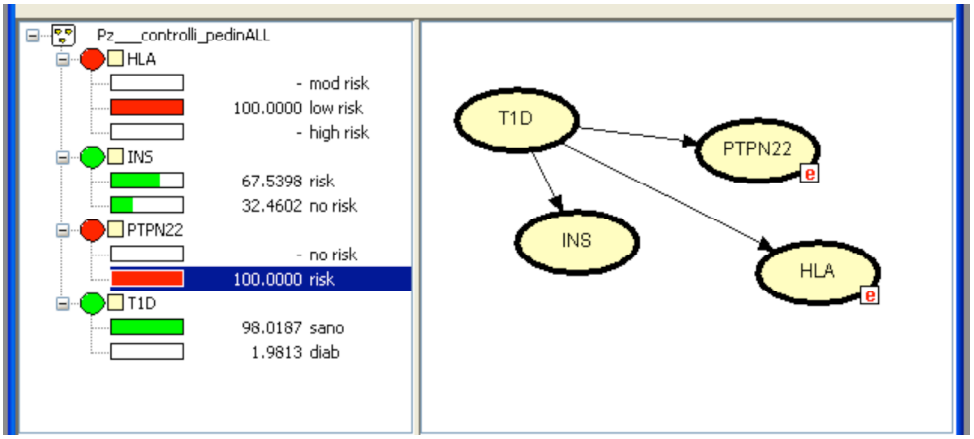
P(HLA Moderate Risk, PTPN22 Risk)



P(HLA Moderate Risk, PTPN22 no risk)



P(HLA Low Risk, PTPN22 Risk)



P(HLA Low Risk, PTPN22 no risk)

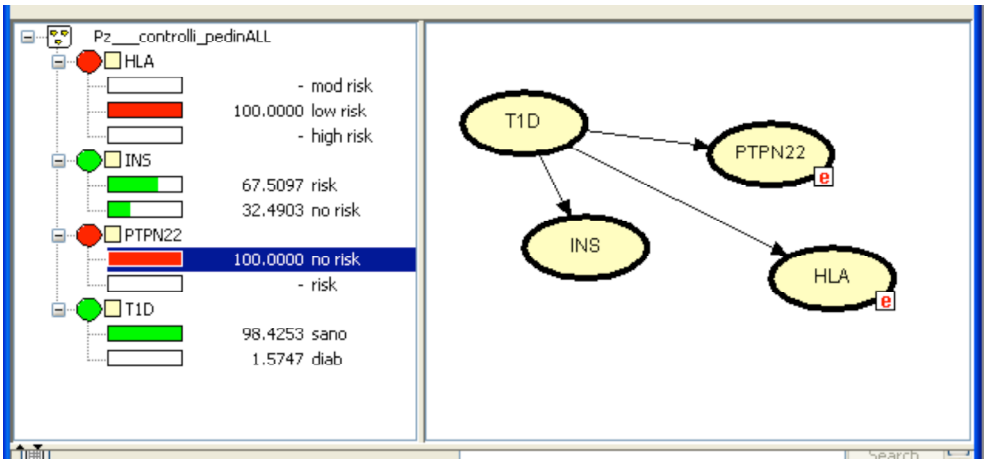


Table 8. Interaction between HLA-PTPN22 in the case-control data set using OR and BN algorithm

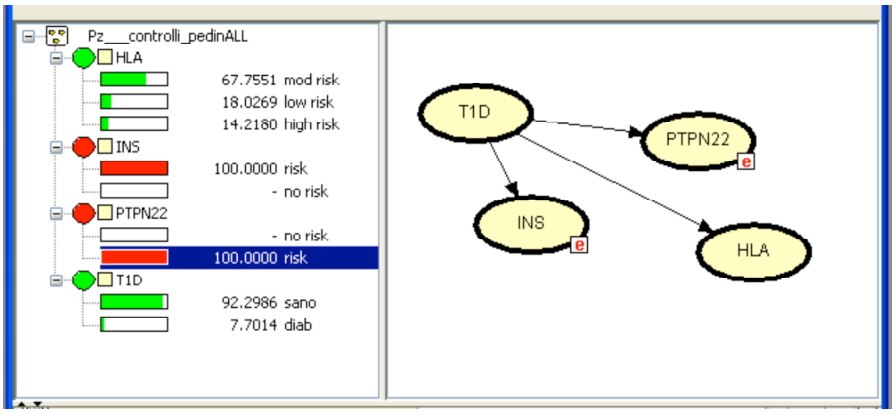
HLA category	PTPN22 alleles	Case N (%)	Controls (%) N (%)	OR	95% CI	BN risk %
High risk	**Risk**	197 (11.7)	79 (3.3)	1.2	0.9-1.7	18.2
	No risk	423 (25)	210 (9)	1	-	14.9
Moderate risk	**Risk**	335 (19.8)	442 (19)	1.35	1.1-1.6	6
	No risk	652 (38.5)	1162 (49.6)	1	-	4.8
Low risk	**Risk**	18 (1)	124 (5.3)	0.7	0.3-1.1	1.9
	No risk	69 (4)	323 (13.8)	1	-	1.5
		1694	2340			

Joint effect of INS and PTPN22

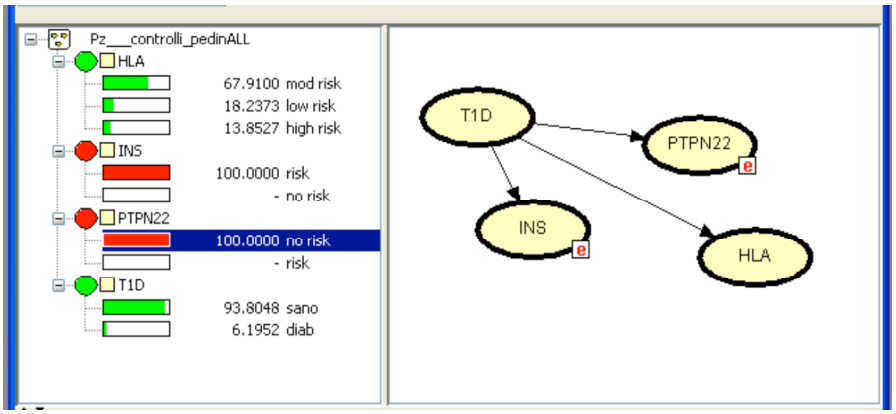
When INS and PTPN22 risk alleles were evaluated in order to establish risk of T1D onset, the risk values were equal to 7.7% for subjects with both INS & PTPN22 risk alleles. The risk decreased to 4.3 % when both these genes were present with non-risk alleles.

The following figures show the conditional probability (P) to develop T1D if an individual carrying susceptibility or non-susceptibility PTPN22 and INS genotypes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

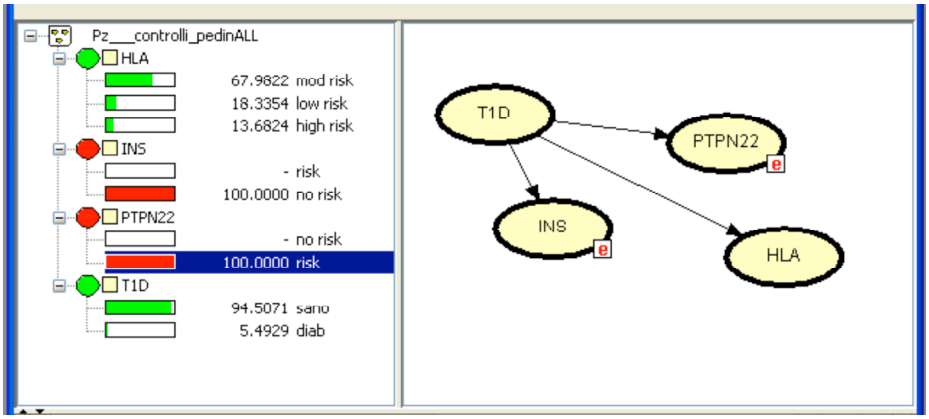
P(D | INS Risk, PTPN22 risk)



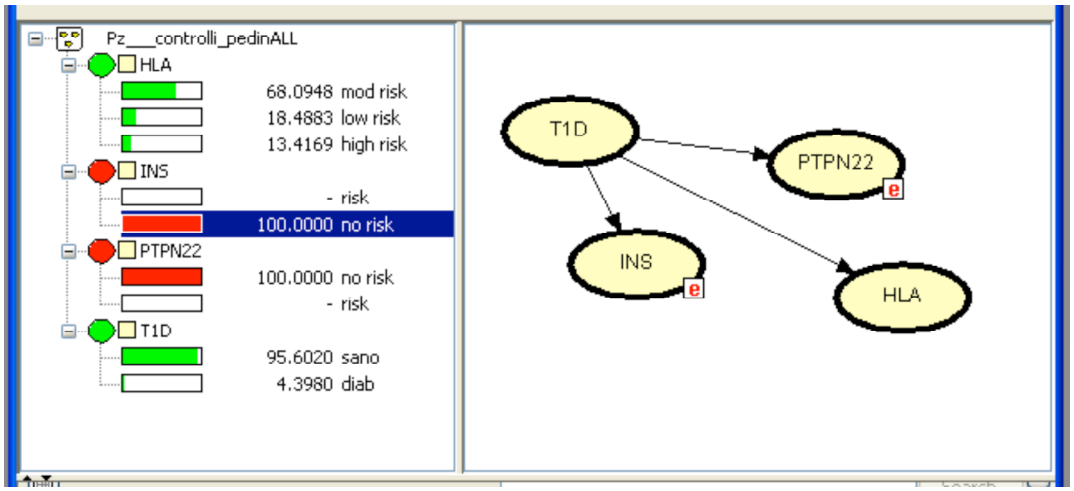
P(D | INS Risk, PTPN22 no risk)



P(D | INS no risk, PTPN22 risk)



P(D|INS no risk, PTPN22 no risk)

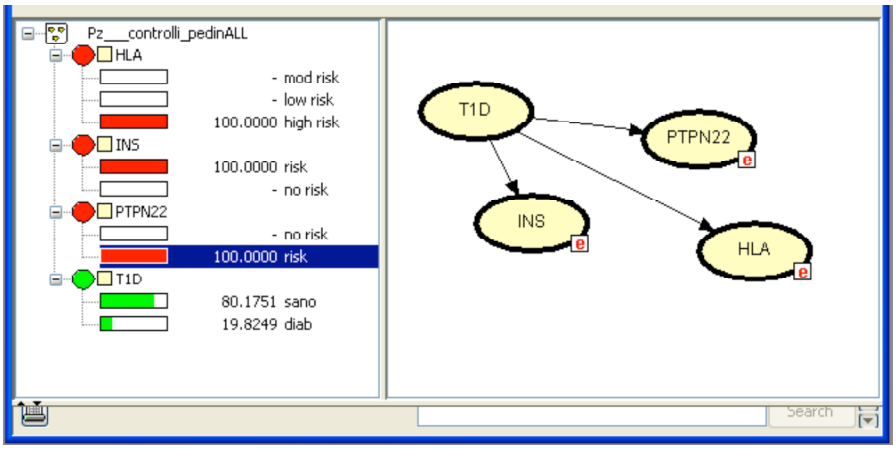


Joint effect of HLA, INS and PTPN22

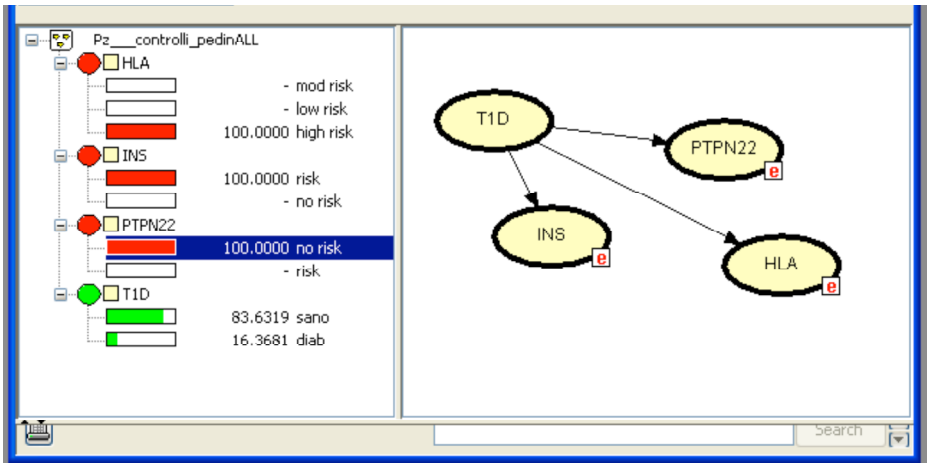
We also tested models with all three-way interactions involving the three susceptibility loci. The results showed that the more risk loci an individual carries, the higher the absolute risk (19.8%), but the presence or absence of HLA risk loci influences the absolute risk much more than the other loci, as expected. Carrying risk genotypes at both non-HLA loci but not at HLA (as shown on page 50, P(D|HLA low, INS risk, PTPN22 risk)) is associated with a much lower risk than HLA risk genotypes together with low-risk genotypes at the other two loci. The absolute risk (BN risk shown in table 9) conferred by simultaneously carrying high, moderate or low risk HLA and risk genotypes at the other two loci, compared with non-risk-associated genotypes at all three loci, was 19.8%, 6.6% and 2.2%, respectively.

The following figures show the conditional probability (P) to develop T1D if an individual carries risk or protective genotypes of HLA, INS and PTPN22 genes. In red is shown the genetic condition based on which the absolute risk (T1D) was calculated.

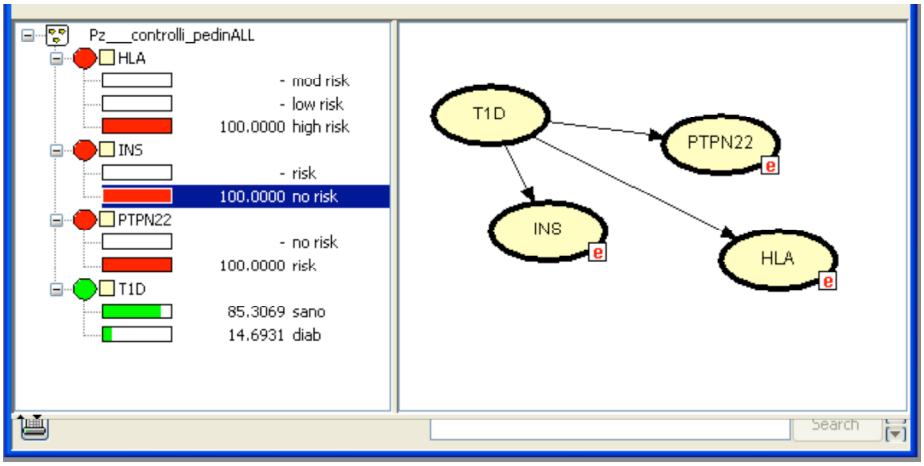
P(D | High Risk, INS Risk, PTPN22 risk)



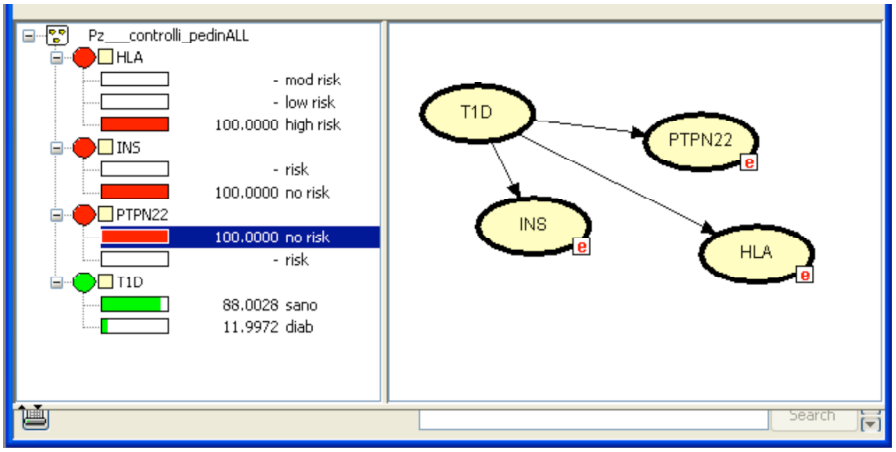
P(D | High Risk, INS Risk, PTPN22 no risk)



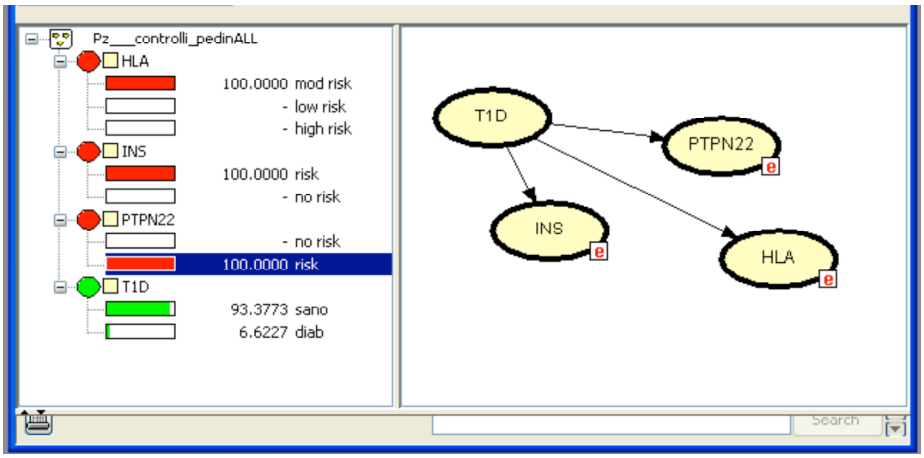
P(D | High Risk, INS no Risk, PTPN22 risk)



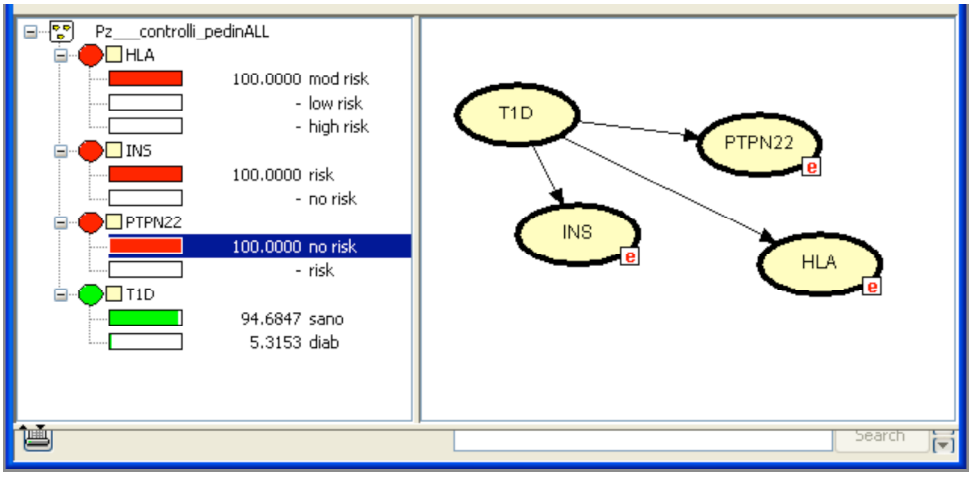
P(D | High Risk, INS no Risk, PTPN22 no risk)



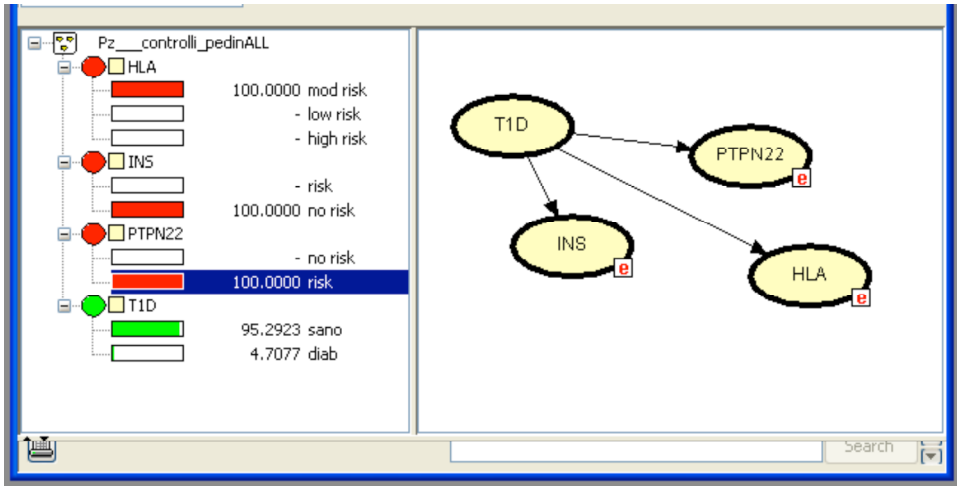
P(D | Moderate Risk, INS Risk, PTPN22 risk)



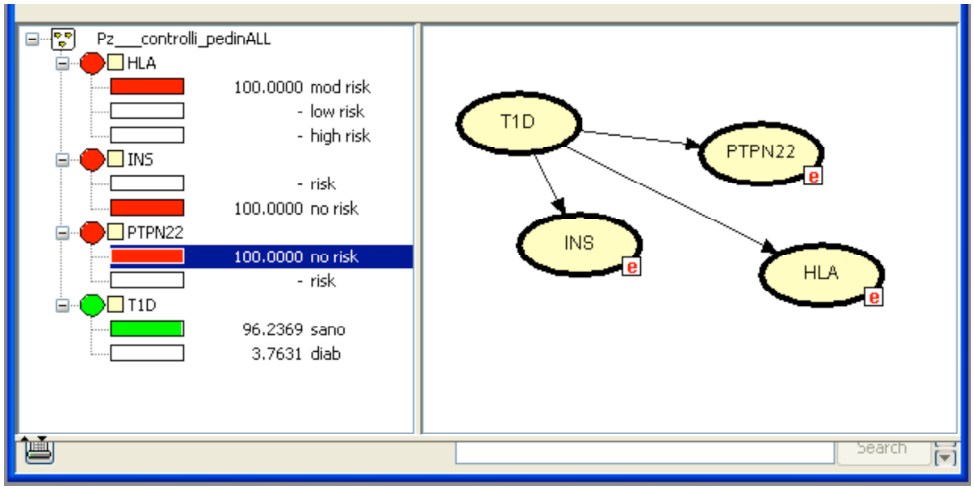
P(D | Moderate Risk, INS Risk, PTPN22 no risk)



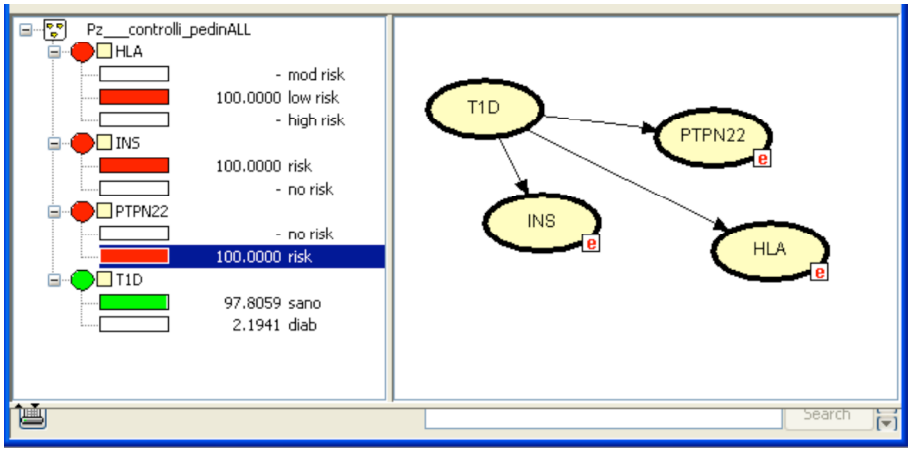
P(D | Moderate Risk, INS no Risk, PTPN22 risk)



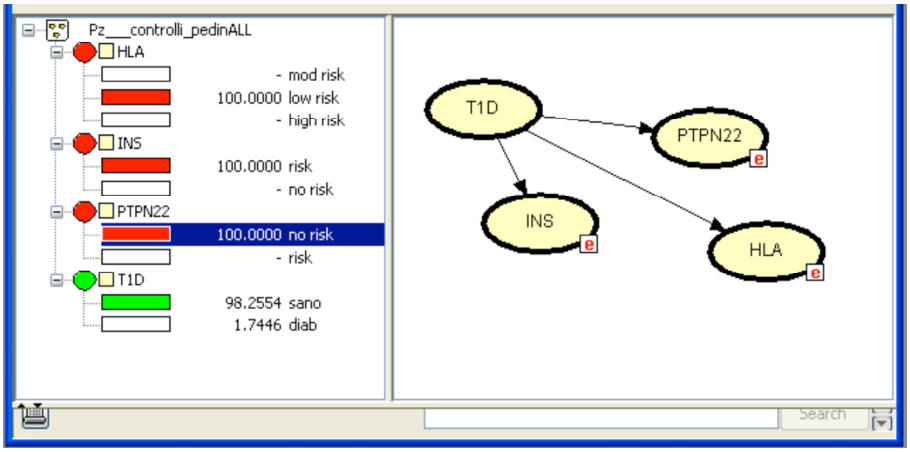
P(D | Moderate Risk, INS no Risk, PTPN22 no risk)



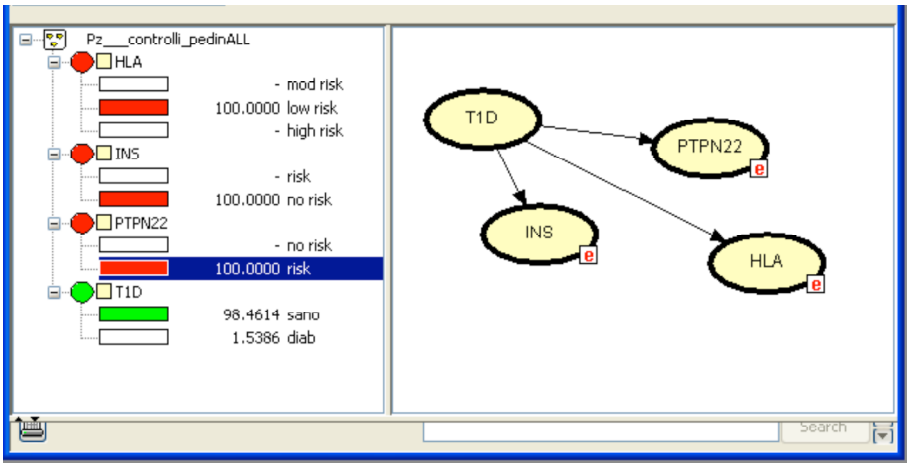
P(D | Low Risk, INS Risk, PTPN22 risk)



P(D | Low Risk, INS Risk, PTPN22 no risk)



P(D | Low Risk, INS no Risk, PTPN22 risk)



P(D | Low Risk, INS no Risk, PTPN22 no risk)

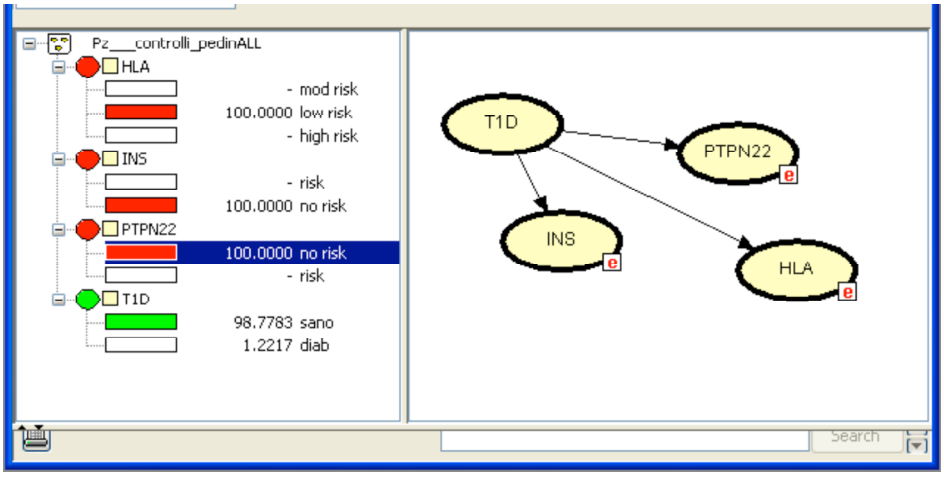


Table 9. Distribution of risk genotypes at all four susceptibility loci among cases and controls

HLA classes	INS1	PTPN1	Cases	Controls	OR (95%CI)[2]	BN risk
High	R	R	145	50	23.4(12-45)	19.8
High	R	NR	303	138	17.7(9.6-32)	16.3
High	NR	R	52	29	14.5(6.9-30)	14.7
High	NR	NR	192	72	21.5(11.4-40)	12
Moderate	R	R	263	285	7.4(4-13)	6.6
Moderate	R	NR	487	799	4.9(2.7-8.8)	5.3
Moderate	NR	R	72	157	3.7(1.9-7)	4.7
Moderate	NR	NR	165	363	3.6(2-6.7)	3.7
Low	R	R	13	87	1.2(0.5-2.7)	2.2
Low	R	NR	56	218	2.07(1.08-3.9)	1.7
Low	NR	R	5	37	1.09(0.3-3.3)	1.5
Low	NR	NR	13	105	1 (ref)	1.2

[1] "R"genotype at given locus associated with risk; "NR", genotype at given locus not associated with risk; [2] Odds ratio vs single reference group without risk genotype at any of the four loci.

3.3.5 Conclusion

This study classified T1D risk on the basis of HLA, INS and PTPN22 genes combinations in a large group of French Caucasian families. When the Bayesian network was implemented the prevalence value of 6% was considered for its training, as confirmed from literature for the T1D risk in siblings (Steck A, 2011). When the relationship between HLA, INS and PTPN alleles was evaluated separately, a significant association between genes and T1D was found ($p=0.01$ for HLA, $p= 1*10^{-3}$ for INS gene and $p= 3*10^{-4}$ for PTPN22). Considering the specific genetic compositions in patients and controls from families group, the marked risk in the offspring with DR3/4 was consistent with the high prevalence of this genotype found in our data set (37% and 12% in patients and controls, respectively). When proportions of high risk HLA genotype of family data set were compared with the values present in case-control data set analyzed before (30% and 2.7% for patients and controls, respectively; see 4.3.3 for details), a significant difference was found ($p\text{-value}<0.0001$). Moreover, analysis of INS and PTPN22 alleles was done in order to determine whether their addition to HLA genotypes might improve prediction of individuals who will develop T1D. This study stratified individual HLA-DR genotypes conferring different risks for T1D. It also confirmed the main contribution of the HLA-DR locus to T1D risk and demonstrated that the INS and PTPN22 genes provided only marginally an additional risk for T1D in subjects carrying the high, moderate and low risk HLA alleles. Furthermore, our results showed that, when HLA and INS genes were considered in patients and controls, any significant heterogeneity in the distribution of the INS genotype (susceptibility/nonsusceptibility), according to the HLA risk classes, was found.

This study remarked that, especially in genetic studies on families, HLA susceptibility genotypes confer the major part of risk for T1D onset. The presence of additional predisposing factors at non-HLA loci seemed to have only a moderate effect on the genetic predisposition for developing disease.

REFERENCES

Heckerman D, Probabilistic Similarity Networks, Cambridge, Mass.: MIT Press, 1991 ,

Akaike, H. (1974)Information Theory and an Extension of the Maximum Likelihood Principle." Automatic Control, IEEE Transactions on 19; 716-723

Bain SC, Prins JB, Hearne CM, Rodrigues NR, Rowe BR, Pritchard LE, Ritchie RJ, Hall JR, Undlien DE, Ronningen KS (1992) Insulin gene region encoded susceptibility to type 1 diabetes is not restricted to HLA-DR4-positive individuals. Nat Genet 2:212–215

Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176 –83

Bennett ST, Gough SCL et al (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9: 284-292

Bjørnvold M, Joner G, Dahl-Jørgensen K, Njølstad PR, Akselsen HE, Gervin K, Rønningen KS, Stene LC (2008). Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes. Diabetologia 51: 589–596.

Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet36: 337–8

Buzzetti R, Petrone A, Del Buono ML, et al (2004) Genetic prediction of type 1 diabetes in a population with low frequency of HLA risk genotypes and

low incidence of the disease (the DIABFIN study). DMRR 20: 137-144

Cowell RG, Dawid AP, Lauritzen StL, Spiegelhalter DJ. Probabilistic Networks and Expert Systems, Berlin: Springer, 1999

Diabetes. 2004 Dec;53(12):3286-91

Elstein, AL, Shulman LS, Sprafka SA Medical Problem Solving - An Analysis of Clinical Reasoning. Cambridge: Harvard University Press, 1978

Fourlanos S, Tait BD, Morahan G, Honeyman MC, Colman PG, Harrison LC (2008) The rising incidence of Type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. Diabetes Care 31: 1546-1549

Ghandil PCC, Dubois-Laforgue D , Senee V, et al (2005) Crohn's disease associated CARD15 (NOD2) variants are not involved in the sus- ceptibility to type 1 diabetes. Mol Genet Metab 86: 379-383

Hathout EH, Hartwick N, Fagoaga OR, Colacino AR, Sharkey J, Racine M, (2003). Clinical, autoimmune, and HLA characteristics of children diagnosed with type 1 diabetes before 5 years of age. Pediatrics 111:860–3

Henrion, M. Uncertainty in Artificial Intelligence: Is Probability Epistemologically and Heuristically Adequate? in J.L. Mumpower, L.D. Philipps, O. Renn, V.R.R.

Heterogeneity in the magnitude of the insulin gene effect on HLA risk in type 1 diabetes

Jensen, FV. Bayesian Networks and Decision Graphs, Statistics for Engineering and Information Science, Berlin: Springer, 2001

Lee YH, Choi SJ, Ji JD, Song GG, et al (2007) The PTPN22 C1858T functional polymorfism and autoimmune disease--a meta-analysis. Rheumatology 46: 49-56

Metcalfe KA, Hitman GA, Fennessy MJ, McCarthy MI, Tuomilehto J, Tuomilehto-Wolf E (1995) In Finland insulin gene region encoded susceptibility to IDDM exerts maximum effect when there is low HLA-DR associated risk: DiMe (Childhood Diabetes in Finland) Study Group. Diabetologia 38:1223–1229

Motzo C, Contu D, Cordell H.J, Lampis R, Congia M, Marrosu MG,Todd JA, Devoto M, Cucca F

Mueller PW, Cleary PA, Zhao Y, Smiles AM, Steffes MW, Bucksa J, Gibson TB, Cordovado SK, Krolewski AS, Nierras CR, Warram JH (2006) Genetics of Kidneys in Diabetes (GoKinD) study: a genetics collection available for identifying genetic susceptibility factors for diabetic nephropathy in type 1 diabetes. Am Soc Nephrol. 17: 1782-1790

Nistico L, Buzzetti R, Pritchard LE, VanderAuwera B, Giovannini C, Bosi E, et al (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Hum Mol Genet 5:1075– 80

Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA (1996) The role of HLA class II genes in insulin-dependent diabetes mellitus: Molecular analysis of 180 Caucasian, multiplex families. Am J Hum Genet. ; 59: 1134–1148

Onkamo P, Vaananen S, Karvonen M, Tuomilehto J.(1999) Worldwide increase in incidence of Type I diabetes–the analysis of the data on published

incidence trends. Diabetologia42:1395–403

Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (Revised Second Printing), Morgan Kaufman Publishers, San Mateo, CA., 1998

Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference, Morgan Kaufmann.

Pearl, J. (1993)Belief Networks Revisited, in: Artificial Intelligence, 59, 49-56

Shafer, G. The Art of Causal Conjecture, Cambridge, Massachusetts: MIT Press, 1996 Castillo, E., Gutiérrez, J.M., Hadi, A.S.: Expert Systems and Probabilistic Network Models, Berlin: Springer, 1997,

Steck A and Rewers M (2011) Genetics of Type 1 Diabetes. Clinical Chemistry 57;176-185 (2011).

Thomson G, Valdes AM, Noble JA et al (2007) Relative predispositional effects ofHLAclass II DRB1-DQB1 haplotypes and genotypes on type 1 diabetes: a meta-analysis. Tissue Antigens 70:110–127

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al (2007) Robust associations of four new chromosome regions from genomewide analyses of type 1 diabetes. Nat Genet;39:857–864

Uppuluri (eds), Expert Judgements and Expert Systems. Berlin: Springer (NATO ASI Series F: Computer and Systems Science), 1987, 106-129.

van der Auwera B, Schuit F, Lyaruu I, Falorni A, Svanholm S, Vandewalle CL, Gorus FK (1995) Genetic susceptibility for insulin-dependent diabetes mellitus in Caucasians revisited: the importance of diabetes registries in disclosing

interactions between HLA-DQ- and insulin gene-linked risk. Belgian Diabetes Registry. J Clin Endocrinol Metab 80:2567–2573

Vehik K, Hamman RF, Lezotte D, Norris JM, Klingensmith GJ, Rewers M, Dabelea D (2008) Trends in high-risk HLA susceptibility genes among Colorado youth with type 1 diabetes. Diabetes Care 31:1392– 6

Walter M, Albert E, Conrad M, Keller E, Hummel M, Ferber K, Barratt BJ, Todd JA, Ziegler AG, Bonifacio E (2003) IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for development of type 1 diabetes and associated autoimmunity. Diabetologia 46:712–720

Waldmann MR, Holyoak, K.J.(1992) Predictive and Diagnostic Learning Within Causal Models: Asymmetries of Cue Competition, in: Journal of Experimental Psychology: General 121; 222-236

3.4 Second aim. Do susceptibility genes have any influence on age of T1D onset?

Rationale

There is a steady increase worldwide of the incidence of autoimmune T1D especially in the age group <20 years (ADA, 2005; Karvonen M, 2000). This disease often manifests itself during childhood with a peak between the ages of 10 and 14 years (Hyttinen V, 2003). Previous research indicates that progression to disease and age at onset are directly linked to the HLA genotype (Erlich H, 2008). In the case of PTPN22 no convincing support for interaction between PTPN22 and age of T1D onset was found for rs2476601/Arg620Trp SNP (Smyth DJ, 2008), whereas rs2542151*G allele seems to confer risk to an earlier onset of T1D (L. Espino-Paisan, 2011). In order to examine the effects of HLA-DR, INS and PTPN22 genotypes on the age of T1D onset three different samples of T1D patients from French, American and Italian Caucasian population were investigated, using the BN probabilistic approach.

3.4.1 Materials and Methods

At first, a case-only interaction was performed for age at diagnosis. For this purpose, three different samples consisting of HLA, INS and PTPN22 alleles on T1D patients from French (called data set A, consisting of 741 patients), American (called data set C, consisting of 1739 patients) and Italian (called data set B, consisting of 391 patients) Caucasian populations, were considered

separately (for details on the data sets, see section 4.2). Age of disease onset was classified into two groups, (<=15– and >15 years).

Chi-squared statistics. Chi-squared statistics were used to analyse the three different samples with the aim of comparing genotype frequency in relation to the age category of disease onset.

BN analysis. Three different BNs, one for each data set were implemented, to assess the influence of each gene on the age of T1D onset. Genes, which were represented in the network by nodes, were considered in relation to their risk classes: high, moderate and low HLA risk genotypes, susceptibility and non-susceptibility INS and PTPN22 alleles. The variable "age of onset" was classified into two different classes: <=15 and >15 years.

The same approach was used in order to assess the absolute risk of having T1D before or after the age of 15 years, based on all different combinations of HLA-DR, INS and PTPN22 genes. For this reason, two different samples were used: the case-control data set A and the family clusters of the data set (for details on data sets used see paragraph 4.2).

Timeline analysis. A comparison between data sets was carried out using a timeline analysis, to investigate the impact of the single gene on the age of disease onset. Only HLA-DR and INS genes were considered in this last analysis.

For each data sets all combinations of HLA and INS-VNTR genes risk classes were selected (namely "HLA-high vs. HLA-low", "HLA-high/INS-susceptibility vs. HLA-high/INS-protection", "HLA-moderate/INS-susceptibility vs. HLA-moderate/INS-protection" "HLA-low/INS-

susceptibility vs. HLA-low/INS-protection″) and the effects on the age at T1D onset of each genetic combination were evaluated.

A two-sided Wilcoxon-rank test (alpha < 0.01) was used to evaluate the significance of this effect.

3.4.2 Results

Data set A. Frequency of genotypes and age of T1D onset.

In the **French** sample, the moderate-risk genotypes (HLA-DR4/X, HLA-DR4/4, HLADR3/X and HLA-DR3/3) constituted the most abundant category, together accounting for 61% of all genotypes. The second most abundant category was the high-risk genotype, HLA-DR3/4, at 30%. The lowest risk classmade up 9% of the sample. The high-risk HLA group comprised 61%of the subjects <=15 years and 39% of patients >15 years. In the low-risk group, 35% subjects were <=15 years and 65%of subjects were >15 years. The frequency of genotypes analyzed is presented in Table 1 and Figs. 1-4.

Moreover, INS risk and no risk alleles were represented in 75% and 25% of the cases, respectively. PTPN22 susceptibility and non-susceptibility genotypes characterized 30% and 70%, respectively. The age of disease onset was significantly related to HLA-DR genotype (p-value=0.01). A closer look at age/genotype associations within the two categories, <= 15 years and >15 years age classes, revealed that both high and low risk HLA genotypes were significantly associated with age at disease manifestation. In total, 36% of subjects in the age category <=15 years and 23% in the age category >15 years carried the high risk HLA genotype. The low risk HLA genotype was carried by 6% of the subjects in the age category <=15 years and by 11% among patients with the age of >15 years (Fig. 5). Chi-squared analysis showed a significant difference in the age of disease onset for subjects with high and low risk HLA genotypes (Fig. 5). Moreover, INS and PTPN22 genes did not have any significant influence on the age of T1D onset (p-value=0.7) (Figs. 6-7). The same analysis was performed using BN approach and similar results were obtained (Fig. 8).

Figure 1. Distribution of age of T1D onset in the French patients data set (x axis = age of onset and y axis = number of patients).

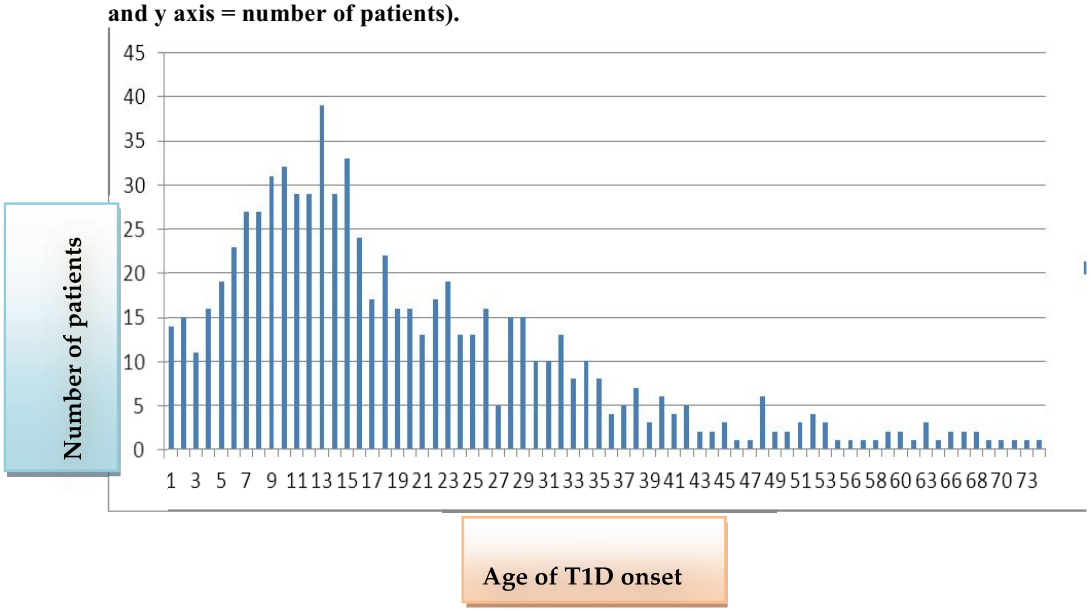


Figure 2. Distribution of HLA risk classes based on the age of T1D onset in the French patients data set (x axis = age of onset and y axis = number of patients).

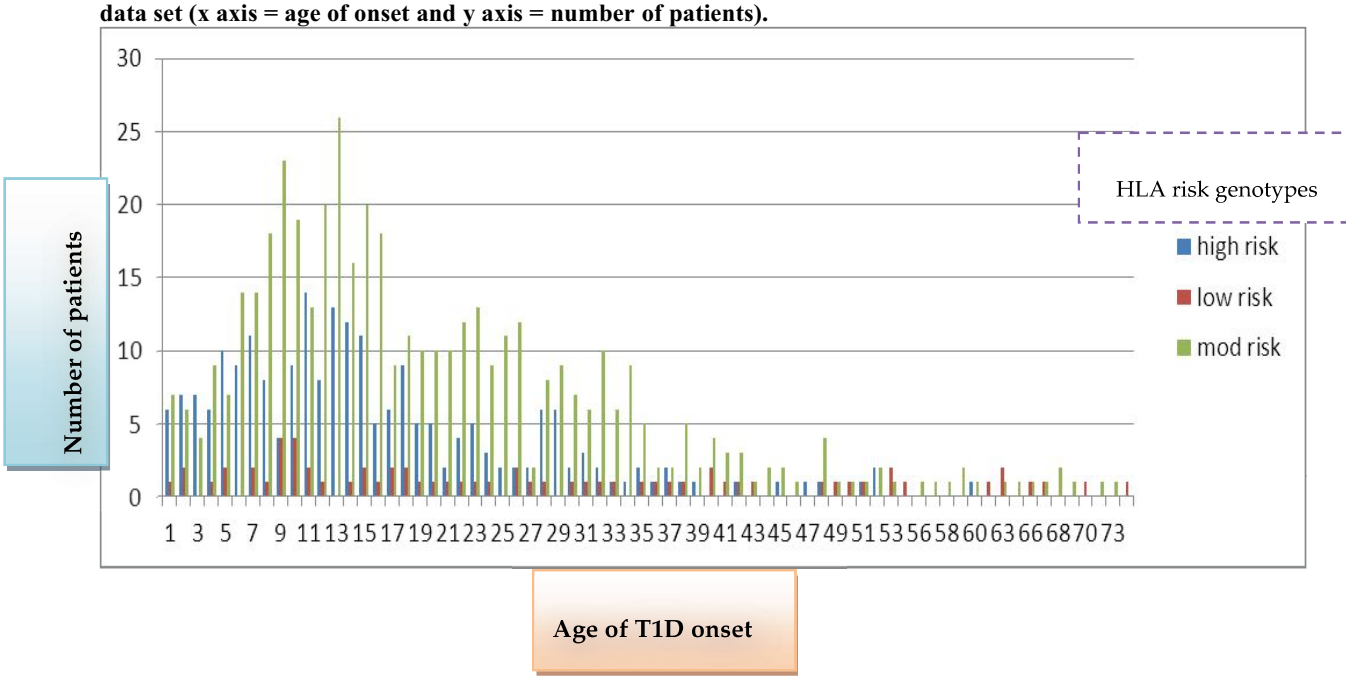


Figure 3. Distribution of INS risk classes based on the age of T1D onset in the French patients data set (x axis = age of onset and y axis = number of patients).

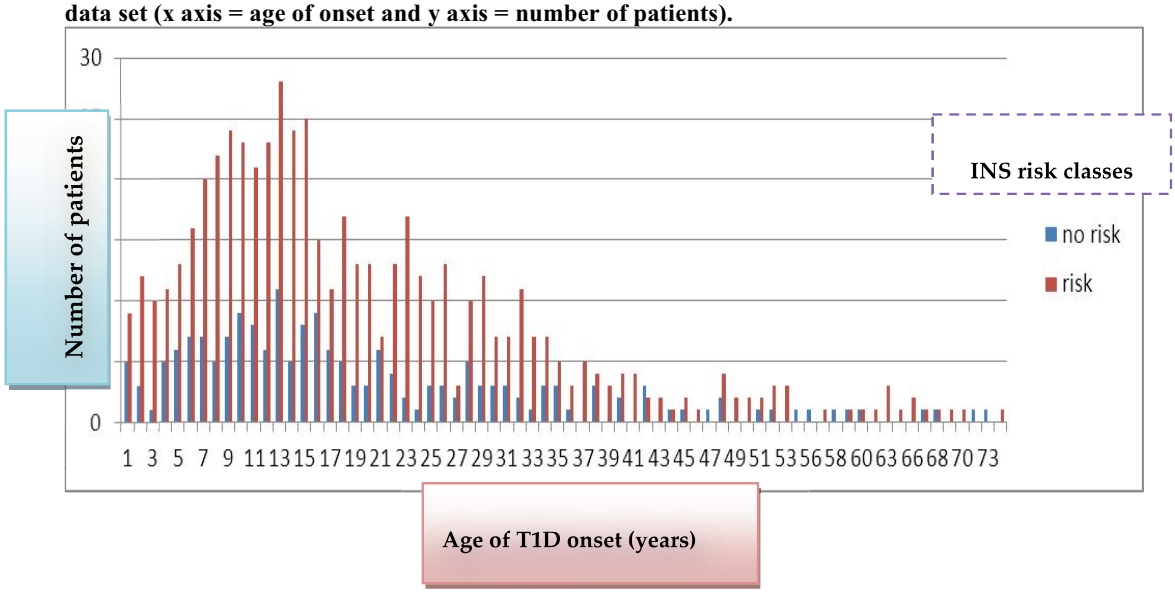


Figure 4. Distribution of PTPN22 risk classes based on the age of T1D onset in the French patients data set (x axis = age of onset and y axis = number of patients).

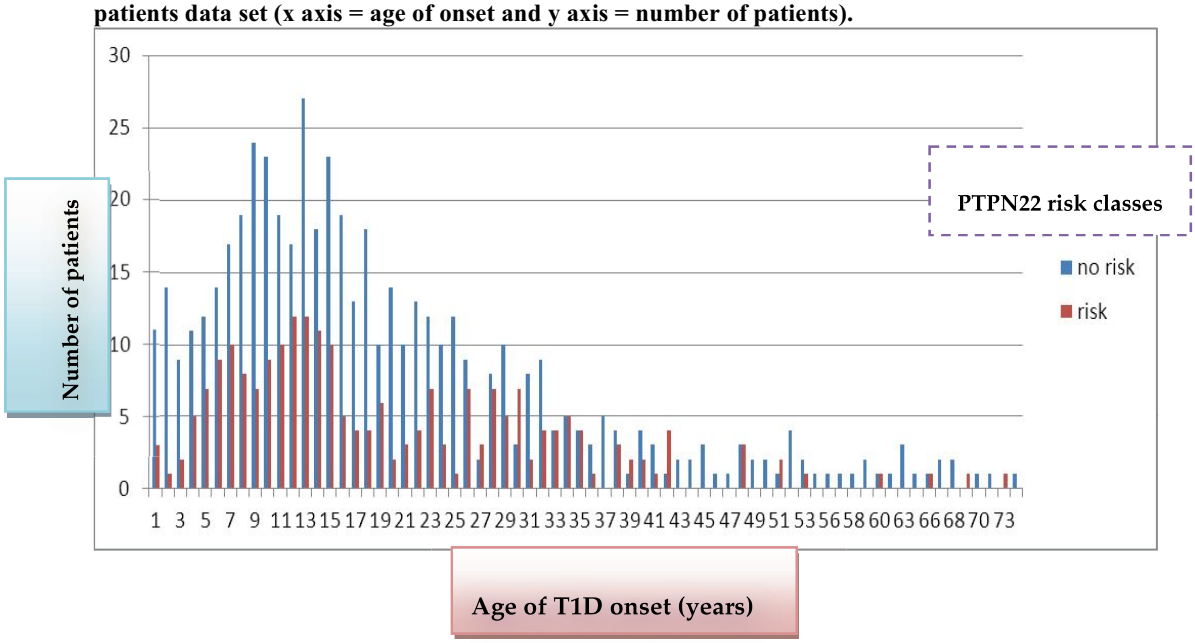


Figure 5. Age at disease onset. Frequency of HLA-DR risk classes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

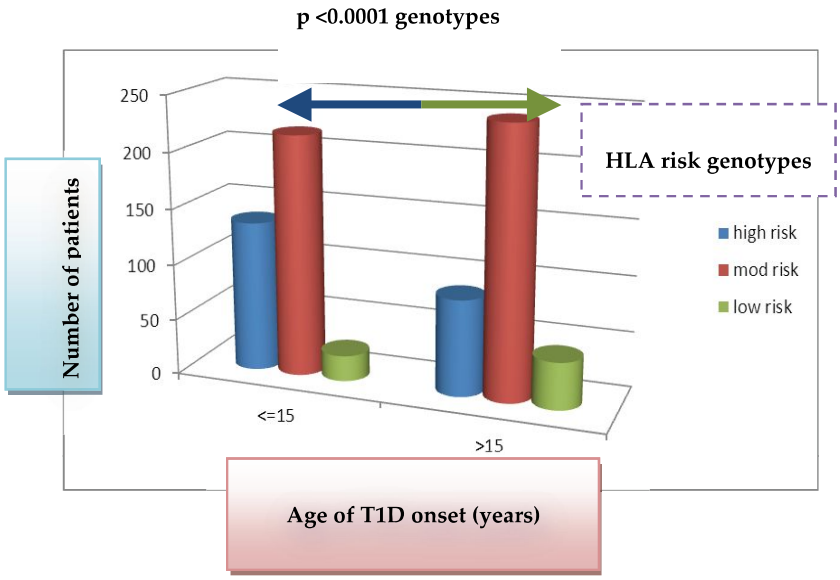


Figure 6. Age at disease onset. Frequency of INS risk genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

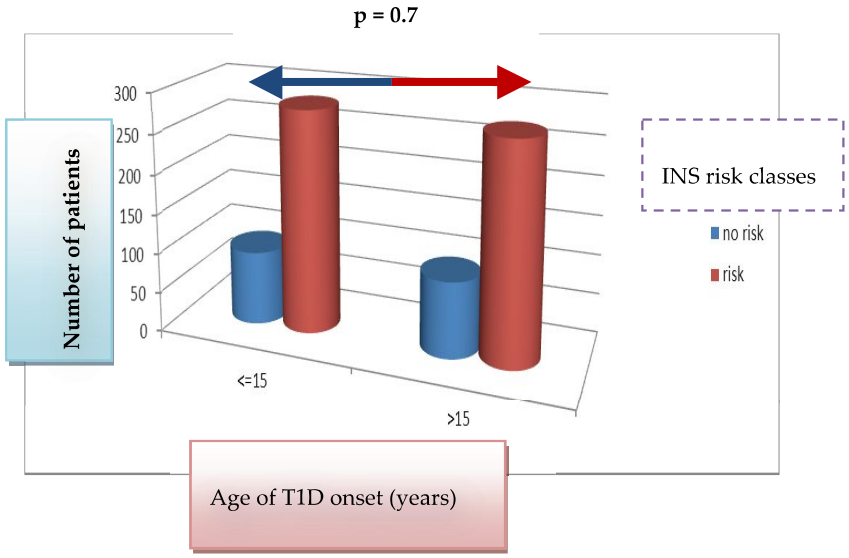


Figure 7. Age at disease onset. Frequency of PTPN22 risk genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

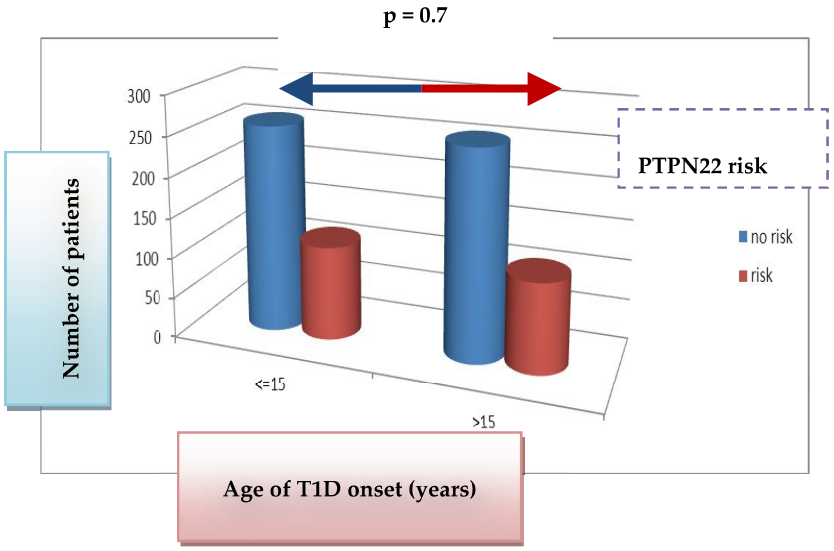
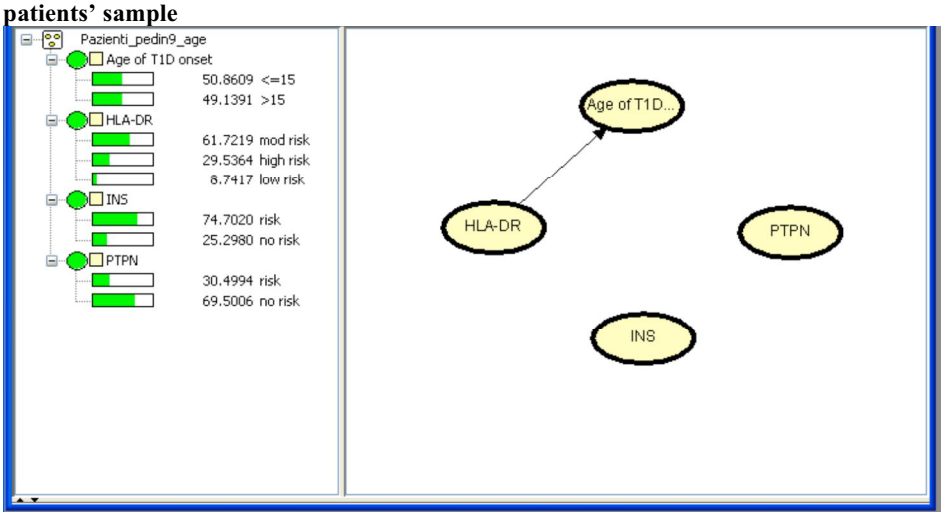


Table 1. Age at disease onset. Frequency of HLA-DR, INS and PTPN22 genotypes in two age categories (<=15 and >15 years)

	Age of onset					
HLA-DR	**<=15**		**>15**		**Total**	
High risk	135	36%	86	23%	221	30%
Mod risk	216	58%	239	65%	455	61%
Low risk	23	6%	42	11%	65	9%
Total	374		367		741	
	Age of onset					
INS	**<=15**		**>15**		**Total**	
Risk	281	75%	272	74%	553	75%
No Risk	93	25%	95	26%	188	25%
Total	374		367		741	
	Age of onset					
PTPN22	**<=15**		**>15**		**Total**	
Risk	116	31%	110	30%	226	30%
No Risk	258	69%	257	70%	515	70%
Total	374		367		741	

Figure 8. BN showing the relationship between HLA-DR gene and age of T1D onset in French patients' sample



Frequency of genotypes and age of T1D onset.

In the **American** sample, high, moderate and low-risk HLA genotypes were carried by 32 %, 59% and 8% of the patients. The frequency of genotypes analyzed is presented in Table 2 and Figs. 9-12

Moreover, the risk and no risk INS alleles were represented in 28% and 72% of the cases respectively. In this sample genetic data on PTPN22 was missing. When the relationship between the age of T1D onset and HLA risk classes was considered, both high and low risk HLA genotypes were significantly related to the age of onset (p-value=$6*10^{-6}$). In total, 77% of subjects in the age category <=15 years and 33% in the age category >15 years carried the high risk HLA genotype. The low risk HLA genotype was carried by 58% of the subjects in the age category <=15 years and by 42% from patients aged >15 years. (Fig. 13 and Tab. 2). The chi-squared test showed a significant difference in the age of disease onset between subjects with high and low HLA genotypes ($p<0.0001$), as illustrated in Fig. 13. The results were also confirmed when BN approach was used (Fig. 16).Our results showed that INS and PTPN22 genes did not have any significant influence on the age of T1D onset (p-value=0.1)(Figs. 14-15).

Figure 9. Distribution of age of T1D onset in the American patients data set (x axis = age of onset and y axis = number of patients).

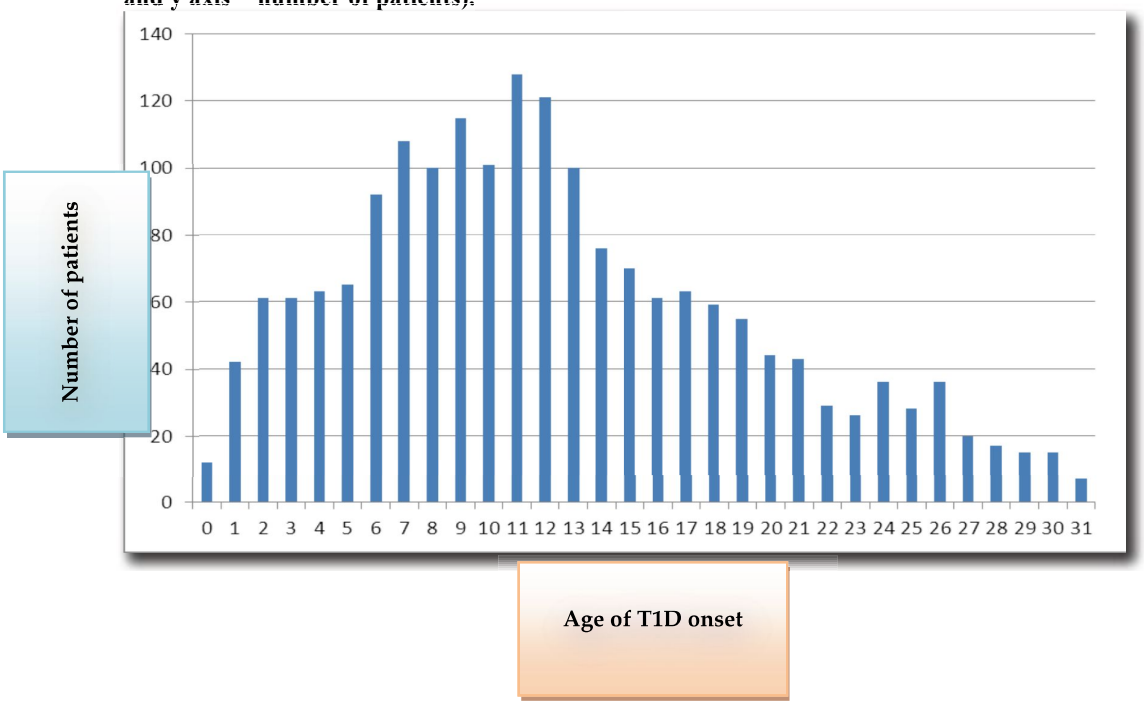


Figure 10. Distribution of HLA risk classes based on the age of T1D onset in the American patients data set (x axis = age of onset and y axis = number of patients).

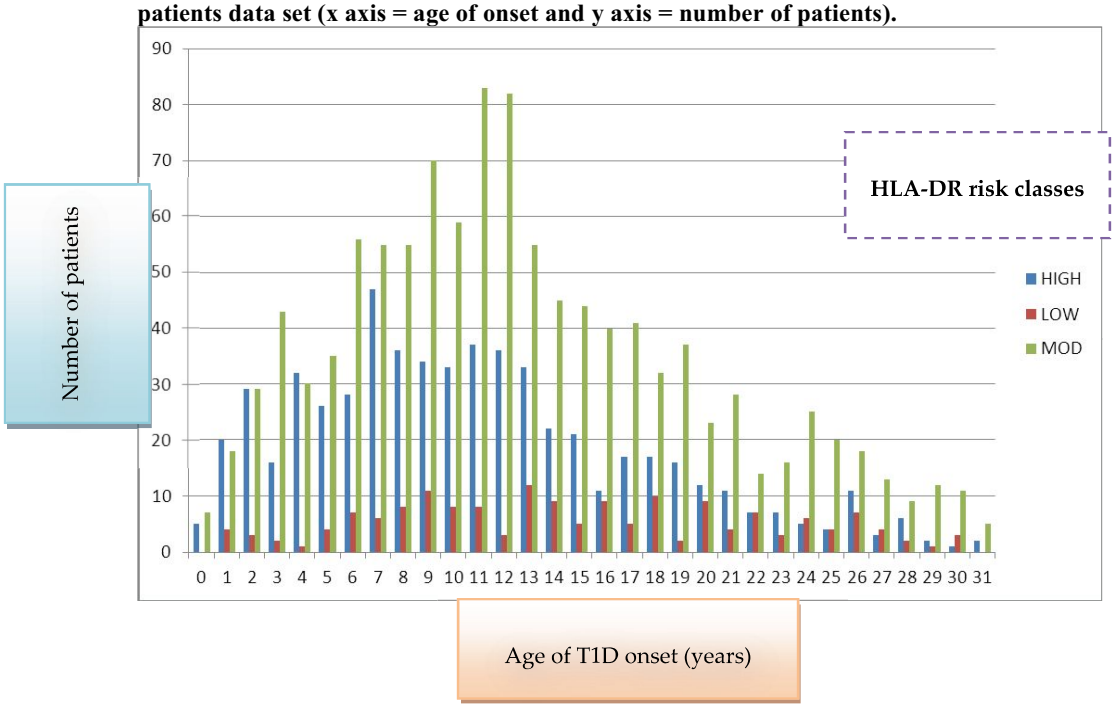


Figure 11. Distribution of INS alleles based on the age of T1D onset in the American patients data set (x axis = age of onset and y axis = number of patients).

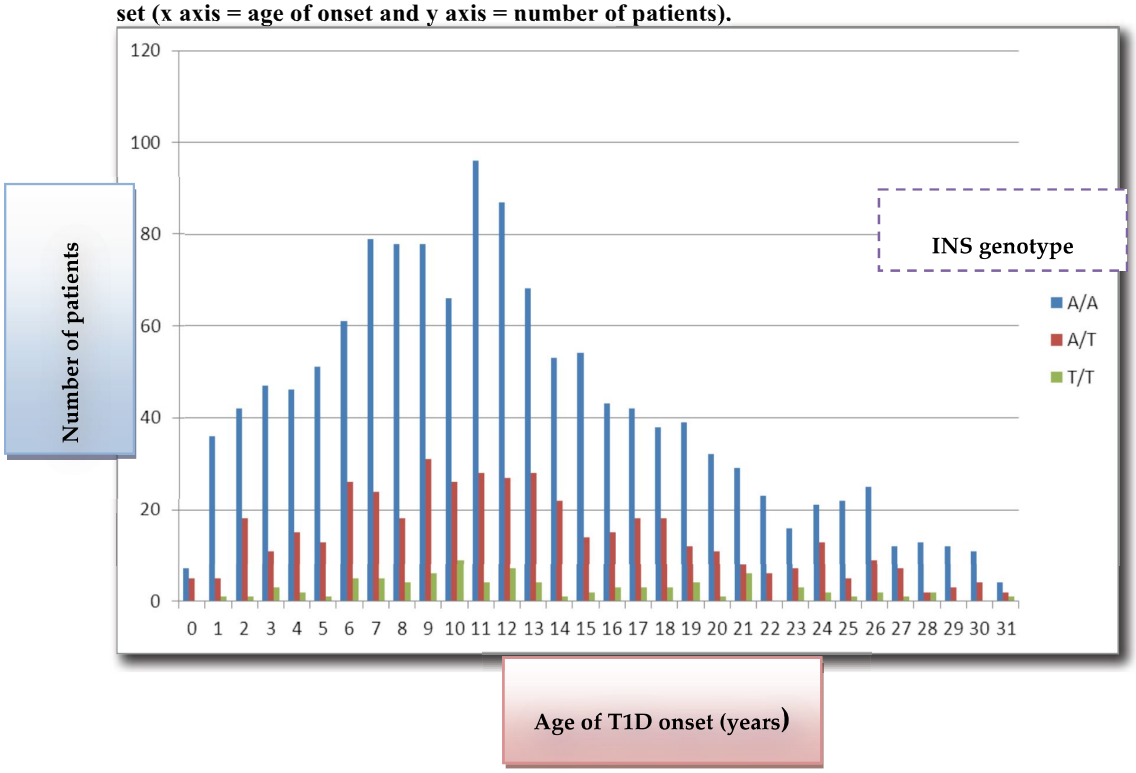


Figure 12. Distribution of INS risk classes based on the age of T1D onset in the American patients data set (x axis = age of onset and y axis = number of patients).

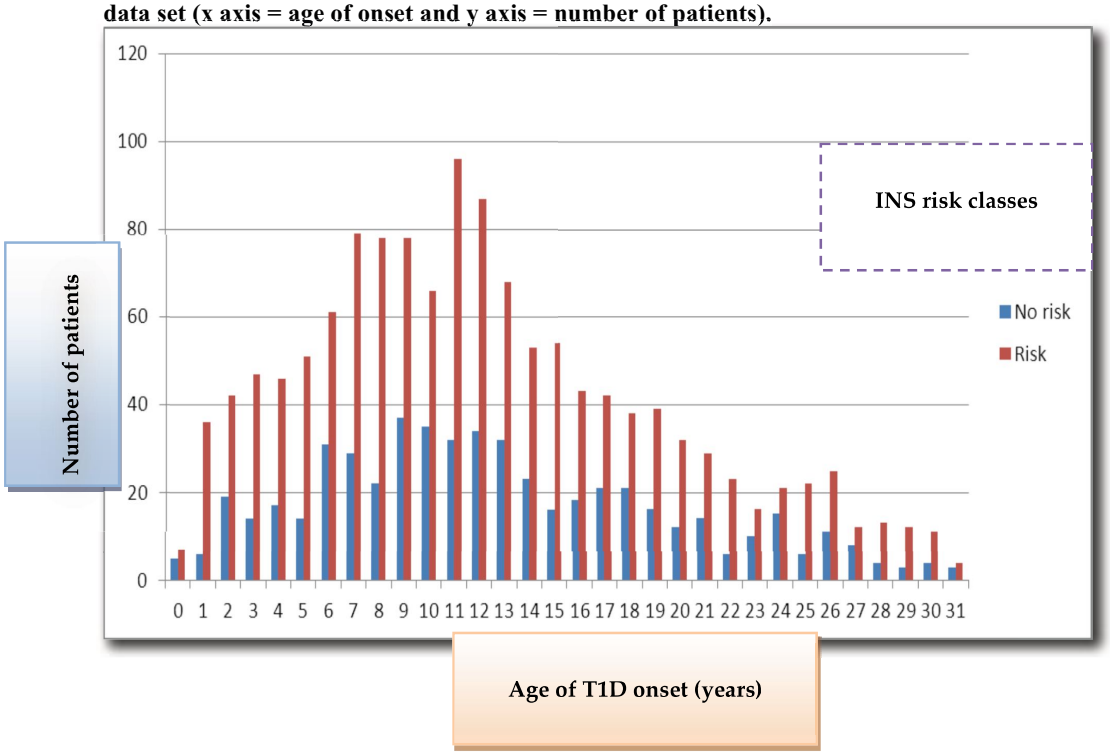


Figure 13. Age at disease onset. Frequency of HLA-DR risk classes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

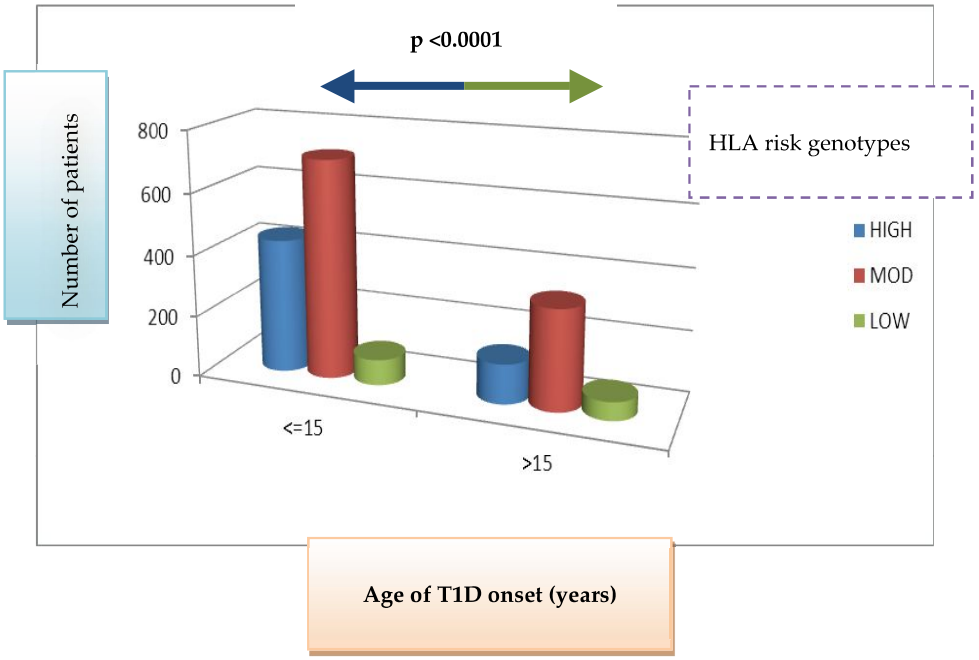


Figure 14. Age at disease onset. Frequency of INS genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

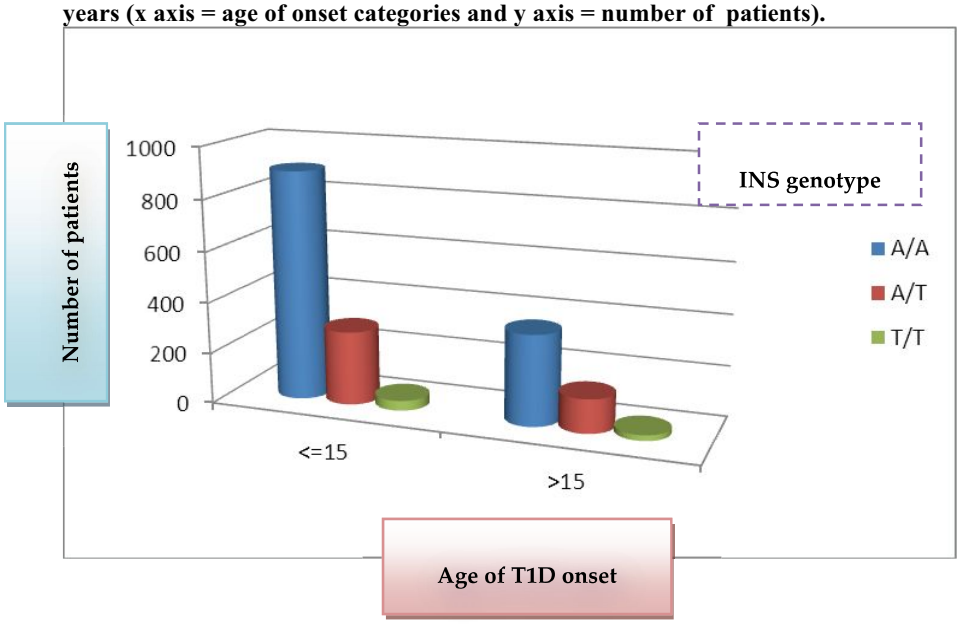


Figure 15. Age at disease onset. Frequency of INS risk genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

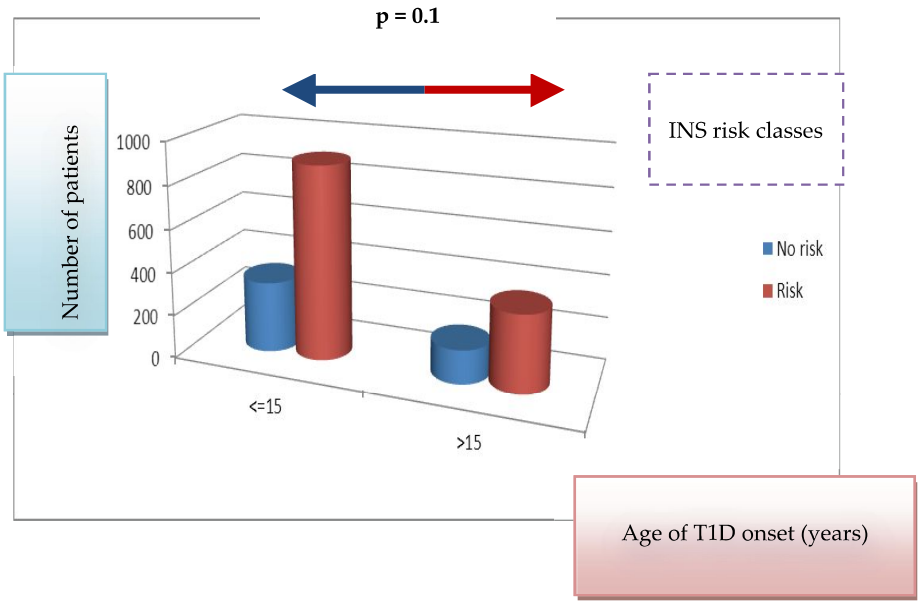
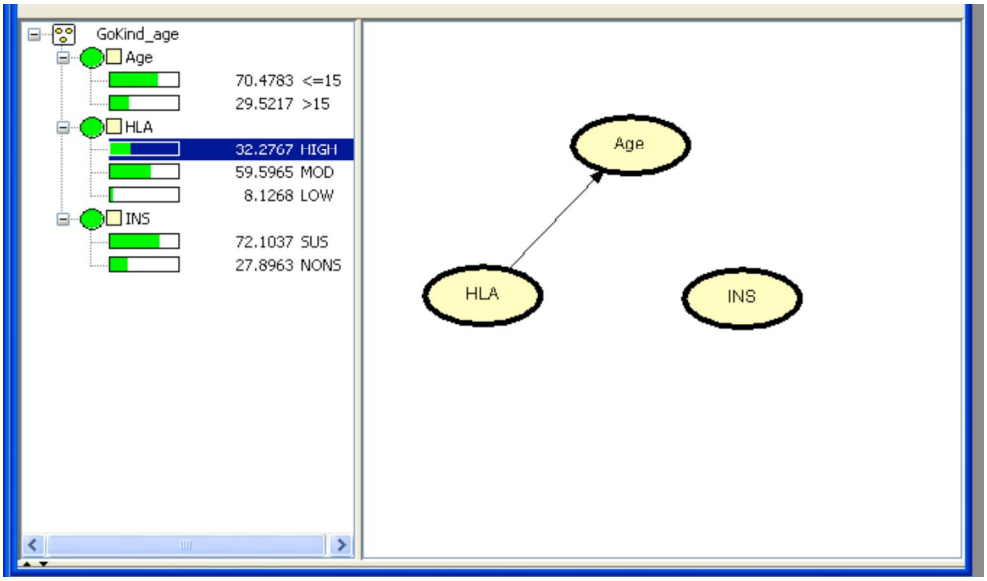


Table 2. Age at disease onset. Frequency of HLA-DR and INS genotypes in two age categories (<=15 and >15 years)

	Age of onset					
HLA-DR	**<=15 (N)**	%	**>15 (N)**	%	Total	%
High risk	436	35%	128	25%	564	32%
Mod risk	710	58%	323	63%	1033	59%
Low risk	83	7%	59	12%	142	8%
Total	1229		510		1739	
	Age of onset					
INS	**<=15 (N)**	%	**>15 (N)**	%	Total	%
Risk	329	27%	155	30%	484	28%
No risk	900	73%	355	70%	1255	72%
Total	1229		510		1739	

Figure 16. BN showing the relationship between HLA-DR gene and age of T1D onset in American patients sample (p-value=$6*10^{-6}$).



Data set B. Frequency of genotypes and age of T1D onset.

In the **Italian** sample, high, moderate and low risk HLA genotypes were carried by 20%, 66% and 13% of the patients. The frequency of genotypes analyzed is presented in Table 3 and Figs. 17-20.

Moreover, INS risk and no risk alleles were represented in 76% and 24%, respectively. PTPN22 susceptibility and non-susceptibility genotypes characterized 18% and the 82%, respectively. The age of disease onset was significantly co-related to HLA-DR genotype (p-value=$6*10^{-5}$). Moderate and low risk HLA- genotypes were significantly associated with the age at disease manifestation. In total, 66% of the subjects in the age category <=15 years and 52% in the age category >15 years carried the moderate risk HLA-DR genotype. Analogously, the low risk HLA-DR genotype was carried by 13% of subjects in the age category <=15 years and by 25% from patients with age of

>15 years. (Fig. 21). When chi-squared test was used, a significant statistical difference in age of disease onset between subjects carrying moderate and low risk HLA-DR genotypes, occurred, as illustrated in Figs. 16 and 24. INS and PTPN22 genes did not have any significant influence on the age of T1D onset (p-value>0.8 and 0.9, respectively) (Figs. 22-23).

Figure 17. Distribution of age of T1D onset in the Italian patients data set (x axis = age of onset and y axis = number of patients).

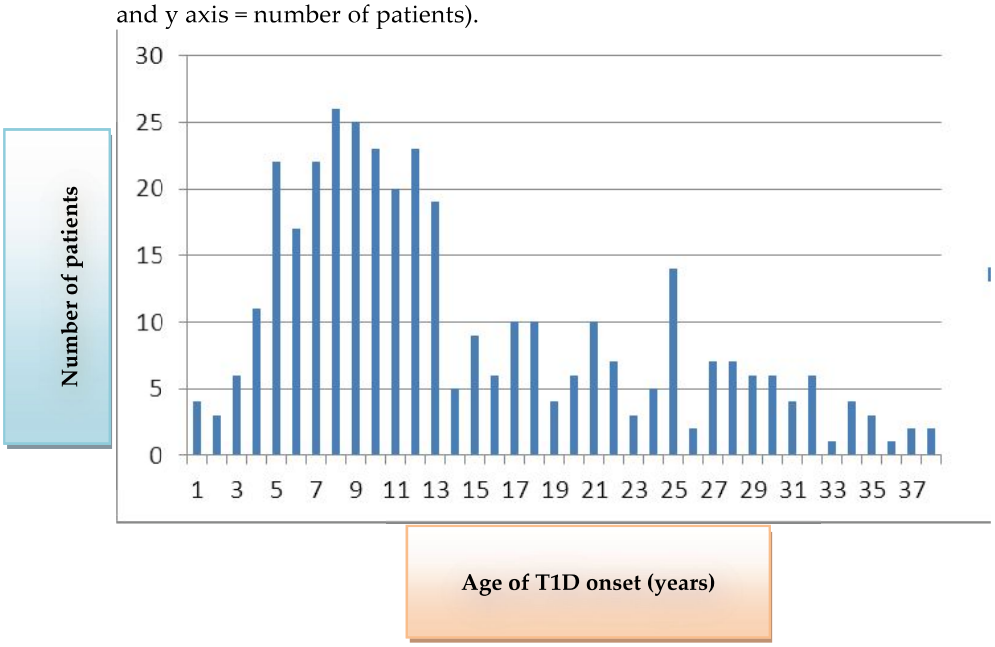


Figure 18. Distribution of HLA risk classes based on the age of T1D onset in the Italian patients data set (x axis = age of onset and y axis = number of patients).

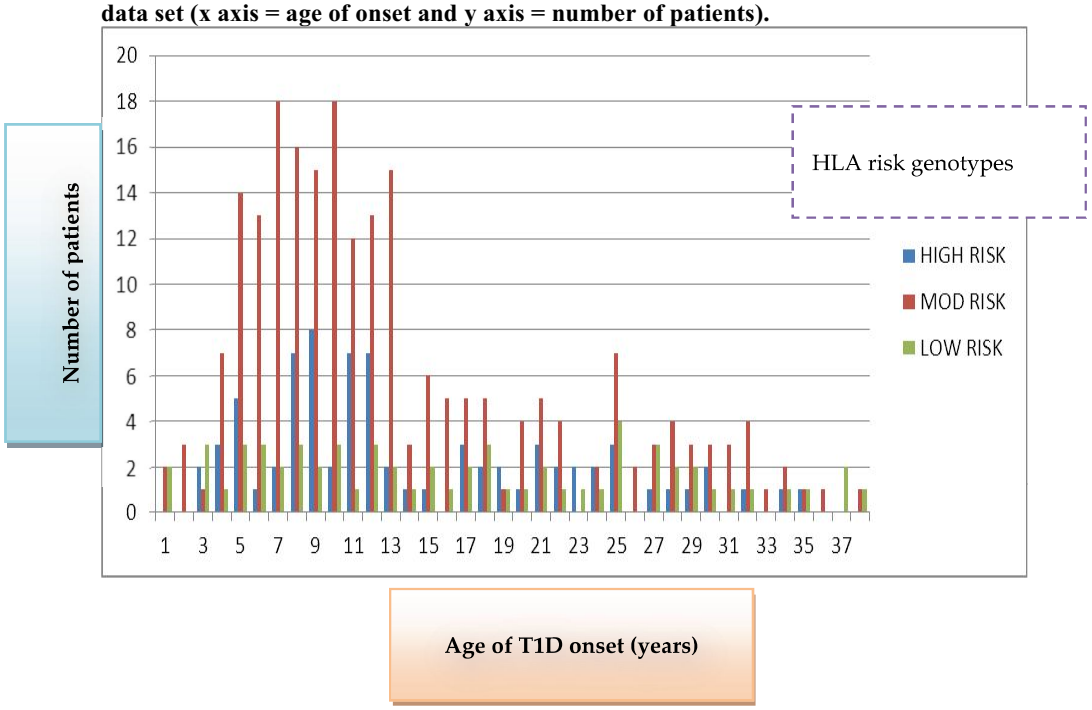


Figure 19. Distribution of INS risk classes based on the age of T1D onset in the Italian patients data set (x axis = age of onset and y axis = number of patients).

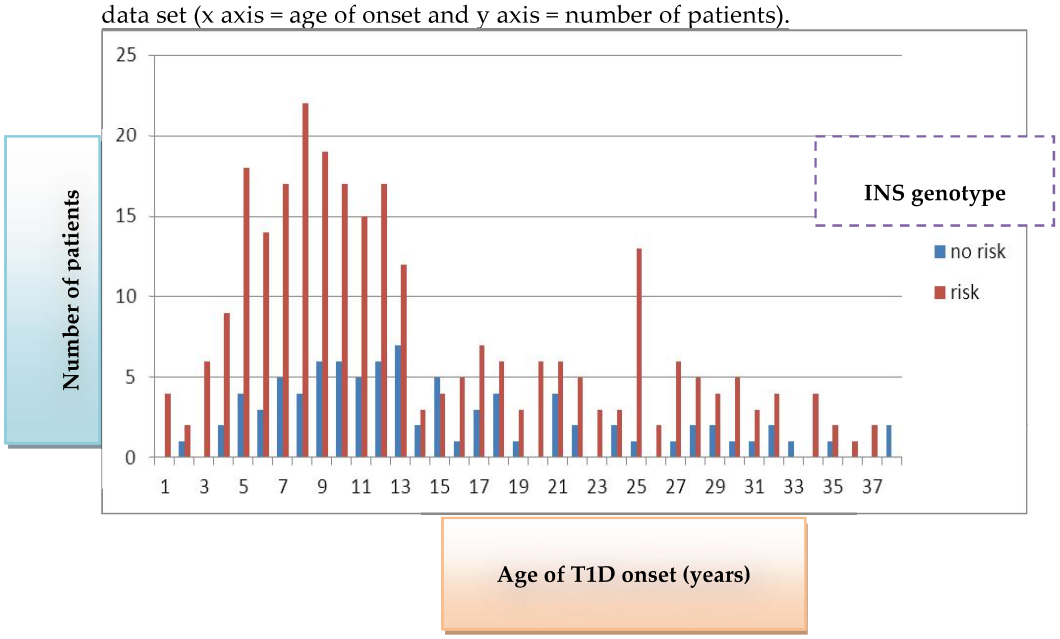


Figure 20. Distribution of PTPN22 risk classes based on the age of T1D onset in the Italian patients data set (x axis = age of onset and y axis = number of patients).

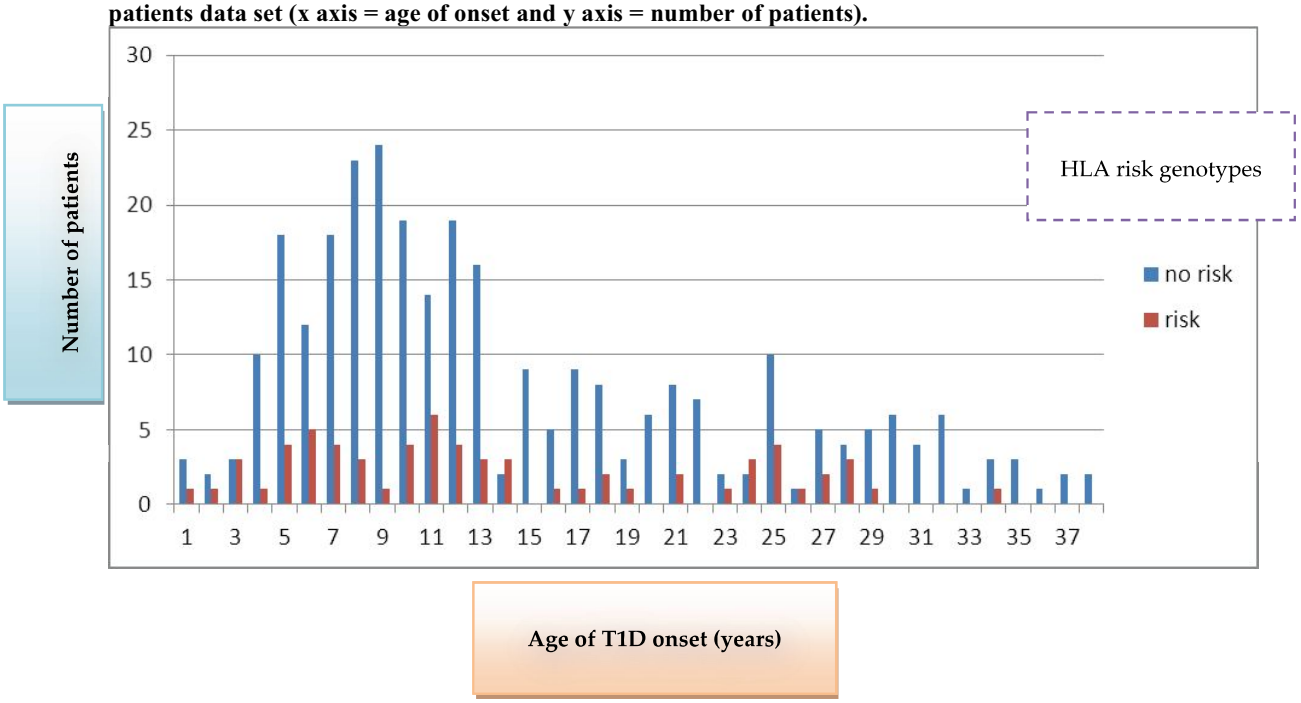


Figure 21. Age at disease onset. Frequency of HLA-DR risk classes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

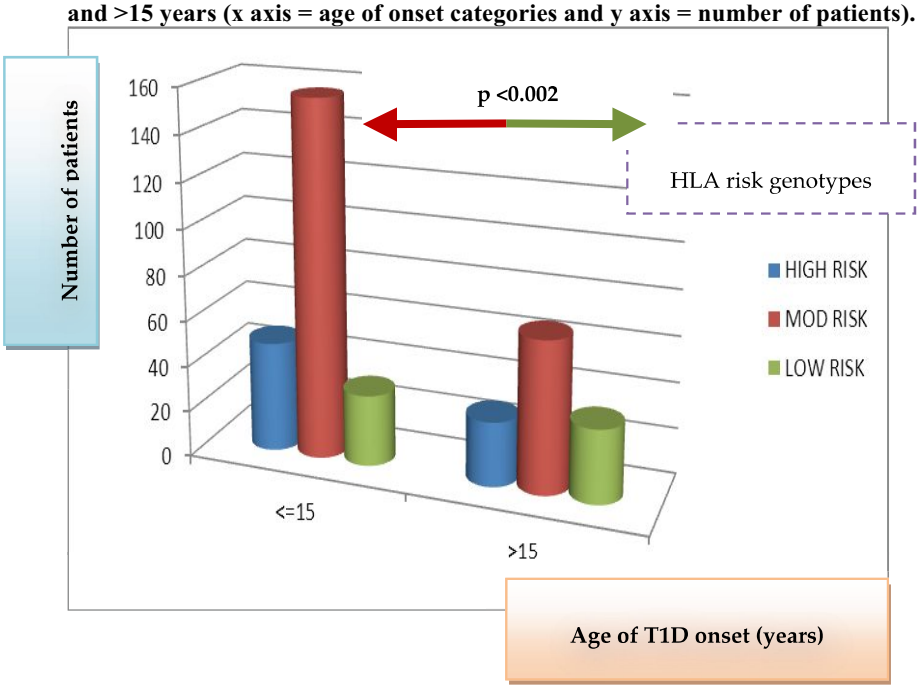


Figure 22. Age at disease onset. Frequency of INS risk genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

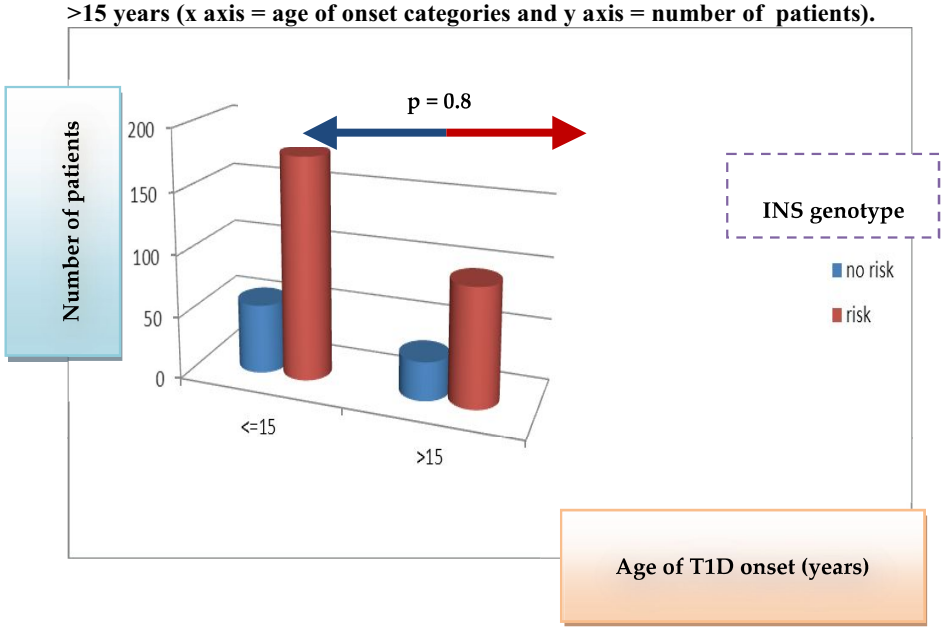


Figure 23. Age at disease onset. Frequency of PTPN22 risk genotypes in two age categories, <=15 and >15 years (x axis = age of onset categories and y axis = number of patients).

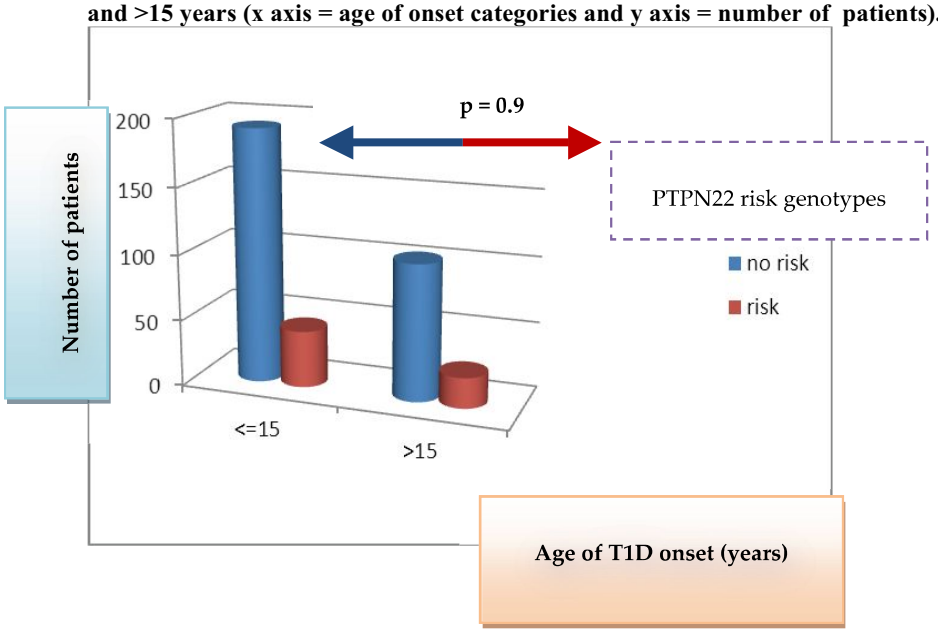
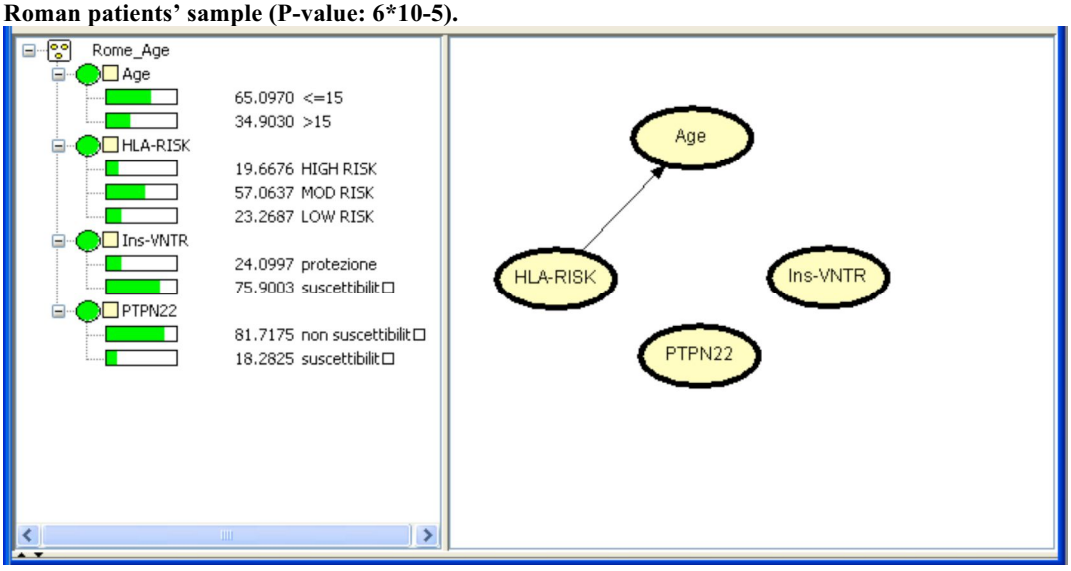


Table 3. Age at disease onset. Frequency of HLA-DR and INS genotypes in two age categories (<=15 and >15 years)

	Age of onset					
HLA-DR	**<=15**		**>15**		**Total**	
High risk	48	20%	28	22%	76	21%
Mod risk	156	66%	66	52%	222	61%
Low risk	31	13%	32	25%	63	17%
Total	235		126		361	
	Age of onset					
INS	**<=15**		**>15**		**Total**	
Risk	179	76%	95	75%	274	76%
No risk	56	24%	31	25%	87	24%
Total	235		126		361	
	Age of onset					
PTPN22	**<=15**		**>15**		Total	
Risk	43	18%	23	18%	66	18%
No risk	192	82%	103	82%	295	82%
Total	235		126		361	

Figure 24. BN showing the relationship between HLA-DR gene and age of T1D onset in the Roman patients' sample (P-value: 6*10-5).



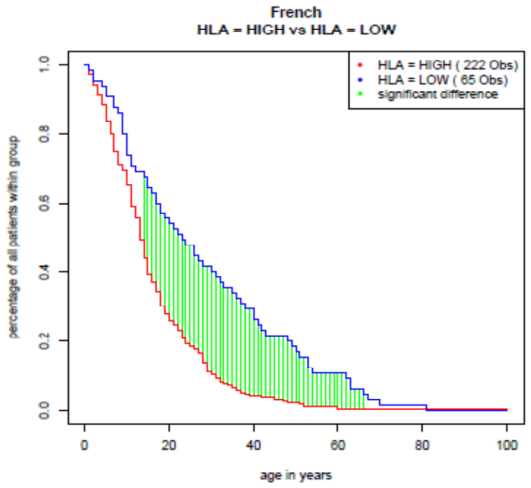
Timeline analyses. Four different comparisons were made to estimate which genes could have some role in influencing age at T1D onset: HLA high risk patients vs. HLA low risk patients, HLA-high/INS-susceptibility vs. HLA-high/INS-protection, HLA-moderate/INS-susceptibility vs. HLA-moderate/INS-protection and HLA-low/INS-susceptibility vs. HLA-low/INS-protection.

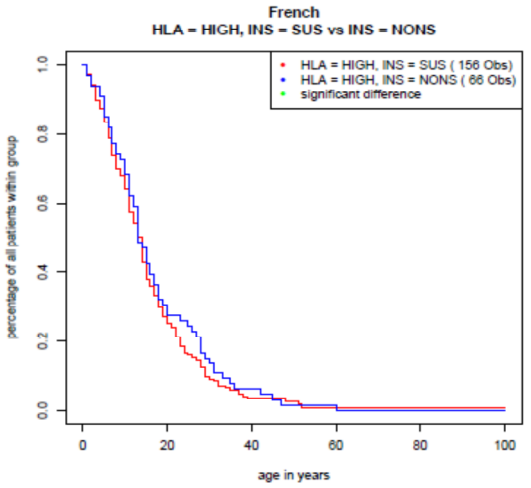
The influence of the combination of genetic risk factors (HLA-DR and INS-VNTR) on the age at onset is shown in Fig. 3 for the French data set, in Figure 4 for the Italian data set and in Fig. 5 for the American one. In all data sets the curves describing patients with high risk HLA are lower than for patients with low risk HLA. Thus, for all three populations our results, supporting previous research, confirmed that high risk HLA-DR genes play a key role in the early onset of diabetes.

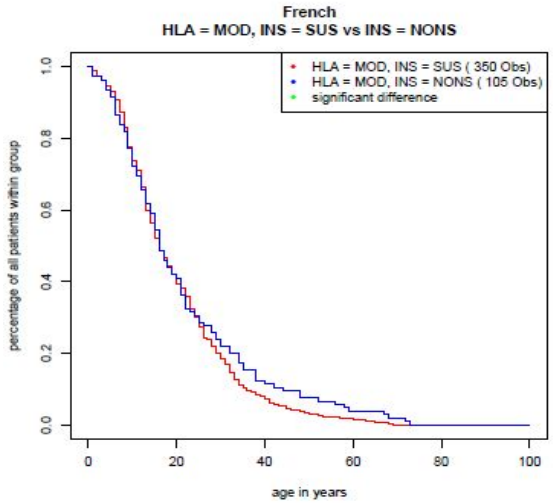
In each sample analysed the curves describing "HLA-high/INS-susceptibility vs. HLA-high/INS-protection", "HLA-moderate/INS-susceptibility vs. HLA-moderate/INS-protection" and "HLA-low/INS-susceptibility vs. HLA-low/INS-protection" gene combinations, are more or less asymptotic to each other. We can confirm that INS-VNTR gene cannot change the effects of HLA high, moderate and low risk alleles on age at T1D onset.

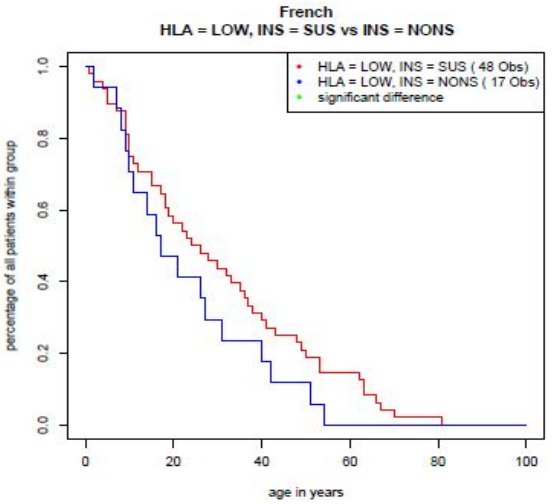
Starting at the age of about 15 years, the curves for "HLA low risk-INS susceptibility" are lower than the curve of "HLA low risk-INS protection" in the Italian and French data sets, showing a tendency of the INS-VNTR protection allele to delay the age at T1D onset (p-value=NS). This effect cannot be seen in the American data set.

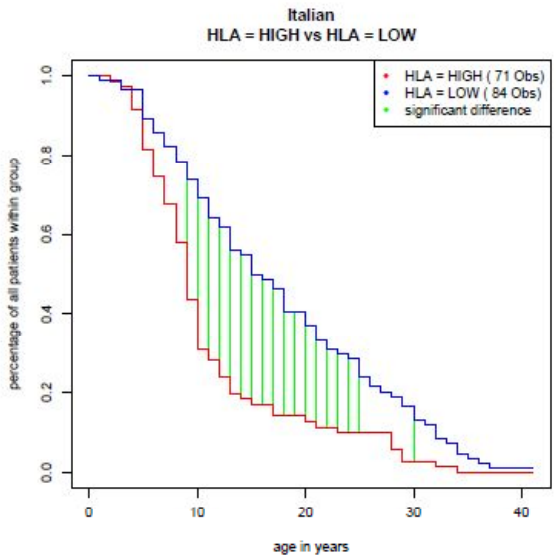
Figure 21. Timeline Analysis. The curves are empirical evaluations of the data sets for genetic risk factors combinations (patients only); a certain point of the curve shows the probability of getting R1D after a certain age. A green line denotes that the two curves are significantly different.

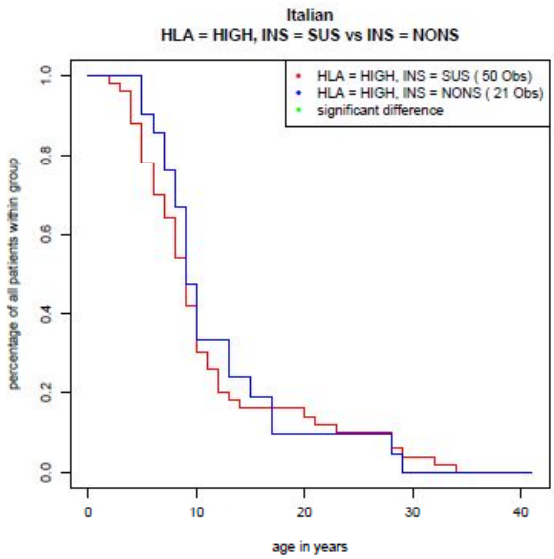


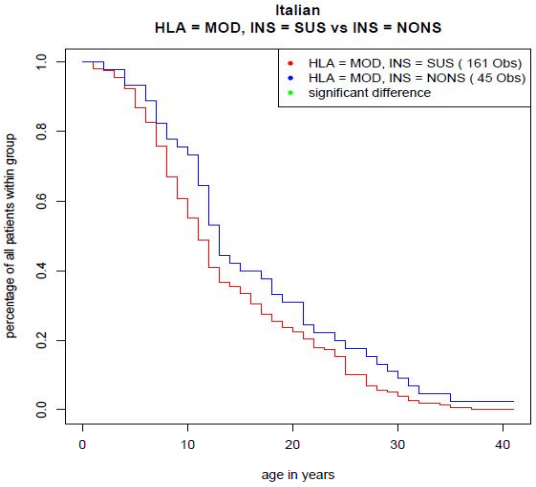


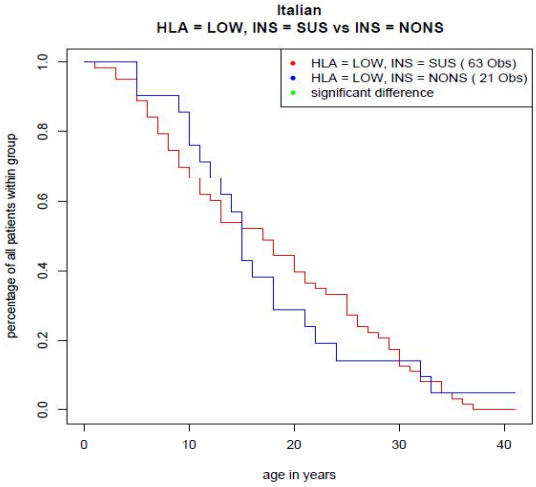


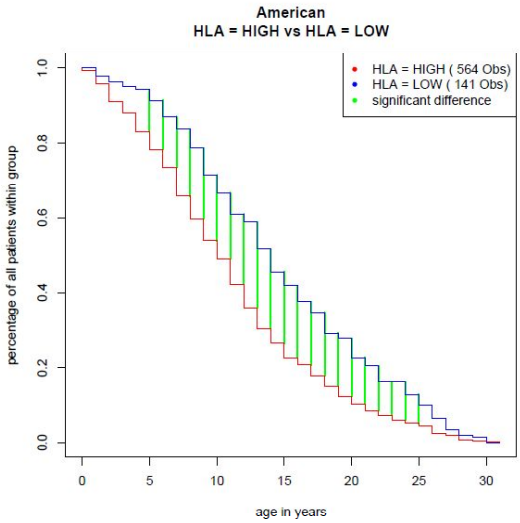


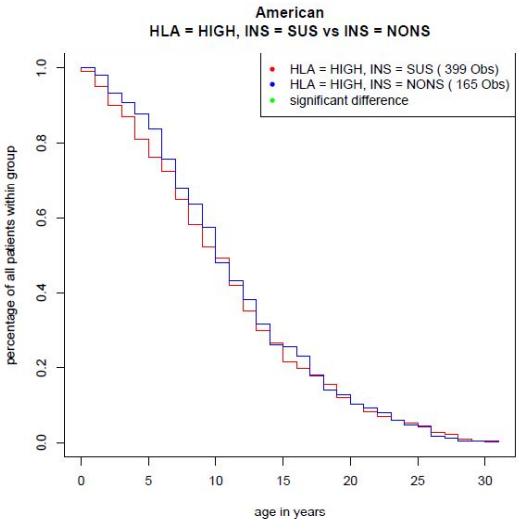


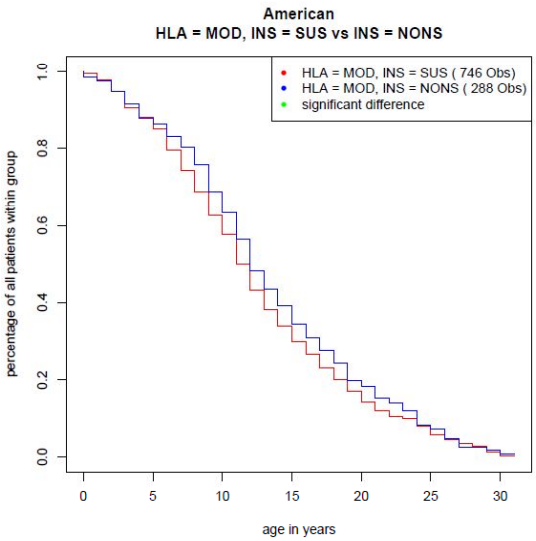


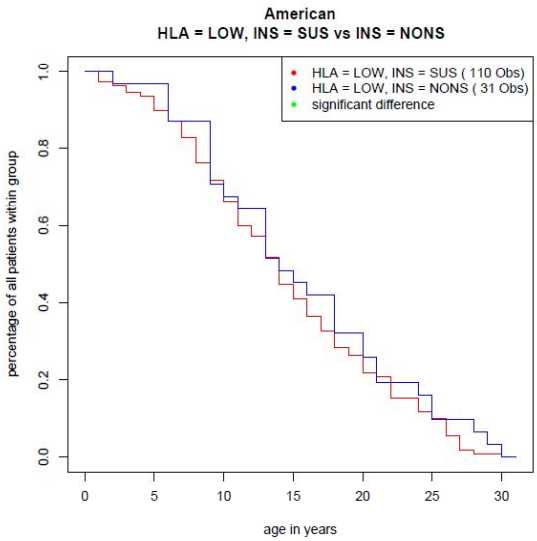












Assessment of the absolute risk of having T1D before or after the age of 15 years. BN analysis was used in order to assess the absolute risk of having T1D before or after the age of 15 years, based on all possible combinations of the risk classes of HLA-DR, INS and PTPN22 genes. For this purpose, two different samples were used: the case-control sample of data set "A" and the family clusters of data set "D" (for details on data sets, see paragraph 4.2).

Case-control study

Single-locus analysis. As shown in Fig. 25, considering the absolute risk of 0.4% of having T1D in the general Caucasian population, the BN analysis showed that the distribution of the risk was 0.2% and 0.19% in the age <=15 years and >15 years, respectively. Our results showed that the HLA gene determined the main change in the risk of having T1D in the two different classes of age <= or >15 years. The absolute risk conferred by the high, moderate and low risk HLA genotypes was 2.9%, 0.3% and 0.02% in the the age group <=15 years and 2%, 0.3% and 0.04% in the age group >15 years, respectively.

When the effects of non-HLA genes were considered individually, the absolute risk of having T1D in the age of <= 15 years or >15 years did not change.

Joint effect of INS and PTPN22 genes. INS and PTPN22 did not influence the risk of having T1D before or after the age of 15 years. This result was consistent with previous findings about the influence of non-HLA genes on the age of T1D onset as explained earlier (for details, see 4.3.3)

Joint effect of HLA and non-HLA genes. Considering all of the three HLA risk categories, INS and PTPN22 genotypes did not influence the age of the disease

onset and the absolute risk was equally distributed in the two age groups. Our results showed that absolute risk for T1D before the age of 15 years for an individual carrying risk alleles at the three susceptibility genes was 7%. The risk decreased to 4.5% for the age >15 years.

If the no risk genotypes for HLA, INS and PTPN22 were considered jointly, the absolute risk of T1D onset was 0.009% and 0.01%, before and after the age of 15 years, respectively.

Family study

Single-locus analysis. Age of disease onset was significantly related to HLA-DR risk categories, conferring an absolute risk of 12%, 3.2% and 1.1% in the age group <=15 years and 3.8%, 1.9% and 0.5% in the age group >15 years, respectively, if high, moderate and low risk HLA genotypes were considered. As showed in Fig. 26, considering the T1D absolute risk of 6% in siblings of T1D patients, the distribution of the risk was 4% and 2% in the age <=15 years and >15 years, respectively.

When the effects of non-HLA genes were considered individually, the absolute risk of having T1D did not change in respect of the age of <= 15 years or >15 years.

Joint effect of INS and PTPN22 genes. INS and PTPN22 genotypes, when jointly considered, did not show any influence on the age of onset. Namely, the absolute risk values were 0.6% and 0.07% for risk and non-risk INS and PTPN22 alleles, independently of the age of onset considered.

Joint effect of HLA and non-HLA genes. The absolute risk of having T1D before or after the age of 15 years, for all of the three HLA risk categories, did not change when the INS and PTPN22 genotypes were considered in the analysis. The absolute risk of having T1D was therefore equally distributed in the two age groups. Our results showed that a subject carrying risk alleles at HLA, INS and PTPN22 susceptibility genes, had an absolute risk of 14.4% for the onset of T1D before the age of 15 years. The risk decreased to 5.2% by the age >15. However, for a subject carrying non-risk alleles, the probability of having disease before and after the age of 15 years was 0.8% and 0.4%, respectively.

Figure 25. BN showing the distribution of the absolute risk by the two age classes, <=15 and >15 years in the general population, based on the case-control sample.

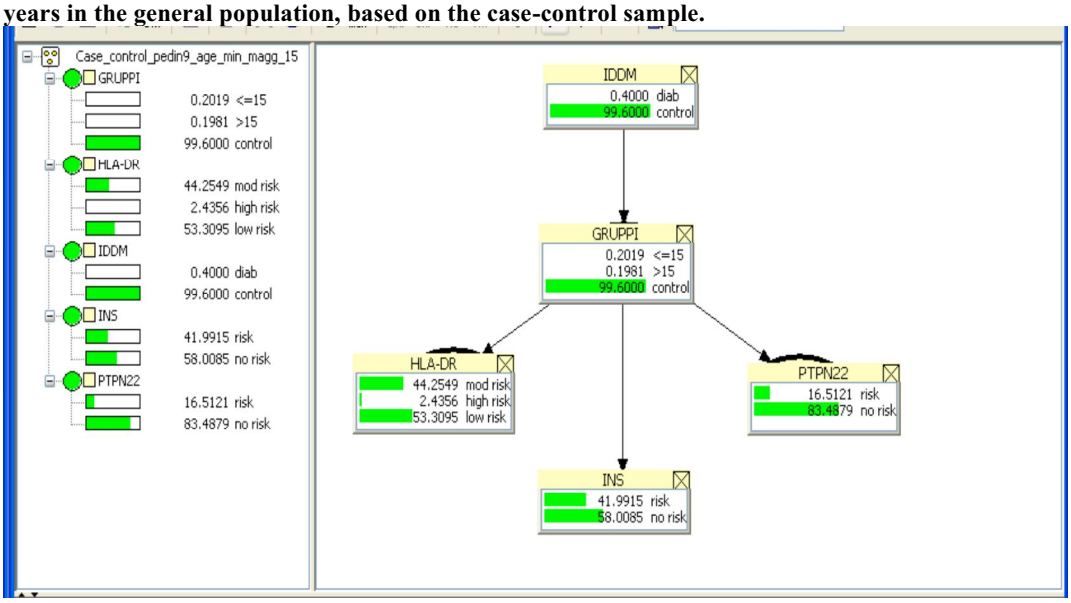
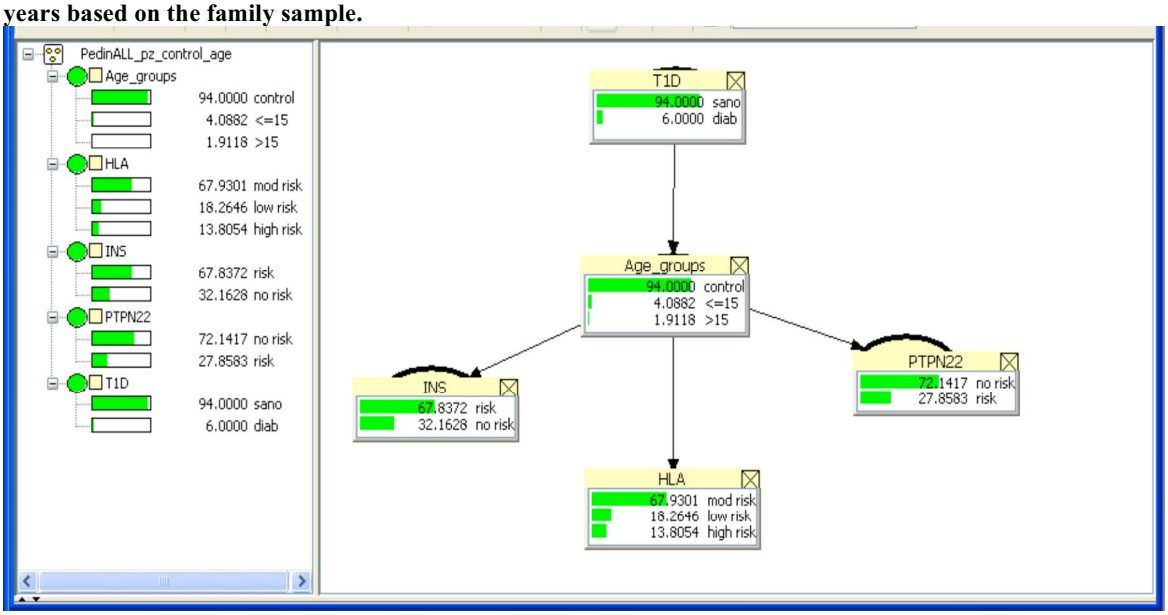


Figure 26. BN showing the distribution of the absolute risk by the two age classes, <=15 and >15 years based on the family sample.



3.4.3 Conclusion

In this study the effects of HLA-DR, INS and PTPN22 genotypes on the age of T1D onset were investigated in three different samples of T1D patients from French, American and Italian Caucasian populations. The joint absolute risk of having T1D before or after the age of 15 years was also examined using the BN method.

We observed that the distribution of HLA genotypes within risk categories in the case-only study, considering the three data sets analyzed, was as follows: high risk genotype (30%), moderate risk genotypes (60%) and low risk genotype (10%). This distribution was similar to that of other Caucasian populations (Pozzilli P, 2009).

Previous studies on age-dependent HLA genetic heterogeneity of T1D, indicate that HLA-DR3/DR4 strongly predisposes to T1D and is associated with a low age at disease onset (Gillespie KM, 2004; Caillat-Zucman S, 1992;Tait BD, 1995).

In this study, disease in patients with the HLA-DR3/4 genotype manifested at a significantly younger age with the high risk HLA-DR genotype manifested at a significantly younger age (15.7±11 years) than in patients with the low HLA-DR genotype, who had the highest mean age at disease onset in the sample (28±19 years).Our results thus confirmed the findings of previous studies, whereby T1D manifests at lower ages in individuals with the highly predisposing HLA-DR3/4 genotype and significantly later in patients with the low risk HLA genotype (Caillat-Zucman S, 1992). Association between age at disease onset and the moderate-risk genotypes was only significant in the Italian data set analysis.

The fact that the relative risks associated with both risk and low-risk genotypes seems to diminish with age above 15 years (Sabbah E, 2000), raises the question whether gene–gene interactions may also differ in different age-groups.

Moreover, in the light of the well-known influence that HLA plays on the age at T1D onset (Awa W 2010) and confirmed by this study, we also investigated whether INS and PTPN22 alone was associated with an effect on the age at disease onset. In our study no association between non-HLA genotype and age of T1D onset was observed in the three samples consisting of case-only (p-value=NS).

In this study the gene-related risk of having T1D before or after the age of 15 was also analysed, using the BN method. Our results in case-control data set analysis, showed that the absolute risk conferred by the high, moderate and low risk HLA genotypes was 2.9%, 0.3% and 0.02% in the age group <=15 years and 2%, 0.3% and 0.04% in the age group >15 years, respectively. Whereas, in the family study, HLA-DR risk categories, conferred an absolute risk of 12%, 3.2% and 1.1% in the age class <=15 years and 3.8%, 1.9% and 0.5% in the age group >15 years respectively, if high, moderate and low risk HLA genotypes were considered. In both groups studied, when the effects of non-HLA genes were considered both individually and jointly, the absolute risk of having T1D in the age of <= 15 years or >15 years did not change. This work represents, to the best of our knowledge, the first study aimed at assessing the absolute risk of having T1D based on the effects of the three main susceptibility genes for T1D on the age at disease onset. Moreover, the main strength it has is that the data analysis has been performed in three different samples, in a large and heterogeneous sample of patients characterized by an

age of onset in youth and in adult years. Nonetheless, similar studies over case-control data set are needed in order to back-up the results of this analysis.

REFERENCES

American Diabetes Association (2005) Diagnosis and classification of diabetes mellitus. Diabetes Care 28 S37–S42

Awa W, Boehm B, Kapellen T, Rami B, Rupprath P, Marg W, Becker M, Holl R. (2010). "HLA-DR genotypes influence age at disease onset in children and juveniles with type 1 diabetes mellitus." Eur J Endocrinol. 163:97–104

Caillat-Zucman S, Garchon HJ, Timsit J, Assan R, Boitard C, Djilali-Saiah I, Bougneeres P & Bach J. (1992) Age-dependent HLA genetic heterogeneity of type 1 insulin-dependent diabetes mellitus. Journal of Clinical Investigation 90;2242–2250.

Deborah J. Smyth, Jason D. Cooper, Joanna M.M. Howson, Neil M. Walker, Vincent Plagnol, Helen Stevens, David G. Clayton, and John A. Todd (2008) PTPN22 Trp620 Explains the Association of Chromosome 1p13 With Type 1 Diabetes and Shows a Statistical Interaction With HLA Class II Genotypes. Diabetes, 57; 1730-1737

Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A & Moonsamy P for the Type 1 Diabetes Genetics Consortium (2008) HLADR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes 57;1084–1092

Gillespie KM, Bain SC, Barnett AH, Bingley PJ, Christie MR, Gill GV & Gale EA (2004) The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. Lancet 364; 1699–1700.

Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M & Tuomilehto J (2003)

Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes 52:1052–1055

Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R & Tuomilehto J (2000) The Diabetes Mondiales (DiaMond) Project Group: incidence of childhood type 1 diabetes worldwide. Diabetes Care 23:1516–1526

L. Espino-Paisan, H de la Calle, M Fernández-Arquero, MÁ Figueredo, EG de la Concha, E Urcelay and JL Santiago (2011) A polymorphism in PTPN2 gene is associated with an earlier onset of type 1 diabetes. Immunogenetics 63; 255-258

Pozzilli P, Spoletini M, Zampetti S, Capizzi M, Kyanvash S, Venditti C & Buzzetti R IMDIAB Group (2009) Similar frequencies of HLA risk genotypes in type 1 diabetes over the past six decades in a Southern European population. Diabetologia 521–550

Sabbah E, Savola K, Ebeling T et al (2000) Genetic, autoimmune, and clinical characteristics of childhood- and adult-onset type 1 diabetes. Diabetes Care 23:1326–1332

Tait BD, Harrison LC, Drummond BP, Stewart V, Varney MD & Honeyman MC. (1995) HLA antigens and age at diagnosis of insulin-dependent diabetes mellitus. Human Immunology 42;116–122).

3.5 Third aim. Do HLA-DR, INS-VNTR and PTPN22 interact with each other?

Rationale

In the last years the search for genetic factors influencing multifactorial diseases has been revolutionised by genome-wide association studies. Most of them have used a single-locus analysis approach, whereby each locus was tested individually for association with a particular disease. However, the existence of interactions between susceptibility loci (defined as epistasis) was the reason for the lack of success in genetic studies of complex diseases. Epistasis means that something different happens when a particular set of alleles, from different loci, are found in combination than when they are apart. So that the joint effect must be different from what we would expect if the effects of the two loci were combined independently. From statistical viewpoint, epistasis refers to an interaction between multiple genetic loci such that the disease cannot be predicted by simply combining the effects of the single susceptibility gene. Fisher defined epistasis as a deviation from the additive model of allelic effects[1] (Fisher RA, 1918) and probably he chose this definition because additive linear models are tractable from a statistica[l] point of view. In the late 1960s, epistasis was defined as deviation from a

1 For a haploid model, this could be represented as Wxy = ax + ay + b, where "a" is the individual effects of each allele at loci x and y, "b" is the deviation that is due to epistasis and W is the observed phenotype. Relationships for diploids are more complex because of the possibility of one locus interacting with the dominance state of the other locus (Musani SK, 2007).

multiplicative model of gene action, instead of deviation from the additive model.[2]

Previous studies have tried to evaluate if and how T1D susceptibility genes interact with each other. Different methods have been applied and several studies have been performed (Julier C 1991; Bain SC 1992; van der Auwera BJ 1993; Dizier MH 1994; She JX 1994; Cordell HJ 1995; Metcalfe KA 1995; Laine AP 2004; Motzo C 2004; Bojornvold, 2008). Nonetheless, the true nature of statistical and biological interaction between HLA and the other T1D susceptibility genes still remains unclear.

Our previous risk analysis, reported in "Results" section (paragraph 4.3) showed that the joint risk conferred by HLA, INS and PTPN22 genes on T1D was far from a simple combination of risks due to the effects of the single susceptibility genes. Based on this finding, the aim of this work was to investigate the possible statistical interaction between HLA-DR, INS and PTPN22 genes to gain insight into the genetic basis of the disease, using BN probabilistic approach.

3.5.1 Materials and methods

Two different data sets consisting of genetic data from T1D patients (HLA-DR, INS-VNTR and PTPN22), were analyzed (namely, data set A and B; for details on samples see paragraph 4.2). BN algorithm was implemented to test the statistical interactions between genes, for each data set. The definition of

2 A multiplicative haploid model would be represented as $W_{xy} = ax*ay + b$.

statistical interaction here used is a correlation where the effects of one factor go in the same direction at different levels of the other, but differ in magnitude (Duncan Thomas. 2010). To figure out how accurately BN models define the gene interaction map, we used the Akaike Information Criterion (AIC) (Akaike 1974), an estimator of predictive accuracy. According to this criterion the network with the highest AIC score was selected as the *best network*.

The validity of the edges, indicating the statistical correlation (interaction), was measured by testing the mutual information between the parent node and the corresponding child and was compared to a chi-square distribution. The corresponding p-value was used to assess the statistical significance of the arrows and thus the strength of the interaction.

Moreover, the fact that the genetic relative risks diminish from the age of 15 years and above (Sabbah E, 2000), raises the question whether the gene–gene interactions, eventually present, may also differ in different age-groups. To highlight this concept, in the second part of this "Third aim", the statistical interactions between HLA, INS and PTPN22 genes, related to the age at disease onset, were investigated using BN method.

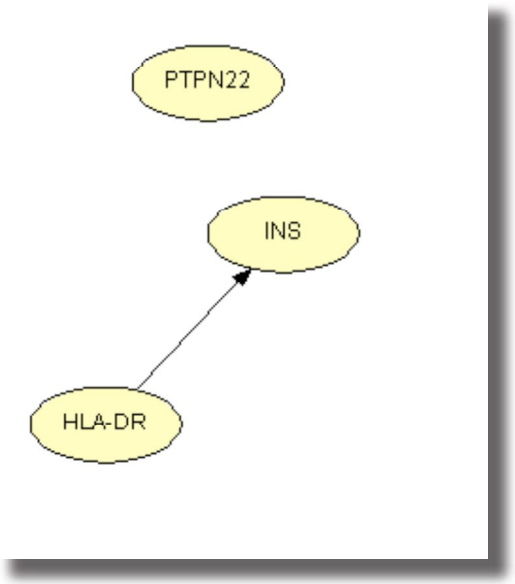
3.5.2 Results

We used genetic data from the three genes, HLA-DR, INS an PTPN22 classified into three and two main risk categories: high, moderate and low risk for HLA-DR genotypes and risk and non-risk for both INS and PTPN22 genes (as described in detail in section 4.2). BN analysis was applied in French and Italian case-only data sets (named data set A and B, respectively). In this

analysis the AIC-scores for all possible BN were calculated and the best network (characterized by the highest AIC-score) was selected.

Based on the data sets A and B, BNs resulting from learning process showed differences in the gene interaction findings. BN representing data set A, showed a statistically significant interaction between HLA-DR and INS-VNTR genes (p-value = 0.01, AIC score = -1811). To highlight the meaning of the interaction, the alleles combinations of HLA-DR and INS, expressed by risk classes, were analysed, and statistically significant differences between the genes associations were found. In particular, INS risk and non-risk genotypes were associated in 32% and 41% of the cases with the high risk HLA. Whereas in the 60% and 49% of the cases, risk and non-risk INS alleles were associated with the moderate risk HLA (chi-squared= 7.98, degrees of freedom = 2, p = 0.01). The interaction was therefore risk allele dependent. No interaction was found between PTPN22 and INS genes nor PTPN22 and HLA-DR genes (Fig. 1).

Figure 1. BN built from case-only data set A. The arrow represents the statistical interaction between HLA-DR and INS genes.



The BN representing data set B, did not show any interaction between the genes.

To go deeper into the differences between the results, further analyses were done on French and Italian data sets, after performing a prior data selection: all data was divided, according to the age at T1D onset, into three age groups: ≤ 10 years, from 10 to 20 years and ≥20 years. The aim of this approach was to investigate if and how gene interactions could change depending on the age at T1D onset. Patient sub-groups with a different age at T1D onset of both data sets were analysed using BN. No interaction between genes was showed by the age-related BNs for both data sets. Furthermore, we did not observe any significant differences in the distribution of HLA-DR and INS genotypes according to age at onset of T1D nor in the relative associations of INS genotypes conditional on HLA genotypes when the patients were divided into the three different groups according to the respective age of disease onset (0–10 vs. 11–19 vs. >20 years).

3.5.3 Conclusion

The aim of this work was to investigate the possible statistical interaction between HLA-DR, INS and PTPN22 genes to gain insight into the genetic basis of the disease, using BN probabilistic approach. For this purpose two different samples consisting of genetic data from T1D Caucasian patients were analysed (data set A and B). BN representing data set A showed a statistically significant interaction between HLA-DR and INS-VNTR genes (p-value = 0.01, AIC score = -1811) (Fig. 1), whereas the BN representing the data set B, did not show any genes interaction. When these results were checked using the chi-squared statistics, in data set A, a significant statistical correlation between the HLA-DR and INS alleles was found (p = 0.01). Namely, INS predisposing genotype was significantly less frequent in high-risk HLA genotype–positive patients than in those with HLA intermediate-and low-risk categories, confirming previous findings in literature (Motzo C, 2004; Laine AP, 2004). Taking into account the evidence that the genetic relative risks could diminish from the age of 15 years and above (Sabbah E, 2000), we hypothesize that the gene–gene interactions, eventually present, may also differ in different age-groups. We then investigated the statistical interactions between HLA, INS and PTPN22 genes, based on the age at disease onset. Our results showed that for both data sets no interaction between genes was found by the age-related BNs .

The use of BN for detecting gene interaction has its advantages when compared with the traditional methods. The advent of high-throughput technologies has enabled genome-wide association studies (GWAS). GWAS involves sampling in a population of individuals about 500,000 representative SNPs. A significant challenge in the analysis of genome-wide data sets is the identification of interacting loci that interact in their association with disease.

Many existing methods for epistasis learning such as combinatorial methods cannot handle a high-dimensional GWAS data set. For example, if we only investigated all 0, 1, 2, 3 and 4-SNP combinations when there are 500,000 SNPs, we would need to investigate 2.604 × 1021 combinations. In this scenario, BN approach represents a valid and alternative method to manage this huge amount of genetic information with the aim of detecting the interaction between them. Moreover, BN provide a graphical representation of the independence between the modelled variables, which allows for transparency and ease of interpretation of the models and their parameters. The presence of interaction between susceptibility genes can explain why the study of a single susceptibility gene in polygenic disorders, such as T1D, produced fewer results than expected.

The weakness of our study was the rather low number of patients and the absence of a representative sample consisting of control subjects. The comparison between patients and controls could be useful for achieving this purpose. Further studies are needed to verify our results and to better understand the genetic interaction between the main susceptibility genes for T1D. However, in this study we had two different cohorts of patients, so that one of them was used as a replication data set. Thanks to this, the interaction hypothesis was tested in two samples consisting of T1D patients from Italian and French populations. The reason for the diverging results obtained in this study, could be based on the heterogeneity in HLA-associated risk of T1D Although no preventive intervention is available for T1D today, prediction of disease is an important part of prevention strategies, both for recruitment of participants for research studies and for identification of target populations for future preventive interventions. Understanding the interaction between the established T1D susceptibility genes will favour this possibility. Further

substantiation is needed for the identification of the statistically significant conditional dependency of INS on HLA.

REFERENCES

Akaike H (1974) Information Theory and an Extension of the Maximum Likelihood Principle." Automatic Control, IEEE Transactions on 19; 716-723

Bain SC, Hearne CM et al (1992) Insulin gene region-encoded susceptibility to type 1 diabetes is not restricted to HLA-DR4-positive individuals. Nature Genetics 2: 212-215

Bjørnvold M, U. D., Joner G, Dahl-Jørgensen K, Njølstad PR, Akselsen HE, Gervin K, Rønningen KS, Stene LC (2008) Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes Diabetologia 51: 589–596.

Cordell HJ, Bennet ST, Kawaguchi Y, Farral M (1995) Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes." Am J Hum Genet 57: 920-934

Dizier MH, B. M., Clerget-Darpoux F (1994) Interactive effect of two candidate genes in a disease: extension of the marker-association-segregation chi square method. Am J Hum Gener 55: 1042-1049.

Duncan Thomas (2010) Gene-Environment-Wide Association Studies: Emerging Approaches. Nat Rev Genet. 11:259–272

Fisher RA (1918) The correlations between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb;52:399–433

Julier C, H. R., Davies J, Merlin F, Soularue P, Briant L, Cathelineau G et al (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature 354: 155-159

Laine AP, Hermann R, Knip M, Simell O, Akerblom HK, Ilonen J (2004) The

human leukocyte antigen genotype has a modest effect on the insulin gene polymorphism-associated susceptibility to type 1 diabetes in the Finnish population. Tissue Antigens 63:72–74

Metcalfe KA, HG, Fennessy MJ, McCarthy MI, Tuomilehto J, Tuomilehto-Wolf E (1995) In Finland insulin gene region encoded susceptibility to IDDM exerts maximum e ect when there is low HLA-DR associated risk. Diabetologia 38: 1223-1229

Motzo CCD, Cordell HJ et al (2004) Heterogeneity in the magnitude of the insulin gene effect on HLA risk in type 1 diabetes. Diabetes 53: 3286 3291

Musani SK, et al. (2007) Detection of gene × gene interactions in genome-wide association studies of human population data. Hum Hered;63:67–84

Ng AY, Jordan M (2002) On discriminative vs. generative classifiers: a comparison of naive Bayes and logistic regression. In: T.G. Dietterich, S. Becker, Z. Ghahramani (editors). Advances in Neural Information Processing Systems, 14, pp. 605 – 610. MIT Press, Cambridge, Mass

Sabbah E, Savola K, Ebeling T et al (2000) Genetic, autoimmune, and clinical characteristics of childhood- and adult-onset type 1 diabetes. Diabetes Care 23:1326–1332

She JX, Tian XH et al (1994) Additive susceptibility to insulin-dependent diabetes conferred by HLA-DQB1 and insulin genes Autoimmunity 18: 195-203

Twardy, C.R., Nicholson, A.E., Korb, K.B., McNeil, J. (2006) Epidemiological data mining of cardiovascular Bayesian networks. Electronic Journal of

Health Informatics, vol. 1, no. 1.

van der Gaag LC (2008) Aligning Bayesian Network Classifiers with Medical Contexts. Technical Report. http://www.cs.uu.nl/research/techreps/repo/CS-2008/2008-015.pdf).

CHAPTER 4. INTRODUCTION TO SYSTEM BIOLOGY

('The aim of System Biology is to understand biological processes as integrated systems instead of as isolated parts')

(http://is.gd/tQK0)

4.1 The Systems Biology: an answer to the complexity of biological systems

The panorama of new challenges and research opportunities has to do with *predictions of physical reality* by computer models. Today the phenomena and processes we ask computer models to predict are of enormous importance to critical decisions that affect our welfare concerning for example the biology of diseases and the outcome of medical procedures.

In recent years the *interdisciplinary efforts,* developed so far, aimed at elucidating structures and functions of living systems. The main challenges in computational modeling and analysis were to understand, analyze and predict the complex mechanisms of biological systems. Research investigations in computational biology and physiology have addressed important issues across many applications from, biological signaling pathways, cellular biology and communication, organ function and performance, all the way up to lifestyle and environmental influences and behavioral responses. Over the last few years, this kind of research work is being extended to explore translational biomedical research, to better understanding the mechanisms of disease and its treatment, thus helping to establish diagnostic biomarkers, physiology-based patient selection criteria, and strategies for choosing, personalizing and optimizing therapeutic options.

Computational modeling promises to become a fundamental contributor to future biomedical sciences and technologies, and personalized predictive healthcare.

In this scenario, Systems Biology represents one of the most fashionable and modern approaches to complexity of biological systems.

Starting from "system" definition, as an entity that maintains its existence through the mutual interaction of its parts (von Bertalanffy, 1968), we can say that Systems Biology may represent an approach (as detailed below), rather than a discipline (such as Biology). However, classic physiology largely lacked the ability to pursue the quantitative integration of observed behaviors with a few notable exceptions, such as the *Hodgkin–Huxley equations for the nerve impulse* (Hodgkin and Huxley 1952), their application to the *heart* (Noble, 1962), or to the *pancreatic beta-cell* (Chay and Keizer, 1983) or the early ideas of Guyton for a quantitative model of the *circulation* (Guyton et al, 1972). Systems biology, on the other hand, thrives on the revolutionary improvement of experimental techniques to investigate system components and their interactions and it is based on significant advances in computational power, tools, and techniques, which allow quantitative modeling.

The essence of Systems Biology lies in providing a framework for the re-unification of biological studies with 'the other' sciences, it represents the approach on which quantitative descriptions of parts and their interactions give rise to an understanding of the maintenance of biological entities across all relevant levels of structural and functional integration (Fig. 1).

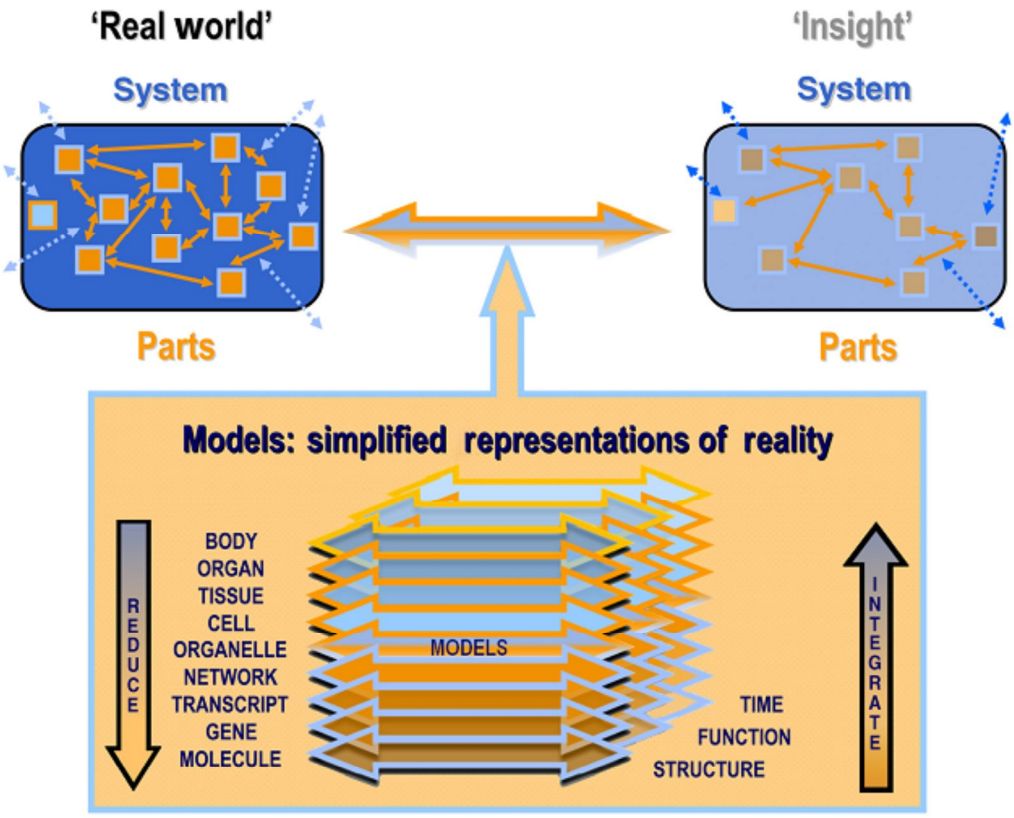
The best results are often obtained when theoretical work is pursued in close and continuous iteration with experimental and/or clinical investigations. From the point of view of experimentalists, ignoring computations often

seems justified, since computations tend to be highly idealized, with respect to the "real things" we suppose to obtain in laboratory. Furthermore, ignoring experiments often seems justified from the point of view of theorists, since experiments also do not often replicate well or they pertain to such unrepresentative special cases that one hardly knows what general lesson to draw from their outcomes.

The utility of both seems greatly enhanced when the same individual or laboratory does both computations and experiments, taking both seriously to refine the other in the next iteration and when the prominence of artifacts in both methods is given due attention.

The second purpose of this PhD project is based on Systems Biology approach and its aim is to evaluate the effect of progressive lymphocytic islet infiltration (insulitis) on beta-cells signalling. Before presenting the results obtained, we explain the following concepts: a) the structure of pancreatic islet of Langerhans, b) the beta-cell electrophysiology and c) the mathematical models of bursting electrical activity of beta-cells.

Figure 1. Our understanding of 'real world systems' (top left) usually is translated in a simplified representation (top right) of that reality. The progressive development of this understanding is based on the application and analysis of experimental and theoretical models. For biological systems research, these models allow the exploration of relevant structural levels between body and molecule. M Like their experimental counterparts, computational models are - by the very definition of the term 'model'- simplified representations of reality. Like items in a toolbox, models for biomedical research, have a range of applications for which they are suited.



Before dealing with the simulated autoimmune damage process of beta-cells, the following is a brief description of the anatomic structure of islet of Langerhans in the endocrine pancreas and the physiological process involved in the production and release insulin.

4.2.1 Structure of Pancreas and the Islet of Langerhans

The **pancreas** is a mixed exocrine and endocrine gland. It is an elongated conical organ about 12-15 cm in length, lying laterally to the rear of the upper right hand side of the abdominal cavity. The exocrine and endocrine functions of the gland are carried out by two histologically distinct subunits. The exocrine portion is organized as a tubuloalveolar gland, acini empty into centroacinar ducts which conduct to progressively larger intralobular ducts, interlobular ducts and eventually to the main pancreatic duct. Both acini and the ductal epithelium contribute to pancreatic secretion. The ductal cells are stimulated by the hormone secretin from the gastro intestinal tract to secrete bicarbonate ions and water. This alkaline secretion neutralizes the acidic chime from the stomach, bringing the pH into the optimal range for the action of the pancreatic enzymes within the duodenum. The acinar cells secrete an extensive range of digestive enzymes which break down specific components of the chyme as part of the digestive process.

The endocrine function is performed by clusters of cells called islets of Langerhans, which are separated from the tubuloalveolar components by a loose reticular capsule. The islet are richly vascularized by a network of fenestrated capillaries. The course of capillaries allows a degree of autoregulation of the pancreas in that small intralobular arteries first supply the capillary network of the islet, then subsequently invest the nearby acini, allowing the autocrine control of pancreatic function. The part of pancreas with endocrine function is made up of a million cell clusters called **islets of Langerhans**. They consist of four main cell types that can be classified by their secretion: alpha cells (secreting glucagone), beta-cells (secreting insulin), delta cells (secreting somatostatin and gastrin) and PP cells (secreting pancreatic polypeptide).

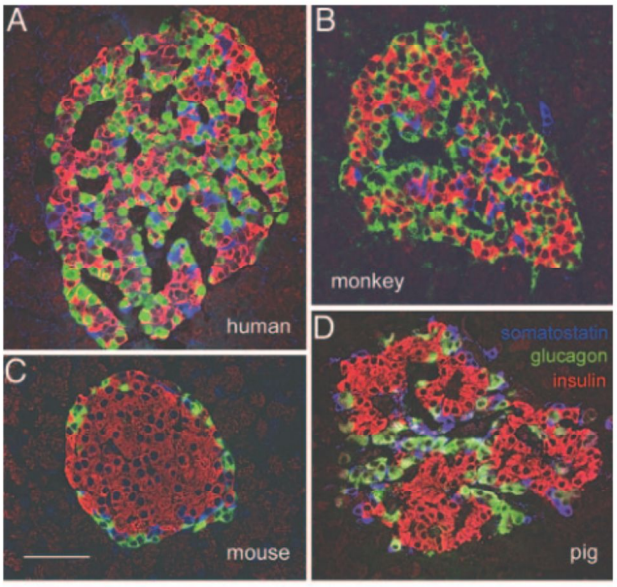
There is a regional heterogeneity in the pattern of endocrine cells in islet, for example, the glucagon-secreting cells are more abundant in the body and tail of the pancreas in contrast to the more frequent PP-secreting cells present in the head of the pancreas (Fig. 2). There have been extensive studies about these different cells, particularly beta-cells, which constitute the major represented cell type (occupying 70–80%) in an islet (Brissova M, 2005). Beta-cells are located adjacent to blood vessels and can easily respond to changes in blood glucose concentration by adjusting insulin production. Insulin facilitates uptake of glucose, the main fuel source, into cells of tissues such as muscle. So, the islets of Langerhans play a key role in maintaining the appropriate level of glucose in the blood; their malfunction is closely related to diabetes. In human being there are approximately 1 million of islets of Langerhans scattered in the pancreas. Quantitative studies assessing the cell composition of human islets are sparse (Stefan Y, 1982; Rahier J, 1983; Clark A, 1988) and these studies have reported that islets are composed of ≈ 70% beta-cells, alpha 20% cells, delta <10% cells, and <5% PP cells. More recent studies, however, have provided lower numbers of beta-cells and higher numbers of alpha-cells (60% and 30%, respectively) (Street CN, 2004; Brissova M, 2005). Contrary to the pancreatic islet structure in mouse human islets do not show anatomical subdivisions. Human cells are not clustered and most part (71%) shows associations with other endocrine cells, suggesting unique paracrine interactions. In fact, as the last findings have shown (Cabrera O, 2006), alpha, delta and beta-cells were found scattered throughout the human islet. Most of beta, alpha and delta cells were aligned along blood vessels with no particular order or arrangement, indicating that islet micro-circulation most likely does not determine the order of paracrine interactions. Moreover, studies performed by Cabrera et al. have shown striking species differences

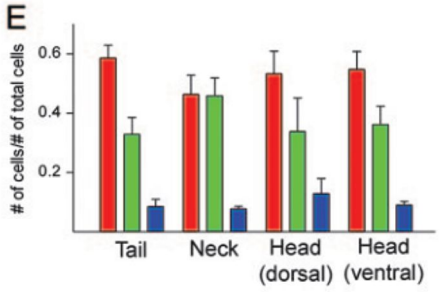
with regard to both cyto-architecture and function and emphasized that human islet structure and function need to be re-evaluated.

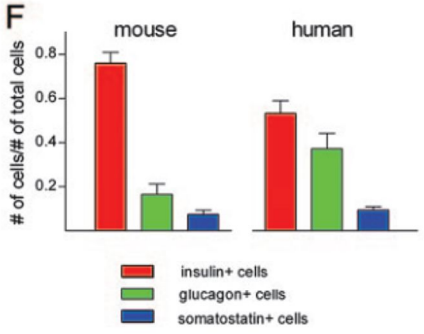
4.2.2 Islet structure in NOD mouse

In NOD (Non-Obese Diabetic) mice, the development of clinical diabetes is preceded by an inflammation of the pancreatic islets. It is generally thought that an initial event in triggering the development of insulitis and cell destruction is the processing of cell antigens by macrophages and dendritic cells residing in the pancreatic islets (Jun HS, 1999). The antigen-presenting cells are then drained to the pancreatic lymph nodes where they will present the antigen to auto reactive T-cells. In the absence of proper peripheral tolerance mechanisms, this leads to activation and insufficiently controlled expansion of these T-cell clones and eventually to their migration back to the pancreatic islets, where they mediate beta-cell destruction (Hoglund P, 1999). It has been suggested in experimental models that infiltration of lymphocytes (insulitis) of the islet begins at 3-5 weeks of age, and becomes evident in diabetic mice at around 8 weeks of age, which leads to an active destruction of beta-cells and consequently to the progressive decrease in beta-cell volume (Alanentalo T, 2010; Jansen A, 1994). Alanentalo T. et al (Alanentalo T, 2010) demonstrated that in NOD mice 20% of islets between 8-12 weeks and 50% of islets between 8-16 weeks are severely damaged. The smallest islets are the first to be destroyed during insulitis, whereas a reduction of large and intermediate size islets appear after 12 weeks of age. The authors also showed that when approximately 85% of beta-cells volume is destroyed, mice developed clinical diabetes.

Figure 2.Interspecies variability in cell composition in the islet structure. Pancreatic islet interspecies differences are shown in *A–D*. Confocal micrographs show representative immune-stained pancreatic sections containing islets of Langerhans from human being *A*, monkey *B*, mouse *C*, and pig *D*. Insulin-immuno-reactive (red), glucagon-immuno-reactive (green), and somatostatin-immuno-reactive (blue) cells were all found randomly distributed in *human* and monkey islets. By contrast, insulin-containing cells were located in the core and glucagon and somatostatin-containing cells in the mantle of *mouse* islets. Pig islets seemed to be formed of smaller units (three, in this case) showing a core–mantle organization (scale bar, 50 m). *E* Quantitative enumeration of the contribution of insulin, glucagon and somatostatin-immuno-reactive cells to the composition of islets in four different regions of the human pancreas. (*F*) Comparison of the cell composition of human islets with that of mouse islets. Human islets had more glucagon-immuno-reactive cells and fewer insulin-immuno-reactive cells.







4.3 Molecular mechanisms underlying insulin secretion

There are three main features of cytosolic glucose metabolism in the beta-cell. First, glucose can equilibrate rapidly across the beta-cell membrane due to the expression of the high capacity, low affinity glucose transporter-2 (Newgard CB, 1995). Second, after glucose has entered the beta-cell, it is phosphorylated to glucose-6-phospate by the high K^+. Glucokinase, which constitutes the flux determining step for glycolysis (Iynedjian 1993; De Vos, 1995; Newgard CB, 1995; Matschinsky 1996) and is considered as the 'glucose sensor' in the pancreatic beta-cell (Matschinsky 1996). Third, once phosphorylated, glucose is metabolized by glycolysis to produce pyruvate, NADH and ATP.

Glucose depolarizes the pancreatic beta-cell membrane potential and initiates firing of action potentials (Dean PM, 1968).

In the absence of stimulatory glucose (generally less than 5 mM), rodent beta-cells are electrically silent, with a resting membrane potential of approximately - 70 mV due to a high resting K^+ conductance in these cells. Reduction of the resting K^+ conductance by stimulatory glucose leads to membrane depolarization and initiation of electrical activity characterized by slow wave depolarization with superimposed bursts of action potentials. As outlined in fig. 4, ATP-sensitive K^+ channels set the beta-cell membrane potential and closure of these leads to depolarization. Membrane depolarization triggers action potential firing and opening of voltage-dependent Ca^{++} channels (VDCCs), leading to Ca^{++} influx which triggers exocytosis. Action potentials are terminated by the opening of voltage-dependent K^+ channels which limit Ca^{++} entry and thus insulin release.

The ability of pancreatic beta-cells to respond to glucose has long been known to depend on extracellular Ca^{++} (Curry C, 1968; Hales CN, 1968). Release of

intracellular Ca^{++} stores is also thought to be involved in regulating insulin secretion. Intracellular Ca^{++} stores may be released by the influx of extracellular Ca^{++}, called Ca^{++} -induced Ca^{++} release (Graves TK, 2003), or by other external signals in order to enhance or prolong insulin secretion rather than trigger it directly. Glucose-stimulated electrical activity in pancreatic beta-cells serves the general purpose of activating VDCCs and delivering the Ca^{++}stimulus to sites of insulin granule exocytosis. While the membrane potential is exquisitely controlled by glucose, the distribution and exocytotic competence of the insulin granules themselves must be equally well controlled to ensure an appropriate physiological response to glucose. Insulin granules, similar to secretory vesicles in numerous other cell types, exist within the cell in various functional pools (Rorsman P, 2003). As shown in figure 3, these include an intracellular reserve pool (90%), a morphologically docked pool (ca 10%), and a readily releasable pool (RRP) that is chemically 'primed' for release (0.3–2.2%)(Bratanova-Tochkova M, 2002; Olofsson O, 2002).

Figure 3. Chain of events in pancreatic beta-cell glucose-sensing

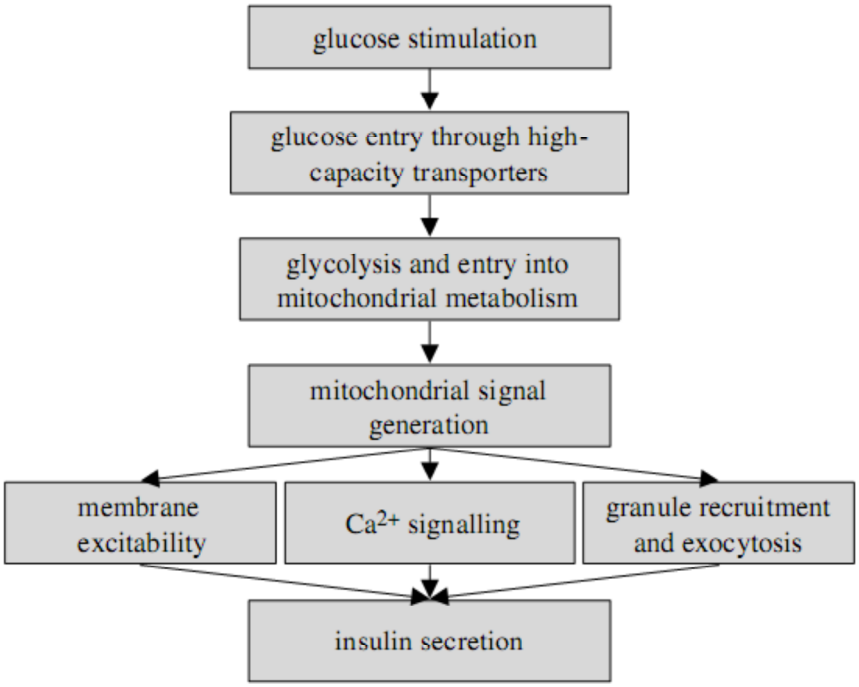
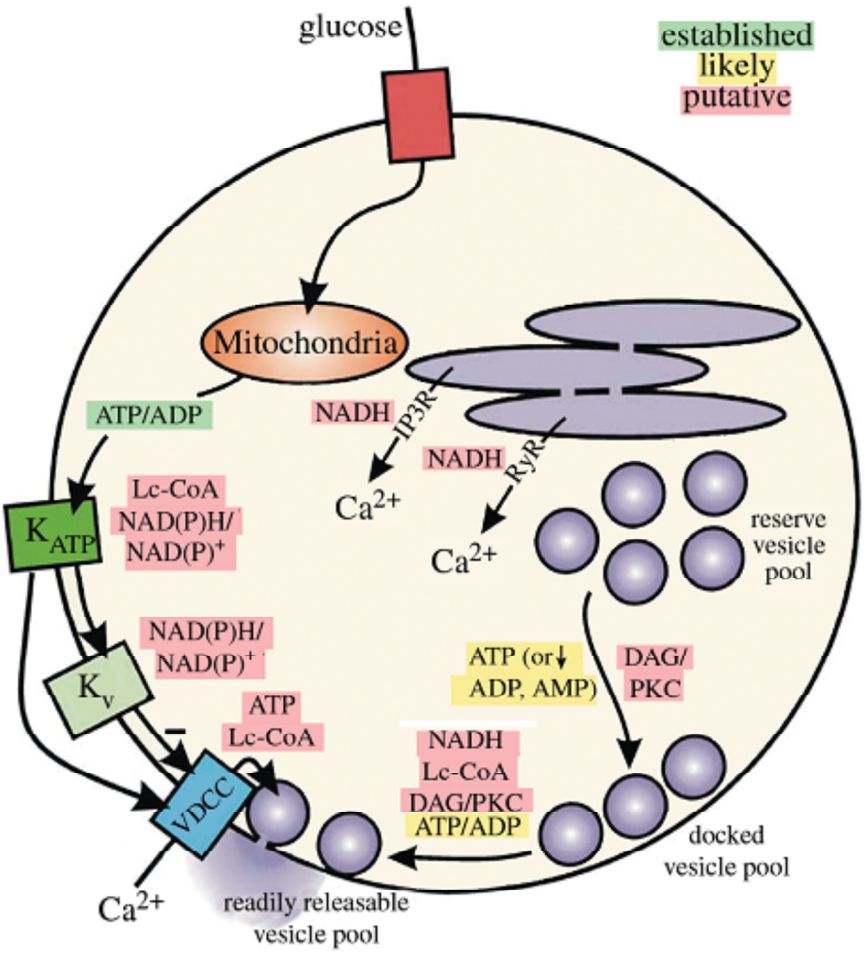


Figure 4. Stimulus–secretion coupling in the pancreatic beta-cell. Glucose entry and mitochondrial metabolism increases the intracellular ATP-to-ADP ratio which leads to closure of K+ channels and membrane depolarization. This activates VDCCs, allowing influx of Ca++ which triggers exocytosis of insulin granules. Kv channels also activate upon depolarization to mediate action potential repolarization, limiting Ca++ entry and insulin secretion. Established and putative metabolic signalling molecules may regulate insulin secretion at a number of sites. These include the activity of ion channels, release of intracellular Ca++ stores, mobilization and priming of secretory vesicles, and exocytosis. Here, we have shown the established (green), likely (yellow) and putative (red) regulatory interactions.



4.4 Role of beta-cells coupling

The primary task of the beta-cell is to secrete insulin in response to glucose, and it is desirable for the output to increase in a graded way with the stimulus. This is achieved mainly by raising Ca^{++} concentration. Empirically, insulin secretion increases linearly with the bulk $[Ca^{++}]_i$ and the rise in $[Ca^{++}]$ is in turn controlled by the extracellular glucose concentration. When [Glucose] is less than the threshold level of about 5 mM, corresponding to fasting conditions in vivo, $[Ca^{++}]$ is low (<100 nM) and there is almost no secretion. After a meal, [Glucose] exceeds the threshold, and, in vitro, $[Ca^{++}]_I$ rises hyperbolically with [Glucose], saturating at about 200 nM when [Glucose] reaches about 10 mM (Heart E, 2006; Henquin JC, 2006).

Many cell types exhibit complex electrophysiological behaviour, characterized by brief bursts of oscillatory activity interspersed with quiescent periods during which the membrane potential changes only slowly. This behaviour is called bursting.

Bursting in pancreatic beta-cells does not usually occur in isolated cells, but only in intact islets or in groups of beta-cells coupled by gap junctions.

In vitro, beta-cell electrical and Ca^{++} activity is generally oscillatory when [Glucose] is in the range of 5–15 mM; above that range, a constant high $[Ca^{++}]$ is seen, corresponding to continuous *spike activity*. These oscillations occur on a variety of time scales, *spiking* (<1 s), *bursting* (10–60 s), and slow oscillations (1–6 min). Similar to the neuronal cell, secretion is driven by Ca^{++}entry through voltage-dependent Ca^{++} channels. This is unlike non-electrically excitable cells, such as the pancreatic acinar cell, in which secretion is governed by release of Ca^{++} from internal stores (endoplasmic reticulum or ER) (Fig. 4).

Several studies have revealed that the electrophysiological properties of beta-cells within the intact islet differ from those of isolated cells in several respects such as channel density and voltage dependence of their activation and inactivation, etc. (Gopel et al. 1999).

To better understand these findings we need to consider the natural environment in which the beta-cell is immersed: it naturally exists in an electrically-coupled ***network***, the islet of Langerhans.

The first electrophysiological studies on the insulin-secreting islet cells with intracellular microelectrodes were first reported by Dean and Matthews in 1968, whereas the first papers by Dean and Matthews (Dean PM, 1972) as well as Petersen and Matthews (Petersen OH, 1972), describing membrane potential measurements from the acinar cells during rest and activation appeared in 1972. Thus electrophysiological studies on the pancreas started late considering that the first intracellular microelectrode studies on the salivary glands were published in 1955 and 1957 by Lundberg (Lundberg A, 1955; Lundberg A, 1957). A few years later the first single-channel current recording studies on the beta-cells from the pancreatic islets were published (Cook DL, 1984; Ashcroft FM , 1984).

The application of the patch-clamp technique to pancreatic islet cells has revolutionized our concepts about the stimulus–secretion coupling of the insulin-producing beta-cells.

Resulting various oscillatory behaviors of beta-cells, together with the insulin secretion and glucose regulation, have been examined (Kang H,2005). In particular, it has been reported that calcium oscillations appear more regular in cell clusters, compared with those in isolated cells (Jonkers FC, 1999).

As a consequence, it was demonstrated that *single cells* do *not burst*, but rather exhibit only irregular, apparently random, *spiking* (Bangham , 1986) (Figs. 5-6).

Figure 5. Spiking activity of an isolated beta-cell. If an isolated beta-cell is exposed to an adequate [glucose], a spontaneous spiking activity will occurred. (Taken from Rorsman and Trube, 1986)

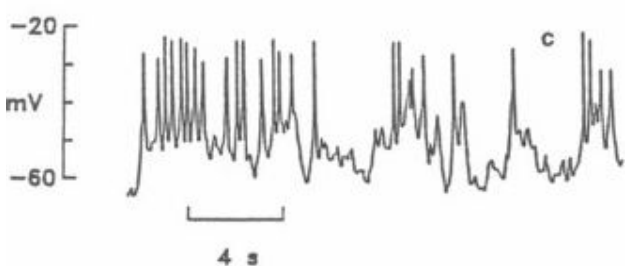
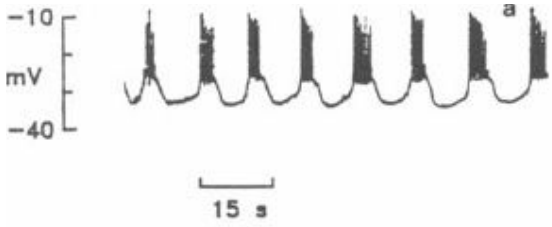


Figure 6. Bursting activity of an intact islet in 11 mM glucose. If a pancreatic beta-cell in an intact islet is exposed to an adequate [glucose], a spontaneous bursting activity will be registered (Taken from Atwater and Rinzel, 1986).



The bulk of experimental evidence indicates that a cell-bursting electrical activity depends on beta-cell coupling (Rorsman P, 1986; Falke LC, 1989)

The intra-islet coordination of beta-cells involves a variety of mechanisms of direct and indirect beta-cell -to beta-cell communication Recent experiments

indicate that, in this complex setting, the beta-cell -to beta-cell communication is mediated by gap junctions consisting of a family of proteins called connexins (Willecke K, 2002). In particular, the most recent evidence has suggested that islet cells express connexin-36 (Calabrese A, 2003).

Evidence for electrical coupling via gap junctions between islet cells has previously been provided by electron microscopy (Orci L, 1975), recordings of membrane potential and currents using sharp intracellular electrodes (Meissner 1976; Eddlestone et al. 1984; Mears et al. 1995), patch-clamp experiments on isolated islet cell pairs (Perez-Armendariz et al. 1991) and dye coupling (Michaels & Sheridan 1981; Charollais et al. 2000).

The gap junctions probably serve to synchronize electrical activity and secretion in beta-cells in different parts of pancreatic islets (Eddlestone et al. 1984; Santos et al. 1991). Then it has been proposed that the beta-cells function as a 'secretory syncytium' (Santos et al. 1991).

The functional significance of intercellular signalling is suggested by the fact that isolated beta-cells exhibit a much poorer secretory capacity than beta-cells in the intact islet (Pipeleers et al. 1982; Luther et al. 2006).

Coupled beta-cells, in fact, secrete insulin more effectively than single beta-cells do, as reflected by the observation that coupled cells can produce bursting action potentials, in contrast with single cells generating spiking action potentials (Sherman, A., J. Rinzel, and J. Keizer. 1988. Emergence of organized bursting in clusters of pancreatic beta-cells by channel sharing. Biophys. J. 54:411–425). Experimental evidence suggests that the integrity of the intercellular communication via gap junctions and the 3D architecture of beta-cells in islets are critical aspects for insulin production (Rocheleau JV, Walker GM, Head WS, McGuinness OP, Piston DW. Microfluidic glucose

stimulation reveals limited coordination of intracellular Ca2+ activity oscillations in pancreatic islets. Proc Natl Acad Sci U S A 2004;101: 12899-12903). In fact, during the beta-cell disruption process, and by losing connections between beta-cells, the development of hyperglycaemia occurred and an impaired insulin level is associated (Dombrowski F et al, 2006). However, it is less clear whether beta-cells are also in electrical contact with neighbouring non-beta-cells. Evidence for heterologous coupling was originally provided by dye-injection experiments (Michaels & Sheridan 1981; Meda et al. 1986), but functional studies (Gopel et al. 1999a; Nadal et al. 1999) suggest that coupling between beta-cells and non-beta-cells is weak.

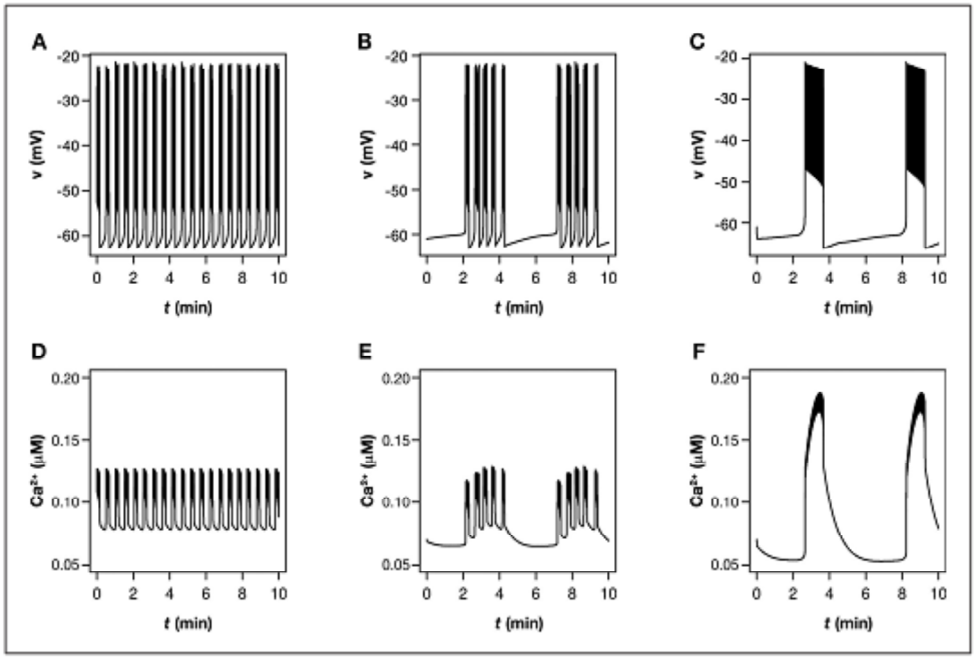
Gap junctional coupling profoundly shapes and modifies that electrical activity. This is seen in the emergence of bursting through mix in beta-cells with disparate properties, in the modification of spike patterns to increase burst period, and in the cooperation of differentially glucose-sensitive cells to produce a well-defined threshold for the glucose response. Thus, the dynamic repertoire of cells is enriched and rendered more robust and, perhaps, better suited to the biological role of regulating insulin secretion.

Taking in consideration all of these recent results, it is raising the possibility that abnormal Cx36-dependent communications between islet cells are involved in the pathophysiology of the early islet changes that contribute to the pancreatic damage.

4.5 From experimental data to mathematical models to explain the pancreatic beta-cell electrical behavior

The central role of the electrophysiology of the beta-cells has led also mathematicians to study this particular part of the triggering pathway. Due to the complex bursting patterns exhibited by the membrane potential (**Fig. 7**), electrical activity in the beta-cells has provided interesting mathematical problems to be analyzed in addition to understanding its role in glucose-stimulated insulin secretion.

Figure 7. Typical patterns of (A–C) bursting electrical activity and (D–F) calcium oscillations observed in beta-cells located in islets. These patterns were simulated using the model by Bertram et al., which can produce (from left to right) either intermediate bursting (fast Ca++ oscillations), compound bursting (mixed Ca++ oscillations), or slow, glycolytic bursting (slow Ca++ oscillations).



4.6 From a Deterministic to a Stochastic model

The release of insulin is pulsatile and it is correlated with rhythmic oscillations in membrane potential. In fact, if a beta-cells is impaled in an isolated islet and it is exposed to a glucose concentration with a microelectrode, a slow oscillation from a low voltage to a rapid spiking occurred (bursting). Calcium flows into the cell during spiking and increase is believed to cause insulin release (Rubin 1982). In order to explain the electrical aspects to this behaviour, Atwater et al. (1980) proposed a mechanism for bursting based on voltage-gated potassium and calcium channels and a calcium-activated potassium (K-Ca) channel. This mechanism was based on extensive experimental data, incorporating the important cellular mechanisms that were thought to underlie bursting and, based on this idea, a quantitative and dynamic mathematical model was later developed by Chay and Keizer (1983). In absence of voltage clump data for the beta-cell, this model was based on modifications of Hodgkin-Huxley kinetics for the squid giant axon. *Chay and Keizer model was a deterministic model, representing both single-cell and cluster behaviour*. Although the mathematical model included only those processes believed to be essential to the bursting process and thus omits many features of the cell, it is able to reproduce many of the basic properties of bursting. Few years later, in 1983, Atwater et al. performed further studies and proposed a qualitative mechanism to account for the difference between the single-cell behaviour and the behaviour of the cell in a cluster (Figs 8-9). They noted that the conductance of a single K-Ca channel is large compared with the total K^+ conductance of the beta-cell and that the K-Ca channels are open very rarely under physiological voltages and intracellular calcium concentration (Findlay et al 1985). They hypnotized that the resting potential would be unstable unless the cells shared channels by electrical coupling

through gap junctions. If few of these channels are open at any given time, then random channel events may perturb the membrane potential of an isolated cell enough to disrupt the burst pattern and produce the observed irregular behaviour of single cells. Instead, in coupled beta-cells cluster, regular bursting may occurred thanks to the contribution of each channel to the current density in the shared membrane. Each K–Ca channel, in fact, has a much smaller effect on the potential of each individual cell, as the channel current is spread over the network of cells. Each cell integrates the effects of a large number of K–Ca channels, each of which has only a small influence.

Figure 8 - Recording from an isolated beta-cell (Rorsman and Trube, 1986)

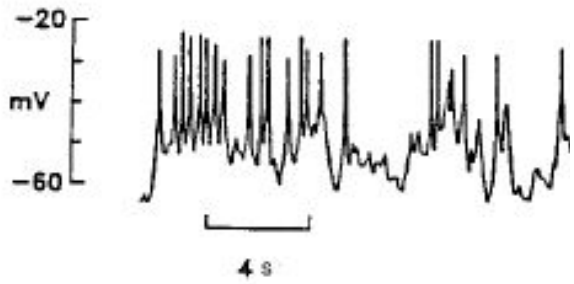
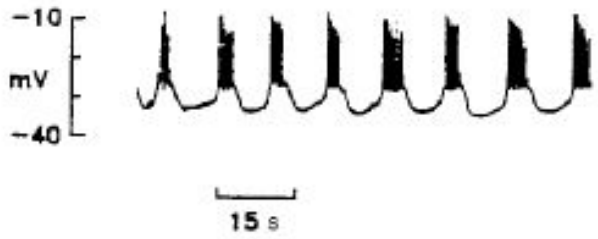


Figure 9. Recording from a beta-cell in an intact cluster (Atwater and Rinzel, 1986)



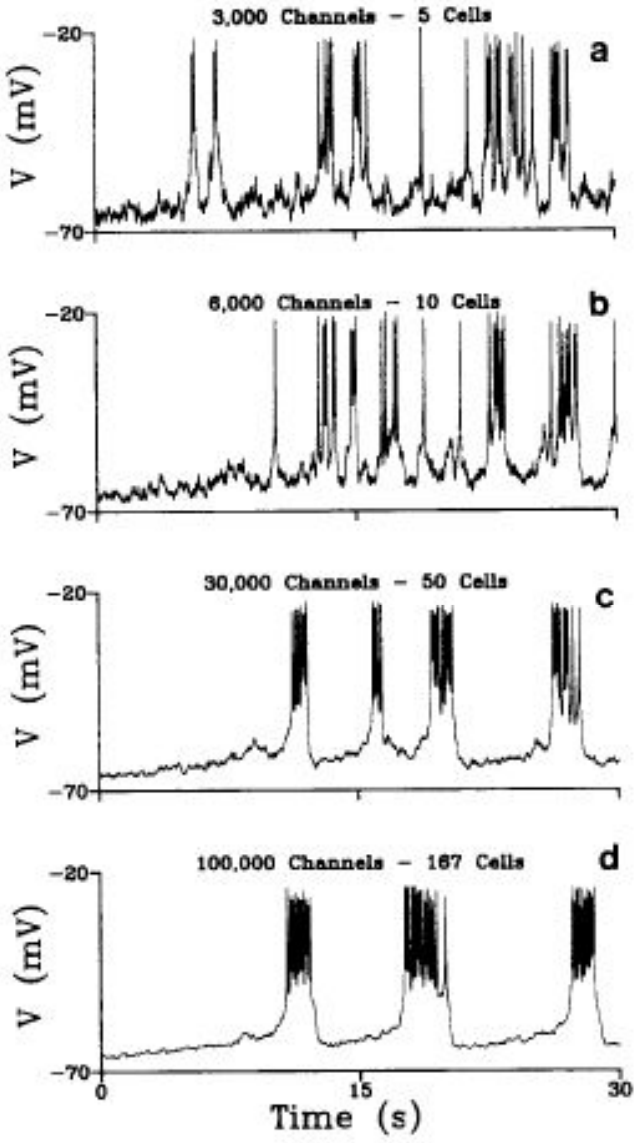
It is worth noting how the need for a specific layer of organization arises from the presence of "biological material", with its constitutive and unavoidable "noise". Consequently it is necessary to introduce explanations that overcome the purely reductionist style that considers the most microscopic layer as the ultimate point of causation from where the reasoning chain can, at least theoretically, follow "all-the-way-down".

In fact thermal agitation (noise) imposes the organization of a coherent structure in which the combined action of averaging and coupling stabilizes the system. Otherwise, any small event initiated on the molecular scale would rapidly expand and dominate the entire system (deterministic approach).

On a more general level, the maintenance of a delicate balance between uncoupling due to noise and coupling due to self-organization was found to be at the origin of any "pathology" of systems. On a more qualitative level, physicians are well aware that diseases arise from a too strict coupling between elements or from decoupling (i.e. atrial fibrillation, T1D, etc). From the above considerations, the need to consider a stochastic approach in modeling appears to be a necessary step.

Two different researcher groups, Sherman et al – 1988 and Chay and Kang – 1988, used theoretical models having a finite number of ***stochastic*** channels to explore if this "channel sharing" hypothesis could reconcile the regular bursting of islet with the irregular spiking of isolated cells. Both considered the entire cluster of cells as a single giant cell with an enlarge channel population and they called it the "supercell" (Fig 6.3).

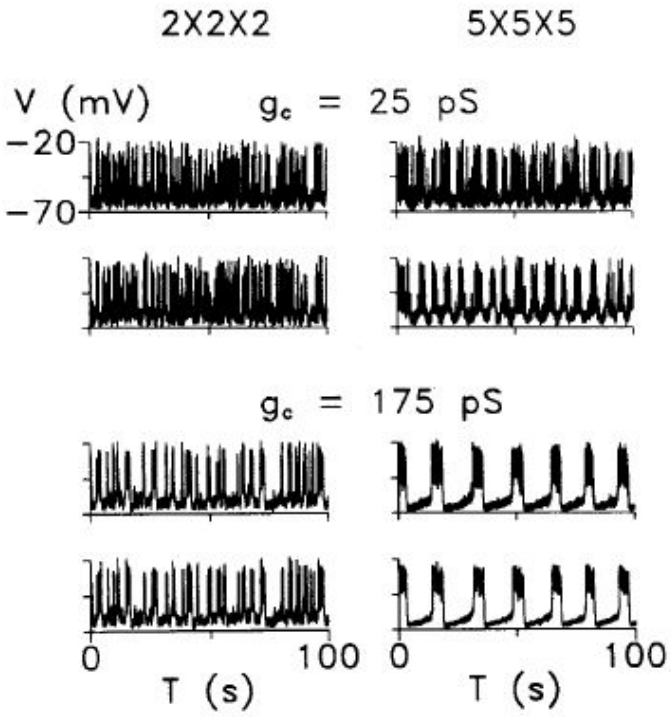
Figure 10. Numerical simulations of the "supercell" model for a cluster of cells ranging in size from 5 to 167 cells. As the size of the cluster increases, more organized bursting appears. (Sherman et al., 1988).



As expected, synchronized bursting appears as the number of cells in the cluster increases and the coupling strength increases. Both strong enough coupling and a large enough cluster size are required to achieve regular bursting due to the beta-cells self-organization.

This is illustrated in Fig. 11, where we show numerical simulations for two different cluster sizes and two different coupling strengths.

Figure 11. Numerical simulations of the multicell model, in which the cells in the cluster are coupled by gap junctions with finite conductance. Results are shown for two cells (upper and lower traces in each pair) from two different cluster sizes (2 × 2 × 2 cells and 5 × 5 × 5 cells) and two different junctional conductances. (Sherman and Rinzel, 1991).



A basic ingredient of this self-organization dynamics is the role played by material noise caused by physical laws. It is precisely thermal agitation that, contrary to the naïve thinking, is necessary for the stability of structures defending the systems from the otherwise catastrophic effects of any microscopic perturbation in a totally deterministic setting.

Mathematical modeling has significantly contributed to the understanding of the mechanisms underlying the various patterns of bursting electrical activity in response to different glucose stimuli. Beta-cell modeling will likely move closer to clinical applications, where it can be expected to continue play an important role, as it has up to now, in the understanding of the complex oscillatory phenomena observed in pancreatic beta-cells and islets. The next chapter deals with the mathematical models developed to better understand the effects of progressive lymphocytic islet infiltration (insulitis) on beta-cells signaling.

REFERENCES

Alanentalo T, Hörnblad A, Mayans S, Karin Nilsson A, Sharpe J, Larefalk A, Ahlgren U, Holmberg D (2010) Quantification and three-dimensional imaging of the insulitis-induced destruction of beta-cells in murine type 1 diabetes. Diabetes. 59:1756-1764

Ashcroft FM, Harrison DE, Ashcroft SJH (1984) Glucose induces closure of single potassium channels in isolated rat pancreatic B-cells. Nature Lond. 312: 446-448

Atwater I, Dawson CM, Scott A, Eddlestone G, Rojas E. The nature of the oscillatory behaviour in electrical activity from pancreatic beta-cell. Horm Metab Res Suppl. 1980;Suppl 10:100–7

Bosco D, Orci L, Meda P (1989) Homologous but not heterologous contact increases the insulin secretion of individual pancreatic B-cells. Exp. Cell Res. 184:72–80

Bratanova-Tochkova T K et al (2002) Triggering and augmentation mechanisms, granule pools, and biphasic insulin secretion. Diabetes 51(Suppl. 1), S83–S90.

Brissova M, Fowler M J, Nicholson W E, Chu A, Hirshberg B, Harlan D M and Powers A C (2005) J. Histochem. Cytochem. 53:1087–1097

Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, Harlan DM and Powers AC (2005) Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J. Histochem. Cytochem.

53:1087–1097

Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A (2006) The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 103:2334–2339.

Calabrese A, Zhang M, Serre-Beinier V, Caton D, Mas C, Satin LS, Meda P (2003) Connexin 36 controls synchronization of Ca2+ oscillations and insulin secretion in MIN6 cells. Diabetes. 52(2):417-24.

Chay TR, Kang HS (1988) Role of single-channel stochastic noise on bursting clusters of pancreatic beta-cells. Biophys J. 54(3):427– 35

Chay TR, Keizer J (1983) Minimal model for membrane oscillations in the pancreatic beta-cell. Biophys J. 42:181–90

Chay TR, Keizer J Minimal model for membrane oscillations in the pancreatic beta-cell. Biophys J. 1983:42(2):181–90

Clark A, Wells C A, Buley I D, Cruickshank J K, Vanhegan R I, Matthews D R, Cooper G J, Holman R Rand Turner R C (1988) Diabetes Res. 9:151–159

Cook DL, Hales CN (1984) Intracellular ATP directly blocks K+ channels in pancreatic B-cells. Nature Lond. 311: 271-273

Cook DL, Ikeuchi M, Fujimoto WY (1984) Lowering of pHi inhibits Ca2+-activated K+ channels in pancreatic B-cells. Nature Land. 311: 269-271

Curry DL, Bennett LL, Grodsky GM (1968) Requirement for calcium ion in insulin secretion by the perfused rat pancreas. Am. J. Physiol. 214:174–178

De Vos A, Heimberg H, Quartier E, Huypens P, Bouwens L, Pipeleers D,

Schuit F (1995) Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. J. Clin. Invest. 96:2489–2495

Dean PM, Matthews EK (1968) Electrical activity in pancreatic islet cells. Nature. 219:389–390

Dean PM, Matthews EK (1972) Pancreatic acinar cells: measurement of membrane potential and miniature depolarization potentials. J. Phys. L. Land. 225: 1-13

E Heart, R F Corkey, J D Wikstrom, O S Shirihai, B E Corkey (2006) Glucosedependent increase in mitochondrial membrane potential, but not cytoplasmic calcium, correlates with insulin secretion in single islet cells, Am. J. Physiol. Endocrinol. Metab. 290 (1) E143

E Heart, R F Corkey, J D Wikstrom, O S Shirihai, B E Corkey, Glucosedependent increase in mitochondrial membrane potential, but not cytoplasmic calcium, correlates with insulin secretion in single islet cells, Am. J. Physiol. Endocrinol. Metab. 290 (1) (2006) E143.

Falke L C,K D Gillis,D M Pressel and S Misler (1989) Perforated patch recording'allows long-term monitoring of metabolite-induced electrical activity and voltage-dependent Ca^{++} currents in pancreatic islet beta-cells. FEBS (Fed.Eur.Biochem.Soc.) Lett. 251:167-172

Gannon M, Ray MK, Van Zee K, Rausa F, Costa RH, Wright CV (2000) Persistent expression of HNF6 in islet endocrine cells causes disrupted islet architecture and loss of beta cell function. Development. 127:2883–2895

Graves TK, Hinkle PM (2003) Ca2+-induced Ca2+ release in the pancreatic b-cell: direct evidence of endoplasmic reticulum Ca2C release. Endocrinology.

144:3565–3574

Guyton AC, Coleman TG, Granger HJ (1972) Circulation: overall regulation. Ann Rev Physiol. 34: 13–46

Hales CN, Milner RD (1968) Cations and the secretion of insulin from rabbit pancreas in vitro. J. Physiol. 199:177–187

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117: 500–544

Hoglund P, Mintern J, Waltzinger C, Heath W, Benoist C, Mathis D. Benoist, and DMathis (1999) Initiation of autoimmune diabetes by developmentally regulated presentation of islet cell antigens in the pancreatic lymph nodes. J. Exp Med.189:331–339

Iynedjian PB (1993) Mammalian glucokinase and its gene. Biochem. J. 293:1–13

J C Henquin, M Nenquin, P Stiernet, B Ahren (2006) In vivo and in vitro glucoseinduced biphasic insulin secretion in the mouse: pattern and role of cytoplasmic Ca^{++} and amplification signals in b-cells. Diabetes. 55: 441-452

J C Henquin, M Nenquin, P Stiernet, B Ahren, In vivo and in vitro glucoseinduced biphasic insulin secretion in the mouse: pattern and role of cytoplasmic Ca^{++} and amplification signals in beta-cells, Diabetes. 55 (2006) 441

Jansen A, Homo-Delarche F, Hooijkaas H, Leenen PJ, Dardenne M and Drexhage HA (1994) Immunohistochemical characterization of monocytes-macrophages and dendritic cells involved in the initiation of the insulitis and

beta-cell destruction in NOD mice. Diabetes. 43:667-675

Jonkers FC, Jonas JC, Gilon P, Henquin JC (1999) Influence of cell number on the characteristics and synchrony of Ca21 oscillations in clusters of mouse pancreatic islet cells. J. Physiol. 520:839–849

Jun HS, Yoon CS, Zbytnuik L, van Rooijen N, Yoon JW (1999) The role of macrophages in T cell-mediated autoimmune diabetes in nonobese diabetic mice. J. Exp. Med. 189:347–358

Kang H, J Jo, Kim HJ, Choi MY, Rhee SW, Koh DS (2005) Glucose metabolism and oscillatory behavior of pancreatic islets. Phys. Rev. E. 72:051905.

Lundberg A (1955) The electrophysiology of the submaxillary gland of the cat. Acta Phisiol Scand. 35: l-25

Lundberg A (1957) Secretory potentials in the sublingual gland of the cat. Acta Phisiol Scand. 40: 21-34

Matschinsky, FM (1996) Banting Lecture 1995. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. Diabetes. 45:223–241

Mears D, Sheppard NF, Atwater I and Rojas E (1995) Magnitude and modulation of pancreatic b-cell gap junction electrical conductance in situ. J. Membr. Biol. 146:163–176

Meissner G, McKinley D. Permeability of sarcoplasmic reticulum membrane. The effect of changed ionic environments on Ca2+ release. J Membr Biol. 1976 Dec 25;30(1):79-98

Newgard CB and McGarry JD (1995) Metabolic coupling factors in pancreatic

b-cell signal transduction. Annu. Rev. Biochem. 64: 689–719

Noble D (1962) A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pacemaker potentials. J. Physiol. 160: 317–352

Olofsson C S, Gopel S O, Barg S, Galvanovskis J, Ma X, Salehi A, Rorsman P and Eliasson L (2002) Fast insulin secretion reflects exocytosis of docked granules in mouse pancreatic b-cells. Pflug. Arch. 444:43–51

Orci L, Unger RH (1975) Functional subdivision of islets of Langerhans and possible role of D cells. Lancet. 2:1243–1244

Perez-Armendariz M, Roy C, Spray DC and Bennett MVL (1991) Biophysical properties of gap junctions between freshly dispersed pairs of mouse pancreatic b-cells. Biophys. J. 59:76–92

Petersen 0H, Matthews EK (1972) The effect of pancreozymin and acetylcholine on the membrane potential of the pancreatic acinar cells. Experimentia Based. 128:1037-1038

Rahier J, Goebbels R M and Henquin J C (1983) Diabetologia. 24:366–371

Rorsman P, Renstrom E (2003) Insulin granule dynamics in pancreatic beta cells. Diabetologia. 46:1029–1045

Rorsman P,and GTrube (1986) Calcium and delayed potassium currents in mouse pancreatic,beta-cells under voltage clamp conditions. J.Physiol.(Lond.).374:531-550

S G‡opel, T Kanno, S Barg, J Galvanovskis and P Rorsman (1999) Voltagegated and resting membrane currents recorded from B_cells in intact

mouse pancreatic islets. Journal of Physiology. 521.3: 717—728

Sakuraba H, Mizukami H, Yagihashi N, Wada R, Hanyu C and Yagihashi S (2002) Diabetologia. 45: 85–96.

Samols E, Stagner JI (1988) Islet somatostatin: microvascular, paracrine, and pulsatile regulation. Metabolism. 1990:39:55–60

Sherman A and Rinzel J (1991) Model for synchronization of pancreatic b-cells by gap junction coupling. Biophys. J. 59:547–559

Sherman A, Rinzel J, Keizer J (1988) Emergence of organized bursting in clusters of pancreatic beta-cells by channel sharing. Biophys J. 54(3):411–25

Stefan Y, Orci L, Malaisse-Lagae F, Perrelet A, Patel Y and Unger RH (1982) Diabetes. 31:694–700

Street C N, Lakey J R, Shapiro A M, Imes S, Rajotte R V, Ryan E A, Lyon J G, Kin T, Avila J, Tsujimura Tand Korbutt G S (2004) Diabetes. 53: 3107–3114

Von Bertalanfyy (1968) Organismic psychology.Clark University Press

CHAPTER 5. A STOCHASTIC MATHEMATICAL MODEL TO STUDY THE AUTOIMMUNE PROGRESSION TOWARDS TYPE 1 DIABETES

5.1 Introduction

Islets of Langerhans are small highly vascularised endocrine structures playing a key role in glucose homeostasis. Having their own complex anatomy including a core consisting mainly of insulin producing beta-cells, which are tightly interconnected by gap junctions and surrounded by a mantle of other endocrine cells secreting different hormones, such as glucagon, somatostatin and pancreatic polypeptide (Orci L, 1973; Eddlestone GT, 1984; Michaels RL, 1981).

While large ($>5\cdot10^6$ μm^3) and intermediate ($1\cdot10^6$- $5\cdot10^6$ μm^3) islets can be located centrally, near the large blood vessels and pancreatic duct, whereas small ($<1\cdot10^6$ μm^3) islets can be observed in the periphery of pancreatic endocrine tissue (Alanentalo T, 2010).

Beta-cells act as glucose-sensors regulating the release of insulin according to blood glucose levels. Increased levels of extracellular glucose cause a cascade of events inside the beta-cells that, by membrane depolarization and consequent influx of Ca^{2+} ions through voltage-dependent membrane channels, lead to pulsatile insulin release (Rocheleau JV, 2004; Bertram R, 2010). Moreover glucose-stimulated islets exhibit a pattern of bursting electrical activity coupled with Ca^{2+} oscillations. This complex behaviour is synchronized throughout the islet so that the latter works as a functional syncytium (Santos RM, 1991). Studies on pancreatic islets revealed that

communication between beta-cells via gap junctions is crucial for proper biosynthesis, storage and release of insulin, whereas uncoupling leads to altered beta-cell function (Dombrowski F, 2006; Meda P, 1990; Nittala A, 2008). Moreover normal blood glucose levels are maintained if the beta-cell critical number is guaranteed (Jansen A, 1994; Jonkers FC, 1999).

Type 1 diabetes is an autoimmune disease characterized by the progressive destruction of beta-cells in the islets of Langerhans, associated with insulin deficiency and hyperglycaemia.

Experimental models of diabetic mice suggested that lymphocytic infiltration of the islets (insulitis) begins at 3-5 weeks of age, and becomes evident at around 8 weeks. This scenario leads to an active destruction of beta-cells and consequently to a progressive decrease in beta-cells population (Alanentalo T, et al, 2010; Jansen A et al, 1994). Alanentalo et al demonstrated that in the pancreas of NOD mice between 8-12 weeks of age a severe damage occurred, namely the disruption of 20% of islets, while between 8-16weeksof age the 50% of islets was destroyed. The smallest islets are the first to be destroyed during insulitis, whereas a reduction of large and intermediate sized islets appears after 12 weeks of age. These authors also showed that when approximately 85% of beta-cell volume is destroyed, mice developed overt diabetes. During beta-cell disruption, the development of hyperglycaemia is associated with impaired insulin levels (Dombrowski F, 2006).

By losing connections between beta-cells, insulin production is dramatically decreased. Experimental evidence suggests that the integrity of intercellular communication via gap junctions and 3D architecture of beta-cells inside the islets are critical aspects for insulin production (Rocheleau JV, 2006). Electrophysiological studies have shown that the number of beta-cells

influences the synchrony and amplitude of Ca^{2+} oscillations in clusters (Rocheleau JV, 2006; Santos RM, 1991).

The autoimmune process associated with beta-cell loss in type 1 diabetes can be considered a typical scenario where the complex oscillatory phenomenon usually observed in single beta-cell and in the whole intact islet, appears compromised.

The role of gap junctions in the connection among beta-cells and the effect of a reduced number of beta-cells in a damaged cluster on the associated electrophysiology can be quantitatively analyzed by using mathematical models.

The first computational model by Chay and Keizer (Fall CP, 2002; Chay TR, 1983), simulating the electrical behaviour of a single beta-cell in an intact islet, was able to describe the electrical bursting activity and the Ca^{2+} oscillatory dynamics. This theoretical model of beta-cell behaviour was implemented on the lines of the classical Hodgkin-Huxley's work (Hodgkin A, 1952). Inspired by its purely theoretical predictions, these important oscillations in Ca^{2+} were eventually recorded for the first time in vitro eight years later.

In the same period, (Atwater I, 1983) suggested that the stochastic opening and closing of ion channels represented the key mechanism to take into account for the difference between the single beta-cell and the bursting behavior of coupled beta-cells in a cluster. In the latter case, in fact, through an effective coupling due to a channel sharing dynamics, the noise intensity is reduced by gap junction electrical connections.

Further studies, performed by Sherman et al (Sherman A, 1988; Sherman A, 1991), clarified the irregular spiking behavior of isolated beta-cells as well as the collective ordered bursting dynamics of coupled cells.

In the light of these results, the aim of the present study was to evaluate the effect of progressive lymphocytic islet infiltration (insulitis) on beta-cells signalling. The destruction of cluster architecture consisting of pancreatic beta-cells population, as consequence of insulitis, was modelled in terms of apoptotic events. For a quantitative comprehension of such a complex phenomenon, mainly involving cellular synchronization, we have investigated the activity of beta-cells in a 3D cells' lattice with defined neighbourhood relations and altered glucose concentrations. We numerically simulated a small cluster of mouse beta-cells via an extended stochastic electrophysiological model (Sherman A, 1988; Sherman A, 1991). This formulation was originally developed in order to deal with an intact and regular cluster configuration only. Here we have induced, however, a gradually break of the cluster's spatial symmetry by decreasing the number of communicating beta-cells in the cluster, creating a pathological environment inspired by experimental histological sections of diseased diabetic mice. We point out that a minimal physiologically inspired stochastic model can be an efficient tool to obtain original and quantitative insights into the complex dynamics of beta-cell destruction. This work was specifically focused on the clinical implications of the simulation results together with possible future tests in vitro performed through recently developed bioengineering devices built to experimentally study the spatio-temporal electrophysiological dynamics of mice islets.

5.2 Material and Methods

The stochastic model for beta-cell electrical excitability developed by Sherman et al. (Sherman, 1988; Sherman A, 1991), was adopted to simulate (XPP-Aut and C++) a virtual portion of a pancreatic islet (cluster) with progressive apoptotic events (death of beta-cells). The performed system consisted of identical beta-cells coupled through gap junctions, with morphological and physiological characteristics following experimental data of murine beta-cells (Atwater I, 1983; Rorsman, 1986).

The pancreatic autoimmune progressive damage occurring in type 1 diabetes was modeled at different stages of lymphocytic infiltration in the experimental model of the non-obese diabetic (NOD) mouse model. One aim was to investigate the main electrophysiological changes in beta-cell membrane potential and free intracellular Ca^{2+} ions concentration [Ca^{2+}]. Furthermore, many experimental studies have shown that both the slow and fast Ca^{2+} oscillations correspond to synchronized pulsatile insulin secretion of beta-cells (Bergsten P, 1995; Gilon P, 1993), also proposing that glucose stimulation of a prompt and sustained insulin release is mediated by increase of the cytoplasmic Ca^{2+} concentration (Wollheim CB, 1978).

Based on such a correspondence, our analysis aimed to investigate this dynamics (Wollheim CB, 1978).

In type 1 diabetes the death of beta-cells is induced by an autoimmune response underlying the disease process. In our model, we assumed that beta-cells apoptosis corresponded to its disappearance from the cluster. This was achieved by setting to zero both the electrical dynamics of the dead beta-cell and the gap junctions conductance g_c (representing the connection between beta-cells). This process simulates, in a first approximation, the progressive

lymphocytic infiltration which, by substituting the dead beta-cells, destroys their gap junction coupling.

Previous electrophysiological studies have shown that a single ion channel can have two conductance states only (i.e. open and closed). Fluctuations between closed and open states appear to occur randomly. Consequently, in pursuing our goal of understanding bursting in the single beta-cell and the synchronized activity of the beta-cells in a cluster, we considered such a stochastic (probabilistic) mechanism (Atwater I, 1983;Rorsman P, 1986).

In this model a stochastic dynamics of calcium-activated potassium (K-Ca) channels occurs. It is well known that a 50 beta-cells cluster exhibits organized bursting (Sherman A, 1988; Sherman A, 1991). Therefore, we performed numerical simulations, starting for simplicity from an intact cubic shaped cluster of identical 4x4x4 cells, each of them having 600 of these stochastic channels.

For a system of N interacting beta-cells, the differential equations for the j-th cell are:

$$C_m \frac{dV_j}{dt} = -I_K(V_j, n_j) - I_{Ca}(V_j) - \overline{g}_{K-Ca} x_j (V_j - V_K) - g_c \sum_{k \in \Omega_j} (V_j - V_k)$$

$$\frac{dn_j}{dt} = \lambda \left[\frac{n_\infty(V_j) - n_j}{\tau_n(V_j)} \right]$$

$$\frac{dCa_j}{dt} = f\left[\alpha I_{Ca}(V_j) - k_c Ca_j\right]$$

$$\langle x_j \rangle = \frac{Ca_j}{K_d + Ca_j}$$

(Eqs. 1-4)

where

V_j is the membrane potential,

n_j is the open fraction of voltage gated K^+ channels,

Ca_j is the free intracellular Ca^{2+}.

Ω_j in Eq.(1) indicates the interactions of the j-th cell through gap-junctions (Von Neumann neighbourhood).

I_K and I_{Ca} represent the currents through the voltage-gated K^+ and Ca^{2+} channels respectively.

x_j is a random variable indicating the fraction of open K-Ca stochastic channels of the j-th cell, whose transition probabilities are described extensively in ref (Pérez-Armendariz M, 1991; Smolen P, 1993) while its mean value is given by equation 4.

Explicit expressions and parameters of Eqs (1-4) are reported in Appendix.

The parameter k_c plays a central role in the system dynamics. We assumed specifically the simplest linear monotonic relation which links k_c with the external glucose concentration $[Glu]$ measured in mmol:

$$k_c = 7 \cdot 10^{-3}\, mmol^{-1}[Glu] - 0.019$$

This relationship shows that our modelling considers bursting dynamics as being directly dependent on glucose concentration. Other more accurate choices could be proposed, but for the aims of our analysis this form allows a model calibration in accordance with experimental data on clusters stimulated by different glucose concentrations.

We then considered the specific values of k_c inducing different behaviours in the bursting dynamics in agreement with experiments on clusters discussed later in more detail.

In this study, the value of gap junction conductance $g_c = 215pS$ was chosen corresponding to the mean value taken from experimental data, as discussed elsewhere (Pérez-Armendariz M, 1991; Smolen P, 1993).

The increase of the percentage of beta-cell damage in the cluster was considered at corresponding levels of glycaemia as follows:

- 0% and 31% of dead cells at 7mmol of glucose concentration (k_c=0.03), normal value for a physiological state.
- 69% of dead cells at 11.6 mmol (k_c=0.062), typical value for a mild hyperglycaemic state [Dombrowski F, 2006].
- 84%, 94% and 98% of dead cells at 17 mmol (k_c=0.1), typical value of a severe hyperglycaemic state.

The cell loss is not random. It starts in fact at one corner of the cubic cluster configuration and spreads symmetrically towards the opposite corner of the system.

Regarding performed simulations, X-PPAUT software was adopted to integrate the model equations (Ermentrout B, 2002).

5.3 Results

In Figure 1, the experimental evidence in mice islet of the autoimmune beta-cells' destruction leading to type 1 diabetes at different stages of insulitis is presented. In panel (a) we have shown specifically a grade I insulitis with large portion of beta-cells surrounded a modest infiltration of lymphocytes. In panel (b) we presented a grade II infiltratum with a very large portion of the islet surrounded by the lymphocytes. Only a small region of the insula is still recognizable by beta-cells. Finally in panel (c) a grade III case is presented; here the islet has practically almost disappeared.

This pathological scenario has been implemented in the mathematical model of beta-cells electro-physiology previously introduced. The numerical simulations performed for an intact cubic cluster of 4x4x4 cells showed that for $k_c > 0.02$ (corresponding to a threshold of $[Glu]$= 5.5 mmol) a regular bursting occurred.

The bursting becomes continuous for a value of $k_c \geq 0.97$ (corresponding to $[Glu]$= 16.6 mmol). These values are in agreement with experimental data in which at $[Glu]$= 5.5 mmol bursting is not observed, while at 16.6 mmol a continuous bursting starts (Meissner H P et al, 1974).

The numerical results allowed us to understand the relationship between the progressive beta-cell disruption in the pancreatic cluster and the system's electrophysiological response to the damage.

In figure 2 we showed:

a) the cluster (initially consisting of 64 beta-cells) characterized by a progressive degree of beta-cells damage (left panel);

b) the respective membrane potential versus time (central panel);

c) the associated intracellular Ca^{2+} concentration (right panel), for a generic beta-cell in the cluster.

Starting from the intact cluster (0% damage), we considered a normoglycaemic state (k_c=0.03 equivalent to a glucose concentration of 7 mmol) [30].

The model exhibited bursts of action potential and Ca^{2+} oscillations with a qualitative trend similar to those experimentally observed [Rorsman P, 1986], namely, the bursts occurred during the peak of the Ca^{2+} oscillations (figure 2a). This was due to the gap junctions coupling effect underlying the beta-cell activity that determined electrical bursting and oscillatory metabolic rhythms. In particular, the intracellular calcium concentration oscillates approximately around the average value of $0.5\,\mu$M, with a frequency of 3-4/min. These values are comparable with experimental results on intact clusters [Jonkers FC, 1999].

As reported by Atwater et al. [31], in a murine intact islet immersed in 11mmol of glucose, 6/min oscillations can be observed; whereas in the case of a 70 micron diameter cluster of murine cultured beta-cells exposed to 10 mmol of glucose, 3/min oscillations are detectable (Rorsman P, 1986).

Other simulations performed in this study, aimed at calibrating the model, showed that in an intact beta-cell cluster a progressive increase of $[Glu]$ led to

Ca^{2+} oscillations, characterized by an increased duration without changes in amplitude, as reported in ref. (Barbosa RM, 1998).

We then introduced the damage simulation protocol (31% dead cells) (fig. 2b) for a normoglycaemic state (7 mmol). Even if a portion of the cluster has disappeared, the system is biologically robust enough to maintain regular Ca^{2+} oscillations (insulin is expected to be released effectively) as reported elsewhere (Rocheleau JV, 2006; Barbosa RM, 1998).

In particular, the frequency of oscillations slightly increases although the amplitude of action potential and Ca^{2+} remains unaffected.

When the damage progression further increased (69% of damage) (fig 2c), we imposed a glucose concentration typical of mild hyperglycaemia (k_c=0.062 equivalent to 11.6 mmol of glucose) (Dombrowski F, 2006). In this case, Ca^{2+} oscillations occurred, however they were characterized by an increased frequency (~ 8/min) and a smaller amplitude (a reduction of 33%), with an associate effective insulin pulsatility (Bergsten P, 1994). The Ca^{2+} mean value around 0.52 μmol, is in agreement with experimental data in the case of an intact cluster condition.

The final steps of the simulation protocol assumed an advanced regime of cellular disruption (84%, 94% and 98% of dead cells) in the beta-cell cluster (fig. 2d, 2e and 2f). They were characterized by severe hyperglycemia (k_c = 0.1 equivalent to 17 mmol of glucose) (Meissner HP, 1974; Visser J, 2009).

Simulations showed that the complex periodic bursting pattern disappeared and that Ca^{2+} oscillations were absent. The mean level of intracellular Ca^{2+} was lower than in the previous cases (0.45 μmol, 0.4 μmol and 0.3 μmol, respectively). In these conditions insulin pulsatility is not expected to occur.

Following experimental data, the results obtained in the last three simulations set a framework characterized by a greatly reduced insulin release.

In Figure 3 we focused instead on the effects of different values of glycaemia on the electrophysiology of beta-cells for the specific case of a cluster damaged up to 84 %.

The simulations showed that in case of normoglycaemia (7mmol) (fig 3a), an extremely damaged pancreatic tissue could leads to Ca^{2+} oscillations via cluster synchronization, although characterized by small amplitude and large frequency. Moreover, unlike the situation described in fig. 2d, for the same cluster damage (84%) in hyperglycaemic state, a normoglycaemic state is robust enough to support quasi-regular bursting phenomenology. On the other hand, a mild hyperglycaemic state (11.6 mmol of glucose) (fig. 3b) seems to permit oscillations, while marked hyperglycaemia (17 mmol of glucose) (fig. 3c) shows a dramatic drop of Ca^{2+} pulsatility and no organized bursting behavior with consequent compromised beta-cell secretive pattern. This result theoretically confirms the importance of maintaining normal value of blood glucose when hyperglicaemia is diagnosed in order to protect residual beta-cells. Descriptive characteristics of our analysis are reported in Table 1.

5.4 Conclusion

In this study, we implemented a mathematical model to describe the relationship between lymphocytic infiltration, beta-cell death and the insurgence of hyperglycaemia. We aimed to prove the robustness of pancreatic islets even in conditions of severe damage giving a quantitative view of the alterations in the electrochemical rhythms.

Due to the key role of the gating probabilistic phenomenon of experimentally observed ion-channel, the stochastic model of beta-cell electrical excitability developed by Sherman et al. (Sherman A, 1988, Sherman A, 1991) was here adopted to simulate the electrochemical behaviour of a virtual cluster of beta-cells under the typical autoimmune process occurring in type 1 diabetes.

We considered an heterogeneity of beta-cell dysfunction by requiring that a beta-cell can communicate with the surrounding alive cells but not with the cluster of dead ones. A simulation of progressive non-symmetric damage, which mimics an autoimmune progression towards type 1 diabetes, was performed. We monitored the loss of electrophysiological integrity of the cluster during reduction of beta-cell number for altered glucose levels. We found that the cluster is robust enough because its bursting activity was present even when 69% of contiguous cells were dead.

Furthermore, our simulations have shown that in a cluster with 84% of dead beta-cells, a normoglycaemic state guarantees Ca^{2+} oscillation patterns and action potential bursting activity near the physiological condition, due to cellular network synchronization. For the same degree of damage, but in an hyperglycaemic state, the Ca^{2+} pulsatility was absent and no organized bursting behavior occurred.

We should emphasize that insulitis is a complex phenomenon in which the immune attack plays a key role than the simple destruction of contiguous beta-cells with consequent interruption of cell communication. Focusing on the purely electrophysiological problem, we are aware in fact that the entire ionic dynamics of the pancreatic cells should be altered in the pathological state leading to strong heterogeneity for each cellular component of the cluster. Our analysis in this sense is a starting point for future investigations in which the electrical parameters of each cell could be different. To this aim, more complicated electrophysiological models should be adopted in order to include the missing information of the existing models; moreover the inclusion of insulin secretion feedback on larger sized clusters is mandatory for more detailed representation of the complex phenomenon here addressed. Previous studies performed in animal models have shown that gap junction channels coupling beta-cells are made of the Connexin36 protein and that the loss of this protein desynchronizes beta-cells, leading to secretory defects in terms of recruitment of cells into insulin biosynthesis and release (Serre-Beinier V, 2000). It has been recently demonstrated that Connexin36 is also a protein of human pancreatic islets, which mediates the coupling of the insulin–producing beta-cells. Previous studies in rodents have demonstrated that the loss of the Connexin36 results in an alteration of insulin secretion resembling the pre-diabetic state (Ravier MA, 2005; Speier S, 2007). Serre-Beiner et al (Serre-Beinier V, 2000) hypothesized that if the effects observed in rodents are to be extended to humans, diabetes subjects could express lower levels of Connexin36 protein and/or decreased beta-cell coupling than normoglycaemic individuals.

In this study we have implemented a revised version of the Chay-Keizer model (Sherman A, 1988; Sherman A, 1991) in order to study the behavior of a

murine pancreatic islet in physiological and pathological conditions and to investigate further the beta-cells' synchronized pacemaker activity. The significant information emerging from this modelling study is that good glucose control allows a better functionality of beta-cell communication and therefore a more efficient insulin secretion. This remains true even in a cluster containing residual amount of beta-cells following the autoimmune process occurring in type 1 diabetes. When blood glucose concentrations tend to rise, the communication within beta-cells becomes altered and insulin release is gradually impaired. With persistent increase of glucose levels, a severe impairment of insulin secretion takes place and diagnosis of type 1 diabetes is made. The animal models, in which single beta-cell cannot longer establish intercellular communication via Connexin36 channels, show alterations in secretion and insulin gene expression that can be corrected after restoration of beta-cell contacts (Speier S, 2007; Serre-Beinier V, 2009). These findings applied to humans underline that the implementation of insulin therapy and a restoration of normoglycaemia as soon as possible after diagnosis is of great relevance potentially allowing to re-establish an endogenous insulin secretion, leaving the patient moving into the well-known phase of clinical remission, taking off exogenous insulin for some time.

Our results suggest that at the diagnosis of type 1 diabetes a very strict normalization of blood glucose levels should be a key action to be achieved in order to obtain a better preservation of beta-cells. In this respect, the use of a continuous insulin infusion pump and a strict glucose monitoring should be implemented as soon as a patient is diagnosed with type 1 diabetes to control hyperglycaemia. In the absence of glucotoxicity any adjuvant therapy to cure this disease including immunomodulation or the regeneration of beta-cells (Porat S, 2011; Cernea S, 2010) can be more beneficial and, hopefully,

contribute to prevent the destruction of any residual beta-cell still functioning (Michon L, 2005).

Moreover the predictions obtained through the proposed mathematical model can be directly tested through the experimental setup described elsewhere (Rocheleau JV, 2004). By means of a specifically built biomedical device controlling an islet microfluidics, the insulin secretion produced by extracellular glucose stimuli was measured by NAD(P)H autofluorescence. Future studies are aimed to specifically insert in such a device islets of animals at different infiltratum, quantified ex vivo or in vivo following the magnetofluorescent nanoparticles techniques (Denis MC, 2004). Moreover, similar fluorescence techniques, based on voltage or Calcium sensitive dyes, could be adopted to monitor the underlying electrophysiology of the islets (Benninger RKP, 2008).

Eventually, because of the importance of the right timing in shutting down insulin secretion during type 1 diabetes, beta-cells regeneration could be implemented into the in silico model in order to further highlight how this is affected by hyperglicemia and inflammation.

REFERENCES

Alanentalo T, Hörnblad A, Mayans S et al (2010) Quantification and three-dimensional imaging of the insulitis-induced destruction of beta cells in murine type 1 diabetes. Diabetes 59:1756-1764

Atwater I, Rinzel J (1986) The beta cell bursting pattern and intracellular calcium, in Ionic

Atwater I, Rosario L, Rojas E (1983) Properties of calcium-activated potassium channels in the pancreatic beta cell Cell. Calcium 4:451-461.

Barbosa RM, Silva AM, Tomé AR, Stamford JA, Santos RM, Rosario LM (1998) Control of pulsatile 5-HT/insulin secretion from single mouse pancreatic islets by intracellular calcium dynamics, J. Physiol 510.1:135-143

Benninger RKP, Zhang M, Head WS, Satin LS, Piston DW (2008) Gap Junction Coupling and Calcium Waves in the Pancreatic Islet. Biophys J 95: 5048–5061.

Bergsten P (1995) Slow and fast oscillations of cytoplasmic Ca2+ in pancreatic islets correspond to pulsatile insulin release. Am J Physiol 268:282-287

Bergsten P (1998) Glucose-induced pulsatile insulin release from single islets at stable and oscillatory cytoplasmic Ca2+. Am J Physiol 274:796–800

Bergsten P, Grapengiesser E, Gylfe E, Tengholm A, Hellman B (1994) Synchronous oscillations of cytoplasmic Ca2+ and insulin release in glucose-stimulated pancreatic islets. J Biol Chem 269:8749-8753

Bertram R, Sherman A, Satin LS (2010) Electrical bursting, calcium oscillations, and synchronization of pancreatic islets. Adv Exp Med Biol

654:261-279

Cernea S, Dobreanu M, Raz I (2010) Prevention of type 1 diabetes: today and tomorrow. Diabetes Metab Res Rev 26(8):602-605

Channels in Cells and Model Systems (R. Latorre, Ed.), Plenum Press New York and London, 353-362

Chay TR, Keizer J (1983) Minimal model for membrane oscillations in the pancreatic beta cell. Biophys J 42:181-190

Cicirata F, Parenti R, Spinella F et al (2000) Genomic organization and chromosomal localization of the mouse Connexin36 (mCx36) gene. Gene Jun 27;251(2):123-130

Denis MC, Mahmood U, Benoist C, Mathis D and Weissleder R (2004) Imaging inflammation of the pancreatic islets in type 1 diabetes. PNAS 101:12634-12639

Dombrowski F, Mathieu C, Evert M (2006) Hepatocellular Neoplasms Induced by Low-Number Pancreatic Islet Transplants in Autoimmune Diabetic BB/Pfd Rats. Cancer Res 66:1833-1843

Eddlestone GT, Goncalves A, Bangham JA, Rojas E (1984) Electrical coupling between cells in islets of Langerhans from mouse. J Membr. Biol 77:1–14

Ermentrout B (2002) Simulating, Analyzing, and Animating Dynamical Systems: A Guide to X-ppaut for Researchers and Students Society for Industrial Mathematics; 1st edition.

Fall CP, Marland ES, Wagner JM, Tyson JJ (Editors) (2002). Computational

Cell Biology. Springer, Berlin

Gilon P, Shepherd RM, Henquin JC (1993) Oscillations of secretion driven by oscillations of cytoplasmic Ca2+ as evidences in single pancreatic islets. J Biol Chem 268:22265-22268

Hellman B (2009) Pulsatility of insulin release – a clinically important phenomenon. Upsala J Medical Sciences 114:193-205

Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544

Jansen A, Homo-Delarche F, Hooijkaas H, Leenen PJ, Dardenne M and Drexhage HA (1994) Immunohistochemical characterization of monocytes-macrophages and dendritic cells involved in the initiation of the insulitis and beta cell destruction in NOD mice. Diabetes 43:667-675

Jonkers FC, Jonas JC, Gilon P, Henquin JC (1999) Influence of cell number on the characteristics and synchrony of Ca2+ oscillations in clusters of mouse pancreatic islet cells. J Physiol 520.3:839-849

Meda P, Bosco D, Chanson M et al (1990) Rapid and reversible secretion changes during uncoupling of rat insulin producing cells. J Clin Invest 86:759 –768

Meissner H P, Schmelz H (1974) Membrane potential of beta cells in pancreatic islets.

Michaels RL, Sheridan JD (1981) Islets of Langerhans: dye coupling among immunocytochemically distinct cell types. Science 214:801– 803

Michon L, Nlend Nlend R, Bavamian S et al (2005) Involvement of gap junctional communication in secretion. Biochim Biophys Acta 1719:82-101

Nittala A, Wang X (2008) The hyperbolic effect of density and strength of inter beta cell coupling on islet bursting: a theoretical investigation. Theor Biol Med Model 5:1-13

Orci L, Unger RH, Renold AE (1973) Structural coupling between pancreatic islet cells. Experientia 29:1015–1018

Pérez-Armendariz M, Roy C, Spray DC, Bennett MV (1991) Biophysical properties of gap junctions between freshly dispersed pairs of mouse pancreatic beta cells. Biophys J 59:76–92

Pflgers Archiv European Journal of Physiology 351:195-206

Porat S, Weinberg-Corem N, Tornovsky-Babaey S et al (2011) Control of Pancreatic β Cell Regeneration by Glucose Metabolism. Cell Metab Apr 6;13(4):440-9.

Ravier MA, Güldenagel M, Charollais A et al (2005) Loss of connexin36 channels alters beta cell coupling, islet synchronization of glucose-induced Ca2+ and insulin oscillations, and basal insulin release. Diabetes 54:1798-1807

Rocheleau JV, Walker GM, Head WS, McGuinness OP, and Piston DW (2004) Microfluidic glucose stimulation reveals limited coordination of intracellular Ca2+ activity oscillations in pancreatic islets. PNAS 101: 12899–12903.

Rocheleau JV, Walker GM, Head WS, McGuinness OP, Piston DW (2004) Microfluidic glucose stimulation reveals limited coordination of intracellular Ca2+ activity oscillations in pancreatic islets. Proc Natl Acad Sci U S A 101:

12899-12903

Rorsman P, Trube G (1986) Calcium and delayed potassium currents in mouse pancreatic beta- cells under voltage clamp conditions. J Physiol 374: 531-550

Santos RM, Rosario LM, Nadal A, Garcia-Sancho J, Soria B, Valdeolmillos M (1991) Widespread synchronous [Ca2+]i oscillations due to bursting electrical activity in single pancreatic islets. Pflugers Arch 418:417-422

Serre-Beinier V, Bosco D, Zulianello L et al (2009) Cx36 makes channels coupling human pancreatic beta cells, and correlates with insulin expression. Hum Mol Genet 18:428-439

Serre-Beinier V, Le Gurun S, Belluardo N et al (2000) Cx36 preferentially connects beta-cells within pancreatic islets. Diabetes 49:727-734

Sherman A, Rinzel J , Keizer J (1988) Emergence of organized bursting in clusters of pancreatic beta cells by channel sharing. Biophys J 54:411-425.

Sherman A, Rinzel J (1991) Model for synchronization of pancreatic beta cells by gap junction coupling. Biophys J 59:547-559

Smolen P, Rinzel J and Sherman A (1993) Why pancreatic islets burst but single beta-cells do not. The heterogeneity hypothesis. Biophys J 64:1668-1680

Speier S, Gjinovci A, Charollais A et al (2007) Cx36-Mediated Coupling Reduces beta cell Heterogeneity, Confines the Stimulating Glucose Concentration Range, and Affects Insulin Release Kinetics Diabetes 56:1078-1086

Visser J, Hillebrands JL, Walther Boer M, Bos NA, Rozing J (2009) Prevention

of diabetes by a hydrolysed casein-based diet in diabetes-prone Bio Breeding rats does not involve restoration of the defective natural regulatory T cell function. Diabetologia 52:1445-1447.

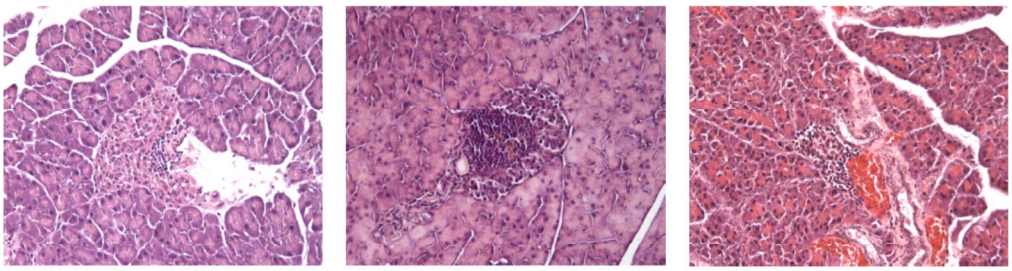
Wollheim CB, Kikuchi M, Renold AE, Sharp GWG (1978) The Roles of Intracellular and Extracellular Ca++ in Glucose-Stimulated Biphasic Insulin Release by Rat Islets. J Clin Invest 62:451-458

Table 1. Numerical details of the progressive damage in the modelled pancreatic cluster. The percentage of cluster damage, glucose concentration associated with an approximate mean value of [Ca^{2+}] during oscillations is here illustrated. Sustained oscillations in [Ca^{2+}] and a discrete number of isolated bursting complexes, even in highly damaged systems, are present with a mean values of [Ca^{2+}] over 0.5 µmol. When the damage is > 84 % and hyperglycaemia occurred, such oscillations and bursting patterns disappear and [Ca^{2+}] mean values decrease. In this scenario insulin secretion is severely compromised.

% Cluster damage	0%	31%	69%	84%	94%	98%
Glucose concentration (mmol)	7	7	11.6	17	17	17
[Ca^{++}] mean value (µmol)	0.55	0.55	0.52	0.46	0.39	0.29
Number of [Ca^{++}] approximated oscillations during 80 s	4	5	9	-	-	-
Number of action potential approximated bursting events during 80 s	5	5	8	-	-	-

Figure 1. (Color online)

Experimental evidence of autoimmune progression towards type-I Diabetes on mice islets at different stages of insulitis. In panel (a) we have a grade I insulitis with large portion of beta-cells (rosed cells on the centre of the picture) attached by a modest infiltration of lymphocytes (whose nuclei have very compact structure in comparison with the pancreatic tissue). In panel (b) we have instead a grade II situation with a very large portion of the islet invaded by the lymphocytes. Only a small region of the insula is still formed by beta-cells with a falcate shape. Finally in panel (c) we have a grade III case; here the islet has practically almost disappeared, invaded totally by lymphocytes.



(a) (b) (c)

Figure 2.

Effect of autoimmune progression on action potential (central panel) and intracellular Ca^{2+} (right panel) assuming a progressive percentage of dead beta-cells (0%, 31%, 69%, 84%, 94% and 98%) in the modelled pancreatic cluster, on the lines of the experimental situation presented in Figure 1. Signals from the 64th cell are shown, although the other cells manifest a similar behavior. Images of simulated progressive damage in a pancreatic cluster are shown on the left panel. In central and right panels the action potential and its associated intracellular calcium dynamics are represented, respectively. Action potential is measured in mV versus seconds while Ca^{2+} concentration is measured in µM versus seconds. At different degree of tissue damage a gradual increase of glucose concentration is associated: a) and b) represent 0% and 31% of dead beta-cells, respectively, in normoglycaemia (7 mmol, k_c=0.03). This scenario demonstrates an intrinsic biological robustness of the pancreatic cluster. c) 69% of dead beta-cells, at value of mild hyperglycaemia (11.6 mmol, k_c=0.062). The action potential and calcium dynamics related pictures still show oscillations typical of a severely damaged system. In fig. 1d), 1e) and 1f) an increasing damage of 84%, 94% and 98% respectively is simulated for a severe hyperglycaemia (17 mmol of glucose equivalent to k_c=0.1). In these three simulations the action potential and intracellular calcium plots show that the periodic bursting complex pattern has disappeared. In particular Ca^{2+} oscillations are absent so that insulin pulsations are not expected to occur.

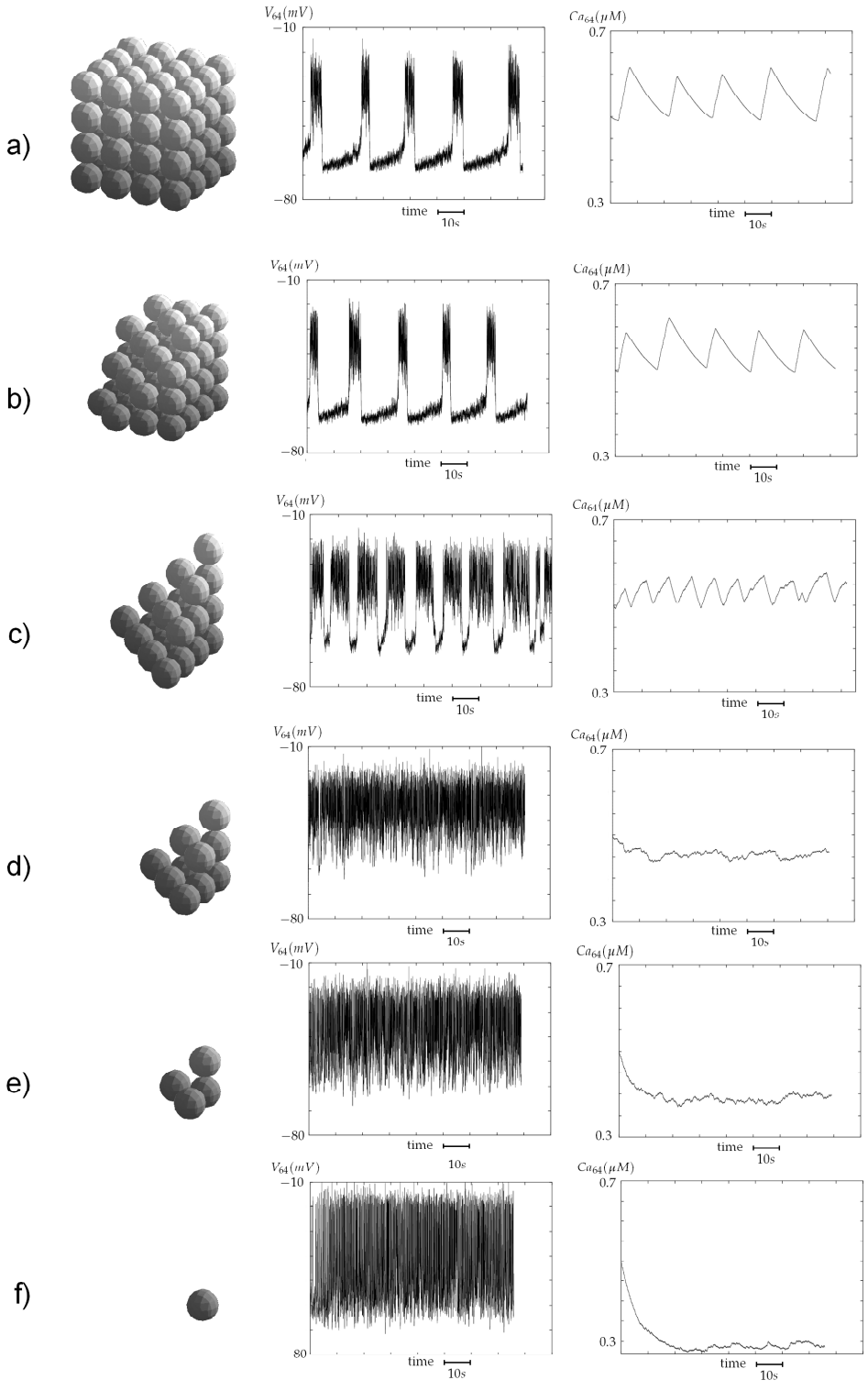
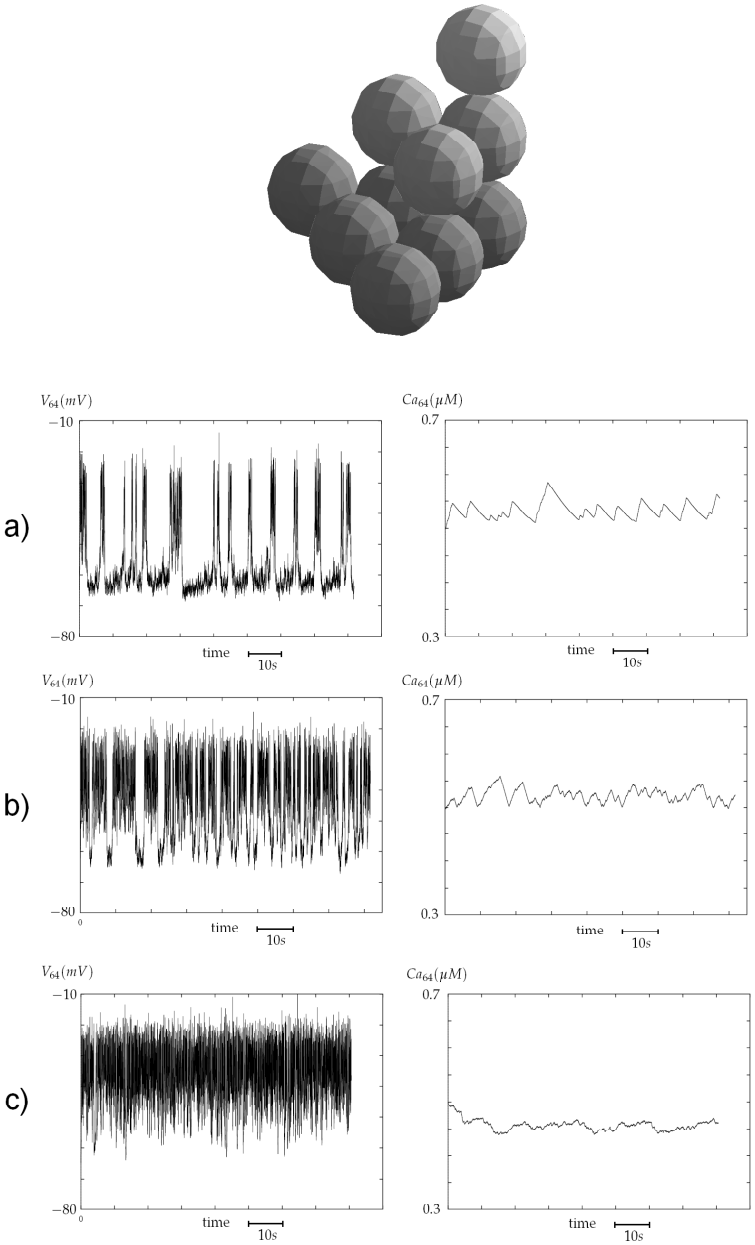


Figure 3. Effect on action potential (left panel) and intracellular Ca2+ (right panel) of an increasing level of glycaemia in a cluster of 64 pancreatic beta-cells with 84% of dead beta-cells.

Figure 3a shows that if a state of normoglycaemia is guaranteed (7 mmol), it is possible to detect Ca2+ oscillations due to synchronization of the remaining beta-cell intact, even in a damaged pancreatic tissue. Figures 3b) and 3c) show the effects on action potential and on intracellular Ca2+ oscillations of a mild and severe hyperglycaemic states (11.6 mmol and 17 mmol, respectively): fluctuations occur in a mild hyperglycaemic state (Fig.3 b) but no in severe hyperglycaemia condition (Fig.3 c). In this last case a dramatic decrease of Ca2+ pulsatility and a consequently severe impaired function should occur.



Appendix

In this Appendix we report the expressions used in the stochastic mathematical model and the parameters' choice fine-tuned on murine experiments [Sherman A et al, 1988]. We applied this model to a cubic cluster of 64 cells. The coupling term in the membrane potential equation (1) couples the j-th cell with the adjacent one through a Von Neumann neighborhood type. The dynamics of the deterministic equations are coupled to a stochastic process governing the K-Ca channels kinetics. Specifically the total conductance gk-Ca is multiplied by the factor pj which represents the number of open channel for the j-cell. Here each channel transition is the result of a Markovian process where the opening mean time is a function of Cai concentration with a fixed channel closing mean time.

$$C_m \frac{dV_j}{dt} = -I_{ion}(V_j, n_j) - \bar{g}_{k-Ca} p_j (V_j - V_k) - g_c \sum_{k \in \Omega_j} (V_j - V_k) \qquad (1)$$

$$\frac{dn_j}{dt} = \lambda \left[\frac{n_\infty(V_j) - n_j}{\tau_n(V_j)} \right] \qquad (2)$$

$$\frac{dCa_j}{dt} = f\left[-\alpha I_{Ca}(V_j) - k_{Ca} Ca_j \right] \qquad (3)$$

$$I_K(V_j, n_j) = \bar{g}_K n_j (V_j - V_K) \qquad (4)$$

$$I_{Ca}(V_j) = \bar{g}_{Ca} m_\infty(V_j) h(V_j)(V_j - V_{Ca}) \qquad (5)$$

$$m_\infty(V_j) = \frac{1}{1+exp\left[(V_m - V_j)/S_m\right]} \quad (6)$$

$$h(V_j) = \frac{1}{1+exp\left[(V_j - V_h)/S_h\right]} \quad (7)$$

$$n_\infty(V_j) = \frac{1}{1+exp\left[(V_n - V_j)/S_n\right]} \quad (8)$$

$$\tau_n(V_j) = \frac{S_c}{exp\left[(V_j - V_b)/S_a\right]} + \frac{S_c}{exp\left[-(V_j - V_b)/S_b\right]} \quad (9)$$

$$C \underset{1/\tau_o}{\overset{1/\tau_c}{\rightleftarrows}} O \quad (10)$$

$$\tau_o = \tau_c \frac{Ca_i}{k_d}. \quad (11)$$

We have considered N= 600 k-Ca channels per cell and adopted a Monte Carlo simulation to evolve such a stochastic dynamics.

Model Parameters

Parameter	**Model Value**	**Parameter**	**Model Value**
f	0.001	$\bar{g}_{K-Ca}$	30000 pS
λ	1.7	$\bar{v}_{Ca}$	110 mV
$\bar{v}_K$	$-75\ mV$	$\bar{g}_K$	2500 pS
$\bar{g}_{Ca}$	1400 pS	C_m	5310 fF
V_b	$-75\ mV$	V_n	$-15\ mV$
V_m	4 mV	S_m	14 mV
V_h	$-10\ mV$	S_h	10 mV
S_a	65 mV	S_b	20 mV
S_c	60 mV	S_n	5.6 mV
K_d	100 μmol	α	$1/(2V_{cell}F)$
V_{cell}	1.150 μm^3	F	96487 $C/mmol$

CHAPTER 6. FINAL REMARKS

In the last 10 years, the idea that an integrated approach is required to understand what comes into play in the development of a multifactorial disease has been steadily consolidating. For this reason, Barabasi et al have introduced a very promising approach that consists of identifying the disease as a complex network, composed of several interacting factors that communicate at multiple levels. Most of the realities that surround us (technological, social and biological systems) are characterized by an interactive multilevel network organization and they are governed by quantifiable organizing principles.

The growing interest in interconnectedness has brought into focus an often ignored issue: "networks pervade all aspects of human health" (Barabási A-L. Linked. New York: Plume, 2003), consequently networks may affect all aspects of medical research and practice, from disease mechanisms to drug discovery.

The existence of intricate molecular links between sub-cellular components and disease genes raises the probability that risk factors may not be as independent of each other as medical practitioners currently consider them to be. It is well known that genetic interactions and environmental pressure play a key role in the pathogenesis of multifactorial diseases. In this scenario Bayesian Network and Mathematical models represent an interactive multilevel approach for a complex disease such as T1D.

The analyses performed in this research project showed that a feasible and accurate risk assessment can be performed by applying the BN method. Here the effects of only three genes were evaluated and compared, but the BN method is able to analyze a huge amount of variables with different risk

categories for each variable, at the same time. This feature could be crucial in the study of multifactorial diseases, in which the triggers involved in the complex mechanisms underlying disease pathophysiology are multiple. By introducing "prior" knowledge from literature, we can also analyze small data sets while maintaining accuracy. Prevalence value was used here as "prior" and, based on that, the network was able to learn the correct rate of genotype combinations characterizing both the general population and patients' group and to elaborate data, giving coherent results as discussed before. The present work represents, to the best of our knowledge, the first study based on both case-control and familiar data sets, showing the joint effect of HLA, INS and PTPN22 in T1D in a Caucasian population with a heterogeneous age of T1D onset, generalizing previous findings regarding data sets consisting of patients and controls < 15 years as demonstrated by Bjørnvold M. et al.

Our results showed that BN represents an alternative way to assess the joint risk of developing T1D, by considering different disease markers at the same time. This method allows us to estimate the risk for each individual carrying a specific gene combination, even if a small data set is available. By collecting data regarding more genes and environmental risk factors, the BN approach gives us the opportunity to elaborate data sets with many variables, to create a population-based risk network and also to establish the relationship between risk markers, with the aim of assessing the global risk for such diseases and to define a personalized risk. On this basis, new predictive trials on the general population could be assessed. Although no preventive intervention is available for T1D today, prediction of disease is an important part of prevention strategies, both for recruitment of participants for research studies and for identification of target populations for future preventive

interventions. Understanding the joint effect of the established T1D susceptibility genes and environmental factors will enhance this possibility.

Moreover, new challenges and research opportunities have to do with predictions of physical reality by computer models. Today the phenomena and processes we ask computer models to predict are of enormous importance to critical decisions that affect our welfare concerning, for example, the biology of diseases and the outcome of medical procedures.

The second part of the experimental work was carried out in the context of a Systems Biology approach, by considering the disease as a complex system, composed of interacting elements at multiple levels with quantifiable organizing principles.

The interdisciplinary efforts developed in this work, aimed at elucidating structures and functions beta-cells in Islets of Langherans. The main challenges in this computational modeling and analysis were to understand, analyze and predict the complex mechanisms of T1D disease progression. In fact, investigations in computational biology and physiology have addressed important issues across many applications from biological signalling pathways, cellular biology and communication, organ function and performance, all the way up to lifestyle and environmental influences and behavioral responses.

In the second part of this research project we have implemented a revised version of the Chay-Keizer model in order to study the behavior of a murine pancreatic islet in physiological and pathological conditions and to investigate further the beta-cells' synchronized pacemaker activity. The significant information emerging from this modelling study is that good glucose control allows a better functionality of beta-cell communication and therefore a more

efficient insulin secretion. This remains true even in a cluster containing residual amount of beta-cells following the autoimmune process occurring in T1D. When blood glucose concentrations tend to rise, the communication within beta-cells becomes altered and insulin release is gradually impaired. With persistent increase of glucose levels, a severe impairment of insulin secretion takes place and diagnosis of T1D is made. These findings underline that the implementation of insulin therapy and a restoration of normoglycaemia as soon as possible after diagnosis is of great relevance. It potentially allows the re-establishment of an endogenous insulin secretion, leaving the patient moving into the well-known phase of clinical remission, taking off exogenous insulin for some time.

Our results suggest that at the diagnosis of T1D a very strict normalization of blood glucose levels should be a key action to be achieved in order to obtain a better preservation of beta-cells. In this respect, the use of a continuous insulin infusion pump and a strict glucose monitoring should be implemented as soon as a patient is diagnosed with T1D to control hyperglycaemia even of modest in nature. In the absence of glucotoxicity any adjuvant therapy to cure this disease including immunomodulation or the regeneration of beta-cells can be more beneficial and, hopefully, contribute to prevent the destruction of any residual beta-cell that is still functioning.

Moreover the predictions obtained through the proposed mathematical model can be directly easily tested through the experimental setup described elsewhere. By means of a specifically built biomedical device controlling an islet microfluidics, the insulin secretion produced by extracellular glucose stimuli was measured by NAD(P)H autofluorescence. Future degree studies are aimed at specifically inserting in such a device, islets of animals at different infiltratum, quantified ex vivo or in vivo following the magneto-

fluorescent nanoparticles techniques. Moreover, similar fluorescence techniques, based on voltage or Calcium sensitive dyes, could be adopted to monitor the underlying electrophysiology of the islets.

Finally, because of the importance of the right timing in shutting down insulin secretion during T1D, beta-cell regeneration could be implemented into the in silico model in order to further highlight how this is affected by hyperglycaemia and inflammation.

Over the last few years, this kind of research work could be extended to explore translational biomedical research, to better understanding the mechanisms of disease and its treatment, thus helping to establish diagnostic biomarkers, physiology-based patient selection criteria, and strategies for choosing, personalizing and optimizing therapeutic options. Computational modeling promises to become a fundamental contributor to future biomedical sciences and technologies, as well as personalized predictive healthcare.

CHAPTER 7. METHODOLOGICAL APPENDIX

7.1 Predictive methods for multifactorial diseases: the different statistical approaches

Statistical approaches, which use models or computer algorithms to generate predictive probabilities of clinical events, can be used prospectively as clinical decision aids or retrospectively to assist in the evaluation of health care systems. Some models are prognostic, making predictions concerning future outcomes in patients. Other models are diagnostic. Generalization is of crucial importance for predictive models designed on a sample data set of correctly classified cases (training data) (Bishop CM, 1995; Vapnik VN, 1999). It is defined as the capacity of the model to maintain the same predictive performance on data not used for training, but belonging to the same population. It is therefore estimated by testing model performance on a different data set of correctly classified cases (testing data). The model generalizes well when predictive errors in testing and training data sets do not differ significantly. Models have to be designed with efficient control of the training process to improve generalization power. Theoretically, the optimal model is the simplest possible model designed on training data which shows the highest possible performance on any other equally representative set of testing data. Excessively complex models tend to overfit, which means they show an error on the training data significantly lower than on the testing data. Overfit is a sort of data storage precluding the learning of prediction rules. It must be avoided since it causes loss of generalization.

Moreover, the utility of predictive models depends on their generalizability, in terms of reproducibility (internal validity) and transportability (external

validity). A model is reproducible if it maintains its accuracy when applied to patients from the same underlying population as those in the sample used for the model development. It has an external validity if it maintains its accuracy when applied to patients from populations intrinsically different from the development sample, with respect to location, time period or methods used for data collection. Thus, generalization is defined as the capacity of the model to maintain the same predictive performance on data not used for build the model.

Various pattern recognition approaches can be used to implement models to separate and classify patients into different prognostic classes. Many applications, however, require more than simple classification. In particular, probability estimates are central in medical decision-making, allowing decision makers to incorporate costs/benefits for evaluating alternatives.

7.2 Updating

The capacity of a predictive model to learn from new correctly-classified cases, day by day, is an important index of quality, especially in clinical practice where data is usually scarce and training on new data becomes of crucial importance. For any model, the whole model design should theoretically be repeated when adding a new case to a training set.

7.3 Tolerance to missing data and supplementary clinical information

In clinical practice, data can be missing for many reasons. In certain types of models, missing data can be replaced by suitable procedures, with negligible or tolerable loss of predictive performance.

Additional information of clinical interest can also be obtained. For example, some models allow newly classified clinical cases to be simply associated with previous similar cases, providing a useful tool for interpretative and comparative diagnosis. Other models are particularly convenient for simulation purposes, such as for evaluating changes in prognostic probability due, for example, to contemplated therapy which may modify the values of certain predictive variables.

7.4 Models Description

Different popular approaches can be used to build predictive models for multifactorial diseases. Bayes rule, K-nearest neighbor methods, Tree models, Logistic regression, score systems and Artificial neural networks. Although these models are not the only ones which can be used to estimate the probability of a multifactorial disease onset, they are certainly the most popular in medicine. Currently, logistic regression and artificial neural networks are the most widely used models in biomedicine, as measured by the number of publications indexed in Medline: 28500 for logistic regression, 8500 for neural networks, 1300 for k-nearest neighbors, 1100 for decision tree, 2608 for Bayesian Networks and 100 for support vector machines (Dreiseitl S, 2002).

7.4.1 Decision Trees

The decision tree is one of the most popular classification algorithms in current use in Data Mining and Machine Learning. Decision tree is a tool that is used in a variety of different settings to organize and break down clusters

of data. Using a decision tree helps you to move from a hypothesis to a reasonable conclusion. This algorithm repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure.

Generally, decision trees are suitable if the following requirements are satisfied:

The alternative pathways represented by the different decision trees are independent (in the sense that they do not rely on some common test or action that has not been modeled).

There are no more than a small number of variables, since even if each variable had only two outcomes there are 2n different paths for n variables. As a rule of thumb 6 is a reasonable limit.

Each variable has only a small number of outcomes (as a rule of thumb, less than 5). But if the outcome is on a continuous scale, say 0–1000 in millimeters, then it would not be possible to use a decision tree.

There are no additional causes, effects and dependencies between the variables.

If these requirements are not satisfied the use of decision trees can become impractical or impossible.

Compared with the other machine learning methods, decision trees have the advantage that they are not black-box models, but can easily expressed as rules. Black box is defined as a system which can be viewed solely in terms of its input, output and transfer characteristics without any knowledge of its internal workings, that is, its implementation is "black". The opposite of a

black box is a system where the inner components or logic are available for inspection, which is sometimes known as a white box, a glass box, or a clear box.

7.4.2 The k-nearest neighbour model

The k-nearest neighbour (kNN) allocation rule is a method for classifying patients based on non-parametric estimation of class-conditional probability density functions (Fukunaga K, 1990). Briefly, the training phase of the algorithm consists of storing the predictor vectors and class labels of the training samples and mapping the class labels of training cases into multidimensional predictor space. In the classification phase, the same predictors as before are acquired for the test case (the class of which is unknown). Distances from the new vector to all stored vectors are computed and k closest cases are selected. The new case is predicted to belong to the most numerous class in the neighbourhood of k closest cases. Euclidean distance is usually used. Thus multidimensional predictor space can be simply partitioned into regions by assigning each point in the space to the class which is the most popular among the k nearest training cases. Neighbourhood size is essential in building the kNN classifier because it can strongly influence the quality of predictions; larger values of k can reduce the effect of noise on the classification but make boundaries between classes less distinct.

The performance of k-nearest neighbors is generally worse on high-dimensional data because, when the relative importance of dimensions is not weighted, the data from spurious and irrelevant dimensions may negatively influence the distance calculation (Mitchell T, 1997).

A key advantage of this non-parametric approach is that it does not make any statistical assumption about the data, thus enabling an arbitrary decision boundary. kNN models are also very easy to update with new data: each new correctly classified patient can be added to the training database and used to classify subsequent cases. Another strength of kNN algorithms over other approaches is that any new test case can be analyzed and interpreted by comparing it with its k neighbours. This provides useful insights for clinical interpretation of the classification results, helping in comparative diagnosis. Finally, kNN is also sensitive to the presence of variables that are irrelevant for classification purposes. All non-parametric techniques have a tendency to overfit the model when the number of variables used is too large.

7.4.3 Logistic Regression

Logistic regression is used extensively in numerous disciplines, including the medical and social science fields. Logistic regression is a statistical method based on an algorithm used to assess the effect of certain explanatory variables on a dependent variable. The user, based on own experience with the data and data analysis, must choose the right inputs and specify their functional relationship to the dependent variables. Logistic regression can be binomial or multinomial. Binomial or binary logistic regression refers to the instance in which the observed outcome can have only two possible types. Multinomial logistic regression refers to cases where the outcome can have three or more possible types (e.g., "better' vs. "no change" vs. "worse").

Logistic regression is used to predict the odds of being a case based on the predictor(s). Like other forms of regression analysis, logistic regression makes use of one or more predictor variables that may be either continuous or

categorical data. Unlike ordinary linear regression, however, logistic regression is used for predicting binary outcomes rather than continuous outcomes. All the effects modeled are additive; logistic regression does not account for interaction effects except when interaction terms are created as additional variables in the analysis.

A main weakness of logistic regression is that *outliers affect the results significantly*. The researcher should analyse standardized residuals for outliers and consider removing them or modeling them separately. A final remark is that the logistic regression *is not simple to update with new training data* although periodic full retraining may not cause excessive problems.

7.4.4 Artificial Neural Networks

The starting point of artificial neural networks is quite different from that of statistical models. Neural networks are generic learning systems. An Artificial Neural Network, often just called a neural network, is a mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a "connectionist" approach to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data. Artificial neural networks have recently had many successful applications in medicine (den Boer S et al, 2005; Dreiseitl S et al, 2002; Jaimes F et al, 2005; Shahian DM et al, 2004). Key advantages with respect to common statistical models are: no statistical assumption about data distribution is required; no mathematical model has to be defined; not too much complex network architecture has to be designed to suitably

approximate any unknown nonlinear relationships between predictor input variables and output probabilities; the distributed structure of the network may account for correlations among input variables; ANNs can be trained with examples like human brains to use acquired knowledge in decision making; the training process can be controlled to avoid overfitting and loss of generalization capacity.

Considerable difficulties may arise when designing and using ANNs. The training process is difficult and not univocal: the problem of initialization is all but trivial; as in all nonlinear procedures, many different solutions, which are difficult to compare and interpret, may be obtained; the complexity of ANN architecture is only roughly definable in terms of number of neurons, layers and connections.

The high flexibility and sophisticated training procedures of ANN models allow very good customization to data of local institutions, but continuous updating is practically impossible. Periodic retraining is inadvisable, because of the complexity of the training process. ANNs should therefore be trained once and for all, using a sufficiently large number of cases representative of the study population.

7.4.5 Bayesian Networks

A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. Over the last decade, the Bayesian network has become a popular representation for encoding uncertain expert knowledge in expert systems. More recently, researchers have developed methods for learning Bayesian networks from data. The techniques that have

been developed are new and still evolving, but they have been shown to be remarkably effective for some data-analysis problems.

In a predictive problem with a categorical outcome variable, Bayesian Networks identify predictors *qualitatively* through a graphical diagram with nodes (representing variables) and edge (arrows representing relationships between variables of interest), and *quantitatively*, the degree of dependency is expressed with probabilistic terms. Bayesian Networks produce predictive values through a complex application of the well-developed Bayesian probability theory (Bayes' rule) between predictors and outcome variables.

So what do Bayesian networks and Bayesian methods have to offer?

There are at least four answers.

Bayesian networks can readily handle incomplete data sets. For example, consider a classification or regression problem where two of the explanatory or input variables are strongly anti-correlated. This correlation is not a problem for standard supervised learning techniques, provided all inputs are measured in every case. When one of the inputs is not observed, however, most models will produce an inaccurate prediction, because they do not encode the correlation between the input variables. Bayesian networks offer a natural way to encode such dependencies.

Bayesian networks allow one to learn about causal relationships. Learning about causal relationships are important for at least two reasons. The process is useful when we are trying to gain understanding about a problem domain, for example, during exploratory data analysis. In addition, knowledge of causal relationships allows us to make predictions in the presence of interventions. For example, a marketing analyst may want to know whether

or not it is worthwhile to increase exposure of a particular advertisement in order to increase the sales of a product. To answer this question, the analyst can determine whether or not the advertisement is a cause for increased sales, and to what degree. The use of Bayesian networks helps to answer such questions even when no experiment about the effects of increased exposure is available.

Bayesian networks in conjunction with Bayesian statistical techniques facilitate the combination of domain knowledge and data. Anyone who has performed a real-world analysis knows the importance of prior or domain knowledge, especially when data is scarce or expensive. The fact that some commercial systems (i.e., expert systems) can be built from prior knowledge alone is a testament to the power of prior knowledge. Bayesian networks have a causal semantics that makes the encoding of causal prior knowledge particularly straightforward. In addition, Bayesian networks encode the strength of causal relationships with probabilities. Consequently, prior knowledge and data can be combined with well-studied techniques from Bayesian statistics.

Bayesian methods in conjunction with Bayesian networks and other types of models represent an efficient and principled approach for avoiding the over fitting of data. As we shall see, there is no need to hold out some of the available data for testing. Using the Bayesian approach, models can be "smoothed" in such a way that all available data can be used for training.

Bayesian models seem a good compromise between complexity and predictive performance.

7.5 Logistic Regression Versus Bayesian Networks

An increasingly large number of data items are collected routinely, and often automatically, in many areas of medicine. It is a challenge for the field of machine learning and statistics to extract useful information and knowledge from this wealth of data. In any application, the choice of an optimal model is rarely univocal and cannot be made a priori. For clinical decisions, users should prefer simple intuitive models to complex ones, but this preference should be evaluated in the light of model fit to the experimental data. So far, there is no single algorithm that performs better than all other algorithms on any given data set and application area.

For logistic regression (LR), the popularity may be attributed to the interpretability of model parameters and ease of use. LR has less restricted statistical assumptions than those of other conventional statistical methods with normal distributions, linear relationships, or equal variances with each group of predictors (Tabachnick & Fidell, 2001.). However, two assumptions, linearity in the logit and additivity, need to be made (Harrell Lee et al, 1985; Menard S, 1995), which may limit the use of LR in large data sets (Table 1). Although nonlinear relationships can be modeled using LR, complex transformations of independent variables are needed in order to meet the assumption of linearity in the logit. This process is difficult and requires that investigators explicitly search for these relationships, thus requiring that they are either known or suspected a priori. In addition, LR does not operate well in environments where input variables can number in the hundreds, and the relationships concealed within are complex, which are major sources of violation of the additivity assumption. As the number of predictor variables in any LR model increases, the number of possible interactions increases exponentially, resulting in the complex process of specification of interaction

and confounding variables. LR imposes several theoretical and practical limitations when using variable selection processes in situations where a large number of variables are available, as is often the case with very large databases. Traditional LR techniques are somewhat rigid and are unable to assist investigators in uncovering unknown or unexpected relationships in large sets. Often, the valuable "nuggets" of information that exist in large sets of data are nonlinear and unknown. In contrast, Bayesian Networks (BNs) can allow patterns and trends in data to emerge without hypothesizing a priori or constraining input variables, which enables BN techniques to transform overwhelming volumes of data through the discovery of the creation of predictive (or classification) models (Heckerman, 1997.). Although using BNs may incur limitations such as complex computations that may be expensive (Heckerman, 1994), they offer the distinct advantages of operating where LR modeling techniques cannot. Bayesian networks may (a) be used without conventional statistical assumptions such as linearity or additivity; (b) handle a larger number of predictors when the identification of interactions among predictors is less complex; (c) detect important variables and relationships that could be missed by investigators; and (d) produce accurate predictions even in situations where complete data are not available. Finally, BNs are easily understood, because they represent knowledge through a graphical diagram via nodes and arrows, making BNs prominent even among data mining techniques. Although it is expected that a researcher analyzing large amounts of data has a general idea of what patterns may emerge, the complexity of high-volume data makes it impossible to have a truly comprehensive knowledge of all patterns that may be present. Overall, LR is recommended for well-designed research projects, where investigators can expect the potential predictors and possible interactions, based on literature and experience.

Table 1. The comparison between Logistic Regression and Bayesian Networks

	Logistic Regression	Bayesian Network
Approach	Verification-based (Hypothesized)	Nonverification-based
Types of studies	Experimental/observational with less number of variables Theory testing	Observational study with a larger number of variables
Terms used	Variable Independent (predictor) variable Dependent (response) variable	Node Random node or information node Target node
Assumptions	Linearity in the logit and additivity	No specific assumption in using BN with databases
Efficiency	Requires an investigator to specify interactions and confounding variables	Has the ability to assist to find interactions and confounding variables among variables through the graphical method
Variable selection	Need to specify the variables that create the risk of introducing biases in the selection of relevant variables	Iteratively learned rules can assist to select important variables
Error	Can be influenced by investigators errors because of lack of domain knowledge	Can be reduced by assistance using databases and learning algorithms
Missing	Not able to predict with incomplete data	Relatively accurate prediction with incomplete data
Interpretation	Coefficient Odds ratio	Relationships via a diagram Probabilities Information gain (mutual information) Posterior probabilities

7.6 Used Software Packages

The open source software, R (www.r-project.org) and Weka (formally called Waikato Environment for Knowledge Learning, www.cs.waikato.ac.nz/~ml/WEKA) were adopted to establish the models. R and Weka support many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. These programs are open source software issued under the GNU general Public License.

REFERECES

Bishop CM: Neural Networks for Pattern Recognition New York: Oxford University Press Inc; 1995

Vapnik VN: The Nature of Statistical Learning Theory New York: Springer-Verlag; 1999

Dreiseitl S, Ohno-Machado L. Logistic regression and artificial network classification models: a methodology review. Journal of Biomedical Informatics (35) 352-359, 2002

Fukunaga K: Introduction to Statistical Pattern Recognition Boston: Academic Press; 1990

Mitchell T. Machine Learning. New York: McGrow-Hill; 1997

den Boer S, de Keizer NF, de Jonge E: Performance of prognostic models in critically ill cancer patients – a review. Critical Care, 9:R458-R463, 2005

Dreiseitl S, Ohno-Machado L: Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform, 35:352-359, 2002

Jaimes F, Fabiarz J, Alvarez D, Martinez C: Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Crit Care, 9:R150-156, 2005

Shahian DM, Blackstone EH, Edwards FH, Grover FL, Grunkemeier GL, Naftel DC, Nashef SA, Nugent WC, Peterson ED: Cardiac surgery risk models: a position article. Ann Thorac Surg, 78:1868-1877, 2004

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.) Boston: Allyn & Bacon

Harrell, F. E., Lee, K. L., Matchar, D. B., & Reichert, T. A. Regression models for prognostic prediction: Advantages, problems, and suggested solutions. Cancer Treatment Reports, 69, 1071-1077, 1985

Menard, S. (1995). Applied logistic regression analysis. London: Sage Publications.

Heckerman, D. E. Bayesian networks for data mining. Data Mining and Knowledge Discovery, 1, 79-119, 1997

Heckerman, D. E. Learning Bayesian networks is NP-hard (Rep. No. MSR-TR-94-17). Redmond, WA: Microsoft Research, 1994

ABSTRACTS ARISING FROM THIS WORK

1) **Portuesi R**, Lausser L, Spoletini ML, et al. Bayesian network to investigate the dependency and interaction between HLA, INS, PTPN22 and CTLA4 genes in type 1 diabetes. Diabetologia 2009;52:S110-S110

PS 1 Genetics of type 1 diabetes

253

Bayesian network to investigate the dependency and interaction between HLA, INS, PTPN22 and CTLA4 genes in type 1 diabetes

R.A. Portuesi[1], L. Lausser[2], M.L. Spoletini[3], S. Zampetti[3], A. Petrone[3], B. Boehm[2], H.A. Kestler[4], P. Pozzilli[1], R. Buzzetti[3];
[1]University Campus Bio-Medico, Rome, Italy, [2]Internal Medicine I, University Hospital of Ulm, Germany, [3]University "La Sapienza", Rome, Italy, [4]Institute for Neural Information Processing, University of Ulm, Germany.

Background and aims: Type 1 diabetes (T1D) is a multifactorial and polygenic disease with four major susceptibility genes (HLA (DR/DQ), CTLA4, INS, PTPN22) that by interacting with each other and with environmental factors, determine disease onset. These genes are known to be involved in immune regulation whereas INS modulates insulin expression in human thymus. The incidence rate of T1D is increasing uniformly in Italy with an annual average increase of 3.6%. The genetic profile of individuals who develop diabetes appear to change in countries with a high T1D incidence shifting from predominantly high-risk HLA genotypes towards higher percentage of median and low-risk HLA genotypes. The aim of this study was to investigate in T1D.

Materials and methods: Genetic data (HLA, PTPN22, INS, CTLA4) were analysed from a group of 391 T1D patients and 100 controls (M/F 1.06, age 1-41) diagnosed by participating centres of the IMDIAB group in continental Italy. Diagnosis of T1D was based on the ADA classification criteria. The control group was recruited from the Blood Transfusion Service in Rome. We divided HLA alleles in high, moderate and low risk for T1D, CTLA4 and PTPN22 alleles in susceptibility/non susceptibility alleles, and INS gene alleles in susceptibility/protection. We estimated a Bayesian network to study dependencies and interactions between the different alleles in relationship to group status. Bayesian networks, also called belief networks, are probabilistic graphical models that represent a set of variables and their probabilistic dependencies. In the present study the model was trained on genetic variables and group status (T1D / control). A hill climbing algorithm was used for fitting the model.

Results: The frequency of the HLA genotypes with high, moderate and low risk was 18.9%, 55.2% and 25.8% in T1D patients (2%, 5%, 93% in controls). The model shows that group status is directly influenced by HLA ($p = 1.4*10^{-30}$) and INS ($p=3.0*10^{-8}$), and that there is a dependency of INS on HLA ($p = 3.0*10^{-5}$). For high, moderate and low risk HLA corresponding conditional frequencies of 0.74, 0.76 and 0.55 for INS susceptibility alleles were found. Separating the data group wise, leads to corresponding conditional probabilities of 1, 0.4, 0.34 for the control group and 0.73, 0.76, 0.75 for the patient group for high, moderate and low risk respectively. No significant connection between HLA and PTPN22 ($p = 0.07$) and CTLA4 and INS ($p = 0.1$) could be found.

Conclusion: This study confirms that HLA (DR/DQ) and INS susceptibility alleles are the major determinants of the disease. By analysing all 491 subjects a higher probability for INS susceptibility alleles was found in the presence of high and moderate risk HLA alleles. However, if we consider INS susceptibility alleles only in the T1D group, these are less frequent in high risk HLA alleles. Further substantiation is needed for the identification of the statistically significant conditional dependency of INS on HLA.

RP supported by the Italian Ministry for Education, University and Research (MIUR)

2) **R. Portuesi**, L. Lausser, M. L. Spoletini, S. Zampetti, B. Boehm, H. Kestler, P. Pozzilli, IMDIAB Group, R. Buzzetti. Applicazione del metodo probabilistico di Bayes per l'analisi dell'interazione tra geni di suscettibilità al diabete di tipo 1. Congresso nazionale SID, Padova 2010.

3) **R. Portuesi**, C. Cherubini , A. Gizzi, G. Valorani, R. Buzzetti, P. Pozzilli, S. Filippi. A mathematical model to study the autoimmune progression towards type 1 diabetes. American Diabetes Association (2011).

IMMUNOLOGY

2533-PO

A Mathematical Model To Study the Autoimmune Progression towards Type 1 Diabetes

ROSALBA PORTUESI, CHRISTIAN CHERUBINI, ALESSIO GIZZI, GIUDITTA VALORANI, RAFFAELLA BUZZETTI, SIMONETTA FILIPPI, PAOLO POZZILLI, *Rome, Italy, London, United Kingdom*

Islets of Langerhans are small endocrine organs characterized by an complex anatomy including a core of insulin producing beta-cells, tightly interconnected by gap junctions and a mantle of other endocrine cells. The integrity of the interactions and the 3D architecture among beta-cells is critical for proper biosynthesis, storage and the release of insulin. The aim of this study was to evaluate the effect on beta-cells signalling of progressive lymphocytic islet cell infiltration (insulitis), by modelling the disruption of pancreatic islet anatomy as a consequence of insulitis in terms of apoptotic events of beta-cells and altered glucose concentration. We numerically simulated a 3D small cluster of mouse beta-cells via an extended stochastic Sherman-Rinzel-Keizer electrophysiological model. Progressive damage was modelled (0%, 31%, 69%, 84%, 94% and 98% of dead beta-cells) at different glucose concentrations, representing the different glycaemic states in the autoimmune progression towards type 1 diabetes (T1D). At 31% (normoglycaemia) and 69% (hyperglycaemia) of dead beta-cells, the system appeared to be biologically robust to maintain regular [Ca $^{++}$] oscillations guaranteeing an effective insulin release. Simulations at 84%, 94% and 98% of death beta-cells (severe hyperglycemia) showed that the complex periodic bursting pattern disappeared and [Ca $^{++}$] oscillations were absent. In such conditions insulin pulsatility is not expected to occur. Our results suggest that the islet tissue of beta-cells is biophysically robust enough to compensate high rates of cellular loss, a fact which is in agreement with in vitro animal experiments. The model indicates the necessity of maintaining glycaemia within physiological levels as soon as possible after diabetes onset in order to avoid a dramatic drop of [Ca $^{++}$] pulsatility and consequent insulin release. In the absence of glucotoxicity, any adjuvant therapy to cure this disease including immunomodulation or the regeneration of beta-cells can be more beneficial and hopefully contribute to impeding the destruction of any residual beta-cell still functioning.

Supported by: University Campus Bio-Medico, ICRANet & Centro Internazionale Studi Diabete

4) **Portuesi R**, Bohm BO, Julier C, Pozzilli P. HLA and insulin genes: Bayesian networks confirm interaction between the two most important susceptibility genes in type 1 diabetes in a French Caucasian population. Diabetologia 2010;53:S77-S78 (Oral Presentation).

173

HLA and insulin genes: Bayesian networks confirm interaction between the two most important susceptibility genes in type 1 diabetes in a French caucasian population

R. Portuesi[1], B.O. Böhm[2], C. Julier[3], P. Pozzilli[1];
[1]University Campus Bio-Medico, Rome, Italy, [2]University Hospital Ulm, Internal Medicine I, Ulm, Germany, [3]Centre National de Genotypage, Evry, France.

Background and aims: Type 1 diabetes (T1D) is an autoimmune chronic disease resulting from the interaction between more or less favouring environmental factors with multiple susceptibility genes. HLA, Insulin (INS), CTLA4 and PTPN22 are considered the main T1D susceptibility genes. As many epidemiological studies have demonstrated, T1D incidence is increasing worldwide by 3.9% per year, particularly in Caucasian population of

Springer

Northern Europe. Unlike in single gene disorders, in multifactorial diseases, such as T1D, identifying the combination of causative genes is still difficult. Genetic profiles of individuals who are affected by T1D appear to change among different countries shifting from mainly high-risk genotypes towards higher percentages of median and low-risk genotypes. In a previous study we demonstrated the interaction between HLA and INS genes in an Italian population from Lazio region using the Bayesian Network approach. To confirm our previous findings, the aim of the present study was to investigate and verify in T1D the dependency and interaction between HLA and INS genes by investigating another Caucasian population.

Materials and methods: We have analyzed a database of genetic data from a French Caucasian population, the case-control cohort consisted of 868 French T1D patients (M/F 1.13, 19.63 ±14.40 yrs mean age of T1D onset) and 93 French control subjects (M/F 0.7). Diagnosis of T1D was based on the ADA classification criteria. We divided HLA alleles in high, moderate and low risk for T1D, PTPN22 alleles in susceptibility/non susceptibility alleles and INS gene alleles in susceptibility/protection. We created a Bayesian Network model trained on genetic variables and group status (T1D/control). Bayesian networks, also called belief networks, are probabilistic graphical models that represent a set of variables and their probabilistic dependencies. To gain insights into the dependency/interaction between susceptibility genes involved in T1D, we have assessed more than one gene at the time (namely HLA, INS and PTPN22 genes).

Results: We implemented a Bayesian Networks model learning the structure of the specified database, with a fixed level of significance equal to 0.05 to find out the interaction. The model showed that group status was directly influenced by HLA ($p = 1.0*10^{-26}$) and that there was a dependency of INS on HLA ($p = 4*10^{-4}$). In addition to our previous data, having separated the data group wise, the analysis of T1D patients group also highlighted the gene interaction between HLA and INS ($p = 3.7*10^{-4}$). No significant relation between HLA and PTPN22 (p = NS) and PTPN22 and "status group" (p = NS) was found.

Conclusion: The presence of interactions between susceptibility genes can explain why the study of a single susceptibility gene in a polygenic disease such as T1D offers limited information. Bayesian network type of analysis represents a step forward in understanding gene interactions and may offer novel clues for T1D pathogenesis. Further studies are needed to clarify the true nature of the biological interaction between HLA and INS gene alleles.

PUBLICATION AND PAPER SUBMITTED ARISING FROM THIS WORK

1) A stochastic mathematical model to study the autoimmune progression towards type 1 diabetes. Portuesi R, Cherubini C, Gizzi A, Buzzetti R, Pozzilli P, Filippi S. Diabetes Metab Res Rev. 2012 Dec 10. doi: 10.1002/dmrr.2382. [Epub ahead of print]. PMID: 23229223.

DIABETES/METABOLISM RESEARCH AND REVIEWS
Diabetes Metab Res Rev 2012.
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/dmrr.2382

RESEARCH ARTICLE

A stochastic mathematical model to study the autoimmune progression towards type 1 diabetes

R. Portuesi[1]
C. Cherubini[2,3]
A. Gizzi[2]
R. Buzzetti[4]
P. Pozzilli[1,5]*
S. Filippi[2,3]

[1]*Department of Endocrinology and Diabetes, University Campus Bio-Medico, Rome, Italy*

[2]*Laboratory of Non Linear Physics and Mathematical Model, University Campus Bio-Medico, Rome, Italy*

[3]*ICRA, University of Rome 'La Sapienza', Rome, Italy*

[4]*Department of Clinical Sciences, 'Sapienza', Polo Pontino, Rome, Italy*

[5]*Blizard Institute, Center of Diabetes, St Bartholomew and The London School of Medicine, Queen Mary, University of London, London, UK*

*Correspondence to: Prof. Simonetta Filippi, Faculty of Engineering, University Campus Bio-Medico of Rome, Via Álvaro del Portillo 21 – 00128, Rome, Italy.
E-mail: s.filippi@unicampus.it

Received: 30 March 2012
Revised: 19 November 2012
Accepted: 30 November 2012

Abstract

Background The integrity of the interactions and the 3D architecture among beta cell populations in pancreatic islets is critical for proper biosynthesis, storage and release of insulin. The aim of this study was to evaluate the effect on beta cells electrophysiological signalling of progressive lymphocytic islet cell infiltration (insulitis), by modelling the disruption of pancreatic islet's anatomy as consequence of insulitis and altered glucose concentration.

Methods On the basis of histopathological images of murine islets from non-obese diabetic mice, we simulated the electrophysiological dynamics of a 3D cluster of mouse beta cells via a stochastic model. Progressive damage was modelled at different glucose concentrations, representing the different glycaemic states in the autoimmune progression towards type 1 diabetes.

Results At 31% of dead beta cells (normoglycaemia) and 69% (hyperglycaemia), the system appeared to be biologically robust to maintain regular Ca^{2+} ion oscillations guaranteeing an effective insulin release. Simulations at 84%, 94% and 98% grades (severe hyperglycemia) showed that intracellular calcium oscillations were absent. In such conditions, insulin pulsatility is not expected to occur.

Conclusions Our results suggest that the islet tissue is biophysically robust enough to compensate high rates of beta cell loss. These predictions can be experimentally tested *in vitro* quantifying space and time electrophysiological dynamics of animal islets kept at different glucose gradients. The model indicates the necessity of maintaining glycaemia within physiological levels as soon as possible after diabetes onset to avoid a dramatic interruption of Ca^{2+} pulsatility and consequent drop of insulin release.

Keywords type 1 diabetes; autoimmunity; islet cells; beta cells; hyperglycaemia; mathematical modelling; calcium oscillations; insulitis

Introduction

Islets of Langerhans are small highly vascularised endocrine structures playing a key role in glucose homeostasis, having their own complex anatomy including a core consisting mainly of insulin-producing beta cells, which are tightly interconnected by gap junctions and surrounded by a mantle of other endocrine cells secreting different hormones, such as glucagon, somatostatin and pancreatic polypeptide [1–3].

Large ($>5 \cdot 10^6$ μm^3) and intermediate ($1 \cdot 10^6$ to $5 \cdot 10^6$ μm^3) islets can be located centrally, near the large blood vessels and pancreatic duct, whereas small

($<1 \cdot 10^6$ μm^3) islets can be observed in the periphery of pancreatic endocrine tissue [4].

Beta cells act as glucose sensors regulating the release of insulin according to blood glucose levels. Increased levels of extracellular glucose cause a cascade of events inside the beta cells that, by membrane depolarization and consequent influx of Ca^{2+} ions through voltage-dependent membrane channels, lead to pulsatile insulin release [5,6]. Moreover, glucose-stimulated islets exhibit a pattern of bursting electrical activity coupled with Ca^{2+} oscillations. This complex behaviour is synchronized throughout the islet so that the latter works as a functional syncytium [7]. Studies on pancreatic islets revealed that communication between beta cells via gap junctions is crucial for proper biosynthesis, storage and release of insulin, whereas uncoupling leads to altered beta cell function [8–10]. Moreover, normal blood glucose levels are maintained if the beta cell critical number is guaranteed [11,12].

Type 1 diabetes is an autoimmune disease characterized by the progressive destruction of damaging of beta cells in the islets of Langerhans, associated with insulin deficiency and hyperglycaemia.

Experimental models of diabetic mice suggested that lymphocytic infiltration of the islets (insulitis) begins at 3–5 weeks of age and becomes evident at around 8 weeks. This scenario leads to an active destruction of beta cells and consequently to a progressive decrease in beta cell population [4,11]. Alanentalo *et al.* [4] demonstrated that in the pancreas of non-obese diabetic mice between 8–12 weeks of age, a severe damage occurred, namely the disruption of 20% of islets, whereas between 8–16 weeks of age, 50% of islets was destroyed. The smallest islets are the first to be destroyed during insulitis, whereas a reduction of large-sized and intermediate-sized islets appears after 12 weeks of age. These authors also showed that when approximately 85% of beta cell volume is destroyed, mice developed overt diabetes. During beta cell disruption, the development of hyperglycaemia is associated with impaired insulin levels [8].

By losing connections between beta cells, insulin production is dramatically decreased. Experimental evidence suggests that the integrity of intercellular communication via gap junctions and 3D architecture of beta cells inside the islets are critical aspects for insulin production [5]. Electrophysiological studies have shown that the number of beta cells influences the synchrony and amplitude of Ca^{2+} oscillations in clusters [5,7].

The autoimmune process associated with beta cell loss in type 1 diabetes can be considered a typical scenario where the complex oscillatory phenomenon usually observed in single beta cell and in the whole intact islet appears compromised.

The role of gap junctions in the connection among beta cells and the effect of a reduced number of beta cells in a damaged cluster on the associated electrophysiology can be quantitative analysed also by using mathematical models.

The first computational model by Chay and Keizer [13,14], simulating the electrical behaviour of a single beta cell in an intact islet, was able to describe the electrical bursting activity and the Ca^{2+} oscillatory dynamics. This theoretical model of beta cell behaviour was implemented on the lines of the classical Hodgkin–Huxley's work [15]. Inspired by its purely theoretical predictions, these important oscillations in Ca^{2+} were eventually recorded for the first time *in vitro* 8 years later.

In the same period, Atwater *et al.* [16] suggested that the stochastic opening and closing of ion channels represented the key mechanism to take into account for the difference between the single beta cell and the bursting behaviour of coupled beta cells in a cluster. In the latter case, in fact, through an effective coupling due to a channel-sharing dynamics, the noise intensity is reduced by gap junction electrical connections.

Further studies, performed by Sherman *et al.* [17,18] clarified the irregular spiking behaviour of isolated beta cells as well as the collective bursting dynamics of coupled cells.

In the light of these results, the aim of the present study was to evaluate the effect of progressive lymphocytic islet infiltration (insulitis) on beta cell signalling. The destruction of cluster architecture consisting of pancreatic beta cells, as consequence of insulitis, was modelled in terms of apoptotic events. For a quantitative comprehension of such a complex phenomenon, mainly involving cellular synchronization, we have investigated the activity of beta cells in a 3D cells' lattice with defined neighbourhood relations and altered glucose concentrations. We have numerically simulated a small cluster of mouse beta cells via an extended stochastic electrophysiological model [17,18]. This formulation was originally developed to deal with an intact and regular cluster configuration only. Here, we have induced, however, a gradually break of the cluster's spatial symmetry by decreasing the number of communicating beta cells in the cluster, creating a pathological environment inspired by experimental histological sections of diseased diabetic mice. We point out that a minimal physiologically inspired stochastic model can be an efficient tool to obtain original and quantitative insights into the complex dynamics of beta cell destruction. This work was specifically focused on the clinical implications of the simulation results together with possible future tests *in vitro* performed through recently developed bioengineering devices built to experimentally study the spatio-temporal electrophysiological dynamics of mice islets.

Material and methods

The stochastic model for beta cell electrical excitability developed by Sherman *et al.* [17,18]was adopted to simulate (X-PPAUT and C++) a virtual portion of a pancreatic islet (cluster) with progressive apoptotic events (death of beta cells). It consisted of identical beta cells coupled through gap junctions, with morphological and physiological characteristics following experimental data of murine beta cells [16,19].

DOI: 10.1002/dmrr

The pancreatic autoimmune progressive damage occurring in type 1 diabetes was modelled at different stages of lymphocytic infiltration in the experimental model of the non-obese diabetic mouse model. One aim was to investigate the main electrophysiological changes in beta cell membrane potential and free intracellular Ca^{2+} ion concentration [Ca^{2+}]. Furthermore, many experimental studies have shown that both the slow and fast Ca^{2+} oscillations correspond to practically synchronized pulsatile insulin secretion of beta cells [20,21], also proposing that glucose stimulation of a prompt and sustained insulin release is mediated by increase of the cytoplasmic Ca^{2+} concentration [22].

With such a correspondence, our analysis aimed to investigate this dynamics [22–24].

In type 1 diabetes, the death of beta cells is induced by an autoimmune response underlying the disease process. In our model, we assumed that beta cell apoptosis corresponded to its disappearance from the cluster. This was achieved by setting to zero both the electrical dynamics of the dead beta cell and the gap junction conductance (representing the connection between beta cells). This process simulates, in the first approximation, the g_c progressive lymphocytic infiltration that, by substituting the dead beta cells, destroys their gap junction coupling.

Previous electrophysiological studies have shown that a single ion channel can have two conductance states only (i.e. open and closed). Fluctuations between closed and open states appear to occur randomly. Consequently, in pursuing our goal of understanding bursting in the single beta cell and the synchronized activity of the beta cells in a cluster, we considered such a stochastic (probabilistic) mechanism [16,19].

In this model, a stochastic dynamics of calcium-activated potassium (K-Ca) channels occurs. It is well known that a 50-beta cell cluster exhibits organized bursting [17,18]. Therefore, we performed numerical simulations, starting for simplicity from an intact cubic-shaped cluster of identical $4 \times 4 \times 4$ cells, each of them having 600 of these stochastic channels.

For a system of N interacting beta cells, the differential equations for the jth cell are

$$C_m \frac{dV_j}{dt} = -I_K(V_j, n_j) - I_{Ca}(V_j) - \bar{g}_{K\text{-}Ca}\, x_j (V_j - V_K) - g_c \sum_{k \in \Omega_j} (V_j - V_k) \quad (1)$$

$$\frac{dn_j}{dt} = \lambda \left[\frac{n_\infty(V_j) - n_j}{\tau_n(V_j)} \right] \quad (2)$$

$$\frac{dCa_j}{dt} = f\left[\alpha I_{Ca}(V_j) - k_c Ca_j\right] \quad (3)$$

$$\langle x_j \rangle = \frac{Ca_j}{K_d + Ca_j} \quad (4)$$

where

V_j is the membrane potential,

n_j is the open fraction of voltage-gated K^+ channels,

Ca_j is the free intracellular Ca^{2+}.

Ω_j in Eq. (1) indicates the interactions of the jth cell through gap junctions (Von Neumann neighbourhood).

I_K and I_{Ca} represent the currents through the voltage-gated K^+ and Ca^{2+} channels, respectively.

x_j is a random variable indicating the fraction of open K-Ca stochastic channels of the jth cell, whose transition probabilities are described extensively in [25,26], whereas its mean value is given by 4.

In words, the mathematical model describes, for each cell, the time variation of the corresponding action potential for the ith cell V_j, here denoted with the standard calculus symbol dV_j/dt. The latter depends on the sum of the ionic currents and on the gap junctional coupling constant g_c. A feedback is established with the potassium channels variable n_j and with the intracellular calcium Ca_j. Finally, the instantaneous value of the calcium concentration Ca_j promotes the closing or opening of the corresponding ionic channels.

Explicit expressions and parameters of Eqs (1)–(4) are reported in the Appendix.

The parameter k_c plays a central role in the system dynamics. We assumed specifically the simplest linear monotonic relation that links k_c with the external glucose concentration [Glu] measured in mmol:

$$k_c = 7 \cdot 10^{-3}\ \text{mmol}^{-1}[Glu] - 0.019$$

This relationship shows that our modelling considers bursting dynamics as being directly dependent on glucose concentration. Other more accurate choices could be proposed, but for the aims of our analysis, this form allows a model calibration in accord with experimental data on clusters stimulated by different glucose concentrations.

We then considered the specific values of k_c inducing different behaviours in the bursting dynamics in agreement with experiments on clusters discussed later in more detail.

In this study, the value of gap junction conductance $g_c = 215$ pS was chosen corresponding to the mean value taken from experimental data, as discussed elsewhere [25,26].

The increase of the percentage of beta cell damage in the cluster was considered at corresponding levels of glycaemia as follows:

- 0% and 31% of dead cells at 7 mmol of glucose concentration ($k_c = 0.03$), normal value for a physiological state.
- 69% of dead cells at 11.6 mmol ($k_c = 0.062$), typical value for a mild hyperglycaemic state [8].
- 84%, 94% and 98% of dead cells at 17 mmol ($k_c = 0.1$), typical value of a severe hyperglycaemic state.

The cell loss is not random. It starts in fact at one corner of the cubic cluster configuration and spreads symmetrically towards the opposite corner of the system.

Regarding performed simulations, X-PPAUT software was adopted to integrate the model equations [27].

Results

F1 In Figure 1, the experimental evidence in mice islet of the autoimmune beta cells' destruction leading to type 1 diabetes at different stages of insulitis is presented. In panel (a), we have shown specifically a grade I insulitis with large portion of beta cells surrounded a modest infiltration of lymphocytes. In panel (b), we presented a grade II infiltratum with a very large portion of the islet surrounded by the lymphocytes. Only a small region of the insula is still recognizable by beta cells. Finally in panel (c), a grade III case is presented; here, the islet has practically almost disappeared.

This pathological scenario has been implemented in the mathematical model of beta cell electrophysiology previously introduced. The numerical simulations performed for an intact cubic cluster of $4 \times 4 \times 4$ cells showed that for $k_c > 0.02$ (corresponding to a threshold of $[Glu] = 5.5$ mmol), a regular bursting occurred.

The bursting becomes continuous for a value of $k_c \geq 0.097$ (corresponding to $[Glu] = 16.6$ mmol). These values are in agreement with experimental data in which at $[Glu] = 5.5$ mmol, bursting is not observed, whereas at 16.6 mmol, a continuous bursting starts [28].

The numerical results allowed us to understand the relationship between the progressive beta cell disruption in the pancreatic cluster and the system's electrophysiological response to the damage.

F2 In Figure 2, we showed the following:

1. The cluster (initially consisting of 64 beta cells) characterized by a progressive degree of beta cells damage (left panel).
2. The respective membrane potential versus time (central panel).
3. The associated intracellular Ca^{2+} concentration (right panel), for a generic beta cell in the cluster.

Starting from the intact cluster (0% damage), we considered a normoglycaemic state ($k_c = 0.03$ equivalent to a glucose concentration of 7 mmol) [29].

The model exhibited bursts of action potential and Ca^{2+} oscillations with a qualitative trend similar to those experimentally observed [19], namely the bursts occurred during the peak of the Ca^{2+} oscillations (Figure 2a). This was due to the gap junction coupling effect underlying the beta cell activity that determined electrical bursting and oscillatory metabolic rhythms. In particular, the intracellular calcium concentration oscillates approximately around the average value of 0.5 μM, with a frequency of 3–4/min. These values are comparable with experimental results on intact clusters [12].

As reported by Atwater *et al*. [30], in a murine intact islet immersed in 11 mmol of glucose, 6/min oscillations can be observed, whereas in the case of a 70-micron diameter cluster of murine cultured beta cells exposed to 10 mmol of glucose, 3/min oscillations are detectable [19].

Other simulations performed in this study, aimed at calibrating the model, showed that in an intact beta cell cluster, a progressive increase of $[Glu]$ led to Ca^{2+} oscillations, characterized by an increased duration without changes in amplitude, as reported in [31].

We then introduced the damage simulation protocol (31% dead cells) (Figure 2b) for a normoglycaemic state (7 mmol). Even if a portion of the cluster has disappeared, the system is biologically robust enough to maintain regular Ca^{2+} oscillations (insulin is expected to be released effectively) as reported elsewhere [5,31].

In particular, the frequency of oscillations slightly increases, although the amplitude of action potential and Ca^{2+} remains unaffected.

When the damage progression further increased (69% of damage) (Figure 2c), we imposed a glucose concentration typical of mild hyperglycaemia ($k_c = 0.062$ equivalent to 11.6 mmol of glucose) [8]. In this case, Ca^{2+} oscillations occurred; however, they were characterized by an increased frequency (~8/min) and a smaller amplitude (a reduction of 33%), with an associate effective insulin pulsatility [23]. The Ca^{2+} mean value around 0.52 μmol is in

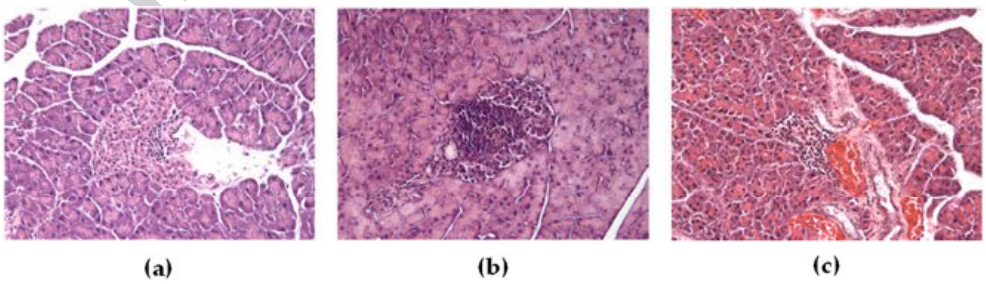


Figure 1. Experimental evidence of autoimmune progression towards type 1 diabetes on mice islets at different stages of insulitis. In panel (a), we have a grade I insulitis with large portion of beta cells (rosed cells on the centre of the picture) attached by a modest infiltration of lymphocytes (whose nuclei have very compact structure in comparison with the pancreatic tissue). In panel (b), we have instead a grade II situation with a very large portion of the islet invaded by the lymphocytes. Only a small region of the insula is still formed by beta cells with a falcate shape. Finally in panel (c), we have a grade III case; here, the islet has practically almost disappeared, invaded totally by lymphocytes

DOI: 10.1002/dmrr

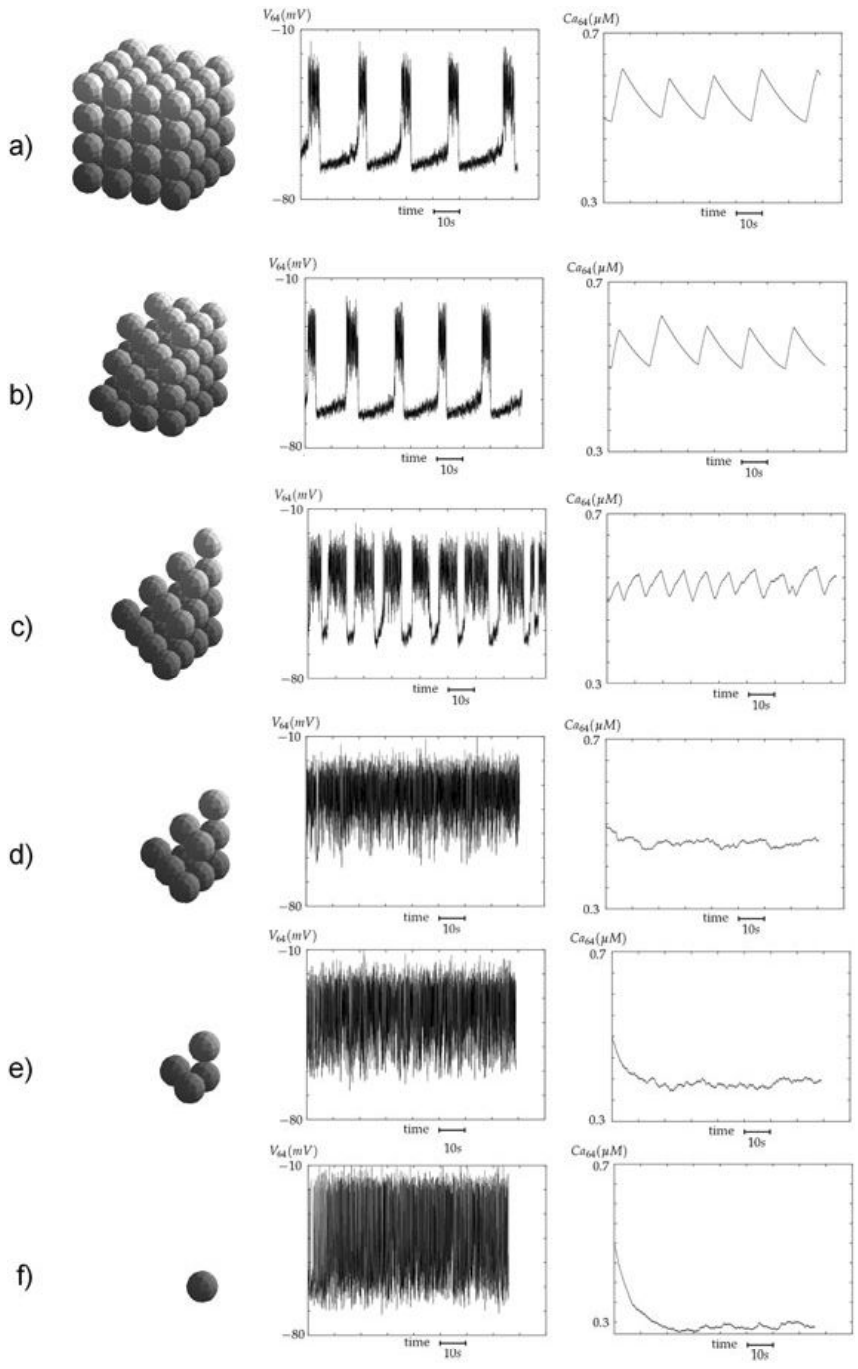


Figure 2. Effect of autoimmune progression on action potential (central panel) and intracellular Ca^{2+} (right panel) assuming a progressive percentage of dead beta cells (0%, 31%, 69%, 84%, 94% and 98%) in the modelled pancreatic cluster, on the lines of the experimental situation presented in Figure 1. Cells are represented as spheres just in contact with the neighbouring ones. This is a simplified representation introduced for pure visualization purposes because in the model, there are no isolated cells. One could equivalently represent each cell as a cube, instead of a sphere. Signals from the 64th cell are shown, although the other cells manifest a similar behaviour. Images of simulated progressive damage in a pancreatic cluster are shown on the left panel. In central and right panels, the action potential and its associated intracellular calcium dynamics are represented, respectively. Action potential is measured in mV *versus* seconds, whereas Ca^{2+} concentration is measured in μM *versus* seconds. At different degree of tissue damage, a gradual increase of glucose concentration is associated: (a) and (b) represent 0% and 31% of dead beta cells, respectively, in normoglycaemia (7 mmol, kc = 0.03). This scenario demonstrates an intrinsic biological robustness of the pancreatic cluster. (c) 69% of dead beta cells, at value of mild hyperglycaemia (11.6 mmol, kc = 0.062). The action potential and calcium dynamics-related pictures still show oscillations typical of a severely damaged system. In (d), (e) and (f), an increasing damage of 84%, 94% and 98%, respectively, is simulated for a severe hyperglycaemia (17 mmol of glucose equivalent to kc = 0.1). In these three simulations, the action potential and intracellular calcium plots show that the periodic bursting complex pattern has disappeared. In particular, Ca^{2+} oscillations are absent so that insulin pulsations are not expected to occur

DOI: 10.1002/dmrr

agreement with experimental data in the case of an intact cluster condition.

The final steps of the simulation protocol assumed an advanced regime of cellular disruption (84%, 94% and 98% of dead cells) in the beta cell cluster (Figure 2d–f). They were characterized by severe hyperglycemia ($k_c = 0.1$ equivalent to 17 mmol of glucose) [28,32].

Simulations showed that the complex periodic bursting pattern disappeared and that Ca^{2+} oscillations were absent. The mean level of intracellular Ca^{2+} was lower than in the previous cases (0.45 μmol, 0.4 μmol and 0.3 μmol, respectively). In these conditions insulin pulsatility is not expected to occur.

Following experimental data, the results obtained in the last three simulations set a framework characterized by a greatly reduced insulin release.

F3 In Figure 3 we focused instead on the effects of different values of glycaemia on the electrophysiology of beta cells for the specific case of a cluster damaged up to 84%.

The simulations showed that in case of normoglycaemia (7 mmol) (Figure 3a), an extremely damaged pancreatic tissue could leads to Ca^{2+} oscillations via cluster synchronization, although characterized by small amplitude and large frequency. Moreover, unlike the situation described in Figure 2d, for the same cluster damage (84%) in hyperglycaemic state, a normoglycaemic state is robust enough to support quasi-regular bursting phenomenology. On the other hand, a mild hyperglycaemic state (11.6 mmol of glucose) (Figure 3b) seems to permit oscillations, whereas marked hyperglycaemia (17 mmol of glucose) (Figure 3c) shows a dramatic drop of Ca^{2+} pulsatility and no organized bursting behaviour with consequent compromised beta cell secretive pattern. This result theoretically confirms the importance of maintaining normal value of blood glucose when hyperglycaemia is diagnosed to protect residual beta cells. Descriptive characteristics of our

T1 analysis are reported in Table 1.

Our modellized cluster percent damages and the associate glucose concentrations, both related to different grades of insulitis, have been chosen to be close to the experimental data aforementioned, in the limits of the model.

Discussion

In this study, we implemented a mathematical model to describe the relationship between lymphocytic infiltration, beta cell death and the insurgence of hyperglycaemia. We aimed to prove the robustness of pancreatic islets even in conditions of severe damage giving a quantitative view of the alterations in the electrochemical rhythms.

Owing to the key role of the gating probabilistic phenomenon of experimentally observed ion channel, the stochastic model of beta cell electrical excitability developed by Sherman *et al.* [17,18] was here adopted to simulate the electrochemical behaviour of a virtual cluster of beta cells under the typical autoimmune process occurring in type 1 diabetes.

We considered a heterogeneity of beta cell dysfunction by requiring that a beta cell can communicate with the surrounding alive cells but not with the cluster of dead ones. A simulation of progressive non-symmetric damage, which mimics an autoimmune progression towards type 1 diabetes, was performed. We monitored the loss of electrophysiological integrity of the cluster during reduction of beta cell number for altered glucose levels. We found that the cluster is robust enough because its bursting activity was present even when 69% of contiguous cells were dead.

Furthermore, our simulations have shown that in a cluster with 84% of dead beta cells, a normoglycaemic state guarantees Ca^{2+} oscillation patterns and action potential-bursting activity near the physiological condition, because of cellular network synchronization. For the same degree of damage, but in a hyperglycaemic state, the Ca^{2+} pulsatility was absent, and no organized bursting behaviour occurred.

We should emphasize that insulitis is a complex phenomenon in which the immune attack plays a key role than the simple destruction of contiguous beta cells with consequent interruption of cell communication. Focusing on the purely electrophysiological problem, we are aware in fact that the entire ionic dynamics of the pancreatic cells should be altered in the pathological state leading to strong heterogeneity for each cellular component of the cluster. Our analysis in this sense is a starting point for future investigations in which the electrical parameters of each cell could be different, both for statistics reasons in healthy cell population and for disease progression. To this aim, more complicated electrophysiological models should be adopted to include the missing information of the existing models; moreover, the inclusion of insulin secretion feedback on larger-sized clusters is mandatory for more quantitative representation of the complex phenomenon here addressed. Previous studies performed in animal models have shown that gap junction channels coupling beta cells are made of the Connexin36 protein and that the loss of this protein desynchronizes beta cells, leading to secretory defects in terms of recruitment of cells into insulin biosynthesis and release [33]. It has been recently demonstrated that Connexin36 is also a protein of human pancreatic islets, which mediates the coupling of the insulin-producing beta cells and therefore contributes to controlling beta cell function by modulating gene expression. Previous studies in rodents have demonstrated that the loss of the Connexin36 results in an alteration of insulin secretion resembling the pre-diabetic state [34,35]. Serre-Beiner *et al.* [33] hypothesized that if the effects observed in rodents are to be extended to humans, diabetes subjects could express lower levels of Connexin36 protein and/or decreased beta cell coupling than normoglycaemic individuals.

We cannot exclude the possibility that the same degree of damage, still with different states of glucose levels, may have different effects on the production of insulin in

DOI: 10.1002/dmrr

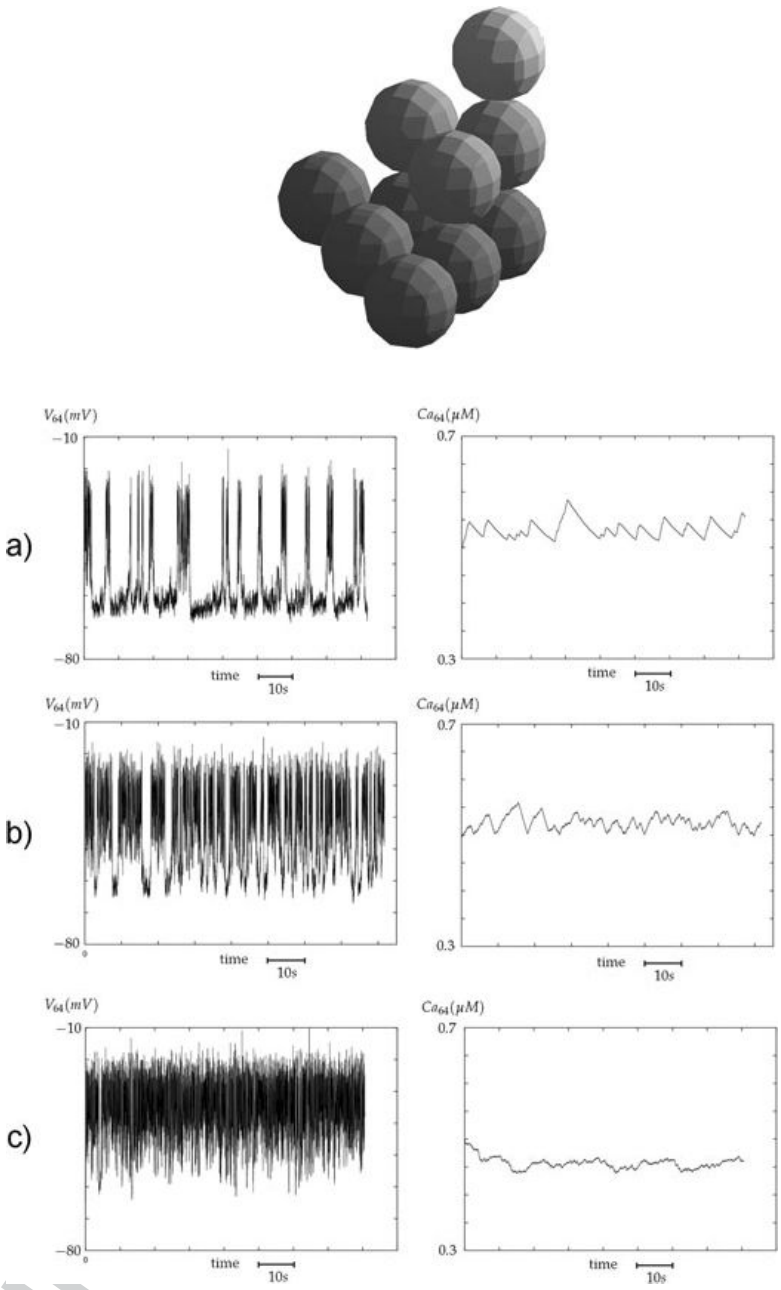


Figure 3. Effect on action potential (left panel) and intracellular Ca^{2+} (right panel) of an increasing level of glycaemia in a cluster of 64 pancreatic beta cells with 84% of dead beta cells. (a) shows that if a state of normoglycaemia is guaranteed (7 mmol), it is possible to detect Ca^{2+} oscillations because of synchronization of the remaining beta cell intact, even in a damaged pancreatic tissue. (b) and (c) show the effects on action potential and on intracellular Ca^{2+} oscillations of a mild and severe hyperglycaemic states (11.6 and 17 mmol, respectively): fluctuations occur in a mild hyperglycaemic state (b) but not in severe hyperglycaemia condition (c). In this last case, a dramatic decrease of Ca^{2+} pulsatility and a consequently severe impaired function should occur

humans and in rodents (as it has been demonstrated in this article) because the levels of Connexin36 are genetically determined [36].

In this study, we have implemented a revised version of the Chay–Keizer model [17,18] to study the behaviour of a murine pancreatic islet in physiological and pathological conditions and to investigate further the beta cells' synchronized pacemaker activity. The significant information emerging from this modelling study is that good glucose control allows a better functionality of beta cell communication and therefore a more efficient insulin secretion. This remains true even in a cluster containing residual amount of beta cells following the autoimmune process occurring in type 1 diabetes. When blood glucose concentrations tend to rise, the communication within beta cells becomes altered, and insulin release is gradually impaired.

DOI: 10.1002/dmrr

Table 1. Numerical details of the progressive damage in the modelled pancreatic cluster

% cluster damage	0	31	69	84	94	98
Glucose concentration (mmol)	7	7	11.6	17	17	17
$[Ca^{++}]$ mean value (μmol)	0.55	0.55	0.52	0.46	0.39	0.29
Number of $[Ca^{++}]$ approximated oscillations during 80 s	4	5	9	—	—	—
Number of action potential approximated bursting events during 80 s	5	5	8	—	—	—

The percentage of cluster damage, glucose concentration associated with an approximate mean value of $[Ca^{2+}]$ during oscillations, is here illustrated. Sustained oscillations in $[Ca^{2+}]$ and a discrete number of isolated bursting complexes, even in highly damaged systems, are present with a mean values of $[Ca^{2+}]$ over 0.5 μmol. When the damage is >84% and hyperglycaemia occurred, such oscillations and bursting patterns disappear and $[Ca^{2+}]$ mean values decrease. In this scenario, insulin secretion is severely compromised.

Q1 **Table 2. Model parameters**

Parameter	Model value	Parameter	Model value
f	0.001	$\bar{g}_{K\text{-}Ca}$	30 000 pS
λ	1.7	$\bar{v}_{Ca}$	110 mV
$\bar{v}_K$	−75 mV	$\bar{g}_K$	2500 pS
$\bar{g}_{Ca}$	1400 pS	C_m	5310 *f*F
V_b	−75 mV	V_n	−15 mV
V_m	4 mV	S_m	14 mV
V_h	−10 mV	S_h	10 mV
S_a	65 mV	S_b	20 mV
S_c	60 mV	S_n	5.6 mV
K_d	100 μmol	α	$1/(2V_{cell}F)$
V_{cell}	1.150 μmol^3	F	96 487 *C*/mmol

With persistent increase of glucose levels, a severe impairment of insulin secretion takes place, and diagnosis of type 1 diabetes is made. The animal models, in which single beta cell cannot longer establish intercellular communication via Connexin36 channels, show alterations in secretion and insulin gene expression that can be corrected after restoration of beta cell contacts [35,37]. These findings applied to humans underline that the implementation of insulin therapy and a restoration of normoglycaemia as soon as possible after diagnosis are of great relevance potentially allowing to re-establish an endogenous insulin secretion, leaving the patient moving into the well-known phase of clinical remission, taking off exogenous insulin for some time.

Our results suggest that at the diagnosis of type 1 diabetes, a very strict normalization of blood glucose levels should be a key action to be achieved to obtain a better preservation of beta cells. In this respect, the use of a continuous insulin infusion pump and a strict glucose monitoring should be implemented as soon as a patient is diagnosed with type 1 diabetes to control hyperglycaemia even of modest in nature. In the absence of glucotoxicity, any adjuvant therapy to cure this disease including immunomodulation or the regeneration of beta cells Q2 [38,39] can be more beneficial and, hopefully, contribute to prevent the destruction of any residual beta cell still functioning [40].

Moreover, the predictions obtained through the proposed mathematical model can be directly easily tested through the experimental setup described elsewhere [41]. By means of a specifically built biomedical device controlling an islet microfluidics, the insulin secretion produced by extracellular glucose stimuli was measured by NAD(P)H autofluorescence. Future degree studies are aimed to specifically insert in such a device islets of animals at different infiltratum, quantified *ex vivo* or *in vivo* following the magnetofluorescent nanoparticle techniques [42]. Moreover, similar fluorescence techniques, based on voltage or calcium sensitive dyes, could be adopted to monitor the underlying electrophysiology of the islets [43].

Finally, because of the importance of the right timing in shutting down insulin secretion during type 1 diabetes, beta cell regeneration could be implemented into the in silico model to further highlight how this is affected by hyperglycaemia and inflammation. Moreover, a further generalization could be achieved by coupling the discrete model here adopted to a continuous reaction–diffusion formulation. This step would be useful to realistically simulate glucose, insulin, glucagon and somatostatin dynamics by possibly including other communication pathways (i.e. direct internalization [44]) together with the here adopted gap junctions, on the lines of previous mathematical studies performed by some of the authors in other different biomedical scenarios [45–52].

Acknowledgements

This work was supported by grants from the University Campus Bio-Medico, International Centre for Relativistic Astrophysics Network and Centro Internazionale Studi Diabete, Rome, Italy.

Conflict of interest

The authors declare that there is no duality of interest associated with this manuscript.

Appendix

In this appendix, we report the expressions used in the stochastic mathematical model and the parameters' choice fine tuned on murine experiments [17]. We applied this model to a cubic cluster of 64 cells. The coupling term in membrane potential equation (1) couples the *j*th cell with the adjacent one through a Von Neumann neighbourhood type. The dynamics of the deterministic equations are coupled to a stochastic process governing the K-Ca channel kinetics. Specifically, the total conductance $g_{K\text{-}Ca}$ is multiplied by the factor p_j, which

DOI: 10.1002/dmrr

represents the number of open channel for the j cell. Here, each channel transition is the result of a Markovian process where the opening mean time is a function of Ca_i concentration with a fixed channel closing mean time.

$$C_m \frac{dV_j}{d_t} = I_{\text{ion}}(V_j, n_j) - \bar{g}_k - c_a p_j (V_j - V_k) \quad (1)$$
$$-g_c \sum\nolimits_{k\in\Omega j} (V_j - V_k)$$

$$\frac{dn_j}{dt} = \lambda \left[\frac{n\infty(V_j) - n_j}{\tau_n(V_j)} \right] \quad (2)$$

$$\frac{dCa_j}{dt} = f\left[-\alpha I_{\text{Ca}}(V_j) - k_{\text{Ca}} Ca_j\right] \quad (3)$$

$$I_{\text{K}}(V_j, n_j) = \bar{g}_{\text{K}} n_j (V_j - V_{\text{K}}) \quad (4)$$

$$I_{\text{Ca}}(V_j) = \bar{g}_{\text{Ca}} m_\infty(V_j) h(V_j) (V_j - V_{\text{Ca}}) \quad (5)$$

$$m_\infty(V_j) = \frac{1}{1 + \exp\left\lfloor (V_m - V_j)/S_m \right\rfloor} \quad (6)$$

$$h(V_j) = \frac{1}{1 + \exp\left\lfloor (V_j - V_h)/S_h \right\rfloor} \quad (7)$$

$$n_\infty(V_j) = \frac{1}{1 + \exp\left\lfloor (V_n - V_j)/S_n \right\rfloor} \quad (8)$$

$$\tau_n(V_j) = \frac{S_c}{\exp\left\lfloor (V_j - V_b)/S_a \right\rfloor} + \frac{S_c}{\exp\left\lfloor -(V_j - V_b)/S_b \right\rfloor} \quad (9)$$

$$C \underset{1/\tau_o}{\overset{1/\tau_c}{\rightleftarrows}} O \quad (10)$$

$$\tau_o = \tau_c \frac{Ca_j}{k_d}. \quad (11)$$

We have considered $N = 600$ K-Ca channels per cell and adopted a Monte Carlo simulation to evolve such a stochastic dynamics.

References

1. Orci L, Unger RH, Renold AE. Structural coupling between pancreatic islet cells. *Experientia* 1973; **29**: 1015–1018.
2. Eddlestone GT, Goncalves A, Bangham JA, Rojas E. Electrical coupling between cells in islets of Langerhans from mouse. *J Membr Biol* 1984; **77**: 1–14.
3. Michaels RL, Sheridan JD. Islets of Langerhans: dye coupling among immunocytochemically distinct cell types. *Science* 1981; **214**: 801–803.
4. Alanentalo T, Hörnblad A, Mayans S, *et al*. Quantification and three-dimensional imaging of the insulitis-induced destruction of beta cells in murine type 1 diabetes. *Diabetes* 2010; **59**: 1756–1764.
5. Rocheleau JV, Walker GM, Head WS, McGuinness OP, Piston DW. Microfluidic glucose stimulation reveals limited coordination of intracellular Ca2+ activity oscillations in pancreatic islets. *Proc Natl Acad Sci U S A* 2004; **101**: 12899–12903.
6. Bertram R, Sherman A, Satin LS. Electrical bursting, calcium oscillations, and synchronization of pancreatic islets. *Adv Exp Med Biol* 2010; **654**: 261–279.
7. Santos RM, Rosario LM, Nadal A, Garcia-Sancho J, Soria B, Valdeolmillos M. Widespread synchronous [Ca2+]i oscillations due to bursting electrical activity in single pancreatic islets. *Pflugers Arch* 1991; **418**: 417–422.
8. Dombrowski F, Mathieu C, Evert M. Hepatocellular neoplasms induced by low-number pancreatic islet transplants in autoimmune diabetic BB/Pfd rats. *Cancer Res* 2006; **66**: 1833–1843.
9. Meda P, Bosco D, Chanson M, *et al*. Rapid and reversible secretion changes during uncoupling of rat insulin producing cells. *J Clin Invest* 1990; **86**: 759–768.
10. Nittala A, Wang X. The hyperbolic effect of density and strength of inter beta cell coupling on islet bursting: a theoretical investigation. *Theor Biol Med Model* 2008; **5**: 1–13.
11. Jansen A, Homo-Delarche F, Hooijkaas H, Leenen PJ, Dardenne M, Drexhage HA. Immunohistochemical characterization of monocytes-macrophages and dendritic cells involved in the initiation of the insulitis and beta cell destruction in NOD mice. *Diabetes* 1994; **43**: 667–675.
12. Jonkers FC, Jonas JC, Gilon P, Henquin JC. Influence of cell number on the characteristics and synchrony of Ca2+ oscillations in clusters of mouse pancreatic islet cells. *J Physiol* 1999; **520.3**: 839–849.
13. Fall CP, Marland ES, Wagner JM, Tyson JJ (eds). *Computational Cell Biology*. Springer: Berlin, 2002.
14. Chay TR, Keizer J. Minimal model for membrane oscillations in the pancreatic beta cell. *Biophys J* 1983; **42**: 181–190.
15. Hodgkin A, Huxley A. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 1952; **117**: 500–544.
16. Atwater I, Rosario L, Rojas E. Properties of calcium-activated potassium channels in the pancreatic beta cell. *Cell Calcium* 1983; **4**: 451–461.
17. Sherman A, Rinzel J, Keizer J. Emergence of organized bursting in clusters of pancreatic beta cells by channel sharing. *Biophys J* 1988; **54**: 411–425.
18. Sherman A, Rinzel J. Model for synchronization of pancreatic beta cells by gap junction coupling. *Biophys J* 1991; **59**: 547–559.
19. Rorsman P, Trube G. Calcium and delayed potassium currents in mouse pancreatic beta- cells under voltage clamp conditions. *J Physiol* 1986; **374**: 531–550.
20. Bergsten P. Slow and fast oscillations of cytoplasmic Ca2+ in pancreatic islets correspond to pulsatile insulin release. *Am J Physiol* 1995; **268**: 282–287.
21. Gilon P, Shepherd RM, Henquin JC. Oscillations of secretion driven by oscillations of cytoplasmic Ca2+ as evidences in single pancreatic islets. *J Biol Chem* 1993; **268**: 22265–22268.
22. Wollheim CB, Kikuchi M, Renold AE, Sharp GWG. The roles of intracellular and extracellular Ca++ in glucose-stimulated biphasic insulin release by rat islets. *J Clin Invest* 1978; **62**: 451–458.
23. Bergsten P, Grapengiesser E, Gylfe E, Tengholm A, Hellman B. Synchronous oscillations of cytoplasmic Ca2+ and insulin release in glucose-stimulated pancreatic islets. *J Biol Chem* 1994; **269**: 8749–8753.
24. Hellman B. Pulsatility of insulin release – a clinically important phenomenon. *Ups J Med Sci* 2009; **114**: 193–205.
25. Pérez-Armendariz M, Roy C, Spray DC, Bennett MV. Biophysical properties of gap junctions between freshly dispersed pairs of mouse pancreatic beta cells. *Biophys J* 1991; **59**: 76–92.
26. Smolen P, Rinzel J, Sherman A. Why pancreatic islets burst but single beta-cells do not. The heterogeneity hypothesis. *Biophys J* 1993; **64**: 1668–1680.
27. Ermentrout B. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to X-PPAUT for Researchers and Students* (1st edn). Society for Industrial Mathematics, 2002. Q3
28. Meissner HP, Schmelz H. Membrane potential of beta cells in pancreatic islets. *Pflug Arch Eur J Phy* 1974; **351**: 195–206.