MISSION-T2D

Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data

Work Package 3

Deliverable 3.1

A probabilistic framework to account for the similarity/difference of the target patient to others in interfering variables related to the likelihood of T2D

Document Information

Grant Agreement	Nº	600803	Acronym	MISSION-T2D
Full Title	Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data			
Project URL	http://www.mission-t2d.eu			
EU Project Officer	Name	Dr. Adina Ratoi		

Deliverable	No	3.1	Title	A probabilistic framework to account for the similarity/difference of the target patient to others in interfering variables related to the likelihood of T2D
Work package	No	3	Title	Stress induced inflammation onset modelling

Date of delivery	Contractual	28.02.2014	Actual	17.03.2014
Status	Version 1.3		Final	1.4

Nature	Prototype		Report	X	Dissemination		Other	

Dissemination level	Consortium+EU	
	Public	X

Target Group (If Public)		Society (in general)	
Specialized research communities	X	Health care enterprises	
Health care professionals		Citizens and Public Authorities	

Responsible Author	Name	Pietro Liò	Partner	UniCAM
	Email	PI219@cam.ac.uk		

Version Log

Issue Date	Version	Author (Name)	Partner
26.02.2014	1.0	Pietro Liò	UniCAM
28.02.2014	1.1	Filippo Castiglione	CNR

03.03.2014	1.2	Gastone Castellani	UniBO
16.03.2014	1.3	Pietro Liò	UniCAM
15.07.2014	1.4	Pietro Liò	UniCAM

Executive Summary	In this deliverable we describe the work done in task 3.1. It consisted in particular, in developing a new ODE model for diabetes incorporating many aspects of published models and includes inflammation and (in a further extension) incorporates gut microbiota. We have also completed a first implementation of a stochastic implementation based on Prism, probabilistic model checker (commonly used to calculate probability of events and verification of conditions; it can return either the exact solution if it operates directly on the structure of the Markov chains, or an approximated solution when it measures statistically the probability to satisfy a property for a set of samples, generated using a Monte Carlo simulation of the system model).
Keywords	Ordinary Differential Equation, inflammation, comorbidity

Contents

1 Introduction and background

Over the past decade it has become apparent that as well as predisposing genetic and environmental factors, the development of diabetes (both types I and II), and other metabolic diseases, is greatly influenced by inflammatory responses. From an evolutionary perspective, it is clear that two of the most crucial functions to species survival are the ability to withstand both: periods of starvation and pathogenic attack. It is therefore unsurprising that common homeostatic molecules and pathways are able to influence both metabolic regulation and immune responses. Whilst the economic cost of diabetes worldwide provides a mounting incentive for research progression, dysfunctions of such homeostatic pathways are only recently being studied via an integrated approach. Their intrinsic complexity necessitates the use of computational models, which can be subsequently validated with experimental data. This emerging approach, known as, `computational medicine' is becoming increasingly popular as we look to create multi-scale models of living systems in disease to help develop improved therapies. One of the aims of our approach is therefore to review existing models of lymphocyte homeostasis and metabolic disorders (within the context of diabetes) and look towards producing a combined model. Diabetes Mellitus is a chronic metabolic disorder characterized by deranged insulin function. This is typically due to either disrupted production or an acquired cellular resistance to the hormone. The resultant impairment in glucose metabolism manifests as postprandial hyperglycemia, which leads to the formation of the distinctive triad of symptoms: polyuria, polydipsia and polyphagia.

Whilst the histology of the pancreas in T2D (non-insulin-dependent diabetes mellitus, NIDDM) usually appears relatively normal, the beta-cells are defective in glucose

detection, which is required for the secretion of insulin. In healthy individuals, a postprandial rise in blood glucose is followed by insulin secretion via a process called glucose-stimulated insulin secretion (GSIS). This process involves glucose transporters within the plasma membranes of the beta cells, such as GLUT-1 and GLUT-2. Insulin receptors in peripheral tissues subsequently respond by increasing their uptake of glucose, thereby normalizing blood glucose levels. T2D is hence associated with a diminished glucose transporter expression. In addition, chronic hyperglycemia (a typical trait in obesity and pregnancy) often leads to insulin resistance in the peripheral tissues and has also been implicated in the reduction of beta cell replication rates. Transition from insulin resistance to diabetes is subsequently caused partial loss of beta-cell function. It is well known too that the development of insulin resistance can be strongly influenced by the immune system. Recent research experiments focus in particular on the adaptive immune system within adipose tissue. Our normal CD4+ T-cell lymphocyte repertoire consists of the pro-inflammatory TH1 and TH17 cells and the anti-inflammatory TH2 and T regulatory cells (Tregs). The relative ratios of these different lymphocytes differ in adipose tissue depending on its location (visceral or subcutaneous) and state (lean or obese). In mice, for example, the TH1:Treg cell ratio in visceral adipose tissue changes from 1.5:1 to 6.5:1 during diet-induced obesity. In absolute numbers, this change corresponds approximately three times more accumulated TH1 cells per gram of fat. The same study observed similar changes in humans with a TH1:Treg ratio of 6:1 in lean people and 12:1 in obese people. This shift in adaptive immune cell composition from anti-inflammatory to pro-inflammatory T-cells results in the recruitment and accumulation of macrophages. The ensuing release of inflammatory cytokines also results in the phenotypic switching of the dominant anti-inflammatory M2 macrophages, to pro-inflammatory M1 macrophages, the presence of which correlates with insulin resistance.

2 Models

Experimental models, such as the non-obese diabetic (NOD) mouse and biobreeding (BB) rat, have cumulatively provided substantial insight into the mechanisms of diabetes development. Today these transgenic mice remain our most important means of testing our hypotheses experimentally. The immunological differences between mouse and man, however, inevitably impose significant limitations in the congruity of their physiologic responses to pathogenic stimuli in humans. Mathematical models, by

contrast, offer a rapid, cost-effective and precise means to collate assorted pieces of information and experimental data. It is hence vitally important that the qualitative properties of a model are congruent with well-established physiology. This can be achieved by focusing on the estimation of a few key parameters, which are crucial to reproducing biologically realistic behaviors. These mathematical models can be divided into two main groups: phenomenological and mechanistic models. Phenomenological models concentrate on empirical observations, which within the context of diabetes includes data from histological and clinical analysis As such they are data-dependent and therefore have limited predictive powers since they are confined to the original training data and prior knowledge. Mechanistic models, by contrast, aim towards a more fundamental understanding of the physicochemical and biochemical processes that underlie a particular observation. They are particularly useful for modeling across a wide variety of scales, which could range from intracellular processes, occurring within the timescales of molecular interactions (10^{-14}s), to organ level interactions occurring over months and years. This makes them incredibly powerful predictive tools for extrapolation, but much more difficult to build. They hence often rely upon complementary phenomenological models in areas where mechanistic aspects are not fully understood. Most comprehensive models of biological systems therefore typically employ some descriptive phenomenological component within the structure of an explanatory mechanistic model.

In the first year we have developed a new ODE model for diabetes that incorporates many aspects of published models [see for instance ref 1-4] and includes inflammation. An extension of this model incorporates gut microbiota. We have also completed a first implementation of a stochastic implementation based on Prism probabilistic model checker [5]; Prism is commonly used to calculate probability of events and verification of conditions; it can return either the exact solution if it operates directly on the structure of the Markov chains, or an approximated solution when it measures statistically the probability to satisfy a property for a set of samples, generated using a Monte Carlo simulation of the system model [6].

Our new model of diabetes starts from integrating different existing models. The first model that we considered was developed by Brian Topp and colleagues [1]. This is also the first model that combines beta-cell mass dynamics with glucose dynamics and insulin dynamics (see below). Glucose and insulin dynamics are fast while beta-cell mass dynamics are slow. Because of that at the beginning of the simulation beta-cell

mass can be considered as a parameter and then later, once glucose and insulin dynamics reach a steady state changes in the beta-cell mass slowly shift that steady state. There are 3 stable points that can be reached depending on the starting state which are: Beta = 300, I = 10, G = 100 stable spiral (physiological _fixed point); Beta = 37, I = 2:8, G = 250 saddle fixed point; Beta = 0, I = 0, G = 600 stable node (pathological fixed point).

If we define diabetes as persistent hyperglycemia, then there are three pathways into diabetes according to this model: 1. Regulated Hyperglycemia (moving the physiological fixed point to hyperglycemic level); there are 2 ways for this to happen:

1.1. A defect in beta-cell mass regulation. A small defect in any of the beta-cell mass parameters can cause this.

1.2. A loss of beta-cell mass regulation combined with a defect in glucose dynamics or insulin dynamics.

2. Bifurcation (eliminating the physiological and saddle fixed points). Any change in parameters that lead to elimination of physiological and saddle fixed points is considered to be bifurcation pathway to diabetes. One example is for parameter r1 to be smaller than 0.0015, then only pathological fixed point exists.

3. Dynamical Hyperglycemia (driving a trajectory across the separating cline).

The equations of this model are:

$$\frac{dG}{dt} = R_0 - (E_{G0} + S_I I)G$$

$$\frac{dI}{dt} = \frac{\beta\sigma G^2}{\alpha + G^2} - k_I I$$

$$\frac{d\beta}{dt} = (-d_0 + r_1 G - r_2 G^2)\beta$$

The variables are: G - Concentration of glucose in the blood; I - Concentration of insulin in the blood; beta is the beta-cell mass.

2.1 The Copenhagen model

It has been suggested that beta-cells are destroyed by cytokine-induced free radical formation before cytotoxic T-helper (Th)-lymphocytes and/or autoantibody- mediated cytolysis. This hypothesis is the base of this model. The model is expressed in rate equations describing the changes in numbers of beta-cells, macrophages, and Th-lymphocytes. Being concerned with the earliest events, it explores the conditions necessary to maintain self-sustained beta-cell elimination based on the feedback between immune cells and insulin-producing cells.

There are two simple versions of mathematical models outlining the interactions between the immune system and the target cells presented in this paper. The first version describes the interactions between macrophages and beta-cells [2]. The second is an extension of the first one, including the effect of Th-lymphocytes [3]. Simple system of equations that reproduce the complete Copenhagen model is as follows (they are very similar to Marinkovic [4] model, furthermore, Marinkovic model is considering the Copenhagen model as the basis and then upgrades it into a better one):

$$\frac{dM}{dt} = a + (k+b)M_A - cM - gMA$$

$$\frac{dM_A}{dt} = gMA - kM_A$$

$$\frac{dA}{dt} = lM_A + pT + qM_AT - mA$$

$$\frac{dT}{dt} = sM_AT - tT$$

The authors considered a autoimmune function of the type f(x) = min(x - min(x;E_1);E2 - min(x;E_2)); M - Amount of macrophages; MA - Amount of activated macrophages; A - Amount of beta-cells antigenic proteins; E - Autoimmune response (E_1 is the lower noud and E_2 is the higher bound value); p - Protective pathway, metabolite.

2.2 The model by Nelson et al.

This model takes into account the dynamics of functional and dysfunctional beta- cells, regulatory T cells, and pathogenic T cells. It assumes that all individuals carrying susceptible HLA haplotypes will develop variable degrees of type 1 diabetes (T1DM)

related immunologic abnormalities. The results provide information about the concentrations and ratios of pathogenic T cells and regulatory T cells, the timing in which beta-cells become dysfunctional, and how certain kinetic parameters affect the progression to T1DM. The model is able to describe changes in the ratio of pathogenic T cells and regulatory T cells after the appearance of islet antibodies in the pancreas. The intention of this model is to study the relationship between immune cells and regulatory T cells by specifically looking at the ratio of pathogenic T cells and regulatory T cells, to determine the level of beta-cell decrease after the appearance of islet antibodies in the pancreas, and to make predictions about the key parameters that are controlling this behavior prior to the clinical onset of T1DM. The key components of this model, which make it unique compared to earlier works, are its ability to track the concentration and functionality of both the beta-cells and the regulatory T cells and to quantify the concentration of beta-cells with the islet marker antibodies. Both of these aspects are critical for finding a way to better control this disease or even reverse it.

The equations of this model are:

$$\frac{dR}{dt} = \frac{a_4 R I_2}{k_2 + I_2}\left(1 - \frac{T_b + R + R_b + B_f + B_n f}{K_p}\right) - d_2 R - \sigma_r R$$

$$\frac{dG}{dt} = R_0 - G(E_{g0} + S_i I)$$

$$\frac{dI}{dt} = \frac{\sigma B_f G^2}{\alpha + G^2} - \delta_I I$$

$$\frac{dB_f}{dt} = (r_1 G - d_0 - r_2 G^2) B_f - \frac{a_1 T_b B_f}{k_1 + T_b} + \epsilon B_n f$$

$$\frac{dB_{nf}}{dt} = \frac{a_1 T_b B_f}{k_1 + T_b} - \gamma_T d_1 B_{nf} - \epsilon B_{nf}$$

$$\frac{dT_b}{dt} = S(t) + \frac{a_3 T_b}{k_4 + T_b}\left(1 - \frac{T_b + R + R_b + B_f + B_n f}{K_p}\right) - d_3 T_b - \delta T_b R$$

$$\frac{dI_2}{dt} = \frac{\rho_1 I_2 R}{k_4 + R} - \mu I_2$$

$$\frac{dR_b}{dt} = \sigma_r R - \delta_R R_b$$

where

$$S(t) = (\alpha_1 H(t - \tau_1) + \alpha_2 H(t - \tau_2) + \alpha_3 H(t - \tau_3) + \alpha_4 H(t - \tau_4))\frac{t^3}{K + t^3}$$

and H(x) is the Heaviside function; R - Normal regulatory T cells; G - Glucose concentration; I - Insulin concentration; B_f - Functioning beta-cells; B_{nf} - Dysfunctional beta-cells; T_b - Pathogenic T cells; I_2 - Cytokine IL-2; Rb - Defective regulatory T cells.

2.3 Combined Model

By applying the equations in the Marinkovic model to a system that distinguishes between functioning and non-functioning beta-cells as well as broadening the immunological scope of the model to incorporate expressions for T-cells as in Nelson the new model is created. Functional and dysfunctional beta-cells rates of change are influenced differently by glucose dynamics, activated macrophages and pathogenic T cells.

2.3.1 Insulin dynamics

Similarly to glucose, when considering time-scale of days to years insulin dynamics are relatively slow and can also be represented by a single-compartment model dependent on the rates of its secretion and clearance. Given the slow time-scale of the model we assume a clearance rate proportional to its concentration, representing its continuous uptake via the kidneys, liver and peripheral insulin receptors. The secretion function of insulin has been demonstrated to be a sigmoidal function of glucose concentration. Topp et al assumed it to be proportional to beta-cell mass [1]. An improvement made by Nelson et al was considering secretion to be proportional to only functioning beta-cells. This led to the equation that we are using:

$$\frac{dI}{dt} = \frac{\sigma B_f G^2}{\alpha + G^2} - \delta_I I$$

Where k is a clearance constant, sigma is the maximal rate at which each of the functioning beta-cells secretes insulin, G^2 /(alpha+ G_2) is a Hill function describing a sigmoid ranging from 0 to 1, reaching half its maximum at G = $alpha^{0,5}$

2.3.2 Glucose dynamics

It has been shown that when considering time-scale of days to years glucose dynamics are relatively slow and hence can be represented by a single-compartment model. This implies that glucose levels depend only upon its production and uptake. The rates of its

production and uptake depend linearly on glucose and insulin levels, which has been well established via the glucose clamp technique and hyperinsulemic clamp studies. All this resulted in a ODE first proposed by [1] that we are using, with addition of insulin resistance parameter.

$$\frac{dG}{dt} = R_0 - G(E_{G0} + S_I \frac{I}{I_R + 1})$$

Where R_0 is the net rate of production at zero glucose, E_{G0} is the total glucose effectiveness at zero insulin, and S_I is the total insulin sensitivity.

2.3.3 Beta-cell mass

Both Topp et al and Nelson et al assume that rate of formation of beta-cells is the same as their replication rate. Their replication rate depends non-linearly on the blood glucose levels. At healthy glucose levels, beta-cell replication rate increases as the glucose levels increase. Above a threshold of 400mg/ml, however, beta-cell replication rate decreases as the glucose levels increase. In the same way apoptotic death of beta-cells have been shown to depend non-linearly on the glucose levels, although it remains low above glucose concentration threshold of 110mg/ml. Another source of loss of functionality of beta-cells is the loss of the ability to produce insulin. This is assumed to be due to Th-lymphocytes (TH1 cell infiltration and their subsequent production of harmful cytokines and cytotoxin. An additional consideration made by Nelson et al, which we will keep using, is that beta-cells may became dysfunctional only temporarily. This is included into the equations by a linear function with parameter epsilon. A new thing that we will consider, that has not been considered in Nelson et al is the effects of the cytokines released by the activated macrophages. This will both increase the rates at which functioning beta-cells lose their functionality and non-functioning beta-cells are destroyed. All this makes up for the following 2 ODEs:

$$\frac{d\beta_f}{dt} = (-d_0 + r_1 G - r_2 G^2)\beta_f - \frac{\delta T \beta_f}{\tau + T} - c_2 M_A \beta_f + \epsilon \beta_{nf}$$

$$\frac{d\beta_{nf}}{dt} = \frac{\delta T \beta_f}{\tau + T} + c_2 M_A \beta_f - (\gamma_T + c_3 M_A) d_1 \beta_{nf} - \epsilon \beta_{nf}$$

Where d_0 is the beta-cell death rate at zero glucose, r_1 and r_2 are beta-cell replication rate constants, T is the concentration of pro-inflammatory pathogenic TH1 cells, delta is a constant representing the rate at which TH1 cells cause functional beta-cells to lose their functionality, c_2 corresponds to the cytokine-induced loss of functionality, d_1 marks the elevated rate at which the non-functioning beta-cells undergo apoptosis, $gamma_T$ represents the rate at which they may be directly destroyed by cytotoxic CD8+ TH1 cells and c_3 describes the rate of cytokine-induced destruction.

2.3.4 Macrophages

Macrophages enter the volume in a steady inflow rate a, and may leave the volume at a rate c proportional to their concentration. Activated macrophages are formed when antigenic proteins are taken up by the pool of resting cells. This process represented by parameter g and is dependent on both the amount of macrophages present and the concentration of antigenic proteins. We assume that the rate of formation is proportional to both quantities. The activated macrophages exist for limited amount of time and then reverse back to inactive form, represented by parameter k. Activated macrophages release cytokines, IL-1 and TNF, that are used as signal molecules during immune responses and will direct other macrophages to enter the volume, represented by parameter b.

$$\frac{dM}{dt} = a + (k+b)M_A - cM - gMA$$

$$\frac{dM_A}{dt} = gMA - kM_A$$

2.3.5 Beta-cell antigenic proteins

The deaths of beta-cells leads to the release of beta-cell antigenic proteins. We will assume that on average, upon its death, each beta-cell releases n proteins which brings us to the equation:

$$\frac{dA}{dt} = -n\beta$$

Where beta represents the total beta-cell population (i.e. beta f + beta nf).

2.3.6 Th-lymphocytes

Th-lymphocytes are stimulated to proliferate when they recognize MHC-peptide complexes. We assume that their formation rate is proportional to both their concentration and the concentration of activated macrophages. The concentration of Th-lymphocites will be reduced in the absence of stimulation, represented by parameter d_2.

$$\frac{dT}{dt} = hM_AT - d_2T$$

Where h represents the Th-lymphocytes proliferation rate.

2.3.7 Insulin Resistance and mTOR

This model also incorporates 2 new equations representing insulin resistance (IR) and mTOR. IR is caused by mTOR, but IR needs to be able to decrease as well, that is why there is "-IR0" i.e. it should be possible to decrease as well. IR's lowest possible value is 0 meaning that there is no resistance in the body to insulin whatsoever. The larger values of IR represent bigger resistance to insulin. mTOR is activated by both insulin (which effect is actually decreased by IR) and glucose.

$$\frac{dI_R}{dt} = -I_{R0} + mM_{TOR}$$

$$\frac{dM_{TOR}}{dt} = -m_0 + c_4\frac{I}{I_R + 1} + c_5G$$

3 Formulation of a new Inflammatory-gut microbiota-diabetes Model

Considering then, a new model, which although is built upon the developments of previous models, can still be derived from first principles. Here we summarize the development of each of the rate equations used in the proposed new model.

The complete list of model equations is:

$$\frac{dI}{dt} = \frac{\sigma B_f G^2}{\alpha + G^2} - \delta_I I$$

$$\frac{dG}{dt} = R_0 - G(E_{G0} + S_I \frac{I}{I_R + 1})$$

$$\frac{d\beta_f}{dt} = (-d_0 + r_1 G - r_2 G^2)\beta_f - \frac{\delta T \beta_f}{\tau + T} - c_2 M_A \beta_f + \epsilon \beta_{nf}$$

$$\frac{d\beta_{nf}}{dt} = \frac{\delta T \beta_f}{\tau + T} + c_2 M_A \beta_f - (\gamma_T + c_3 M_A) d_1 \beta_{nf} - \epsilon \beta_{nf}$$

$$\frac{dM}{dt} = a + (k + b) M_A - cM - gMA$$

$$\frac{dM_A}{dt} = gMA - kM_A$$

$$\frac{dA}{dt} = -n\beta$$

$$\frac{dT}{dt} = hM_A T - d_2 T$$

$$\frac{dI_R}{dt} = -I_{R0} + mM_{TOR}$$

$$\frac{dM_{TOR}}{dt} = -m_0 + c_4 \frac{I}{I_R + 1} + c_5 G$$

The variables are:

G - Glucose concentration

I - Insulin concentration

$beta_f$ - Functioning _-cells

$beta_{nf}$ - Dysfunctional beta-cells

M - Amount of macrophages

M_A - Amount of activated macrophages

A - Amount of _-cells antigenic proteins

T - Pathogenic T cells

IR - Insulin resistance

MTOR - mTOR

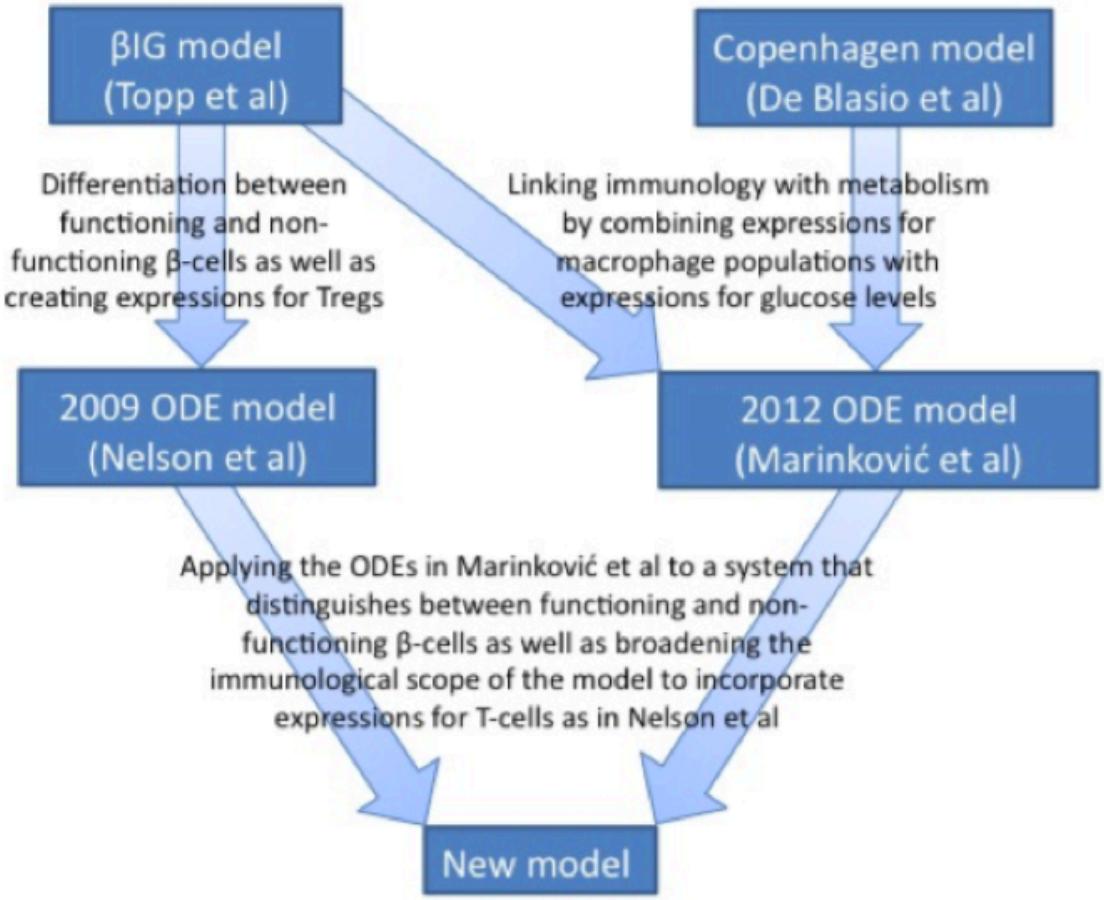

Figure 1. A diagram summarising the development of the discussed models of diabetes and their relationships to one another.

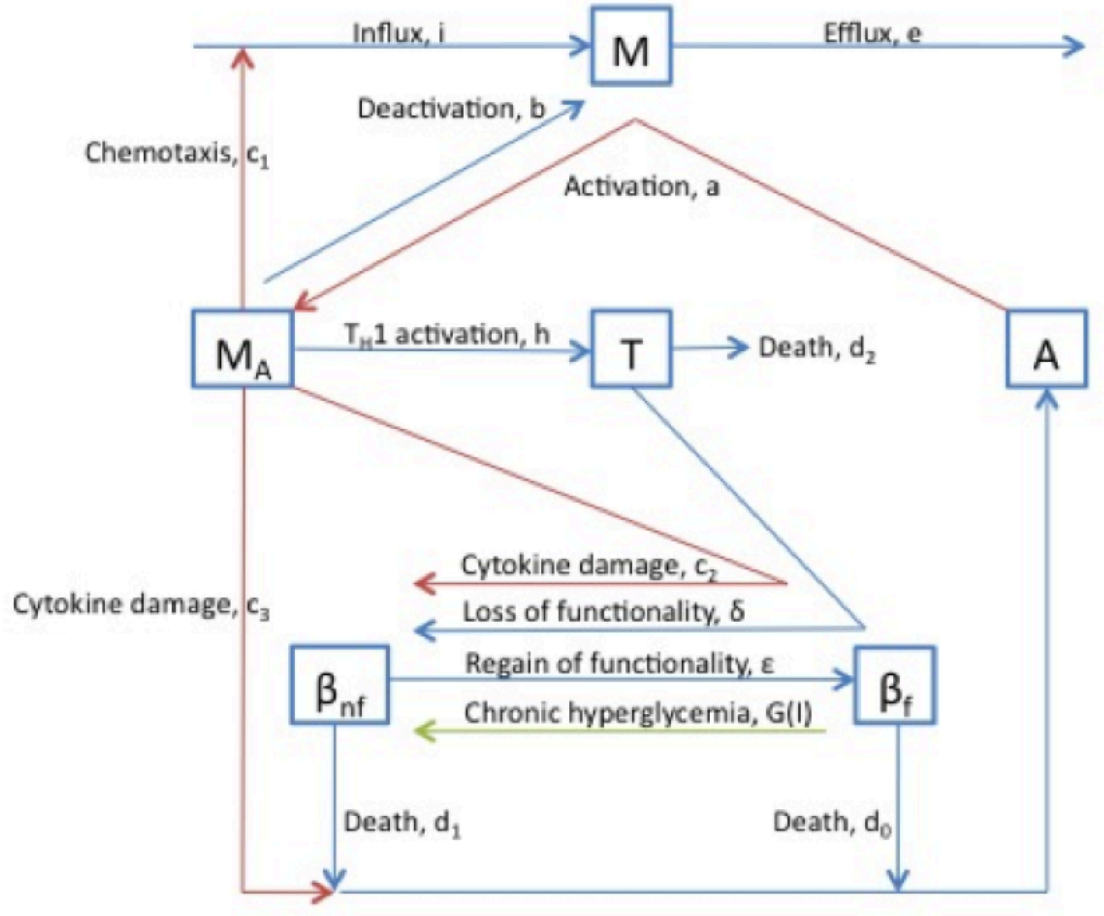

Figure 2. A schematic representation of the new model where the red lines represent one species causing a rate of change in another (i.e. an indirect effect, such as via the action of cytokines), the blue lines represent a rate of change of one species causing a rate of change in another (i.e. a direct effect) and the green line represents the influence of blood glucose and insulin dynamics, as in the betaIG model. The extension will include a gut microbiota – Treg interface.

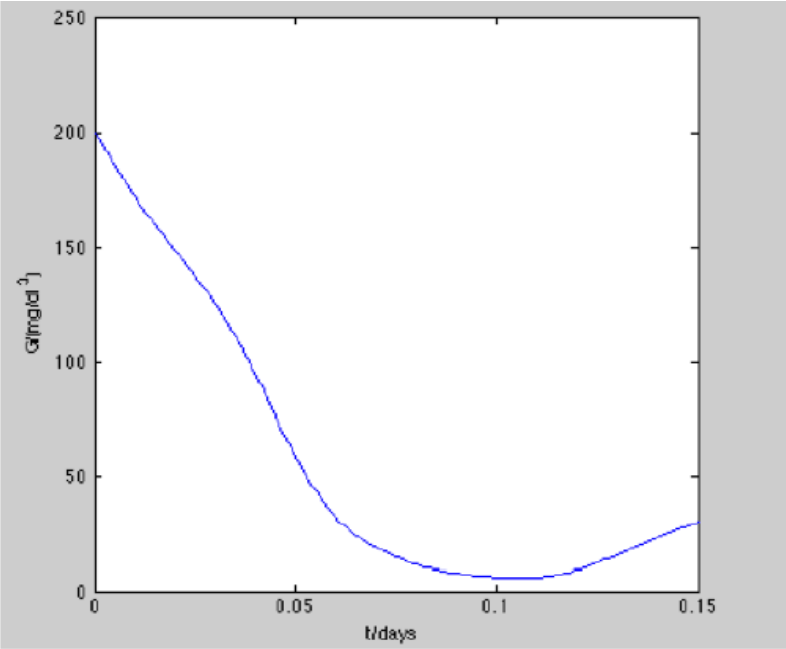

Figure 3. Graphs demonstrating postprandial glucose dynamics.

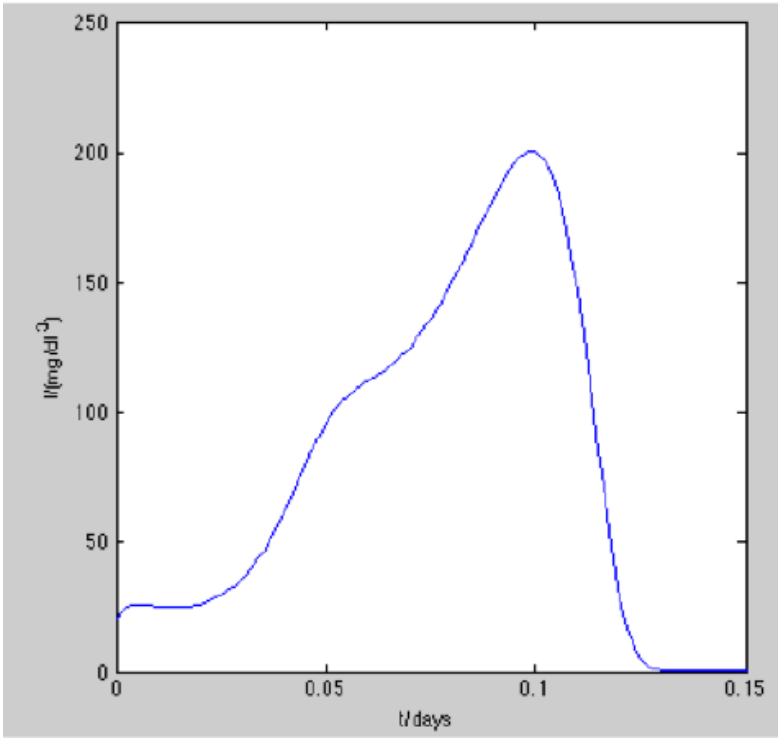

Figure 4. Graphs demonstrating postprandial insulin dynamics, as expected the reduction in glucose level is paralleled by an increase in blood insulin

3.1 Extending the model to gut microbiota

Up to now we have linked the diabetes with the inflammatory (mTOR) and the immune responses (T cells). The immune system has multi scale properties with a wide diversity of effector modalities, whether molecular or cell mediated, innate or adaptive. There is a tight link between inflammatory and immunological responses. The inflammatory response is essential for the recruitment and activation of lymphocytes in order to respond to infection and then the promotion of wound healing and repair. There will always be some amount of collateral tissue damage during a response to a viral or parasite infection; the aim is to limit this damage to just the amount needed to effectively clear (or maintain chronic stasis of) the pathogen. The inflammatory response must be constantly constrained in intensity and duration to prevent molecular, cellular and organ damage. The consequences of unregulated inflammation are associated with many acute and chronic autoimmune diseases and comorbidities. The adaptive immune response is controlled (suppressed) by regulatory T (TReg) cells, mostly directed by the transcription factor Forkhead box P3 expressing (FOXP3+). The expression of FOXP3 is necessary to preserve the Treg cell program. Recent experiments show that both the high levels of FOXP3 and an epigenetically modified FOXP3 locus are associated with suppressive function in both human and mouse CD4+ T cells. Other transcription factors may also be involved. FOXP3+ Treg cells prevent potentially damaging autoimmune and protective immune responses, so the number of Treg cells is a crucial determinant to build an appropriate response and avoid autoimmune (too few Treg cells) or immune suppression (too many). The avoidance of failure to maintain appropriate Treg cell numbers is reached through a dynamic homeostatic processes, through Treg amplification and apoptosis. Although Treg cells seem to have overall homeostatic properties, they adapt to different tissues and to different immune conditions. In particular they can be distinguished between central and tissue polarised. Treg central cell population has circulatory characteristics similar to naive conventional CD4+ T cells, and polarized tissue-resident Treg cell populations, which are present in most organs. The high sensitivity of TRegs to a range of signals makes them readjust and relocate accordingly to changes in the body immune conditions. An important tissue is the gut microbiota. Mammals live in partnership with a rich commensal flora on many of the body epithelial surfaces (such as the skin and the upper and terminal gastrointestinal tract). This partnership is critical for tissue formation, metabolism and both the development and function of innate and adaptive resistance. The gut flora not only has a role in local tissue formation and the development of mucosal immunity, but also the quantitative or qualitative alteration of

its composition (by treatment with antibiotics or other causes of dysbiosis) affects systemic immunity. In mice, the balance between TH17 cells and Tregs cells in the intestine can be determined by the presence of a specific bacterial species, the segmented filamentous bacteria, which are able to associate with the mucosa of the distal small intestine and to specifically induce TH17 cells in the gut.

The extended model considers Treg lymphocytes in the periphery and central body districts. The local and systemic homeostasis are perturbed by the gut microbiota. The gut is modeled by the production of butyrate. The bacteria's behavior is in turn modulated by circadian periodical pattern according to a birth and death model. The model takes into account the dynamics of functional and dysfunctional central and peripheric Treg cells. The intention of this model is to study the relationship between immune cells and regulatory T cells by specifically looking at their local and systemic homeostatic properties. The new model considers different cell populations, endogenous IL-2 and is defined as follows: T(t) is the total inflammation; N(t) is the concentration (cells/L) of Natural Killer (NK) cells per liter of blood; L(t) is the concentration (cells/L) of CD8+ T cells per liter of blood; R(t) is the concentration (cells/L) of CD4+ CD25+ regulatory T (Treg) cells per liter of blood; C(t) is the concentration (cells/L) of lymphocytes per liter of blood, not including NK cells, CD8+ T cells, and regulatory T (Treg) cells; I(t) is the concentration (IU/L) of IL-2 per liter of blood.

The final model will have also an implementation on Prism, a Model Verification software, which is state of art probabilistic model checking technique [5]. (work in progress)

4 Parameter estimation

Parameters from complex multi factorial diseases come from both the clinical observation and measurements and the available genetic data. Here we describe the essence of parameter estimation and provides some description of a software to integrate clinical and bioinformatics data whose related paper has been submitted to BMC clinical Bioinformatics.

Genome-wide association studies (GWAS) have provided further insight into this link between adipose tissue inflammation and insulin resistance. Recent eGWAS studies

are discovering association from the machine learning analysis of a multitude of experimental results. For example the analysis of 130 independent microarray experiments has led to the identification of key genes implicated in the pathogenesis of T2D. The `top candidate' was the immune-cell receptor, CD44, which is highly expressed in the inflammatory cells of obese adipose tissue. Further experimentations in the same study also demonstrated the significant correlation of CD44 serum levels with insulin resistance and glycemic control. This points to a potential future therapy for T2D could hence involve anti-CD44 antibodies.

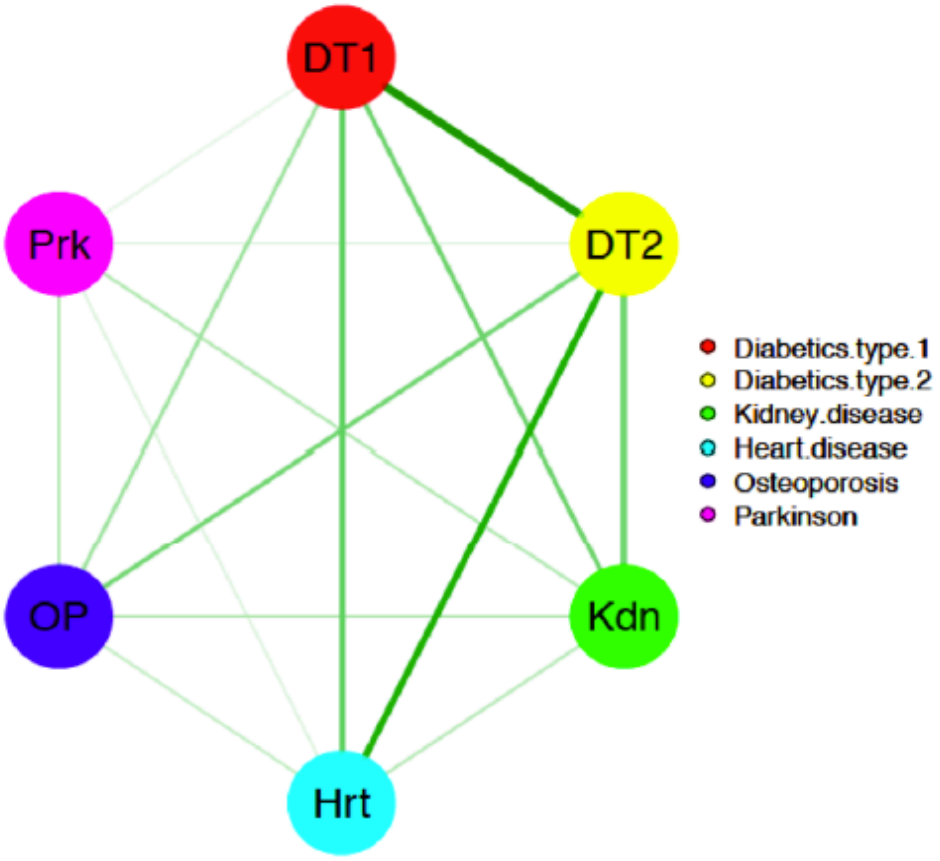

Figure 5. A diagram demonstrating the comorbidity association, i.e. how gene expression data shows how type I and type II diabetes are connected to 6 other diseases: T1D, T2D, Osteoporosis (OP), Kidney disease (Kdn), heart disease (Hrt) and Parkinson's disease (Prk). The thickness of the interconnecting lines represents the correlation (comorbidity). 5 different gene expression data sets were used for this analysis (GSE7158, GSE7621, GSE9006, GSE9128 and GSE 15072), all of which are publicly available from the online GEO database. The gene expression analysis behind this image has led to identify specific up- and down- regulated genes and estimate the parameters.

The tables 1 and 2 below report the parameters used in our study. Where a range has been given for the parameter values in tables 1,2, these can be used flexibly to tailor to

specific needs of a patient. Insulin production, for example, will be at the low end of the range for patients with fully developed diabetes, at the high end for patients without diabetes and mid-range for someone in the process of developing diabetes. It will also be dependent upon the patients size, as well as other factors, such as alcohol consumption. Many of these varying parameters, can hence be clinically defined and tailored to biologically and clinically relevant values. Others, however, such as tau, are simply mathematical constants, which need to be carefully evaluated, as demonstrated in figure below. Matlab was useful for doing this automatically, by adding a small variable to the list of parameters.

Table 1: Summary of Variables

Variable	Definition	Units	Values
G	Glucose	$mgdl^{-3}$	$(0-400)$
I	Insulin	$mgdl^{-3}$	$(0-40)$
M	Macrophage	Cells dl^{-3}	$(0-50)$
M_A	Activated macrophage	Cells dl^{-3}	$M_A(0)=0$
β_f	Functioning β-cell	Cells dl^{-3}	$(0-500)$
β_{nf}	Non-functioning β-cell	Cells dl^{-3}	$\beta_{nf}(0)=0$
β	Total β-cell population	Cells dl^{-3}	$(0-500)$
A	β-cell Antigenic protein	$mgdl^{-3}$	$A(0)=0$
T	T_H1 lymphocyte	Cells mm^{-3}	$T(0)=0$

Table 2: Summary of Parameter Values

Parameter	Definition	Units	Values
R_0	Glucose production	$mgdl^{-1}d^{-1}$	$(500-1000)$
E_{G0}	Glucose effectiveness	d^{-1}	1.44
S_I	Insulin sensitivity	$mgdl^{-3}d^{-1}$	0.72
σ	Insulin production	$mg(\text{Cells } dl^{-3}d)^{-1}$	$(20-50)$
α	Half-saturation constant	mg^2dl^{-2}	20000
k	Clearance constant	d^{-1}	432
i	Macrophage influx	cells d^{-1}	(10^2-10^5)
e	Macrophage efflux	cells d^{-1}	(10^2-10^5)
a	Macrophage activation	cells d^{-1}	$(0-5)$
b	Macrophage deactivation	cells d^{-1}	$(0-5)$
c_1	Cytokine induced macrophage inflow	cells d^{-1}	10^2
c_2	Cytokine induced loss of β-cell functionality	d^{-1}	$(0-5)$
c_3	Cytokine induced β-cell death	d^{-1}	$(0-1)$
d_0	β-cell death	d^{-1}	0.06
d_1	Glucose-dependent β-cell death	d^{-1}	$(0-1)$
d_2	T-cell death	d^{-1}	$(0-1)$
r_1	β-cell replication rate constant	$mg^{-1}dld^{-1}$	0.84×10^{-3}
r_2	β-cell replication rate constant	$mg^{-2}dl^2d^{-1}$	0.24×10^{-5}
δ	T_H1 induced loss of β-cell functionality	d^{-1}	$(0-5)$
ϵ	Regain in β-cell functionality	d^{-1}	$(0-1)$
γ_T	β-cell death by CD8+ T-cells	[no units]	$(0-1)$
τ	Half-saturation constant	Cells	$(100-1000)$
h	T_H1 activation	cells d^{-1}	$(0-5)$
n	Antigenic proteins released by dying β-cell	$d^{-1}\text{cell}^{-1}$	$(0-100)$

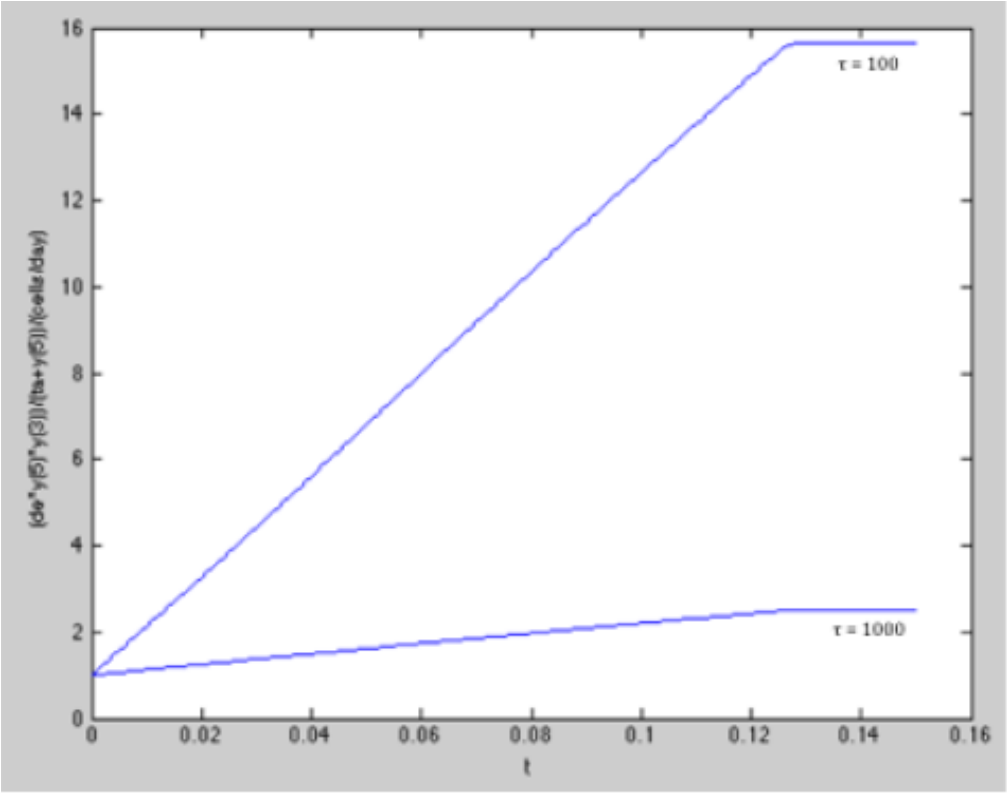

Figure 6. A graph to demonstrate the effect of the estimation of tau.

4.1 Tool for Parameter estimation and Diabetes comorbidities

We have also investigated general methodologies for parameter estimation from clinical and microarray data by developing a software in R that integrates clinical, gene ontology and molecular data, in particular transcriptomics data and outputs a comorbidity profiles (paper in second revision at BMC Clinical Bioinformatics). This software uses network regression and survival analysis data. development of statistical (causality inference through decision trees) and mathematical models able to identify trajectories deviations from healthy aging as well as disease progression.

This software is now in development towards diabetes, inflammation and immunology. At a certain stage of development, through a concertation with the other WPs, it will input specific behavioral information and will output a comorbidity profile with the full comorbidity of a disease, including exercise and diet.

The diagnosis of comorbidities, which refers to the coexistence of different acute and chronic diseases, is di_cult due to the modern extreme specialization of physicians. We envisage that software dedicated to comorbidity diagnosis could result in an effective

aid to the health practice. We have developed an R software comoR to compute novel estimators of the disease comorbidity associations.

Starting from an initial diagnosis of diabetes, genetic and clinical data of a patient the software identi_es the risk of disease comorbidity. Then it provides a pipeline with different causal inference packages (e.g., pcalg, qtlnet etc) to predict the causal relationship of diseases. The input of this software is the initial diagnosis for a patient and the output provides evidences of disease comorbidity mapping.

The comoR provides a number of processing options to find comorbidity of a disease. R bioconductor annotation data packages \org.Hs.eg.db" and \DO.db" are used for the annotation and mapping between gene symbol, Entrez id, OMIM (Online Mendelian Inheritance in Man) id and DO (Disease Ontology) term ([9]). comoR is also dependent on \DOSE" bioconductor package for the mapping of DO and DOLite. A set of di_erential expressed gene symbols/Entrez ids/OMIM id/3 or 5 digit ICD-9-CM code of the disease can be used as input of ComoR functions. Flow diagram of the comoR software is shown in the figure 10.

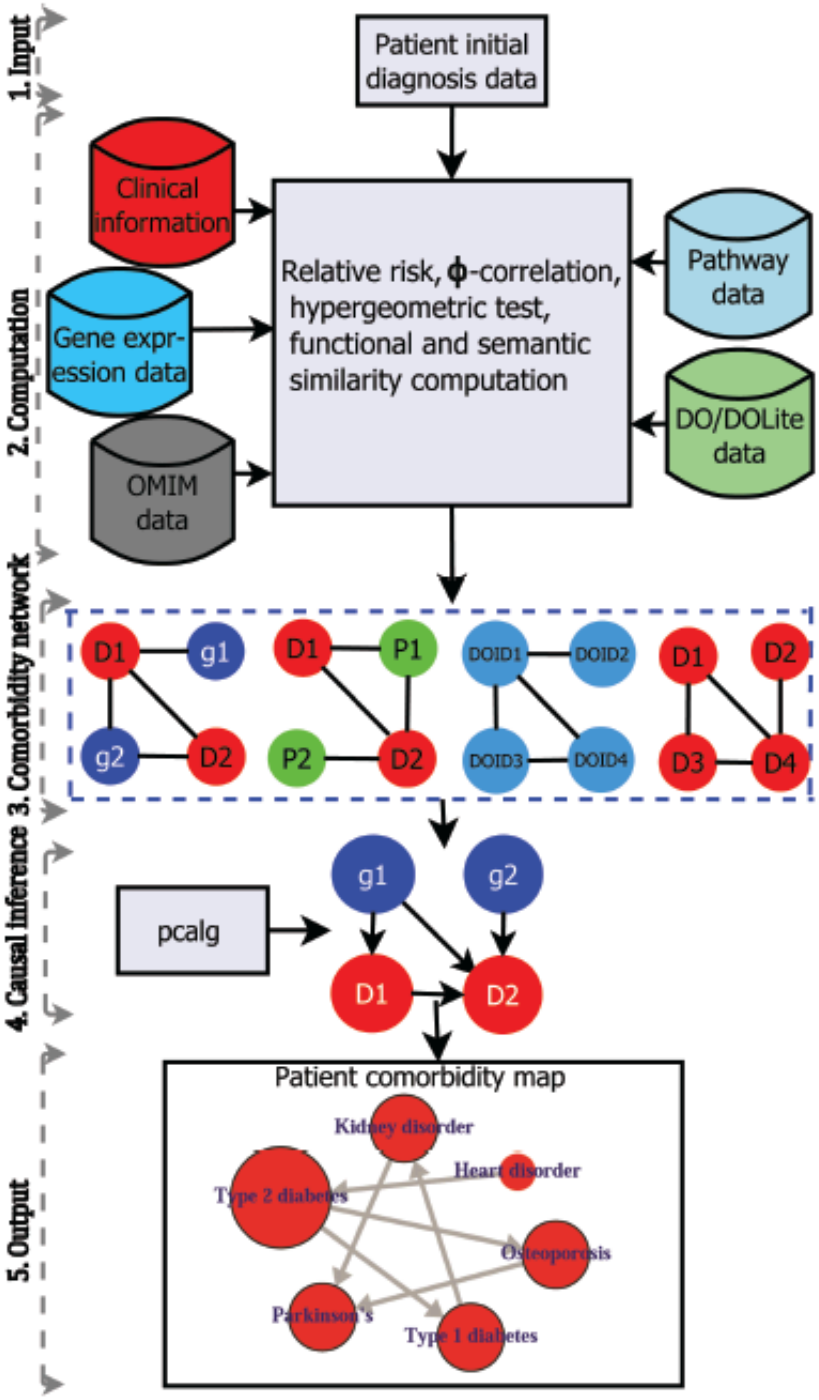

Figure 7. Flow diagram of the comoR software. Step 1: comoR takes as input preliminary diagnosis data of a patient. Step 2: It preprocesses and updates required databases, performs statistical computation (hypergeometric and semantic similarity tests), and calculates relative risks and beta-correlation (Pearsons correlation for binary variables) between diseases. Step 3: Comorbidity scores and disease network are provided as a result to the user. Step 4: Causal inference graphical models with the R package pcalg. Step 5: Visualization of the comorbidity map. This map could be extended to incorporate diet and exercise. Symbols D, g, P and DOID are used to indicate disease, gene, pathway and disease ontology id respectively.

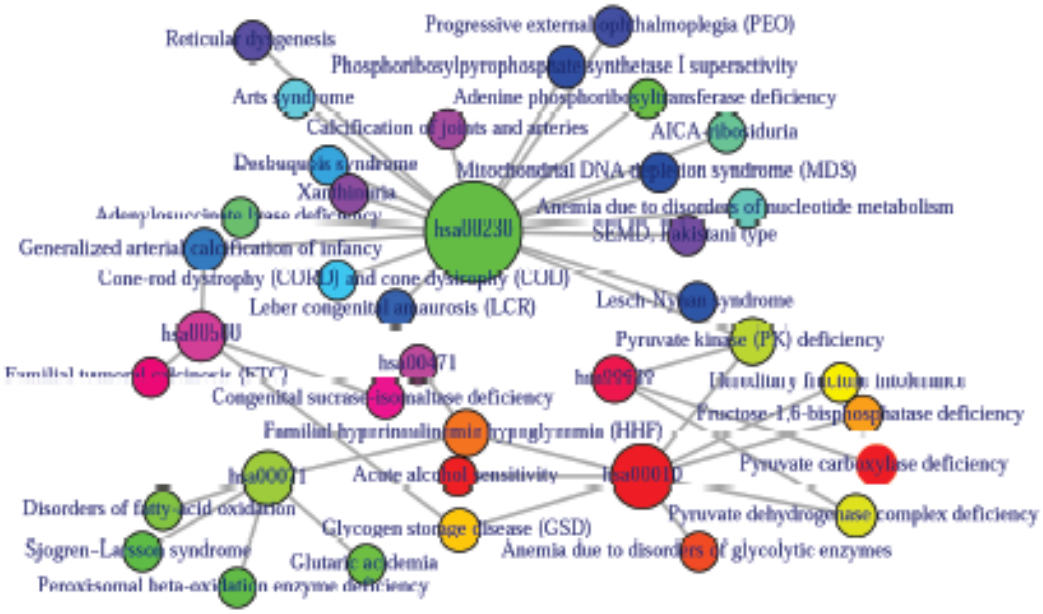

Figure 8. Output _gure of >comorbidityPath("00010", "Pathway"). The kegg pathway id 00010 is used as input to the comorbidityPath. We show disease comorbidity for the pathway "00010" through the pathway disease associations.

```
> comorbidityDO("DOID:9352", "DOID")
                                  type 2 diabetes mellitus
carbohydrate metabolism disease                 0.6083564
acquired metabolic disease                      0.5174153
diabetic peripheral angiopathy                  0.8387095
lipoatrophic diabetes                           0.8387095
gestational diabetes                            0.6730764
prediabetes syndrome                            0.6730764
neonatal diabetes mellitus                      0.6730764
diabetic ketoacidosis                           0.8387095
glucose metabolism disease                      0.7085008
hyperglycemia                                   0.5499992
diabetes mellitus                               0.8333330
type 1 diabetes mellitus                        0.6730764
```

Figure 9. This is a typical output generated by the implemented library. Gene set enrichment analysis are used for predicting the significance of gene-disease and disease-disease associations. The comorbidityDO function operates by using either of the following input: DO id, a list of gene symbols or Entrez gene ids of the patient sample. This function provides disease comorbidity associations and network based on the DO and DOLite. comorbidityDO requires two parameters id list and id type.

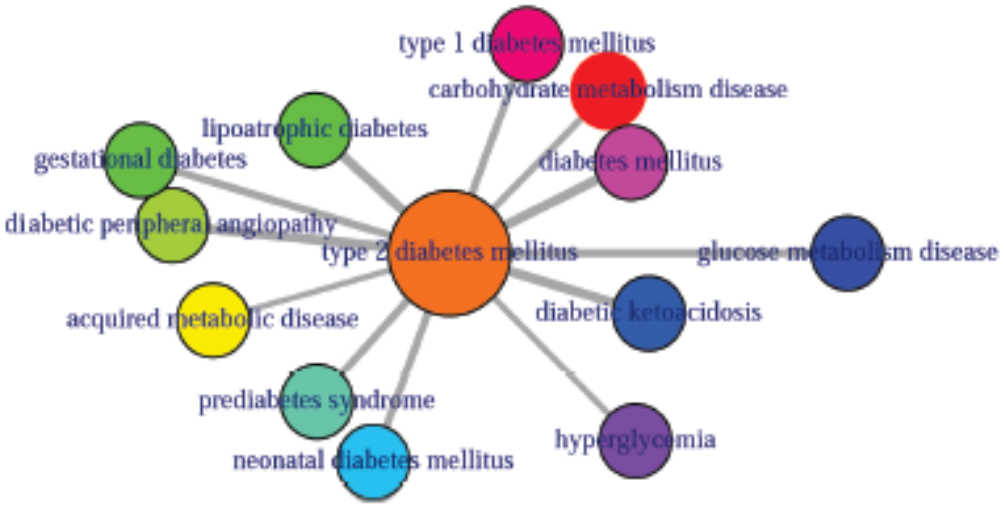

Figure 10. Output of the library comorbidityDO("DOID:9352" , "DOID"). The DO id of the type 2 diabetes mellitus is DOID:9352, which is used as input to the library called comorbidityDO. We show disease comorbidity for the type 2 diabetes mellitus using the disease ontology.

This library contains modules for causality, which could be used to refine medical diagnosis. A case study application on T2D diabetes is being prepared. This will provide further refinement of the model parameters particularly on combining immune, inflammatory, diabetes and gut microbiota comorbidities. A further extension of this library for testing parameters involved in diabetes will include Bayesian decision tree.

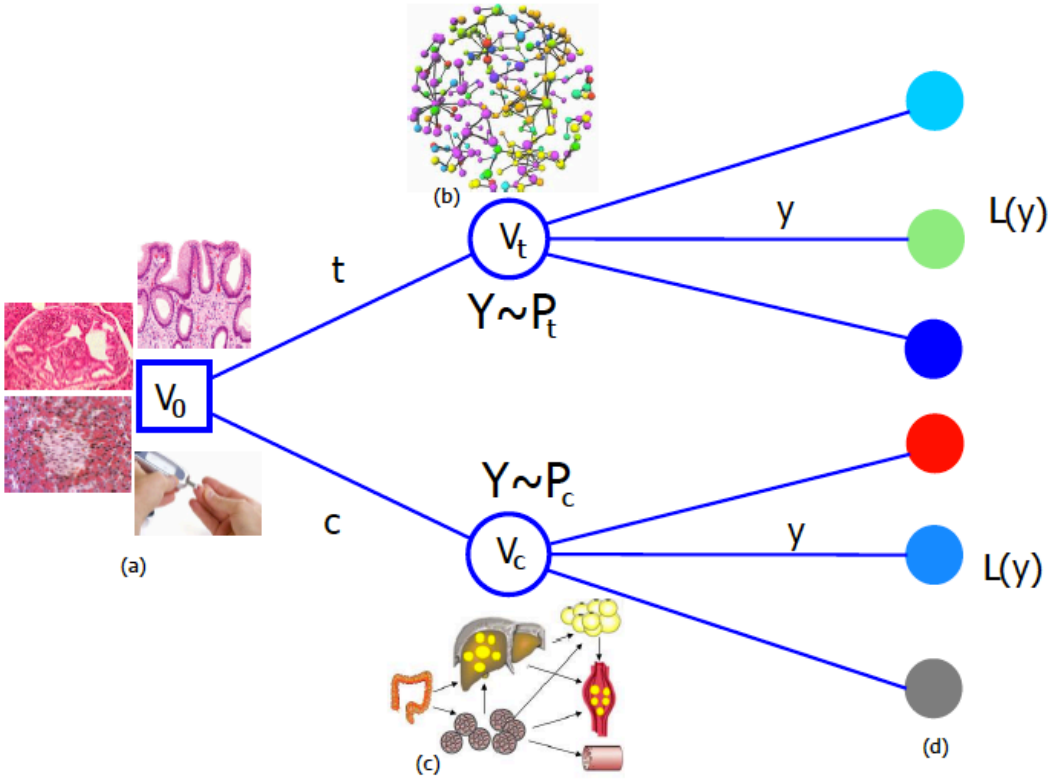

Figure 11. A Bayesian decision tree for the diagnosis (Y) of diabetes comorbidities (L(y)).

5 Estimation of diabetes ODE model parameters depending on others comorbidities (work in progress)

In order to estimate population and longitudinal variances in the parameters related to the ODE models we are devising, we thought to use methods based on networks, MCMC and in general machine learning techniques. Here we provide some explanations of the methodologies we are currently applying.

In complex diseases such as T2D, the signature genes comprise dense protein-protein interaction subnetworks, enriched by extracellular matrix receptors and modulators or by nuclear signaling components downstream of extracellular signal-regulated kinases. Groups of genes are co-expressed under certain conditions or their protein products interact with each other to carry out a biological function. It has been shown that protein-protein interaction network or co-expressions can provide useful prior knowledge to remove statistical randomness and confounding factors from high-dimensional data for several classification and regression models A systems medicine approach overrides the limitation of unifying mechanistic hypothesis and comparisons of single variables, reflecting instead the more realistic situation in which locally and globally connected communication circuits occur between constitutional, genetic, socio-cultural, cognitive, clinical and therapeutic factors. In comorbidity maps, all such entities cause associations and interactions which may be clinically justified but hard to detect when observing the system at a coarse granularity. Through observation, testing, validation, the learning environment can exclude interferences and assess deviations and anomalies of unpredicted size. A significant reduction of complexity can be obtained by assembling data, which embed a variety of signals and patterns seen not in isolation but through ensembles, then testing models on suitably designed training data, and finally validating them on unseen patients. The previous process is highly computational, and requires a multiplexed inference methodology to bridge between observational, experimental and computational fields. The word multiplexing has been first associated with experiments, and its translation into computations has the function of reconstructing complex pathways through the hierarchy of their component activities. In our applications, multiple sources of (semi-) quantitative evidence lead to meta-analysis in view of selected variables defining profiles based on characterized phenotypes. Some variables form the information bulk significant to the analysis and spanning the patient's disease space (a sort of minimal set satisfying model

consistency), while other variables provide risk information useful to assess dynamics (transition between profiles). In particular, stratified patient groups can be generated from the data assemblies based on stylized comorbidity profiles, including distinct phenotypes identified from the available patient cohorts. Generating processes for varieties of data and types of evidence determine the systems' dimensions. Clinical records are of course a main source, likewise other measurements useful for patient profiling. Time drives the inference dynamically based individual (i.e. patient-centered) and ensemble (i.e. patient-class based) comorbidity maps. Once patients' records are processed through standardized measurements and scoring systems, post-processing allows for the quantification of the relationships between comorbidity variables. Networks provide a very natural representation of knowledge underlying complex systems. A disease state can be considered a perturbation to networks depending on the value that is assigned to it. The dissection of the internal characteristics of a network with regard to both topological structure and dynamics, can allow the elucidation of the impact of the perturbation on the complex regulatory interactions occurring during disease pathogenesis. As disease states can vary, and perturbations differentiate their potential impact at network scale, the underlying time-scales can be very different, and often hard to integrate within a unifying frame. Networks have a double signature of complexity, namely emergence and latency. The phenotype intended as the macroscopic response of a system results from the propagation of information across interacting network parts or modules. The complexity that emerges across such coordinated modules is partially disclosed (through observable structures such as protein complexes), and partially latent (nested, convoluted, such as regulative paths). Network resilience ensures the possibility of tackling complexity by integratively coordinated modular functioning. In our scheme network approaches will be implemented with Bayesian decision tree. Decision tree learning is a popular approach for classification and regression in machine learning and statistics, and Bayesian formulations, which introduce a prior distribution over decision trees, and formulate learning as posterior inference given data- have been shown to produce competitive performance. Unlike classic decision tree learning algorithms like ID3, C4.5 and CART, which work in a top-down manner, existing Bayesian algorithms produce an approximation to the posterior distribution by evolving a complete tree (or collection thereof) iteratively via local Monte Carlo modifications to the structure of the tree, e.g., using Markov chain Monte Carlo (MCMC). In classical decision tree learning, a decision tree (or collection thereof) is learned in a greedy, top-down manner from the

examples. Examples of classical approaches that learn single trees include ID3, C4.5 and CART, while methods that learn combinations of decisions trees include boosted decision trees, Random Forests, and many others. A decision tree can be represented more compactly as an influence diagram, focusing attention on the issues and relationships between events. Bayesian decision tree methods cast the problem of decision tree learning into the framework of Bayesian inference (see figure 11). In particular, Bayesian approaches start by placing a prior distribution on the decision tree itself. To complete the specification of the model, it is common to associate each leaf node with a parameter indexing a family of likelihoods, e.g., the means of Gaussians or Bernoullis. The labels are then assumed to be conditionally independent draws from their respective likelihoods. The Bayesian approach has a number of useful properties: e.g., the posterior distribution on the decision tree can be interpreted as reflecting residual uncertainty and can be used to produce point and interval estimates. On the other hand, exact posterior computation is typically infeasible and so existing approaches use approximate methods such as MCMC in the batch setting. The work in progress is making use of the Comor software library (Figure 10) and the decision tree (Figure 11) for inferring how the different parameters used in the models are changed by the presence of gut microbiota, immune systems and inflammatory comorbidities.

6 Phenotype (multi) omics gene ontology (POGO)

Exploring associations among diseases at the molecular and clinical levels could greatly facilitate our understanding of pathogenesis, and eventually lead to better diagnosis and treatment [6.1]. Many researchers have studied the relationships between disease and biological variations such as single nucleotide polymorphisms (SNPs), copy number variations (CNVs), sequence repeats and genetic rearrangement [6.2]. For an instance, several CNVs were significantly associated with the risk of T2DM in a Korean population [6.2]. Gene Ontology (GO) is also helpful to promote the investigation of diseases and disease risk factors [6.3]. Moreover, phenotype and environmental factors should also be considered in order to realize disease association [6.4]. Therefore, development of methods and software tools integrating genetic and clinical data will assist clinical decision making and represent a key step for translation medicine.

We have developed POGO, an R package that implements different statistical approach for the prediction of disease comorbidity using divers set of data (CNVs, GO, SNPs,

miRNA, HPO and environmental). Starting from an initial diagnosis, genetic and clinical data of a patient the software identifies the risk of disease comorbidity. It provides different comorbidity assessment; integration of genetic information with the POGO output data could be used to infer causal relationships among diseases. The input of this software is the initial diagnosis for a patient and the output provides evidences of disease comorbidity mapping. POGO computes disease-disease association by adopting semantic similarity measures and hypergeometric test [6.5]. Semantic values of GO term or diseases were calculated based on the DAG of corresponding diseases. Gene set enrichment analysis are used for predicting the significance of genes/CNVs/SNPs--disease and disease--disease associations. Neighbourhood based benchmark method is used to identify the comorbidity pattern among diseases [6.6]. POGO also build the associated network as a bipartite graph; each common neighbour node is selected based on the Jaccard coefficient method [6]. It works as a pipeline with different causal inference packages (e.g. pcalg, qtlnet etc) to predict the causal relationship of diseases. It also provides a pipeline with network regression and survival analysis tools (e.g. Net-Cox, rbsurv etc) to predict more accurate survival probability of patients.

In the future, clinicians will have to consider genetic/genomic implications to patient care throughout their omics and clinical workflow. Our software will help to gain a better understanding of the complex pathogenesis of disease risk phenotypes and the heterogeneity of disease comorbidity. The identified disease patterns can then be further investigated with regards to their diagnostic/prognostic utility or help in the design of novel personalised therapeutic targets. The functions of the POGO offer flexibility for diagnostic applications to predict disease comorbidities, and can be easily integrated to high–throughput and clinical data analysis pipelines for translation medicine.

6.1 References related to section 6

6.1. Park, J., Lee, D.-S., Christakis, N.A., Barab_asi, A.-L.: The impact of cellular networks on disease comorbidity. Molecular systems biology 5(1) (2009)

6.2. Bae, J.S., Cheong, H.S., Kim, J.-H., Park, B.L., Kim, J.-H., Park, T.J., Kim, J.Y., Pasaje, C.F.A., Lee, J.S., Park, Y.-J., et al.: The genetic e_ect of copy number variations on the risk of type 2 diabetes in a Korean population. PloS one 6(4), 19091 (2011)

6.3. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.-W.W., Mazaitis, M., Felix, V., Feng, G.,

Kibbe, W.A.: Disease ontology: a backbone for disease semantic integration. Nucleic acids research 40(D1), 940{946 (2012)

6.4. Liu, Y.I., Wise, P.H., Butte, A.J.: The etiome: identi_cation and clustering of human disease etiological factors. BMC bioinformatics 10(Suppl 2), 14 (2009)

6.5. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression pro_les. Proceedings of the National Academy of Sciences of the United States of America 102(43), 15545{15550 (2005)

6.6. Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barab_asi, A.-L.: The human disease network. Proceedings of the National Academy of Sciences 104(21), 8685{8690 (2007)

7 Conclusions and perspectives

Up to now, a model of glucose, insulin and beta-cell dynamics (from 2000) [1], which focused on T2D, was linked up with a model of T-lymphocyte and macrophage populations within the context of T1D (from 1999) [2]. This was achieved using the considerations of current biological theories and recent papers, which have developed upon these foundation models, such as [3] and [4]. The fragility of these models, including this new one, is held in the accuracy of the parameter estimation. The model is now been extended to gut microbiota. In the next months we will focus more on linking gut microbiota compartments with the pancreatic processes and on using molecular data to estimate the validity and variances of the parameters. Future work may involve the application of delay differential equations (dde's), through the use of Matlab's dde23 solver, for example. This could enhance the model to better represent biological reality since type II diabetes and their various clinical symptoms present across a broad range of time-scales. It would also allow for finer tuning of parameters at these different timescales, and relates this new model back towards the original IG model, which was purposefully divided into fast and slow subsystems. An additional consideration is the modality of that lymphocyte infiltration into the pancreatic islet. It is possible that this effect not only causes chemical disruption via cytokines, but may also cause mechanical disruption of the tissue, further altering the signaling. Hence this could potentially be incorporated as a functional intercellular signaling modification acting upon extra cytokine or receptor variables. The links between metabolism and immunology, which are highlighted by this model, also imply that perhaps the best new

therapeutic strategies for diabetes may lie in immunological targets. This theory is supported by recent studies into metformin.

The co-morbidity of diabetes with other diseases also suggests that more holistic models connecting themes from both metabolism and immunology will be crucial to our mechanistic understanding of the biochemical processes involved. This new combined model may hence act as a `stepping stone' towards future models that are likely to be process-oriented system models, in which everything is interconnected.

Such complex models are already in the process of being developed in the form of `virtual patients'. Parameterizing and calibrating such extensive interaction networks, however, is naturally slow. The theoretical and practical aspects of diabetic comorbidities have been discussed in a recent review paper [7]. They have been also discussed in the BIO4MED conference at Gulbenkian Institute coorganised by Pedro Fernandes and Pietro Lio' in August 2013 and in a talk given at Dagstuhl workshop on Precision medicine in August 2013.

8 Bibliography

1) Brian Topp, Keith Promislow, Gerda Devries, Robert M Miura, and Diane T Finegood. A model of beta-cell mass, insulin, and glucose kinetics: Pathways to diabetes. Journal of theoretical biology, 206:605-619, 2000.

2) Patrick Nelson, Noah Smith, Stanca Ciupe, Weiping Zou, Gilbert S Omenn, Massimo Pietropaolo, et al. Modeling dynamic changes in type 1 diabetes progression: quantifying beta-cell variation after the appearance of islet-speci_c autoimmune responses. Mathematical biosciences and engineering: MBE, 6(4):753, 2009

3) B Freiesleben De Blasio, Per Bak, Flemming Pociot, Allan E Karlsen, and Jrn Nerup. Onset of type 1 diabetes: a dynamical instability. Diabetes, 48(9):1677{1685, 1999

4) Tijana Marinkovic, Marko Sysi-Aho, and Matej Oresic. Integrated model of metabolism and autoimmune response in beta-cell death and progression to type 1 diabetes. PloS one, 7(12):e51909, 2012.

5) Kwiatkowska M, Norman G, Parker D (2011) PRISM4.0: Verification of probabilistic real-time systems. In: Gopalakrishnan G, Qadeer S, editors, Proc. 23rd

International Conference on Computer Aided Verificaton (CAV'11). Springer, volume 6806 of Lecture Notes in Computer Science, pp. 585{591.

6) Aziz A, Sanwal K, Singhal V, Brayton R (2000) Model-checking continuous-time markov chains. ACM Trans Comput Logic 1: 162-170.

7) Capobianco E. and Liò, P. (2013), Comorbidity: a multidimensional approach. Trends in Molecular Medicine 2013 Sep;19(9):515-21

8) Molloy, M.J., Bouladoux, N. & Belkaid, Y. Intestinal microbiota: shaping local and systemic immune responses. Semin. Immunol. 24, 58-66 (2012). Ivanov, I.I. & Littman, D.R. Segmented filamentous bacteria take the stage. Mucosal Immunol. 3, 209-212 (2010).

9) Beyer T, Busse M, Hristov K, Gurbiel S, Smida M, et al. (2011) Integrating Signals from the T-Cell Receptor and the Interleukin-2 Receptor. PLoS Comput Biol 7(8): e1002121. doi:10.1371/journal.pcbi.1002121.