Helsinki University of Technology
Department of Biomedical Engineering and Computational Science Publications
Teknillisen korkeakoulun Lääketieteellisen tekniikan ja laskennallisen tieteen laitoksen julkaisuja
February, 2010 REPORT A15

COMPUTATIONAL ANALYSIS OF THE METABOLIC PHENOTYPES IN TYPE 1 DIABETES AND THEIR ASSOCIATIONS WITH MORTALITY AND DIABETIC COMPLICATIONS

Ville-Petteri Mäkinen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences, Helsinki University of Technology, for public examination and debate in Auditorium of the F- building at Helsinki University of Technology (Espoo, Finland) on the 5th of February, 2010, at 12 o'clock noon.

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Biomedical Engineering and Computational Science

Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Distribution:
Helsinki University of Technology
Department of Biomedical Engineering and Computational Science
P.O.Box 12200
FI-00076 AALTO
FINLAND

Tel. +358 9 451 3172
Fax +358 9 451 3182
http://www.becs.tkk.fi

Online in pdf format: http://lib.tkk.fi/Diss/2010/isbn9789526030135

E-mail: ville-petteri.makinen@finndiane.fi

© Ville-Petteri Mäkinen

ISBN 978-952-60-3012-8 (printed)
ISBN 978-952-60-3013-5 (pdf)
ISSN 1797-3996

Picaset Oy
Helsinki 2010

Abstract of doctoral dissertation	Aalto University School of Science and Technology P.O.Box 11000, 00076 AALTO http://www.aalto.fi

Author	Ville-Petteri Mäkinen		
Name of dissertation	Computational analysis of the metabolic phenotypes in type 1 diabetes and their associations with mortality and diabetic complications		
Manuscript submitted	2009-09-22	Manuscript revised	2009-12-03
Date of defence	2010-02-05		
☐ Monograph		☒ Compilation (summary + original articles)	
Faculty	Information and Natural Sciences, Aalto University		
Departments	Biomedical Engineering and Computational Science, Aalto University School of Science and Technology; Folkhälsan Research Center, Folkhälsan Institute of Genetics, University of Helsinki; Division of Nephrology, Department of Medicine, Helsinki University Central Hospital		
Research field	Diabetic complications		
Opponent	Professor Kumar Sharma		
Supervisor	Professor Kimmo Kaski		
Instructor	Professor Per-Henrik Groop		

Type 1 diabetes is an autoimmune disease that destroys the secretion of insulin (in the pancreas); insulin is a vital hormone for maintaining normal glucose metabolism. Insulin replacement therapy can prevent the acute symptoms, but is not able to fully match the natural regulation, which puts a metabolic stress on tissues. For some patients, the stress manifests as gradual damage to blood vessels and the nervous system over the next few decades after diabetes diagnosis. The aim of the thesis was to describe the metabolic profiles and to investigate their connections with the spectrum of clinical symptoms. Simultaneously, new techniques were applied to measure the profiles (^{1}H NMR spectroscopy) and to visualize the multivariate statistical associations (the self-organizing map). A total of 4,197 patients with type 1 diabetes were recruited for the thesis by the Finnish Diabetic Nephropathy Study. A quarter of the patients exhibited an obesity-related phenotype (high triglycerides, cholesterol, apolipoprotein B-100, low high-density lipoprotein cholesterol, high C-reactive protein). A third of the individuals had a diabetic kidney disease phenotype (high urinary albumin and serum creatinine). The combination of the two was associated with a 10-fold population-adjusted mortality. Nevertheless, there was no discernible metabolic threshold between the phenotype models, nor were there any single variable that could predict the outcomes accurately. These results suggest a need for multifactorial and multidisciplinary paradigms for the research, treatment and prevention of diabetic complications.

Keywords	type 1 diabetes, kidney disease, NMR spectroscopy, self-organizing map		
ISBN (printed)	978-952-60-3012-8	ISBN (pdf)	978-952-60-3013-5
ISSN	1797-3996		
Number of pages	126		
Publishers	Department of Biomedical Engineering and Computational Science, AU; Folkhälsan Research Center, UH; Division of Nephrology, HUCH		
Print distribution	Department of Biomedical Engineering and Computational Science, AU		

☒ Available at http://lib.tkk.fi/Diss/2010/isbn9789526030135

Väitöskirjan tiivistelmä		Aalto-yliopiston teknillinen korkeakoulu PL 11000, 00076 AALTO http://www.aalto.fi	
Tekijä	Ville-Petteri Mäkinen		
Väitöskirjan nimi	Tyypin 1 diabeteksen laskennalliset aineenvaihduntaprofiilit ja niiden tilastolliset yhteydet kuolleisuuteen ja diabeteksen liitännäissairauksiin		
Käsikirjoitus jätetty	2009-09-22	Esitarkastettu	2009-12-03
Väitöstilaisuus	2010-02-05		
☐ Monografia		☒ Yhdistelmä (yhteenveto + erillisartikkelit)	
Tiedekunta	Informaatio- ja luonnontieteet, Aalto-yliopisto		
Laitokset	Lääketieteellinen tekniikka ja laskennallinen tiede, Aalto-yliopiston teknillinen korkeakoulu; Folkhälsanin tutkimuskeskus, Helsingin yliopisto; Nefrologian klinikka, Sisätaudit, Helsingin yliopistollinen keskussairaala		
Tutkimusala	Diabeteksen liitännäissairaudet		
Vastaväittäjä	Professori Kumar Sharma		
Valvoja	Professori Kimmo Kaski		
Ohjaaja	Professori Per-Henrik Groop		

Tyypin 1 diabetes puhkeaa, kun kehon oma immuunipuolustus virheellisesti tuhoaa haiman insuliinia tuottavat beta-solut. Ilman insuliinia veren glukoosi ei siirry lihaksiin ym. kudoksiin energiantuotantoa varten. Taudin välittömät oireet pystytään ehkäisemään insuliinipistoksilla, mutta ne eivät täysin vastaa luonnollisia säätelymekanismeja, minkä vuoksi jotkut potilaat kärsivät vuosien mittaan vähitellen etenevistä verisuoniston ja hermoston liitännäissairauksista. Väitöskirjatyön tavoite oli selvittää tyypin 1 diabetespotilaiden aineenvaihduntaprofiilit, ja tutkia niiden yhteyksiä liitännäissairauksiin. Tutkimuksia varten kerättiin tietoja 4,197 potilaalta (The Finnish Diabetic Nephropathy Study). Lisäksi profiilien mittauksia ja tulosten analysointia varten sovellettiin uusia menetelmiä (mm. ^{1}H NMR spektroskopiaa ja itseorganisoituvia karttoja). Noin neljäsosalla potilaista havaittiin merkkejä lihavuuteen liittyvistä häiriöistä (korkea triglyseridi, kolesteroli ja apolipoproteiini B-100; matala HDL kolesteroli; korkea C-reaktiivinen proteiini) ja noin kolmasosalla oli merkkejä munuaisvauriosta (korkea virtsan albumiini ja seerumin kreatiniini). Kuolleisuus oli kymmenkertainen koko väestöön verrattuna niillä potilailla, joilla oli piirteitä molemmista profiileista. Laskennalliset mallit eivät kuitenkaan paljastaneen selkeitä raja-arvoja eri aineenvaihduntaprofiilien välillä, eikä mikään muuttuja yksiselitteisesti ennustanut kliinisiä havaintoja. Siksi on tärkeää ottaa huomioon myös useiden tekijöiden yhteisvaikutukset diabeettisten liitännäissairauksien tutkimuksessa, hoidossa ja ennaltaehkäisyssä.

Asiasanat	diabetes, munuaistauti, NMR spektroskopia, itseorganisoituva kartta		
ISBN (painettu)	978-952-60-3012-8	ISBN (pdf)	978-952-60-3013-5
ISSN	1797-3996		
Sivumäärä	126		
Julkaisijat	Lääketieteellisen tekniikan ja laskennallisen tieteen laitos, AY; Folkhälsanin tutkimuskeskus, HY; Nefrologian klinikka, HYKS		
Painetun kirjan jakelu	Lääketieteellisen tekniikan ja laskennallisen tieteen laitos, AY		

☒ Saatavilla osoitteessa http://lib.tkk.fi/Diss/2010/isbn9789526030135

List of publications

I Mäkinen VP, Forsblom C, Thorn LM, Wadén J, Gordin D, Heikkilä O, Hietala K, Kyllönen L, Kytö J, Rosengård-Bärlund M, Saraheimo M, Tolonen N, Parkkonen M, Kaski K, Ala-Korpela M, Groop PH (2008) **Metabolic phenotypes, vascular complications and premature deaths in a population of 4,197 patients with type 1 diabetes.** *Diabetes* 57:2480-2487.

II Mäkinen VP, Soininen P, Forsblom C, Parkkonen M, Ingman P, Kaski K, Groop PH, Ala-Korpela M (2006) **Diagnosing diabetic nephropathy by ^{1}H NMR metabonomics of serum.** *Magn Reson Mater Phy* 19:281-296.

III Mäkinen VP, Soininen P, Forsblom C, Parkkonen M, Ingman P, Kaski K, Groop PH, Ala-Korpela M (2008) **^{1}H NMR metabonomics approach to the disease continuum of diabetic complications and premature death.** *Mol Syst Biol* 4:167.

IV Niemi J, Mäkinen VP, Heikkonen J, Tenkanen L, Hiltunen Y, Hannuksela ML, Jauhiainen M, Forsblom C, Taskinen MR, Kesäniemi YA, Savolainen MJ, Kaski K, Groop PH, Kovanen PT, Ala-Korpela M (2009) **Estimation of VLDL, IDL, LDL, HDL_2, ApoA-I and ApoB from the Friedewald inputs – ApoB and IDL, but not LDL, are associated with mortality in type 1 diabetes.** *Ann Med* 41:451-461.

V Mäkinen VP, Forsblom C, Thorn LM, Wadén J, Kaski K, Ala-Korpela M, Groop PH (2009) **Network of vascular diseases, death and biochemical characteristics in a set of 4,197 patients with type 1 diabetes.** *Cardiovasc Diabetol* 8:54.

Author's contribution

The author conceived the study and designed and implemented the statistical analysis in Publications I-III and V. The author also wrote the manuscript for Publications I, III and V, and partially for Publication II. For Publication IV, the author prepared the data for analysis, reviewed and edited the manuscript and implemented the online component of the final application. The author participated in the data collection for all publications.

List of abbreviations

ADA	The American Diabetes Association
AER	Urinary albumin excretion rate
AGE	Advanced glycation end-product
AHT	Anti-hypertensive treatment
ANN	Artificial neural network
ApoA-I, A-II, B	Apolipoprotein A-I, A-II or B-100
BMI	Body mass index
BP	Blood pressure
CM	Chylomicron
CRP	C-reactive protein
DCCT	The Diabetes Control and Complications Trial
DKD	Diabetic kidney disease
DMDur	Diabetes duration
DRP	Diabetic retinopathy
EGIR	The European Group for the Study of Insulin Resistance
ESRD	Eng-stage renal disease
GFR	Glomerular filtration rate
HbA1c	Hemoglobin A1c
HDL	High-density lipoprotein
IDF	The International Diabetes Federation
IDL	Intermediate-density lipoprotein
k-NN	*k*-nearest neighbor algorithm
KDNEG	Kidney disease negative (normal urinary albumin excretion)
LDA	Linear discriminant analysis
LDL	Low-density lipoprotein
LIPO	Lipoprotein lipid and albumin bound fatty acids (molecular window)
LMWM	Low-molecular-weight metabolites (molecular window)
MBL	Mannan-binding lectin
MetS	The metabolic syndrome

MLP	Multi-layer perceptron
MVD	Macrovascular disease
NCEP ATP	The US National Cholesterol Education Program, Adult Treatment Panel
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PLS	Projection to latent structures
RAGE	Receptor for advanced glycation end-products
ROS	Reactive oxygen species
SOM	Self-organizing map
VLDL	Very low-density lipoprotein
WHO	The World Health Organization
WHR	Waist-hip ratio
WMA	The World Medical Association
24h-uAlb	Excreted albumin per time unit from a 24h urine sample

Contents

1 Introduction

Historically, the autoimmune variant of diabetes was a devastating disease with a short life expectancy after diagnosis. An early clinical report [Twitchell 1907] stated that "Of 64 severe cases (of diabetes), 59 died before the third year after discovery". At the time, it was already known that normal glucose metabolism was absent in these patients, and there was a consensus about the two major forms of diabetes: young people tended to have the severe form, whereas most older individuals could cope with diet adjustments for many years.

The introduction of insulin injections two decades later provided the critical treatment for the acute symptoms [Banting 1922], but already in the mid-thirties it became evident the the insulin replacement therapy was not enough to restore optimal metabolism. It was discovered that some patients with insulin-treated diabetes suffered from high blood pressure and specific lesions in the kidney [Kimmelstiel 1936]. Gradually, a detailed picture of tissue damage in the circulatory and nervous systems emerged in patients with a long-standing juvenile diabetes.

Today, diabetes is classified into two major and a number of minor types according to the mode of development [ADA 2007]. Type 1 is the autoimmune variant, where the body selectively destroys the insulin-producing pancreatic beta-cells through a deranged immune response, typically at a young age. Type 2 is the more common form that develops slowly from a complex set of environmental and genetic causes as people grow older and both the effectiveness and secretion of insulin decline sufficiently [Kahn 2006].

Treatments for most of the diabetic complications have advanced over the last fifty years: dialysis and transplantation can be used to offset kidney failure and laser surgery can prevent vision loss from diabetic eye disease. A decreasing fraction of patients with type 1 diabetes die of failed kidneys [Finne 2005]. On the other hand, heart attacks and strokes have emerged as the major threats to longevity: they are more severe and more common for the patients with diabetic complications [Gross 2005, Daneman 2006].

The traditional risk factors that were known already in the early 1900's are still valid today: obesity, smoking and poor diet seem to predispose to a number of diseases, including type 2 diabetes and the complications of type 1 diabetes. For instance, a contemporary paper on metabolism [Breed 1918] began with the words "in a generously fed community such as may be found anywhere today, a large majority of the people are overfed, and when people are continuously overfed, sooner or later we find that they have diseases due to a changed body

chemistry". A hundred years later, this phenomenon has become even more widespread.

Still, not all obese people suffer from type 2 diabetes or heart disease. Not all patients with type 1 diabetes develop complications. If every patient has the same metabolic problems, why do some develop complications while others do not? There are two possible, non-exclusive answers to these questions: i) the metabolism is, in fact, not the same, but the diagnostic criteria is insufficient to describe the breadth of the phenomenon or ii) the susceptibility to the metabolic stress is genetically determined.

This thesis investigates the first hypothesis from a technological perspective. Publications I and V employ computational approaches to uncover new phenotypic features beyond the conventional clinical classifications. The emphasis is on the multivariate (biochemical) patterns that are associated with death, gender and other qualitative traits. Publications II-IV introduce new technology to obtain more biochemical information and to elaborate the phenotypes even further.

2 Review of the literature

2.1 Diabetes mellitus and its complications

Classification of type 1 diabetes

Diabetes is characterized by the inability of tissues to utilize glucose. Insulin is the critical hormone that activates the normal glucose uptake and metabolism in most cells, and its secretion from the pancreas is stimulated by nutrient absorption in the gut and the subsequent increase in blood glucose concentration. In type 1 diabetes, however, the immune system attacks the insulin producing beta-cells, and rapidly depletes the insulin response capability [Daneman 2006, Knip 2008]. The precise triggers and driving factors for this autoimmune reaction are unknown, but bovine insulin, other dietary agents and environmental toxins [Virtanen 2003, Vaarala 2006], viral infections [Haverkos 2003, Filippi 2008] and an inborn metabolic imbalance [Oresic 2008] have been suggested, in conjunction with an inherited immune system susceptibility [Jahromi 2007, Nejentsev 2007, Todd 2007].

Type 2 diabetes is even more complex than type 1, and gradual in nature: it comprises several defects that reduce the pancreatic insulin secretion, but also make the hormone less efficient in activating the carbohydrate metabolism [Kahn 2006, Lyssenko 2008, Lyssenko 2009]. In addition, type 2 diabetes is heavily influenced by life-style factors that exacerbate the genetic effects. These include the hallmarks of the modern civilization: energy-rich diets and poor physical condition due to the lack of exercise [Lindström 2006, Jeon 2007].

Traditionally, type 1 and type 2 diabetes have been distinguished by the age of onset: type 1 diabetes starts abruptly at an early age, typically within adolescence, whereas type 2 is a disease that gradually develops in adulthood. The current gold-standard is based on the antibodies that are associated with the autoimmune response [Seissler 2006] and also on genetic testing to confirm rarer types of diabetes [Fajans 2001]. On the other hand, most clinical studies employ surrogate classifications based on the age of onset and the type of insulin treatment for practical reasons. In this thesis, type 1 diabetes is defined as an age of onset below 35 and permanent insulin treatment within a year of onset.

Finland has one of the highest incidence rates, rising from 31 in 1980 to 64 per 100,000 per year in 2005 [Harjutsalo 2008]. It is estimated that in 2005 there were 45,000 Finnish patients with type 1, and 193,500 patients with type 2 diabetes. If also those people with subclinical or undetected (type 2) diabetes are included, the total number likely exceeds half a million or close to 10% of the entire population [Reunanen 2006].

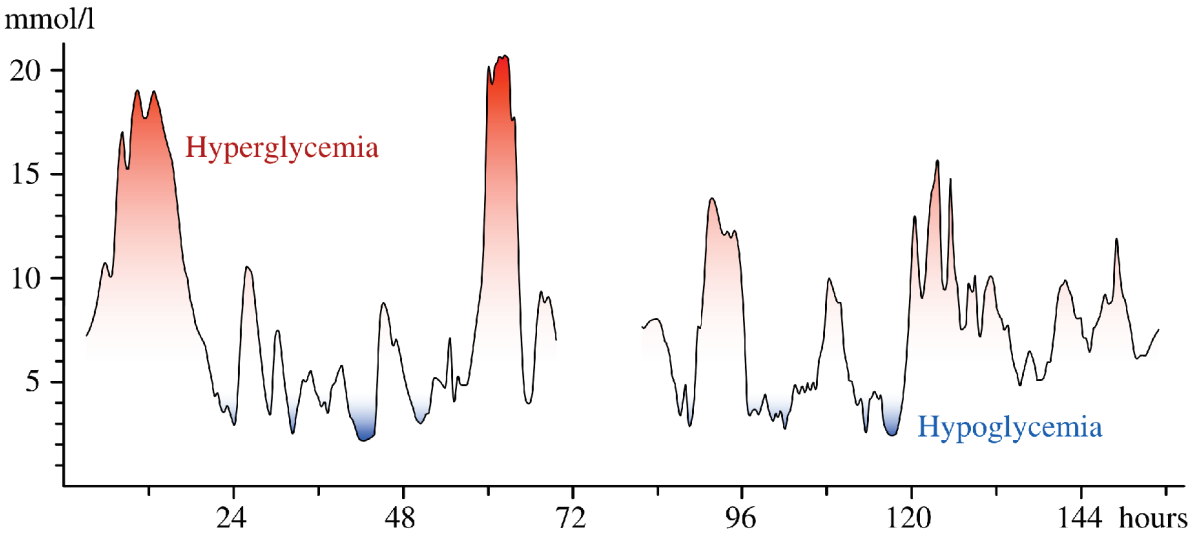

Figure 1: Concentration curves from 72h glucose monitoring experiments for a patient with type 1 diabetes. The continuous data were measured from the interstitial fluid and calibrated by four daily blood tests by finger sticks [Gordin 2008b]. Values below 3 mmol/l are rare in non-diabetic individuals and most of the time blood glucose stays within 4-6 mmol/l, although the peak value may rise close to 10 mmol/l shortly after a meal. Patients with type 1 diabetes must adopt insulin replacement therapies to maintain adequate glucose metabolism in tissues. Insulin pumps can match the natural responses better than manual injections. However, the technology is expensive and does not work for everyone and most type 1 diabetic patients rely on self-administered insulin.

Acute symptoms of glucose imbalance are treatable, but the insulin replacement therapy for type 1 diabetes, for instance, is not able to fully match the natural pancreatic regulation and response. As a result, blood glucose concentration becomes less stable in these patients [Kilpatrick 2007c]. To avoid acute episodes of harmfully low glucose, higher than normal average concentration must be maintained (Figure 1), which drives secondary disease processes (diabetic complications) that cause most of the human suffering and societal burden [Marshall 2006].

Microvascular complications

Persisting high concentration of blood glucose is associated with progressing damage to the vascular system [DCCT 1993, DCCT 2000, Scott 2001, Gross 2005, Lachin 2008]. The kidneys, for example, are filled with small and vulnerable blood vessel structures (hence the term microvascular). They filter out waste products from the blood into the urine and also maintain the overall fluid balance of the body. The filtering units (Figure 2) are composed of a globular mesh of capillary blood vessels (the glomerulus) and tubular structures that re-absorb the spill-over of useful molecules and ions – and most of the water – from the urinary side. Abnormal tissue expansions around the intricate capillaries in the glomerulus are the typical signs of diabetic kidney disease, but lesions can also be seen in the tubular structures and the surrounding tissue [Drummond 2002, Fioretto 2006, Najafian 2006, Perrin 2006]. Reduced survival of the octo-

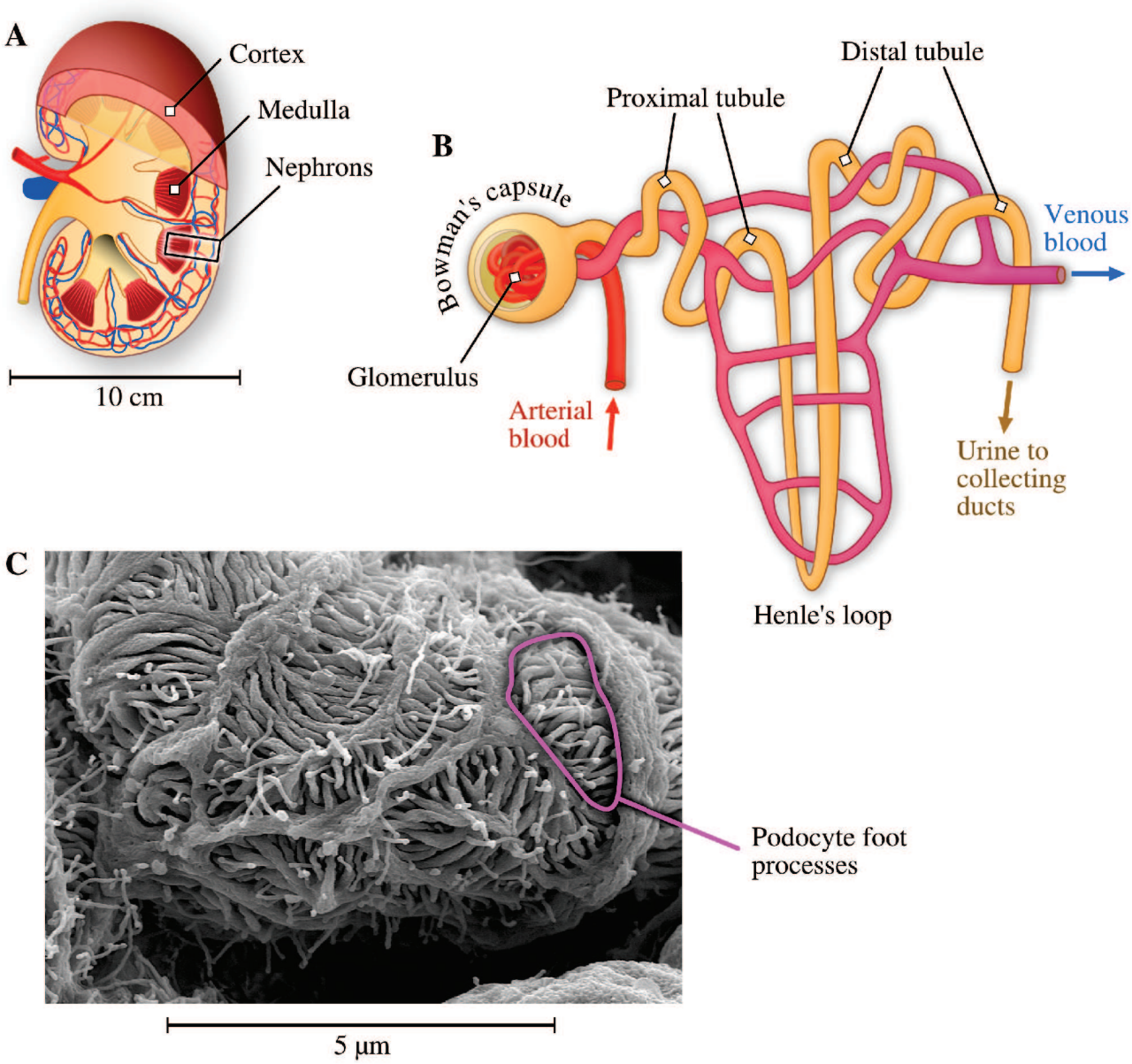

Figure 2: Structural elements of the kidney. **A** The renal cortex is the primary filtration compartment that contains most of the several hundred thousand nephrons in the human kidney. **B** The nephron has two main components. The corpuscle (glomerulus and Bowman's capsule) produces the first filtrate from arterial blood via the small capillaries. Any negatively charged proteins or other macromolecules are blocked from entering the urinary space, but water, small molecules and ions are passed through. The tubules reabsorb essential molecules and ions back from the filtrate, and return them to the blood circulation. Finally, the collecting ducts finish the reabsorption of water and thus maintain the overall fluid balance in the body. **C** Scanning electron microscope image of a mouse glomerular capillary with podocytes (Wikipedia, Creative Commons Attribution ShareAlike 3.0).

pus-like cells (podocytes) that embrace the glomerular capillaries may also accompany the early progressive nephron damage [White 2002, Wolf 2005].

An estimated third of patients with type 1 diabetes will be affected during their lifetime, although recent studies have reported a decline or a delay in incidence [Nordwall 2004, Finne 2005, Rossing 2005]. The disease advances at an individual rate over the course of several years, but ultimately most of the glomeruli will be destroyed and the patients will suffer from insufficient filtration ability [Bloomgarden 2008].

Vulnerable small vessel structures also exist in the back of the eye, where they feed the light-sensitive cells of the retina [DCCT 1995b, Hammes 2002, Alibrahim 2006, Al-Kateb 2007]. Interestingly, both the kidneys and the retina can take up glucose without insulin, which can lead to excess glucose inside certain cell types. In one study, for instance, the endothelial cells on the retinal vessel wall were not susceptible to high glucose, but neighboring pigment cells were, and they released signaling molecules that then drove also the endothelial cells out of balance [Busik 2008]. On the other hand, the brain – which does not need insulin for glucose metabolism either – appears to be protected from increased uptake, and does not exhibit similar changes in small vessels [Badr 2000].

Nearly all patients with type 1 diabetes will have some detectable changes in the retinal vasculature [Skrivarhaug 2006, Klein 2008], but for some patients the small capillaries leak excessively to the surrounding tissue (including the macula, the area of the best sensory accuracy), the small arteries in the back of the eye become stiffer (and less efficient), and new blood vessels start to form to meet the subsequent shortage of oxygen [Davidson 2007b]. These proliferative forms of retinopathy are more common in patients with kidney disease, and lead to severe loss of sight or even blindness if left untreated [Ferris 1999].

Macrovascular disease

Although the microvascular complications pose a serious threat to health, the premature deaths attributed to diabetes come primarily from occlusions in the large arteries that supply oxygen-rich blood to the heart and to the brain [Morrish 2001, Soedamah-Muthu 2004, DCCT 2005, Stadler 2006]. Several studies have reported a three-fold or even higher risk of cardiovascular events (e.g. heart attacks) in the diabetic population. In fact, the risk of the first event for a diabetic patient is comparable to the risk of recurrence for a non-diabetic heart patient [Juutilainen 2005, Juutilainen 2008, Schramm 2008].

The disease process, which is similar to that seen in non-diabetic individuals, involves an inflammatory response within the arterial wall, accompanied by the accumulation of lipids (Figure 3). Gradually, the wall begins to thicken and a plaque of fatty cell debris forms on the inside surface of the affected blood vessel [Pajunen 2000, Beckman 2002, Retnakaran 2008]. When the plaque grows sufficiently, the diameter available for blood flow decreases and oxygen supply to tissues is jeopardized. Or, if the plaque becomes unstable, it may rupture and

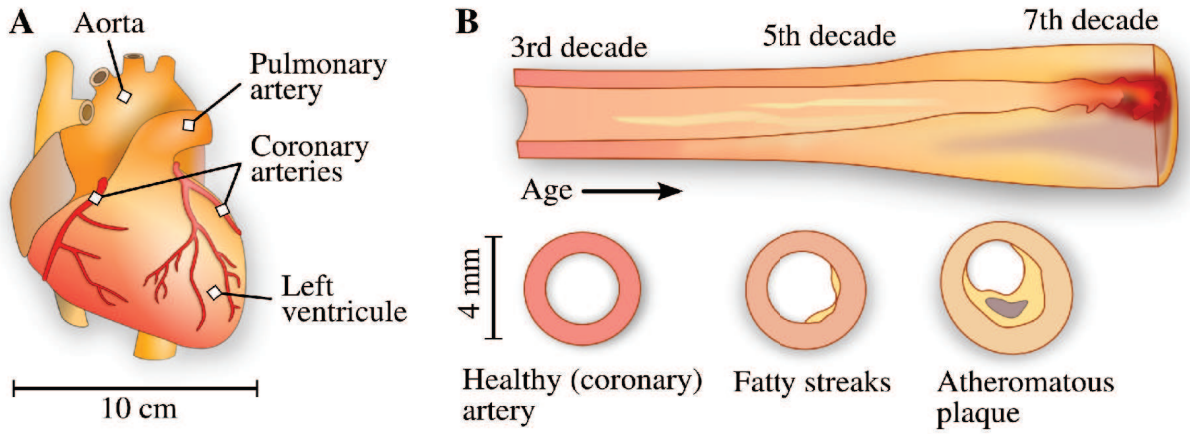

Figure 3: Schematic illustration of coronary atherosclerosis. **A** The heart is highly sensitive to any damage in its vasculature. When the ventricules contract, they create a pressure wave that travels along the arterial tree, which can be detected as the systolic pulse. Simultaneously, most arteries inside the heart muscle are compressed and cannot deliver blood. Therefore, the primary supply of oxygen to the heart itself occurs during the relaxing diastolic phase: some of the systolic energy is reflected from the arterial tree and this reverse wave maintains the flow in the coronary arteries. **B** Thickening of the vascular wall and accumulation of lipid debris with inflammation as a function of time. Eventually, the plaque becomes ulcerated and unstable. A rupture typically leads to a blood clot (platelets adhere to the damaged wall), which can then block the artery or travel further into the arterial tree to cause a blockage elsewhere.

lead to an unanticipated and fatal blockage [Ruggeri 2002, Zgibor 2006, Pundziute 2007].

Plaques can form on any arterial wall. However, the blood vessels that provide oxygen to the heart muscle are of particular importance, since they are easily occluded due to their smaller size, and any disruptions in the blood flow will have catastrophic effects on the whole body [Orchard 2006, Lockhart 2008, Hasin 2009]. The veins are not affected, unless transplanted in replacement for occluded arteries in heart bypass surgery [Desai 2004].

Glucose control is a key modifiable factor in the prevention of complications and there is evidence that the effect extends to arterial plaques, particularly in type 1 diabetes: an intensive treatment program had beneficial long-term effects against arterial degradation [Snell-Bergeon 2003, Cleary 2006, Juutilainen 2008]. Indirect influence is possible: patients with diabetic kidney disease have an increased risk of stroke or heart attack due to an accelerated and wide-spread process of plaque formation [Dahl-Jørgensen 2005, Kim 2007]. The immediate threat to life from kidney failure can be minimized by renal replacement therapies [Finne 2005, Rossing 2005] but, in many cases, these survivors succumb to macrovascular diseases within a few years.

Neuropathy and other complications

Excess glucose is also damaging to the nervous system: direct metabolic effects

make the affected neurons more prone to end their life-cycle prematurely and the impairment of microcirculation through small capillary blood vessels causes further indirect injury [DCCT 1995a, Pittenger 1997, Vincent 2002, Lefrandt 2003, Maguire 2007, Edwards 2008]. The physical manifestations include loss of feeling and chronic pain [Martin 2006], impaired autonomic responses [Lefrandt 1999, Forsén 2004], erectile and other sexual problems [Enzlin 2003, Klein 2005] and inefficient stomach function [Jones 2002, Sogabe 2005]. Cognitive functions may also be affected, and tissue-specific changes can be observed by brain imaging [Wessels 2008]. However, most of the studies can provide only suggestive evidence due to numerous methodological challenges [Van Harten 2006, Weinger 2008].

Diabetic foot ulcers and the amputation of lower extremities causes significant human suffering and disability, and a large portion of overall health care costs [Gordois 2003, Chen 2006]. Foot ulcers are often initiated and exacerbated by diabetic neuropathy [Mueller 2005, Gershater 2009].

Pregnancy for a woman with type 1 diabetes poses additional challenges to metabolic control and insulin treatment. Newborns of diabetic mothers are often larger than normal due to the high-glucose environment, and a phenomenon of run-away blood pressure and leakage of protein into the urine (pre-eclampsia) is more common [Evers 2004]. Perhaps expectedly, those women with a history of pre-eclampsia have a higher prevalence of diabetic kidney disease, although pregnancy itself is not a significant long-term risk factor [Vérier-Mine 2005, Gordin 2007b].

Symptoms of depression are common (10-30%) among patients with diabetes and there is an additional association between mental health and complications [Anderson 2001, De Groot 2001, Li 2008]. Undoubtedly, the quality of life is affected, especially due to loss of eyesight and kidney function, which for young patients may mean permanent inability to work [Kraut 2001]. Adaptation to strict dietary regime and careful glucose monitoring is not easy, and any additional stress and anxiety has an impact on the self-care adherence [McKellar 2004], which perpetuates the vicious cycle of poor mental and physical health.

2.2 Clinical risk factors

Age and diabetes duration

The combination of time and impaired pancreatic function (deranged glucose metabolism) is a necessary (but possibly not sufficient) precondition to diabetic complications. Newly diagnosed type 1 diabetes is often accompanied by increased filtration and enlarged kidneys, probably to counter the rising glucose [O'Donnell 1988, Vallon 2003]. Within a few years, the filtration subsides to normal levels – and declines further [Magee 2009]. Hyperfiltration is not, however, a necessary prerequisite for future kidney complications. The clinical signs of diabetic kidney disease (most notably protein in urine) usually appear in the second decade of diabetes [Mogensen 1995, Österby 1995, Ayodele 2004]. As the functional capacity declines (at a rate of ~10% per year), protein excretion typi-

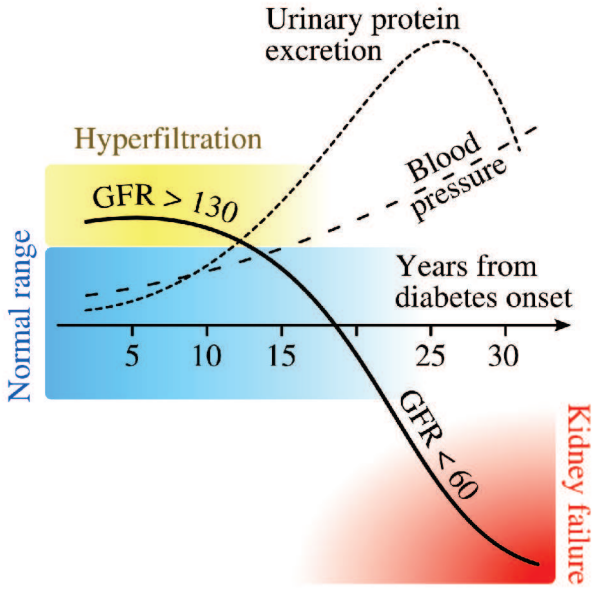

Figure 4: Natural course of diabetic kidney disease. Initially, the glomerular filtration rate (GFR) can be above the typical healthy range from 60 to 130 ml/min/1.73m^2 in newly diagnosed patients with type 1 diabetes. Within two decades the GFR begins to decline and urinary protein excretion increases for one in three patients. Finally, the kidneys fail to filter enough blood and excrete urine to clear the body of accumulating metabolic waste products, and renal replacement therapy is required.

cally increases and the other metabolic effects worsen until the kidneys can no longer maintain normal urine flow; renal replacement therapy is required thereafter (Figure 4). Of note, recent studies suggest that low-level protein excretion may be unrelated to the functional decline in filtering capacity, and can disappear with modern diabetes management [Rosolowsky 2008].

Retinal abnormalities can already be seen shortly after the onset of diabetes in some patients [Malone 2001] and retinal disease is highly prevalent (>50%) in those patients with a long diabetes duration [Agardh 1997, Rossing 2005, Keenan 2007, Klein 2008]. The sight-threatening complications (proliferative retinopathy in particular) affect a subset of patients (approximately 30%), but much of the visual acuity can be preserved with laser surgery [Davidson 2007a, Mohamed 2007]. Curiously, sudden changes in blood glucose control may have unexpected consequences; one study reported an initial worsening of the retinal damage when switching to an intensive insulin treatment, although the long-term prognosis was improved [DCCT 1998, DCCT 2000].

The macrovascular complications accompany the microvascular deterioration of the kidneys and the retina. Put differently, patients with type 1 diabetes (and kidney disease) suffer from heart attacks earlier than their peers, and are less likely to survive from initial events [Fisher 1997, May 2000, Antoniucci 2004, Miller 2009]. It is, however, unclear whether this acceleration extends to the less vulnerable patients with continuing good metabolic control even after decades of diabetes.

Not only diabetes duration, but also the age of onset has an impact on the incidence of complications. Those that have had type 1 diabetes from early childhood (<5 years old) may have less kidney disease [Finne 2005]. On the other hand, there is a strong birth cohort effect thanks to the increased awareness and technological advances in glucose monitoring and insulin preparates [Brange

1997, Siebenhofer 2004, Von Sengbusch 2005, Monami 2009]. Interestingly, age at onset has also a familial association on the risk of type 1 diabetes itself: those patients with an early onset are more likely to have a diabetic sibling [Harjutsalo 2005].

Blood pressure and arterial stiffening

Aging arteries become stiffer, which increases the peak flow resistance and creates stronger pulse fluctuations during a heart beat; in type 1 diabetes arteries age 10-20 years faster than normal, if looking at blood pressure alone [Rönnback 2004, Laugesen 2009]. The continuing increase in systolic blood pressure is also a classical indicator of kidney complications and large vessel diseases – keeping the blood pressure within the normal or even lower range (<130/80 mmHg) with medication slows down the disease progression significantly [Bretzel 1997, Beckman 2002, Tomlinson 2003, Sobolewski 2004, Astrup 2005, Thomas 2006b].

The glomerulus may be particularly vulnerable to intense pulse wave fluctuations: a steeper pressure gradient within the Bowman's capsule (Figure 2) imposes a mechanical stress on the delicate capillaries, which may contribute to the tissue expansions and podocyte abnormalities in diabetic kidney disease [Ishida 1999, Petermann 2002, Gnudi 2007]. The combination of stress and high blood glucose can in some cases lead to further metabolic derangements [Riser 1999, Lewko 2005]. Similarly, damaging effects of cyclic stretching and persistent high pressure on the retinal vasculature have been documented [Suzuma 2001, Beltramo 2006].

High (systolic) blood pressure and stiff arteries have a functional relationship, and the two are associated with an increased risk of heart attack [Schiffrin 2004, Zoungas 2007]. The mechanisms remain under investigation: in type 1 diabetes, for instance, acute glucose load in an experimental setting increased arterial stiffness and prolonged the bioelectric cycle of the heart [Gordin 2007a, Gordin 2008a], both of which are indicators and/or risk factors for heart disease [Veglio 1999]. Another study reported that biochemical markers of glucose exposure are correlated with increased pulse pressure [Schram 2005]. Arterial stiffening increases the heart workload while reducing the supply of fresh blood to the heart muscle, which may partly explain the more severe nature of cardiovascular events in type 1 diabetes [Brooks 1999].

Obesity and the metabolic syndrome

Overweight patients with type 1 diabetes have more complications than lean individuals and obesity-related metabolic changes contribute to the development of diabetic complications [De Block 2005, Stone 2006]. In particular, central obesity (waist circumference) has been identified as a key marker since it reflects the amount of excess fat around internal organs [Després 2006, De Boer 2007]. These visceral fat deposits are associated with reduced insulin action [Macor 1997], which in turn may be a sign of the metabolic defects that lead to the complications [Lorenzo 2003, Groop 2005, Mathieu 2008]. However, the practical predictive value of obesity (waist and hip circumference, weight and height) or indirect measures of the underlying metabolic abnormality are limited in patients

Table 1: A number of definitions for the metabolic syndrome. Each criterion yields a point, and the final diagnosis is confirmed if the total tally exceeds a given threshold. For most definitions, three or more points are required. Abbreviations: IDF (International Diabetes Federation), TG (plasma triglycerides), HDLC (high-density lipoprotein cholesterol), NCEP ATP (the US National Cholesterol Education Program, Adult Treatment Panel), WHO (the World Health Organization), EGIR (the European Group for the Study of Insulin Resistance).

IDF (2006)

Obesity	Wide waist circumference (exact cutoffs depend on ethinicity)
Lipids	TG >1.7 mmol/l or lipid-lowering medication, HDLC <1.0 mmol/l (men) or <1.3 mmol/l (women) or HDL-enhancing medication
Glucose and insulin	Diabetes mellitus or fasting plasma glucose >5.6 mmol/l
Blood pressure	>130 mmHg systolic or >85 mmHg diastolic or anti-hypertensive medication

NCEP ATP III (2001)

Obesity	Waist circumference >102 cm (men) or >88 cm (women)
Lipids	TG >1.7 mmol/l, HDLC <0.9 mmol/l (men) or <1.0 mmol/l (women)
Glucose and insulin	Fasting plasma glucose >6.0 mmol/l
Blood pressure	>130 mmHg systolic or >85 mmHg diastolic or anti-hypertensive medication

WHO (1999)

Obesity	Waist-hip ratio >0.90 (men) or >0.85 (women) or body mass index >30 kg/m^2
Lipids	TG >1.7mmol/l, HDLC <0.9 mmol/l (men) or <1.0 mmol/l (women)
Glucose and insulin	Diabetes mellitus, impaired glucose tolerance, impaired fasting glucose or insulin resistance (methodology not specified)
Blood pressure	>140 mmHg systolic or >90 mmHg diastolic or anti-hypertensive medication

EGIR (1999)

Obesity	Waist circumference >94 cm (men) or >80 cm (women)
Lipids	TG >2.0 mmol/l or HDLC <1.0 mmol/l or under treatment for dyslipidaemia
Glucose and insulin	Top 25% of the fasting insulin values among non-diabetic individuals, fasting plasma glucose >6.0 mmol/l
Blood pressure	>140 mmHg systolic or >90 mmHg diastolic or anti-hypertensive medication

with type 1 diabetes [Giorgino 2004, Kilpatrick 2007b]. Especially, an early indication of diabetic kidney disease (increased urinary albumin excretion) can override the obesity-related risk of macrovascular complications [Pambianco 2007, Thorn 2009a].

Many of the previous studies investigated the "metabolic syndrome" rather than obesity itself. The syndrome was created to assess the underlying insulin resistance and the related cardiovascular risk in the general population without invasive procedures [Eckel 2005]. The classification can be based on several different scoring systems (Table 1) that typically combine measures of obesity, blood pressure, glucose control and a few additional metabolic variables [Alberti 1998, Alberti 2006, Day 2007, NCEP 2002]. For instance, according to the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) recommendations, a waist circumference of >102 cm for men >88 cm for women corresponds to one point. Systolic blood pressure over 130 mmHg yields another and so on. Finally, the points from the individual criteria are tallied and the total score is used for diagnosis.

The prevalence of the NCEP ATP III metabolic syndrome is close to 40% within the Finnish patients with type 1 diabetes [Thorn 2005]. However, the usefulness of the concept has been questioned, not only for type 1 diabetes [Pambianco 2007], but in general as well due to ethnic differences and the lack of additive benefits from combining the individual components [Cossrow 2004, Misra 2005, Vaidya 2007, Sattar 2008]. Also, the cutoffs cause information loss, and indeed the basic idea of qualitative epidemiological classification and insulin resistance at the center of the metabolic syndrome has been criticized [Yudkin 2007]. Nevertheless, the metabolic syndrome continues to enjoy some popularity in the scientific community, thanks to its ease of application.

Gender

Men have higher mortality rates than women (Figure 5). The same trend is also reflected in diabetic complications: there are more type 1 diabetic men than women who are affected by kidney disease, retinopathy or heart attacks [Raile 2007, Villar 2007, Klein 2008]. However, the population-adjusted risk is more complicated, especially for women before and after menopause [Laing 2003a, Laing 2003b]. Unlike the majority of autoimmune disorders, type 1 diabetes itself is more common in men than in women and the male proportion is most pronounced in those patients with an onset during or after puberty [Harjutsalo 2008]. Also, the transmission of type 1 diabetes from the father to a child occurs more often than from the mother, especially if the father has an early disease onset [Harjutsalo 2006].

A number of factors may produce the observed associations between gender and complications. Men tend to lead less healthy lives: data from Statistics Finland, collected in 2007, indicates that 26% of adult men but only 16% of women were smokers, 1,425 men versus 371 women died of alcohol-related causes and 618 men versus 199 women committed suicide. Men have also physical disadvantages: they tend to become

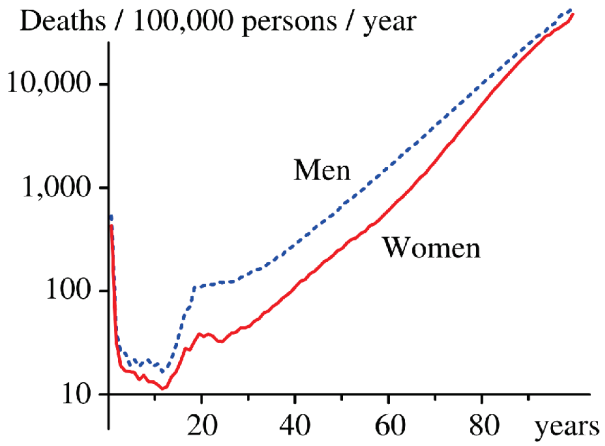

Figure 5: Logarithmic one-year mortality as a function of age. The data were obtained from Statistics Finland and the curves represent all-cause mortality rate estimates within 1980-2006 in the entire Finnish population (a total of 136,466,661 person-years and 1,298,393 deaths).

centrally obese (apple-shape) instead of the less harmful gluteo-femoral type (pear-shape), and have more adverse profiles of lipid and other metabolism [Molarius 1999, Sibley 2006, Ordovas 2007]. An active female reproductive cycle may be protective in this respect. For instance, cardiovascular disease is less common in pre-menopausal women than in men of the same age, but this difference is diminished or even reversed by diabetes and later by the onset of menopause [Kaseta 1999, Marra 2002, Juutilainen 2005]. An increased inherited risk of diabetic microvascular complications in women has also been reported [Monti 2007].

Men are more susceptible than women to non-diabetic kidney diseases. Sex hormones have been proposed as one explanation [Silbiger 2008]: in young men, high concentrations of androgenic hormones such as testosterone (popularly known as "the male hormone") were associated with a reduction in kidney function [Tomaszewski 2009b]. On the other hand, adverse effects of low testosterone have also been observed [Khaw 2007, Carrero 2009] and the overall associations or mechanisms are difficult to ascertain [Reckelhoff 2005, Yeap 2009]. Some reports suggest that estrogen (known as "the female hormone") may offer protection against vascular diseases in a time-dependent manner in women [Lee 2001, Scott 2004, Teede 2007, Collins 2007], whereas other studies have produced contradictory observations [Joles 1998, Von Hertzen 2004, Tomaszewski 2009a].

Diabetes has a profound impact on metabolism and it is likely that the female protection against kidney and other injury is partially lost in type 1 diabetic patients [Ahlgren 2002, Miller 2009]. In fact, conflicting reports on the differential progression of diabetic kidney disease between men and women have been published [Holl 1999, Zhang 2003, Dahlquist 2008, Maric 2008]. However, it is not yet clear what gender-specific mechanisms are at play during diabetic organ damage.

Even though the absolute number of men with complications is larger, a woman who develops diabetic complications may have a worse prognosis than a man in a similar situation. Furthermore, women and men tend to experience diabetes differently. For instance, fear of low blood glucose episodes (hypoglycemia)

makes some patients cautious against injecting sufficient doses of insulin. The cultural pressure to be lean may entice women in particular to reduce insulin intake even further to avoid weight-gain. This behavior can also lead to eating disorders that further exacerbate the diabetic condition [Goebel-Fabbri 2008].

Men and women live in only partially overlapping spheres of cultural and social influence due to the obvious physical features. This fact in itself should be enough to warrant equal – but distinctive – attention to both sexes with respect to the research, treatment and prevention of complications [Legato 2006, Auryan 2008].

Smoking, alcohol and physical activity

Cigarette smoking has multiple adverse effects at tissue level: carcinogenic molecules are released into the inhaled smoke, and lung carcinomas [Hoffman 2000] remain among the most common forms of cancer in both men and women (Statistics Finland). Other bioactive agents, such as nicotine, alter the function of the cells in arterial walls and modify energy and fat metabolism [Jensen 1995, Dewar 2002, Argacha 2008], which probably explains the increase in smokers' susceptibility to macrovascular diseases [Zhang 2001, Karim 2004, Sharrett 2006, Teo 2006].

Aside from its addictive properties, nicotine in tobacco increases the metabolic rate (acutely) and decreases appetite [Jo 2002, Jessen 2003, Bishop 2004]. Smokers tend to be leaner than non-smokers (and smoking cessation increases body mass), which explains the popular belief of the cigarette as a means to lose weight. However, the long-term connection between smoking and body weight is confounded by numerous other environmental and socioeconomic factors and there is no solid evidence to support the therapeutic aspect [Chiolero 2008].

Smoking and nicotine promote insulin resistance and, accordingly, current or past smoking is a risk factor for type 2 diabetes [Rimm 1993, Eliasson 1997, Eliasson 1997]. For the patients that already have diabetes, smoking cessation is an important mode of intervention to reduce the burden of complications [Howard 1990, Boren 2007]. Arterial stiffening is accelerated by smoking [Failla 1997, Barnoya 2005] and for a patient with type 1 diabetes this may manifest as even higher blood pressure and faster progression of kidney disease [Scott 2001, Safar 2004, Cooper 2006, Shahid 2007]. That said, pinpointing the precise actions of smoking is challenging due to the heterogeneity of environmental factors and difficulties in estimating the long-term smoking exposure.

Alcohol consumption has a non-linear relationship with macrovascular diseases: low daily doses are protective, but larger doses can destroy the benefit [Rimm 1996, Bain 2003]. A u-shaped pattern has also been observed in type 1 diabetes and its complications, including kidney disease and neuropathy [Beulens 2008]. Alcohol can increase insulin sensitivity and usually raises high-density lipoprotein cholesterol [Scragg 2004, Joosten 2008, Kim 2009], both of which reduce the risk of heart attacks and strokes [Mukamal 2005]. On the other hand, alcoholism is a common cause for liver failure [Ishak

1991, Rehm 2003]. Estimating the intensity and type of long-term alcohol exposure is difficult and it is uncertain how much of these effects are caused by confounders, as smoking and excessive drinking co-occur with other bad habits [Godsland 1998, Ahmed 2006].

Regular exercise is a key contributor to good health. The benefits of physical activity include improved insulin sensitivity and reduced fat mass [Frank 2005, Wijndaele 2007], improved vascular function [Mason 2006, Heidarianpour 2007, Herbst 2007], strengthening of muscles and bones [Korpelainen 2006, Hofbauer 2007] and reduced risk of mental illness or cognitive decline [Cotman 2002, Masley 2009]. In type 1 diabetes, the usefulness of exercise is undermined by the kidney, eye and foot complications that in many cases prohibit intensive activity and certain sports, and by the fear of harmfully low blood glucose [Dubé 2006, Iafusco 2006, Zhao 2008]. Not surprisingly, low physical activity has been associated with the established risk factors of diabetic kidney disease [Wadén 2008].

Genes and family history

Diabetic kidney and retinal diseases have an inherited component [Harjutsalo 2004, Hietala 2008] although the exact contributions from genes or a shared environment remain obscure [Krolewski 1999, Pitkäniemi 2007]. A number of candidate genes have been investigated [Boright 2005, Ewens 2005, Fröjdö 2005, Vionnet 2006, Al-Kateb 2007, Al-Kateb 2008], but so far the observed effects have been small. One explanation could be the difficult nature of the disease – diabetic complications are gradual and they develop at an individual rate, which hampers accurate phenotyping. Confounding factors such as gender may also prevent the detection of important effects [Pettersson-Fernholm 2006].

Vascular diseases and diabetic complications have a tendency to accumulate within a subset of families [Fagerudd 1998, Thorn 2007]. Direct transmission of diabetic complications from parents to offspring is difficult to detect since type 1 diabetes has a relatively low and sporadic incidence. However, a number of reports suggest that parental type 2 diabetes, high blood pressure and markers of insulin resistance increase the frequency of complications in type 1 diabetic offspring and/or predict a more insulin resistant phenotype [Seaquist 1989, Borch-Johnsen 1992, Parving 1996, Roglic 1998, Thorn 2009b]. The results may be partly explained by socioeconomic factors: children from well-off families tend to have lower blood glucose, and higher education helps to avoid complications in later life [Hassan 2006, Carter 2008, Nádas 2009].

Linkage and candidate gene analyses have found several genotypes that may be associated with type 1 diabetes complications, but it is likely that many of the genetic traits have not yet been located [Mueller 2006, Tarnow 2008]. A recent report highlighted a non-coding susceptibility locus in chromosome 3, and the results were replicated across Finnish, Icelandic and British type 1 diabetic individuals [Österholm 2007, He 2009]. The effect size was modest (odds ratio 1.33 for the high-risk variant), but nevertheless statistically significant. The func-

tional link between the variant and nephron injury remains to be uncovered.

The success of genomic studies on diabetic complications depends on the accurate characterization of patients. This would involve a long-term follow-up or invasive procedures such as kidney biopsies, arterial imaging or high-resolution retinal photography. Unfortunately, data of that quality are extremely difficult to obtain in human studies. Hence an alternate source of phenotypical detail must be exploited: the hidden physiological processes are reflected in biofluid metabolites, the concentrations of which can be measured more easily.

2.3 Metabolic risk markers

Urinary albumin and kidney function

Healthy nephrons are able to block the proteins and other large particles before they enter the collecting ducts (Figure 2). The filtering process is highly efficient: the daily amount of protein that is circulated within the glomerular blood flow is comparable to a person's body weight (50 - 100 kg), but urinary excretion beyond 50 - 100 mg or 0.0001% can already be considered abnormal [Koolman 2005].

Albumin is the most abundant protein species in the blood and small amounts in the urine (microalbuminuria) can be detected in patients with pre-clinical diabetic kidney disease [Comper 2004, Newman 2005]. However, the natural variance of urinary albumin excretion is proportionally high at low concentrations, which makes it difficult to identify the vulnerable patients early enough [Caramori 2006]. For instance, a non-diabetic individual without any kidney disease may have a transient albumin excretion rate in excess of the clinically significant limit, and therefore the classifications are usually based on multiple successive urine samples [Gomes 2001].

At the cellular level, the podocytes that surround the glomerular capillaries have a critical role in preventing the leakage of albumin (Figure 2). The Finnish congenital nephrotic syndrome is the classic example: an inborn defect prevents the formation of the thin slit diaphragms that normally overlay the small gaps between adjacent foot processes, which leads to a massive flow of protein into the urine and a short life expectancy for the newborn [Huttunen 1976, Patrakka 2000]. Nevertheless, the exact mechanisms and functional structures within the glomerular capillary barrier are not yet fully known and other nephron compartments have also been suggested as primary damage sites in diabetic kidney disease [Vallon 2003, Ozdemir 2005].

Creatine is a molecule that works as an energy-buffer mainly in skeletal muscle; a small fraction breaks down non-enzymatically into creatinine as a by-product of normal metabolism. Creatinine, however, is metabolically inactive and subsequently removed by the kidneys [Chiou 1975, Walker 1990a]. Thanks to the constant rate of "production", creatinine clearance and related formulas comprise the standard clinical measures of kidney function [Hogg 2003, Levey 2003]. On the other hand, the concentration in the blood is affected by the individual's body and muscle mass, which reduces the diagnostic accuracy of the clearance esti-

mates. Moreover, the kidneys not only excrete but also secrete small amounts, which can severely bias the results of urine analyses [Kemperman 1999]. A more robust estimate of kidney function (or glomerular filtration rate) can be achieved by injecting an inactive chemical (such as inulin) into the circulation and then measuring its removal from the body, but this approach is laborious and may cause side effects for the patient [Prigent 2008].

Cystatin C has been proposed as a replacement for creatinine clearance [Mussap 2002, Hoek 2003]. It is a signaling molecule of 120 amino acids and mainly inhibits several enzymes from degrading proteins within and outside cells [Abrahamson 1990, Paraoan 2007]. Cystatin C is almost exclusively removed from blood by the kidneys, and any reduction in the filtering capacity leads to an increase in concentration. Compared with creatinine alone cystatin C is preferred, but the limited accuracy beyond the advanced filtration rate formulas may not warrant the high cost of measurement [Buysschaert 2003, Roos 2007, Tidman 2008]. Finally, Cystatin C is metabolically related to homocysteine, which in turn has been associated with increased risk of large vessel disease [Selhub 1995, Bostom 1999, Wijekoon 2007] and type 1 diabetes complications [Soedamah-Muthu 2005].

Lipids and lipoproteins

Lipids include a diverse group of fat-soluble molecules that are essential structural components of cell membranes, necessary for long-term energy storage and also involved in cell signaling and other metabolic processes. In the context of diabetic complications, triglycerides (triacylglycerols) and cholesterol are the two basic lipids that are routinely measured from the blood in the clinics [Hadjadj 2004, Molitch 2006, Tolonen 2008, Kearney 2008]. The lipid molecules are poorly soluble in water (or in blood), so the absorption and subsequent transportation to tissues requires special macromolecular vehicles [Olson 1998, Kwiterovich 2000].

Lipoprotein particles (Figure 6) consist of a lipid-rich core, and a water-soluble surface – they can thus be transported by the circulation [Taskinen 2003, Kumpula 2008]. Food ingestion is the first stage of lipid entry into the body: triglycerides from animal fats and vegetable oils are packaged into large lipoprotein particles (chylomicrons) in the small intestine, released into the blood stream and subsequently absorbed by fat cells (adipose tissue) and muscle [Phillips 2000, Robertson 2003, Williams 2008]. Lipid supply is not solely dependent on dietary sources. For example, almost all cells can synthesize cholesterol and any surplus excreted into the bile can be reabsorbed in the gut (enterohepatic circulation).

The liver is a secondary source of cholesterol and triglycerides: it replaces the chylomicron remnants left over from the ingested lipid absorption with smaller (but still relatively large) lipoprotein particles, referred to as very low density lipoproteins (VLDLs). Specifically, the liver cells produce apolipoprotein B-100, around which the lipids are assembled [Shelness 2005, Parhofer 2006]. The VLDLs are filled with triglycerides, but also contain a higher proportion of cholesterol in the core compared with the

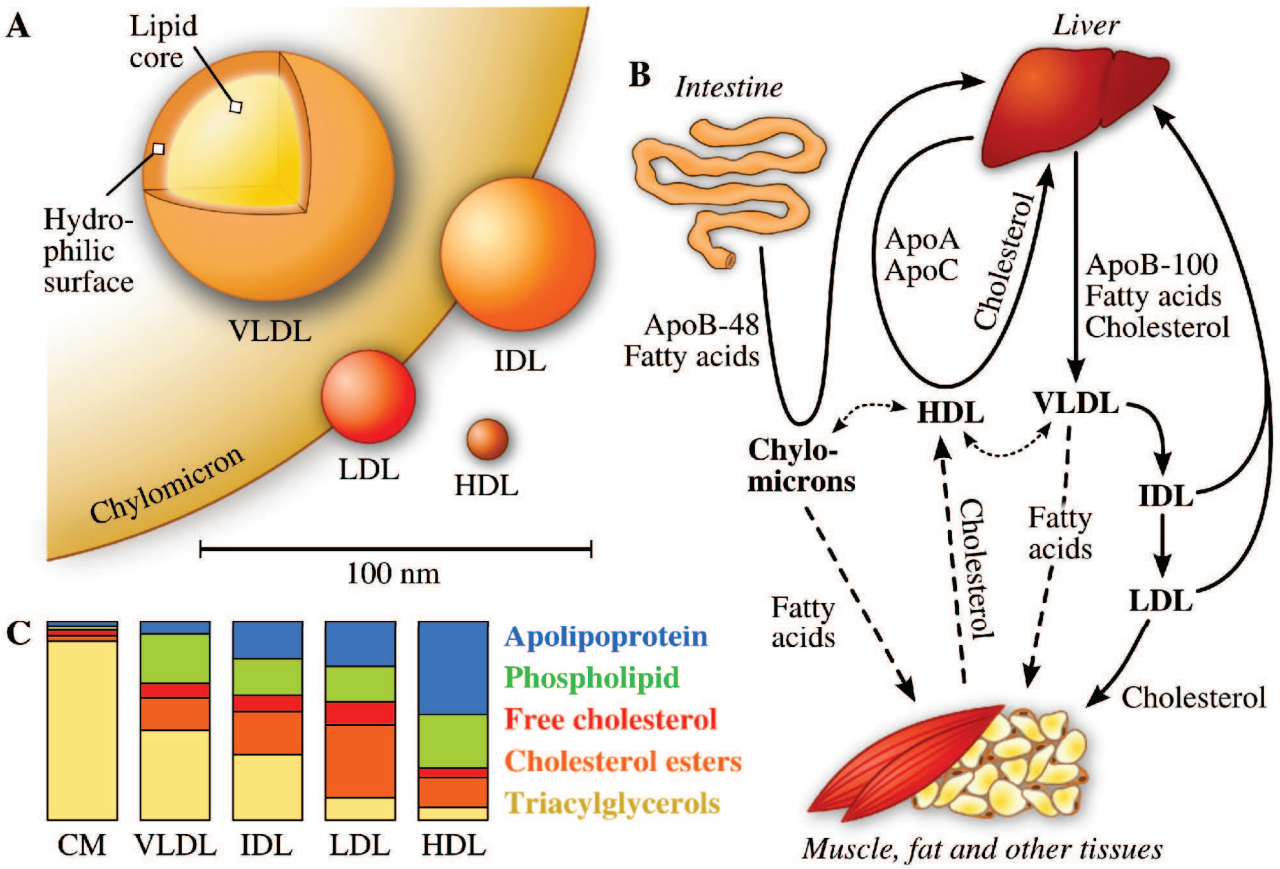

Figure 6: A simplified overview of the human lipoprotein metabolism. **A** Lipoprotein particles vary in size, density and composition: chylomicron diameter is typically within 100-1000 nm and they have the lowest density (<0.95 g/mL), very low density lipoprotein particles (VLDLs, 0.95-1.006 g/mL) measure at 30-80 nm, inter-mediate particles (IDLs, 1.006-1.019 g/mL) at 25-50 nm and low density lipoproteins (LDLs, 1.019-1.063 g/mL) at 18-28 nm. The high-density lipoprotein (HDL, >1.063 g/mL) fraction contains particles 5-15 nm in diameter. In the laboratory, lipoprotein particles are typically isolated into fractions based on their density (buoyancy), which explains the *de facto* taxonomy. The data were obtained from Wikipedia. **B** Lipoprotein metabolism is usually divided into the external pathway (i.e. originating mainly from food intake) and the internal pathway from the liver. In the figure, particle traffic is marked by solid lines, whereas transfer of molecules alone is denoted by dashes. In particular, HDL particles exchange cholesterol, fatty acids and apolipoproteins with the other lipoproteins to render them metabolically active. **C** Molecular components of the main lipoprotein fractions [Koolman 2005].

chylomicrons [Sittiwet 2007]. Again, adipose and muscle tissue are the main destinations for the triglycerides and the particles are gradually reduced to intermediate density (IDL) and then to low density (LDL) particles after the bulk of the core lipid content has been released [Duvillard 2005].

A minority of LDL-like particles contain an apolipoprotein(a) attached to the backbone apolipoprotein B-100, denoted as the Lp(a) class. The concentration of Lp(a) depends heavily on the Apo(a) gene, but it typically comprises less than 10% of the total LDL [Berg 1963, Koschinsky 2003].

So far most of the cholesterol has been preserved inside the lipoproteins that originated from the liver. Rather than merely extract the lipids from the LDLs, the destination cells absorb the entire particles and break them down internally to obtain the cholesterol [Goldstein 1977]. This trait is the reason why LDL and the cholesterol-rich chylomicron remnants are considered the "bad" particles: they can penetrate the wall of large arteries and provide the materials for the abnormal deposits of lipid debris [Proctor 2004, Lyons 2006], which then promote inflammatory responses [Natarajan 2004, Bensinger 2008, Lopes-Virella 2007] and can ultimately lead to the lesions observed in macrovascular diseases (Figure 3). LDL has also been implicated in the progression of kidney disease in type 1 diabetes [Thomas 2006a].

In addition to the VLDL particles, the liver produces the protein components of the smallest particles, denoted by high-density lipoproteins (HDLs), that can take excess cholesterol from peripheral tissues and other lipoproteins back to the liver [Lewis 2006]. Higher levels of HDL-contained cholesterol have been associated with fewer heart attacks and other macrovascular complications, and HDL is often referred to as the "good" particle [Davidson 2007a]. It is also associated with insulin sensitivity, but the causative role of HDL metabolism in the development of type 1 diabetes complications remains uncertain [Chaturvedi 2001, Jenkins 2003, Groop 2007].

Measuring lipoproteins is a challenge [Mora 2009]. They are diverse in qualitative and quantitative characteristics, which makes it difficult to interpret the results [Miljkovic-Gacic 2006, Ala-Korpela 2007, Mora 2007]. In clinical studies, surrogate measures are traditionally used for technical reasons. For instance, the adverse role of LDL is largely based on clinical studies that have measured the cholesterol concentration of the isolated HDL fraction, total cholesterol and total triglycerides. The LDL cholesterol content was then estimated by a mathematical formula [Friedewald 1972]. In fact, this so-called Friedewald formula does not estimate pure LDL by modern standards, but the sum of LDL and IDL cholesterol, which is not appreciated by the majority of clinical literature [Cordova 2004]. One could also argue that the cholesterol content itself is not an optimal biomarker. Ideally, the exact particle size and structure should be used.

Advanced glycation end-products

Mix glucose and protein in water and keep the liquid warm. Eventually, the glucose and protein molecules interact and the affected proteins get irreversibly modified and turn into advanced glycation end-products, or AGEs for short [DeGroot 2004, Horvat 2004]. The process is spontaneous and occurs in every living organism, but in patients with type 1 diabetes the concentration of glucose is higher, which in itself leads to increased opportunity of protein modification in the blood. To make matters worse, cells in the diabetic environment tend to synthesize additional AGEs, which can lead to a further increase of AGE concentrations within the cells [Brownlee 2001]. Glycation is not fast enough to alter most proteins in the body, but for the long-lasting structural components such as collagen the modifications can accu-

mulate with detrimental consequences. Also lipids can be affected by analogous modification processes, and turn into advanced lipoxidation end-products [Baynes 2003].

Hemoglobin in the red blood cells is the classic example of a glycated protein: the proportion of the glycated form A1c is significantly increased in diabetic patients [Bunn 1978]. Since the red blood cells have a limited life span (a few months at most), the hemoglobin A1c proportion is an indirect marker of the total glucose exposure within the last few weeks [Jeffcoate 2004]. Consequently, hemoglobin A1c is often used as a diagnostic tool and a treatment target for adjusting insulin replacement therapy, even if several studies have reported A1c variability beyond mean blood glucose concentrations [Hempe 2002, Kilpatrick 2007a].

In general, AGEs are considered harmful substances in the body [Goldin 2006, Thomas 2005]. A cell detects the presence of AGEs via a set of specific receptor proteins, and the binding of an AGE molecule to a receptor activates secondary metabolic responses inside the cell. For instance, a recent study concluded that an AGE-mediated disturbance to the cellular energy metabolism was hidden until a combination of high glucose and AGE concentrations was introduced, which lead to a significant increase in the intra-cellular oxidative stress or, put simply, an increased risk of harmful molecular reactions [Coughlan 2009]. The receptor for AGE (RAGE) was the key mediator in the process.

With high concentrations of circulating AGEs, the RAGEs proliferate, which may indicate a positive (and possibly uncontrollable) feedback loop of cell signaling [Stern 2002, Mercer 2007]. Increased oxidative stress and other effects of the RAGE have been associated with podocyte cell-death [Chuang 2007], altered signaling in other glomerular cells [Fukami 2004], inflammation [Basta 2002] and reduced adaptation of small arteries to changing blood flow [Linden 2008]. Interestingly, nullifying the cellular receptors with a soluble RAGE (a non-functional decoy receptor) showed a significant reduction in vascular complications in animal models [Hudson 2003].

Signaling molecules and other biomarkers

The cellular processes that drive the tissue damage in diabetic complications are not yet fully understood. Nevertheless, clinical studies have confirmed associations between a number of signaling molecules and kidney disease: C-reactive protein and mannose-binding lectin were elevated (suggesting low-grade inflammation [Østergaard 2005]) in type 1 diabetic patients with kidney complications compared with those patients that had no complications [Saraheimo 2003, Hansen 2004, Saraheimo 2005a], or were lean [Jenkins 2008]. Similar findings have been made with respect to adiponectin [Saraheimo 2005b, Jorsal 2008], a hormone that is correlated with insulin sensitivity and excreted by fatty tissue (adipocytes). The latter finding was unexpected since insulin resistance is considered a co-occurring aspect of diabetic complications, and should have manifested as a low adiponectin concentration [Lara-Castro 2007].

Reactive oxygen species (ROS) are more abundant in diabetic individuals due to the presence of AGEs and altered mitochondrial and metabolic functions in the glucose-rich environment [Nishikawa 2000, Ha 2008]. These molecules are potent damaging agents to DNA, proteins, lipids and carbohydrates and thus pose a threat to tissue integrity and the effects may be further exacerbated by glycation [Jenkins 2004, Forbes 2008]. The ROS can also affect the signaling pathways in the kidney, and this way lead to the typical abnormal growth and extracellular deposits seen in the diabetic nephrons [Lee 2003].

Several biomarker candidates of kidney susceptibility have been proposed, especially in the glomerulus [Wolf 2005]. For instance, the genetic defect behind the Finnish congenital nephrotic syndrome prevents the normal formation of the nephrin protein [Kestilä 1998, Ruotsalainen 1999], which is a critical component of the slit diaphragm between the podocyte foot processes (Figure 2). Altered expression of the protein [Doublier 2003] and increased urinary release [Pätäri 2003] have been observed in patients with diabetic kidney disease.

Also angiotensin II has been implicated in nephron injury; its concentration is increased in the diabetic cellular environment [Zhang 1999, Durvasula 2008], with detrimental effects on nephrin and other podocyte proteins [Hsu 2008, Jia 2008] and complex interactions with increased ROS production [Seshiah 2002, Banday 2008]. In fact, inhibition of the renin-aldosterone system (including angiotensin II) is the pharmacological cornerstone of the current treatments for diabetic small vessel complications and high blood pressure [Bonnet 2001, Parving 2001, Brewster 2004, Balamuthusamy 2008].

At the final stages of insulin production, the hormone is cleaved from a precursor protein (proinsulin), and trimmed down to the final amino-acid sequence and folding structure. C-peptide is the remainder of proinsulin and it is released into the blood in equimolar numbers with the insulin molecule. A direct measurement of blood insulin is difficult due to the highly active nature of the hormone. C-peptide, on the other hand, is less active and can yield more accurate estimates of pancreatic beta-cell function. It was previously considered metabolically inactive, but recent studies suggest that it can have beneficial effects on diabetic complications [Samnegård 2005, Rebsomen 2008]. Furthermore, regulatory effects on several metabolic pathways have been reported [Marques 2004]. However, large-scale longitudinal evidence on the effectiveness of C-peptide replacement therapy is not yet available.

2.4 Metabolite measurements

Metabonomics

The explosion of the “omics” sciences is attributed to the advancements in automated measurement technologies. Geneticists can now measure most of the genomic variation in large populations [Frazer 2007, Novembre 2008], cell biologists have numerous high-throughput methods to detect gene expression and proteins [Young 2000, Omenn 2006] and two technologies have emerged as the primary tools in metabolite research:

nuclear magnetic resonance (NMR) spectroscopy of biofluids is the cheaper, reproducible but relatively insensitive screening platform, while mass spectrometry (with its multiple variants) is the method of choice for new biomarker detection and for tracking minute quantities of signaling molecules [Dunn 2005, Domon 2006, Ellis 2007].

Diabetes is a metabolic disorder, and the study of the metabolite composition from a system-wide perspective has been recognized as a path to discovering the disease mechanisms [Griffin 2006, Kell 2006]. Metabonomics is focused on the characteristics and dynamic responses of metabolism and the subsequent effects on the organism's phenotype [Nicholson 2002, Holmes 2008b]. A more widely used term metabolomics can mean the same thing depending on the context [Gieger 2008], but metabolomic research has traditionally focused on cellular processes and the identification of metabolites, that is, the mapping of the human metabolome. Furthermore, metabonomics is often associated with NMR spectroscopy, whereas the term metabolomics is favored in mass spectrometry studies.

The early applications of proton (^{1}H) NMR spectroscopy in diabetes research date back to the 1980's [Nicholson 1984]. Since then, biofluid ^{1}H NMR has been extensively applied in animal [Serkova 2005, Clayton 2006] and human studies [Kirschenlohr 2006, Holmes 2008a, Tukiainen 2008], and in combination with mass spectrometry [Atherton 2006, Chan 2009]. NMR is a non-invasive non-chemical technique and it can measure lipoprotein particles in their natural state [Ala-Korpela 1994], which has lead to commercial clinical applications [Otvos 1991, Soedamah-Muthu 2003, Klein 2004, LipoScience 2009].

Proton NMR spectroscopy

NMR is a phenomenon at the sub-atomic scale. Atoms consist of a nucleus (which is made of protons and neutrons) and a cloud of orbiting electrons. Some elements, such as hydrogen and fluorine, have an odd number of protons and/or neutrons within their nuclei and in the world of quantum mechanics this means sensitivity to external magnetic fields. Specifically, the nuclei tend to align themselves according to the magnetic forces they are subjected to [Keeler 2002].

The nuclear alignment requires energy, and this energy can be detected indirectly. The principle of an NMR spectrometer is analogous to a church bell. First, a constant stabilizing force is needed: gravity aligns the bell, and an artificial magnetic field does the same for the atomic nuclei. Next, the bell is perturbed by the pulling of ropes – the nucleus is perturbed by radio waves. Some of the energy is consumed in the process as the bell tilts. The quantum unit of the radio wave (the photon) is spent for the nuclear (mis)alignment.

When the ringer stops pulling, the bell swings back and produces an audible sound. Similarly, when the radio signal subsides, the nuclei try to return to their original state. However, the bell keeps on swinging and producing the sound until the extra energy is spent; the nucleus will also continue to precess, and the small

changes in magnetization can be captured and converted to an electric signal.

Church bells have a specific resonance frequency: if the ringer pulls the rope at the wrong time, the rhythm is broken. Exactly the same holds for the nuclear equivalent: only a certain radio frequency will excite a signal from a certain nucleus. The power of NMR comes from the fact that the electron cloud shields the nucleus from electromagnetic fields. The cloud, in turn, is modulated by the surrounding atoms. Consequently, the nuclei of an element behave differently in different molecular environments (the characteristic frequency is shifted) – and these differences can be exploited to resolve the numbers of the target nuclei within specific metabolites. This is equivalent to measuring the metabolite concentration in a known volume of biofluid.

The hydrogen proton (denoted by ^{1}H) is a highly NMR-sensitive stable nucleus. It is also abundant in organic molecules, hence it is the most useful target for NMR spectroscopy in metabonomic studies, although other nuclei can also be used in specific applications [Linden 1997, Komoroski 2000, Eisenreich 2007].

Data processing

The raw signal from an NMR instrument is a time-dependent curve of the observed net magnetization after the excitation of the sample material by pulses of radio waves. To resolve signals from individual metabolites, the data are Fourier transformed to produce a spectral representation. Typical spectra of blood serum and urine are depicted in Figure 7.

The peak areas in the figure represent the total number of target nuclei for the pool of a particular molecular species, but the translation into concentrations is not straightforward. Most NMR experiments are not intrinsically quantitative, and a reference substance is often chosen among the metabolites to specify the measurement unit [Constantinou 2005, Heikkilä 2008]. Another popular technique is to divide the spectral intensities with the total area of the spectrum with the rationale that every sample contains approximately the same amount of metabolic products, albeit at different relative concentrations [Craig 2006]. Quantitative protocols are also possible: an external reference substance can be added and more advanced systems can estimate concentrations directly from the electromagnetic signals in selected applications [Burton 2005, Holzgrabe 2005, Wider 2006].

Peak overlap and misalignment are other challenges for NMR data analysis. Serum contains proteins, lipids and lipoproteins, which produce much wider signals than smaller molecules such as glucose and creatinine [Adosraku 1994, De Graaf 2003]. Inevitably, the signal shapes overlap and form continuous envelopes with no traces of the individual molecules left (Figure 7). Fortunately, the spectra is still useful even if the signals cannot be resolved, since the overall information is preserved in the complex shapes [Petersen 2005].

Each atom within a molecule has a specific chemical shift in the spectrum and the shifts are of crucial importance for the identification and quantification of molecules [Mielke 2005, Jukarainen

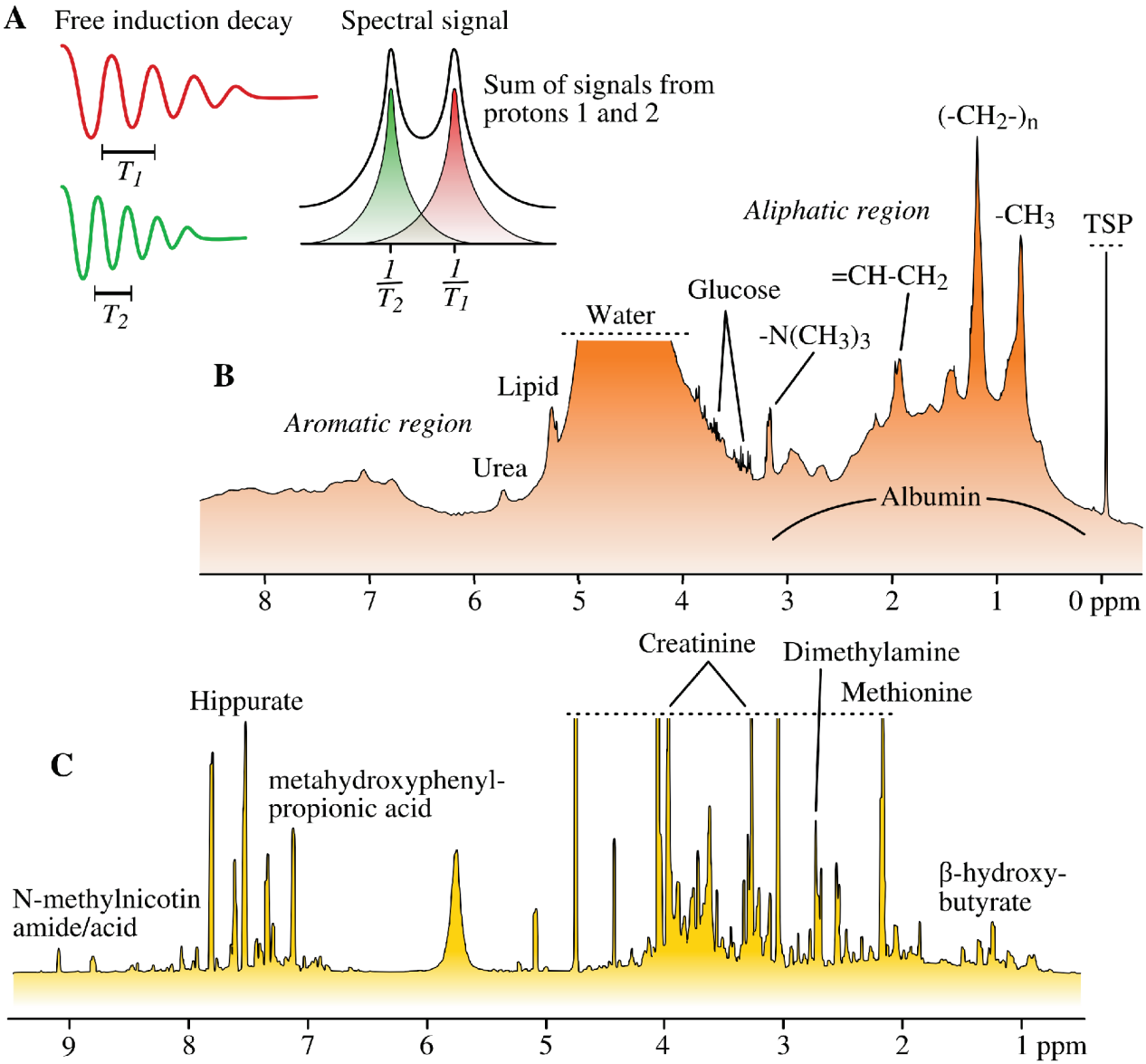

Figure 7: ^{1}H NMR spectra of human biofluids. **A** Each hydrogen proton in a molecule produces a characteristic resonance based on its neighboring atoms. The raw data form a time-series of the net magnetization in the sample. After Fourier transformation, individual protons can be seen as peaks in the spectrum. It is customary to put the lower frequencies on the right on the horizontal axis. **B** Typical serum spectrum, measured at 500MHz magnetic field strength. Note the heavy overlap between signals from lipoprotein particles and macromolecules in the aliphatic region. **C** ^{1}H NMR spetrum of human urine, measured at 600MHz [Salek 2007]. The signals come mainly from small unbound molecules, and the peaks are therefore narrow and easily distinguishable. On the other hand, the acidity/basicity of the sample has a strong effect on the peak positions, which may pose additional problems for urine analyses.

2008]. In biomedical applications, the pH of a sample has a complicated and often confounding effect on the peak shifts, which causes problems particularly in the urine analyses. On the other hand, the extra information can also have biological origins and therefore have a value on its own [Cloarec 2005]. The adverse effects of experimental peak shifts are typically either aligned automatically or circumvented by binning [De Meyer 2008, Veselkov 2009].

2.5 Multi-variate pattern recognition

Statistical modeling

The classical statistical methods in clinical medicine were restrained by the lack of computational power at the time of their conception in the early half of the 20th century [Moore 1965]. Most tools were based on descriptive statistics (the mean and variance are typical examples) that could be calculated with simple numerical procedures [Milton 1995]. The mathematical effort was on the analytical derivation of formulas for statistical significance under a variety of model assumptions so that, given a formula, a researcher could do the calculations by hand. The culture of medical statistics is still rooted in this principle, although the new computationally intensive techniques are becoming more common [Lucas 2004, Petrovsky 2004, Ashby 2006].

The traditional tools of hypothesis testing were naturally suited for reductionist analyses. However, the high-throughput data that became available in the "omics"-sciences required a different approach due to multiple testing problems and the ambiguity of inference targets [Loscalzo 2007b]. Also, univariate statistics were incapable of detecting multivariate effects that are typical for a biological system. Thus there was a need for data condensation: the relevant parts of the observations must be separated from the irrelevant noise to better understand the complex phenomenon under investigation [Jain 2000].

The choice of a pattern recognition method is usually application-specific. The aim of this thesis, however, was not to compare different methods (previous experiences and literature were taken into account when making the modeling choices). Hence only a small subset of techniques are discussed.

Linear projections

Principal component analysis (PCA) is a basic technique for compressing the variation within a multi-variate dataset into a small number of linear components [Pearson 1901, Shlens 2005]. The method can be described by projecting a dataset with three variables onto a two-dimensional canvas. Suppose an irregular 3D object is placed between a light source and a canvas. By rotating the object one can alter the size of the shaded area on the canvas. Similarly, the data points can be represented as dots in three dimensions (one dimension per variable), and the PCA defines the rotation which produces the most dispersed point cloud on just two dimensions (that is, the largest shadow on the canvas).

Mathematically, the principal components (PCs) define the projection from the original data space into a rotated data

space in such a way that the first coordinate axis in the transformed space (the 1st PC in the original space) spreads the samples maximally, the second coordinate axis spreads the samples maximally perpendicular to the first and so on until the number of dimensions is full. Obviously, such structure cannot be directly visualized beyond three dimensions, but the power comes from the fact that the first few axes usually explain nearly all of the variation in the data [Stanimirova 2004, Zhu 2006]. Figure 8 shows the results for a four-dimensional set of socioeconomic data.

Supervised projection methods

PCA is the *de facto* standard of unsupervised analysis in metabonomics [Trygg 2007]. Recent applications include ^{1}H NMR urinary studies of ethnic groups [Holmes 2008a], a comparison of animal and human diabetes [Salek 2007], and various studies on type 2 diabetes, the metabolic syndrome and cardiovascular disease [Lehto 2000, Yang 2004, Wang 2005, Hillier 2006, Kirschenlohr 2006]. In clinical studies, however, the samples are rarely clustered in distinct groups, which makes PCA less useful in characterizing subtle disease states.

To improve detection of weak but biologically relevant signals, PCA is usually accompanied by the supervised projection variants such as the PLS discriminant analysis (PLS-DA). The acronym PLS is derived from "partial least squares projection to latent structures" or it can just refer to the partial least squares algorithm that in most applications iteratively determines the projections [Wold 1973]. Unlike PCA, PLS adapts the rotations so that they best explain the co-variation of the input data and a preselected target variable, even if the total input data variance would be captured inefficiently. Hence the noise (which is related to nothing) or experimental effects (which should not be related to the target variable) are not allowed to mask biologically relevant patterns [Rosipal 2006, Trygg 2007].

Linear discriminant analysis (LDA) is an algebraic approach to estimate the likelihood of a sample belonging to a pre-defined class [Fisher 1936, Martínez 2001, Ye 2007]. The goal is to find a linear combination of the explanatory variables that best separates the samples into two or more pre-defined groups. The dataset is, in effect, compressed much the same way as with PCA and PLS. To cancel out as much noise as possible, the (L)DA can be performed on a limited number of PLS components (hence the acronym PLS-DA), where the target variable for PLS is a binary indicator matrix of the group memberships.

Linear projection methods perform adequately in most situations, and without excessive computational load. The difficulties of classical PCA and PLS stem from the model interpretation. The connections between variables, samples and classification are translated by an *n*-dimensional linear basis, which makes the results less intuitive. In general, the components may not be related to biologically relevant phenomena; especially non-linear effects are often dispersed between a number of PCs. As only three dimensions can be visualized simultaneously, some data is always lost in the presentation. Fortunately the recent adap-

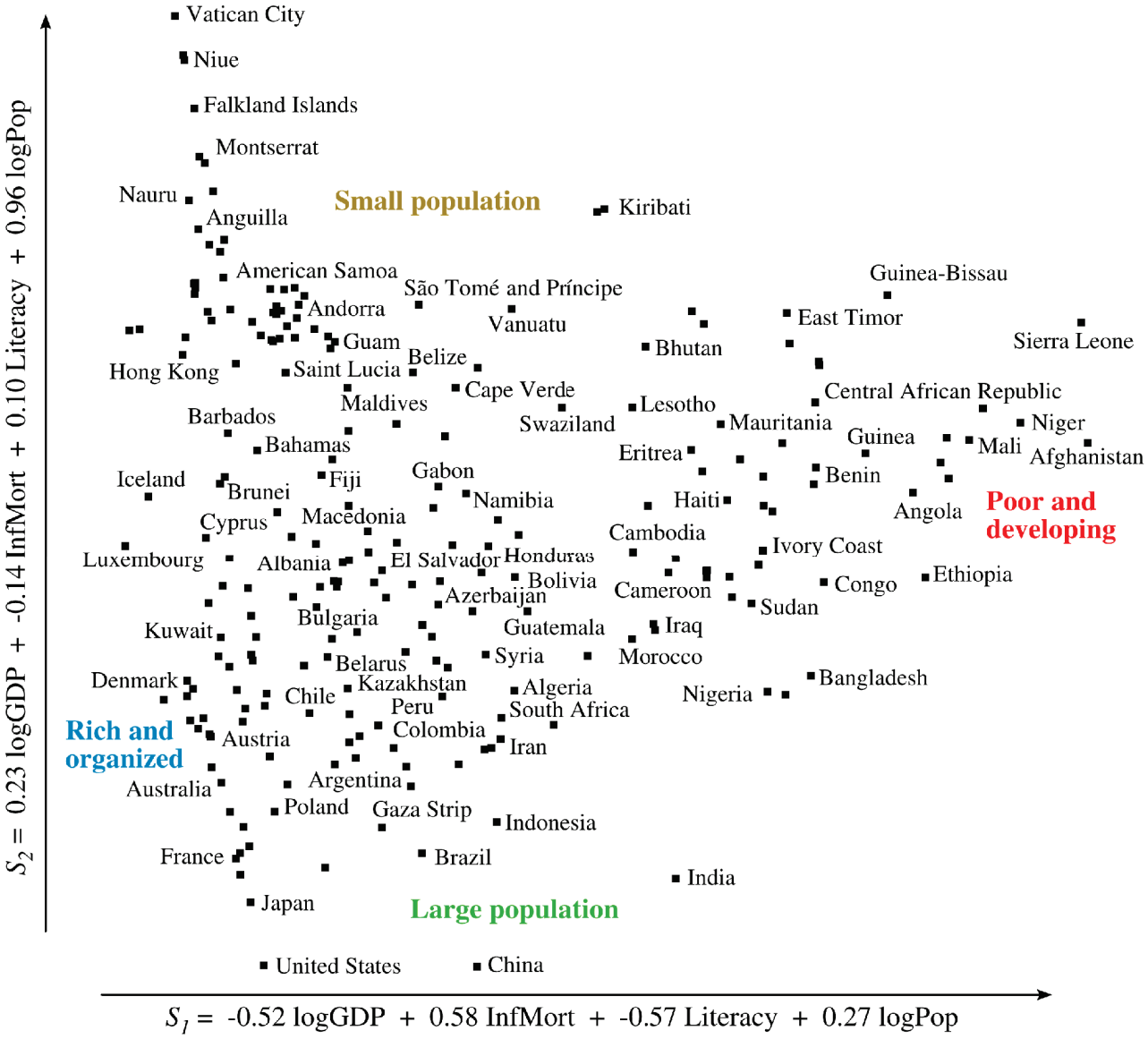

Figure 8: Principal component analysis (PCA) of the countries and autonomous regions of the world. Before analysis, the raw data were normalized to unit variance and shifted to zero mean. The goal of PCA is to show the differences in the overall country characteristics in as few dimensions as possible. The horizontal axis represents the first principal component score S_1 that explains 63% of total dataset variance. The coefficients define the PC vector, and the weighted sum produces the projection onto the PC. The first PC is dominated by the *per capita* gross domestic product (GDP), infant mortality and adult literacy. On the other hand, the second PC score S_2 (explains 23% of variance) is almost exclusively quantified by the population size. The data were obtained from the United Nations (literacy and population), the International Monetary Fund (GDP) and from the US Central Intelligence Agency (infant mortality).

tions of the basic methods (orthogonal PLS in particular) are easier to interpret, especially if tailored for a particular problem [Ergon 2004, Trygg 2002]. Finally, the strict assumption of a linear orthogonal basis for the dataset (combined with Gaussian residuals) limits the model severely and can lead to inefficient solutions in demanding applications [Malthouse 1997, Demiriz 2001, Venna 2001].

Linear regression and kernel models

Excessive number and collinearity of input variables prevent the use of ordinary linear regression in most metabonomics studies, which has lead to the widespread use of the projection methods. However, a regularization technique that effectively assigns a cost to the sum of the regression coefficient magnitudes is another possibility [Tikhonov 1963] and is often referred to as ridge regression. The choice between the two is application-dependent: if the dataset can be reduced to a few orthogonal components efficiently, the projection methods are favored, otherwise the regularized linear regression yields better results [Vigneau 1997, Huang 2002, Burr 2005].

Most noisy and collinear datasets result in a poor recovery of regression coefficients [Farkas 2005, Kiers 2007], even when prediction performance remains strong. Biomarker detection from spectroscopic data, for instance, can thus be compromised. The otherwise reliable projection methods will also fail in this respect if the dataset cannot be described by the first few principal components.

The so-called kernel methods are a way to bypass the collinearity: instead of using the spectra as such, one can decode them into a weighted sum of pre-determined line-shapes or quantities [Opstad 2007, Vehtari 2007, Bylesjö 2008]. The weights for the summation are then determined by linear regression or projection methods. For NMR spectra, a single resonance peak (in the shape of a Cauchy function) is the natural kernel [Ala-Korpela 1995b]. Broadly speaking, the kernel can be almost anything (even wavelets [Liu 2007]) and enables the incorporation of non-linearity within the standard linear framework. Choosing the number and shape of kernels is the critical modeling aspect and often requires detailed knowledge of the phenomenon under study – and should always by followed by computationally intensive procedures to optimize and validate the choices, whatever method was used [Efron 1995, Westerhuis 2008].

Artificial neural networks

The advanced modifications of the projection methods can provide detailed information on the relevance of individual input variables, but if that is of low priority, more powerful "black-box" modeling, such as artificial neural networks (ANNs), can improve the regression estimates for complex non-linear phenomena [Haykin 1994, Alpaydin 1998, Yang 2004]. Biomedical applications of ANNs include lipoprotein quantification from NMR spectra [Ala-Korpela 1995a], the prediction of diabetes [Park 2001], the discovery of signaling proteins in autoimmunity [Honeyman 1998], and the assessment of retinal complications in the eye [Usher 2004] and of plaque formation in the arteries [Ergün 2004].

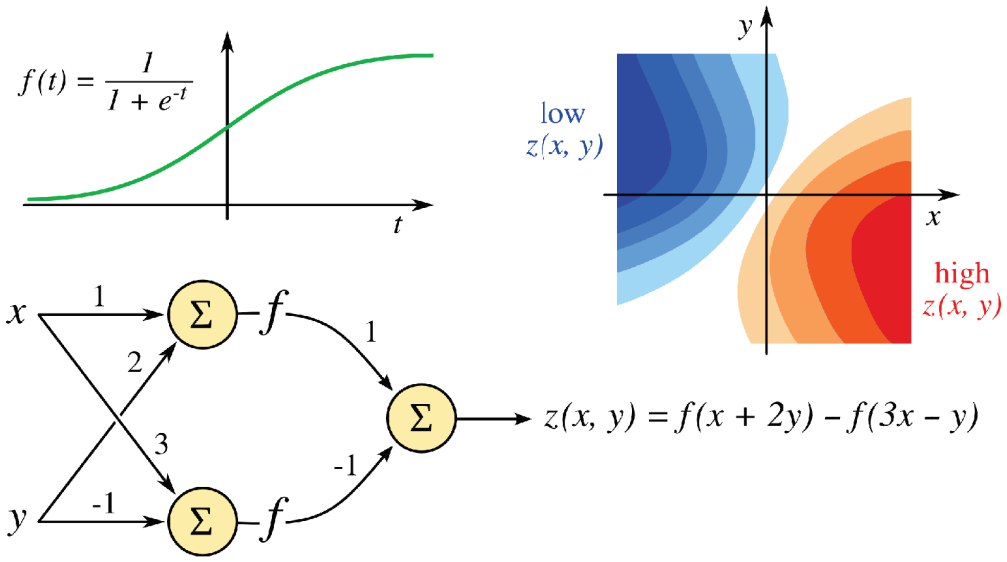

Figure 9: A fully trained multi-layer neural network model (multi-layer perceptron). The first layer consists of a number of input variables that are thought to explain the target variable(s) in the output layer. Here, there are two inputs, denoted by x and y, and a single output variable z. A hidden layer of two additive non-linear neurons (sigmoid response $f(t)$ to the summed signal) is placed in between. The layers are connected by weighted links, where the weights correspond to the coefficients for the summation. Initially, the weights are unknown. The model is "trained" with an iterative algorithm that gradually finds the best set of weigths for the dataset at hand. Finally, the mature neuron is now able to produce an estimate $z(x, y)$ for new values of x and y.

A typical ANN mimics a simple nervous network with a layer of input neurons (one per input variable), connected with a layer of hidden neuronal units, which in turn are connected to a common output neuron that provides the regression estimate (Figure 9). The hidden and output neurons with multiple inputs behave as weighted summing devices, where the weights (the strengths of connections to the previous layer) are tuned so as to produce the least error between the estimated output and observed value [Theodoridis 2003].

Each neuron is a non-linear summing device, which means that the sum of inputs is transformed (a sigmoidal shape is common) before the value is passed on to the next layer. A network with a single hidden unit is equivalent to a generalized linear model, but more hidden neurons result in a highly flexible, albeit complicated, non-linear model. Therefore, the validation and regularization of ANNs is crucial before drawing any inferences [Larsen 1995, Bishop 1996, Amari 1997, Lampinen 2001]. In situations of no added gain from ANNs, simpler models should be chosen [Harrison 2005].

Self-organizing map

The projection methods are built on the correlations between variables, and indirectly model the individual differences between the samples. The reverse approach is also possible: one can start

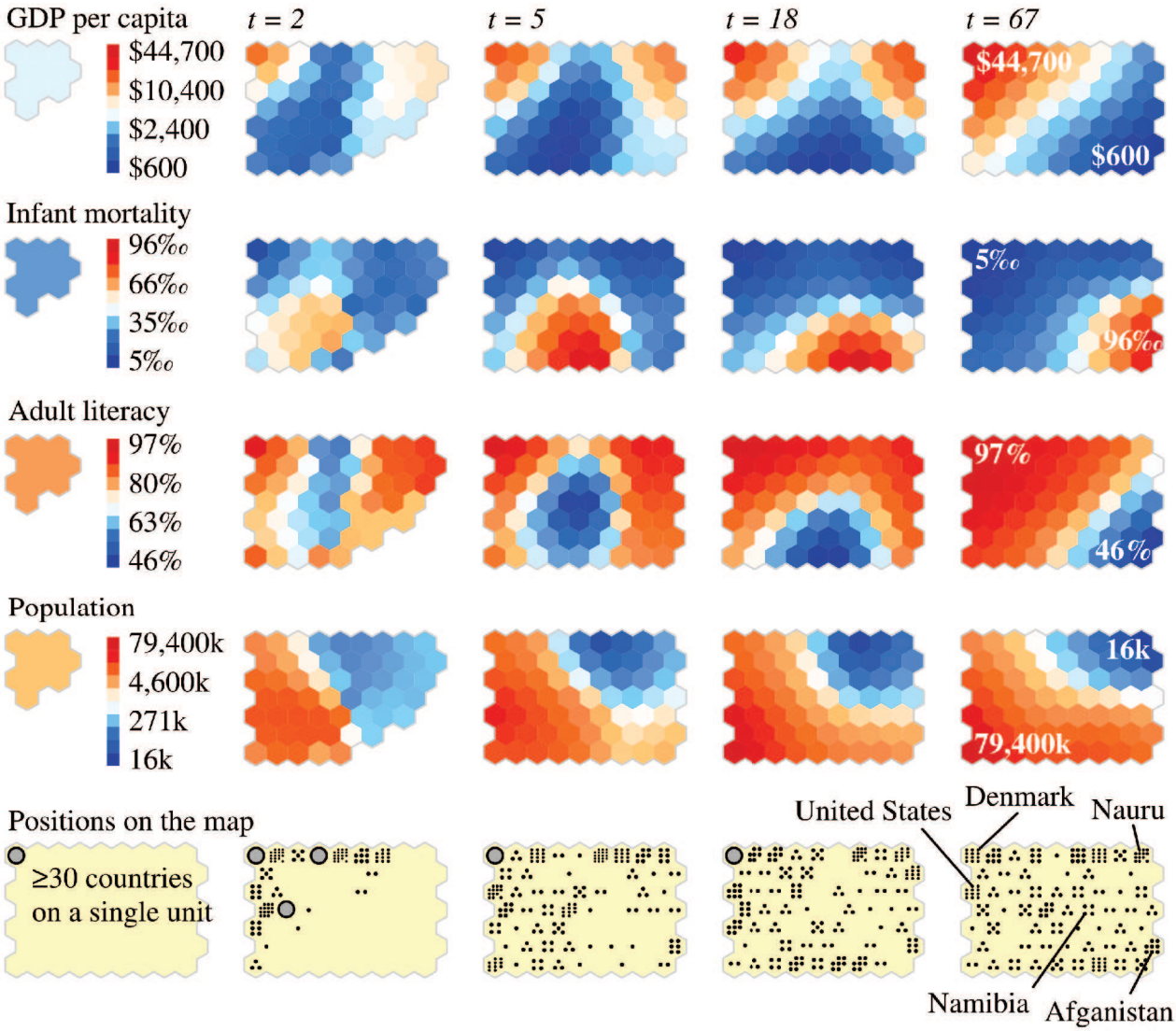

Figure 10: A series of snapshots at different stages of the self-forganizing map (SOM) algorithm. The colored plots indicate the averages in different parts of the SOM for each of the societal and governmental indicators respectively. During each iteration (t), the SOM is updated so as to better reflect the statistical properties of the dataset. This can be viewed as analogous to a plastic neural tissue that learns the regular statistical patterns that connect the stimuli in an unsupervised manner. Simultaneously, each country will have a single unit that best describes it: for illustration purposes only one unit is active in the beginning and thus "collects" all the countries. Later on, the map structure begins to diversify and the countries start to spread on the map. Finally, the SOM stabilizes to an optimized configuration that best separates large and small, poor and rich and developed and developing countries. The dataset is the same that was investigated by PCA in Figure 8.

from the differences between the observed metabolic profiles, and indirectly investigate the associations between the metabolites. The nearest-neighbor (*k*-NN) methods are extreme examples of the latter approach: they try to find a set of model profiles (*k* denotes the number) that divide the samples into groups that are accurately summarized by the models. Put differently, the samples in a group are more similar to the group model than any other model [Cover 1967]. For clustered data, this leads to the detection of distinct sample clouds in the data space, provided that the parameter *k* is chosen appropriately.

The self-organizing map (SOM) shares the ideology of the *k*-NN, but implements an unsupervised neural network to position the samples on a confined lattice [Kohonen 2001]. It was originally conceived to mimic the plasticity and adaptation of human nervous tissue. First, a set of neurons is put on a lattice so that each one is connected to a small number of adjacent neurons. The neurons are considered as memory units that can store a single metabolic profile such as an NMR spectrum, for instance. Initially, the model profiles in memory are arbitrarily determined.

The learning process starts by finding the best match for an observed sample profile among the profiles stored in the neurons. The matching neuron then "responds to the stimulus" by updating its profile even closer to the sample. More importantly, though, the adjacent neurons are affected: they also adapt to the sample profile, albeit not as much as the best-matching neuron. When the aforementioned procedure is repeatedly performed on every observation, the profiles stored in the neurons gradually converge to a stable configuration (Figure 10) that describes the characteristics of the dataset [Bengio 1995].

The SOM is a popular data mining tool: for example, the map illustrated how NMR spectroscopy of breast cancer tissue was able to characterize the metabolic features of tumors [Beckonert 2003]. Similarly, the lipoprotein abnormalities of the metabolic syndrome were visualized by the SOM [Suna 2007]. Other biological applications include the elucidation of clusters in protein-protein networks [Barrios-Rodiles 2005], a committee of SOMs for assessing arterial plaques [Christodoulou 2003] and a clinical study on insulin resistance [Valkonen 2002].

Complex networks

The inter-dependence of variables, weak but multiple effects from unknown molecular processes, and the difficulties in making targeted and yet comprehensive measurements characterize the study of diabetic complications in humans. The gradually progressive diseases produce the typical clinical dataset with a variable number of statistically significant biomarkers and risk factors, but no (obvious) defining indicator that would predict an outcome. Instead, a multitude of subtle defects interact to produce an adverse physical manifestation [Loscalzo 2007a, Pawson 2008].

Complex network theory is built on the notion of emergent complex phenomena from a large number of simplified interactions (Figure 11) between the individ-

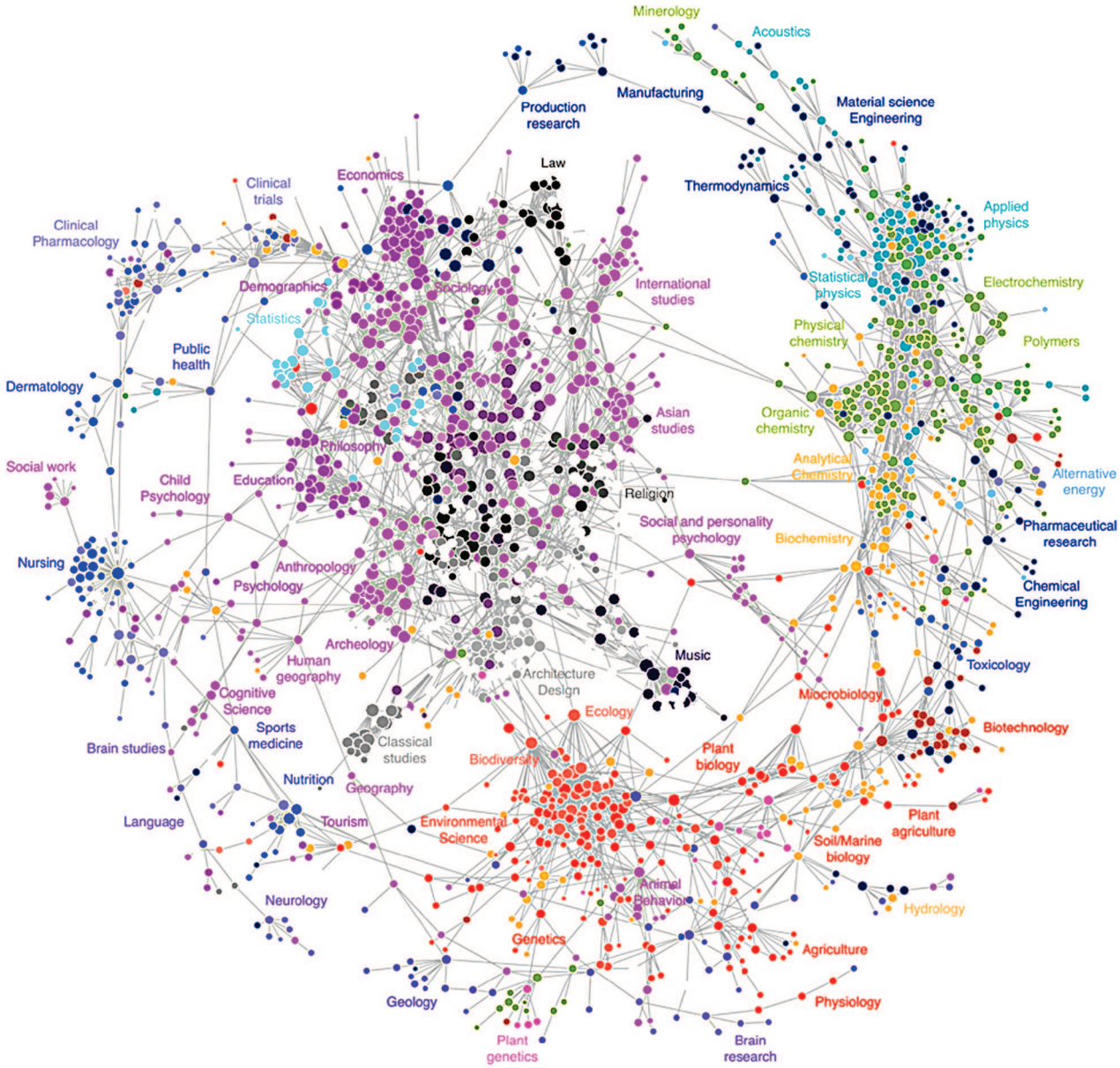

Figure 11: Web of science derived from clickstream data [Bollen 2009]. A website client's usage history was tracked from one journal to another during a browsing session. The resulting clickstream database was analysed to see if there were temporal connections between journals from various disciplines. The network topology was then visualized by drawing the five most strongest out-bound links from a journal. Circles represent individual journals. The lines that connect journals are the edges of the clickstream model. The colors and labels correspond to the Getty Research Institute Art and Architecture Thesaurus classification of the journal.

ual constituents (or factors or nodes) of the system. The methodology has been applied in biology [Almaas 2007], social sciences [Onnela 2007], transportation [Guimerà 2005] and disease spreading [Small 2007], among others. In biosciences, networks arise naturally from gene expression [Benson 2006], protein metabolism [Pieroni 2008] and disease pathology [Goh 2007, Lee 2008]. Interestingly, networks from different fields seem to share similar basic properties: most empirical networks exhibit scale-free, modular and/or so-called small-world characteristics [Barabasi 1999, Girvan 2002].

The term network is widely used in biomedical literature, although not necessarily in reference to graphs. For instance, a structural equation model, conceptualized as a network, revealed the importance of self-efficacy and positive attitude to ensure a good quality of life in the presence of diabetic complications [Rose 2002]. An integrative analysis of cellular studies and publication databases revealed novel "nexus" genes in the development of atherosclerosis [King 2005] and numerous studies have described molecular interactions and network structures relevant to diabetes [Lum 2006, Bergholdt 2007, Ferrara 2008].

Applications of complex network theory are rare in clinical research. This may be due to the bias that favors a "mechanistic" paradigm (characterization of phenotypes from a collection of molecular mechanisms) in the network field, rather than an "observational" (characterization of risk factors from observed patient profiles), which is the natural choice for clinicians. Also, complex networks are more apt at describing the properties of high-dimensional objects – such datasets have only recently become available in large-scale human studies.

3 Aims of the study

I To investigate the multi-variate profiles of serum and urine biochemistry in a large set of patients with type 1 diabetes. The secondary aim was to determine the connections between the representative metabolic profiles, diabetic kidney disease and all-cause mortality.

II To study how to assess the metabolic state of type 1 diabetic individuals by ^{1}H NMR spectroscopy of serum and how this assessment compares to traditional biochemical assays.

III To develop suitable visualization tools for complex clinical datasets and to analyze the statistical relationship between ^{1}H NMR spectra of serum and all-cause mortality in type 1 diabetes.

IV To improve the currently available lipoprotein markers by computational modeling and to verify the clinical relevance of the new estimates in an independent test set of patients with type 1 diabetes.

V To characterize the network of inter-dependencies between clinical and biochemical risk factors for diabetic complications.

4 Materials and methods

4.1 Patients and samples

The Finnish Diabetic Nephropathy Study

The aim of the Finnish Diabetic Nephropathy Study (FinnDiane) is to determine the genetic and environmental risk factors for the complications of type 1 diabetes, especially for the chronic kidney disease. The nation-wide recruitment of patients was launched in 1997 with a specific goal of 25% of all type 1 diabetes cases in the cross-sectional first phase, their first-degree relatives in the second phase and finally re-examination of the original patients in the prospective third phase. At the time of this thesis, 4,200 patients out of 40,000 (10%) have been investigated cross-sectionally, 2,400 relatives have also participated and follow-up data are available on 3,100 patients, although only 1,300 of them have been fully re-examined.

The Division of Nephrology in the Department of Medicine at the Helsinki University Central Hospital was the initial FinnDiane center. Furthermore, all the five university hospitals, all the 16 central hospitals and 56 regional hospitals and health care centers have actively participated in the collection of patient records and biofluid samples. Most of the urban areas in Finland are covered (Figure 12) and initial results from geogenomic analyses suggest that the study set reflects the overall ancestral structure of the Finnish population (unpublished data).

Written informed consent was obtained from every participant and the study protocols were approved by the local ethics committees in the collaborating centers. The FinnDiane study conforms to the Declaration of Helsinki [WMA 2008].

All patients with adequate data available were included in Publications I (n = 4,197), IV (n = 4,084) and V (n = 4,197). A case-control design was adopted in Publication II with a total of 182 age- and sex-matched patients with type 1 diabetes: 73 controls with normal AER, 93 cases with macroalbuminuria, and 16 additional patients with microalbuminuria. The set was expanded to 251 controls, 225 cases and 137 patients with microalbuminuria in Publication III. Basic clinical characteristics for each substudy are listed in Table 2.

Clinical definitions

Diagnostic criteria for type 1 diabetes included age of onset <35 years and transition to insulin treatment within a year of onset. Four patients were excluded because of insufficient biochemical data. The design was cross-sectional (n = 4,197), but with longitudinal records of albuminuria and clinical events before baseline and with all-cause mortality data available after an average of 6.5

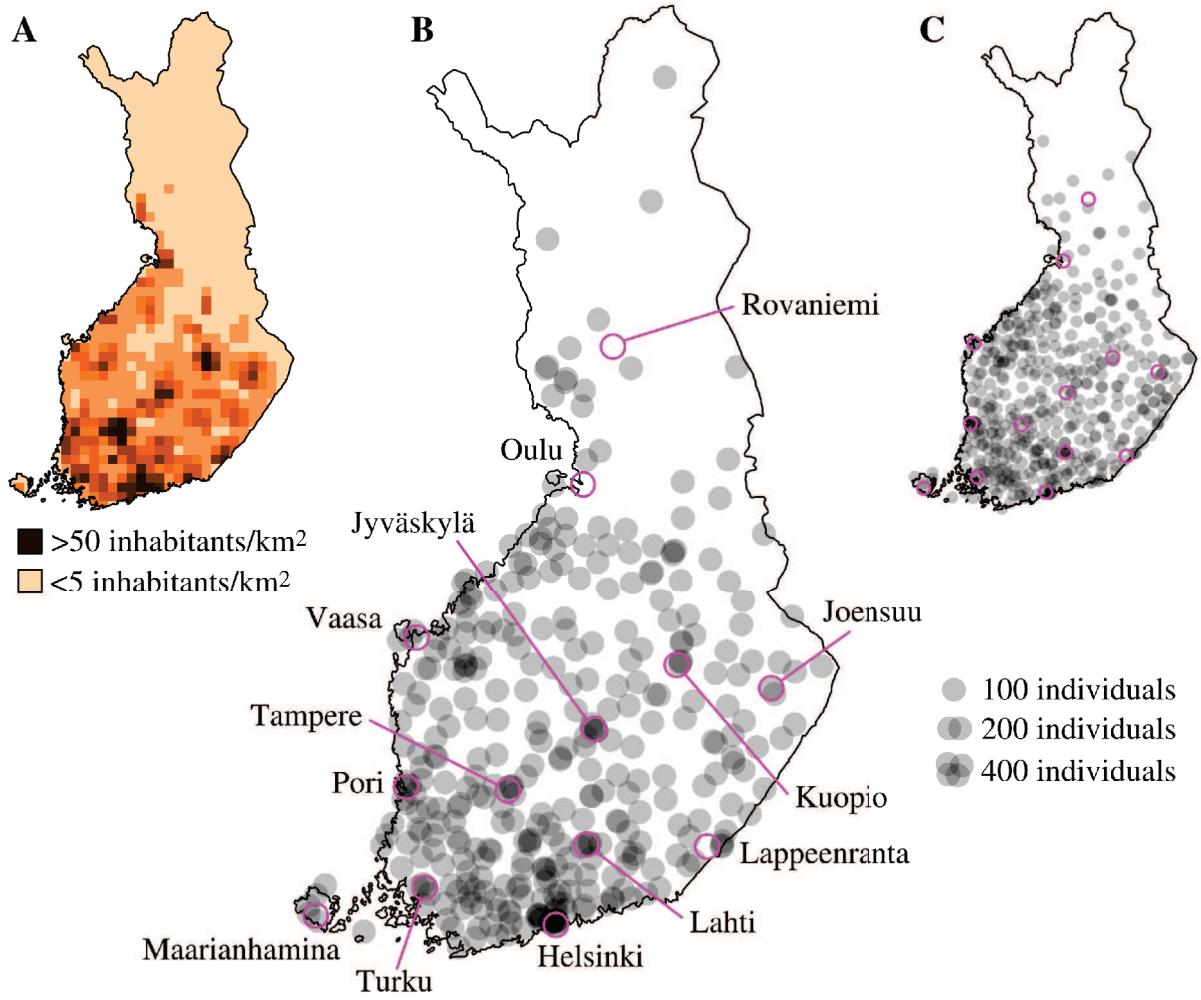

Figure 12: Geographical distribution of the FinnDiane patients. **A** Regional population density in Finland (data from Statistics Finland). **B** Current residence of patients with type 1 diabetes who are participating in the FinnDiane Study (n = 4,130 addresses available). **C** The birthplaces of the patients' parents (n = 5,291 addresses). Each of the circular semi-transparent markers represents a set of 100 local inhabitants.

years of follow-up from the baseline (25,714 patient-years).

Data on medication, cardiovascular status, diabetic complications, education level, smoking dose and alcohol consumption, working status (disabled vs. employed or unemployed), asthma, rheumatoid arthritis and thyroid disease were registered by two standardized questionnaires, one of which was completed by the patient and the other by the patient's attending physician according to the medical file. Vitality status was obtained from the national registry maintained by the Population Register Center of Finland.

The classification of renal status was made centrally according to urinary albumin excretion rate (AER) in at least two out of three consecutive overnight or 24h urine samples. Absence of diabetic kidney disease (DKD) was defined as AER within the normal range (AER <20 µg/min or <30 mg/24h) and at least 15

Table 2: Clinical characteristics of the patient subsets at baseline. Median values ± standard deviation are listed for continuous variables. Blood pressure is abbreviated as BP.

Publication	I & V	II	III	IV
Number	4,197	182	613	4,084
Male	52%	47%	51%	52%
Age [year]	37 ± 12	38 ± 12	39 ± 11	37 ± 12
Diabetes duration [year]	22 ± 12	26 ± 8	25 ± 10	22 ± 12
Waist-hip ratio	0.86 ± 0.08	0.86 ± 0.09	0.88 ± 0.08	0.86 ± 0.08
Systolic BP [mmHg]	131 ± 19	135 ± 20	139 ± 18	131 ± 19
Diastolic BP [mmHg]	80 ± 10	80 ± 11	80 ± 10	80 ± 10
Kidney disease	23%	51%	36%	23%
Retinopathy	35%	56%	51%	35%
Macrovascular disease	8%	8%	9%	8%
Metabolic syndrome	32%	43%	33%	32%
Follow-up time [year]	6.5 ± 5.0	8.5 ± 0.9	8.2 ± 0.6	6.5 ± 5.0
Died within follow-up	7%	10%	9%	7%

years of type 1 diabetes. This kidney disease negative subset is denoted by KD-NEG (Publication V only). Macroalbuminuria or overt kidney disease was defined as AER ≥ 200 μg/min or ≥ 300 mg/24h. The intermediary range was defined as microalbuminuria (20 ≤ AER < 200 μg/min or 30 ≤ AER < 300 mg/24h). Patients on renal replacement therapy (dialysis or transplantation) were classified as having end-stage renal disease (ESRD). An additional subset, denoted by DMDur<15, was formed from patients with less than 15 years of diabetes duration, and normal or unknown AER (Publication V only).

The locally measured AER values were not used for statistical analyses. Instead, the 24h albumin excretion rate was estimated from a single 24h collection, measured by the laboratory of Helsinki University Central Hospital. This continuous variable is abbreviated by 24h-uAlb to distinguish it from the AER estimates made by the local hospitals and health care centers.

The metabolic syndrome was defined as a score of 3 or higher according to the modified National Cholesterol Education Program Adult Treatment Panel III criteria [NCEP 2002, Thorn 2005], where every patient with type 1 diabetes has an initial score of 1 for high blood glucose (hyperglycemia). Diabetic retinopathy was defined as present if a patient had undergone laser treatment of the retina. Macrovascular disease was defined as a pooled end point of coronary heart disease, myocardial infarction, stroke, and peripheral vascular disease. Blood pressure was measured twice with two-minute intervals in the sitting position after a 10-minute rest.

Laboratory measurements

Biochemical data came both from centrally organized measurements (90% of values) and from local health care centers and hospitals (10%). When both were available, the centrally measured value was used. The pattern of missing values was regular (Online appendix 1 in Publi-

cation I), but no significant sampling bias was detected.

Serum lipid and lipoprotein concentrations were measured from fasting blood samples at the research laboratory of Helsinki University Central Hospital, Division of Cardiology. Total cholesterol and triglycerides were determined enzymatically using an auto-analyzer (Cobas Mira or Mira Plus; ABX Diagnostics, Montpellier, France). Total HDL and HDL_3 cholesterol were determined enzymatically using an assay reader (HTS 7000 Plus Bio; Perkin Elmer, Wellesley, MA). HDL_2 cholesterol was calculated by subtracting HDL_3 cholesterol from total HDL cholesterol. LDL_F cholesterol was calculated according to the Friedewald formula (Publication I only). VLDL triglycerides and IDL and LDL cholesterol were estimated by neural network modeling (Publications IV and V). Serum apolipoprotein A-I, A-II, and B concentrations were determined by immunoassays (Orion Diagnostica, Espoo, Finland).

Serum and 24h urinary creatinine (enzymatic), 24h urinary albumin (immunoturbidimetric), C-reactive protein (radioimmunoassay), and C-peptide (radioimmunoassay) were quantified at the Helsinki University Central Hospital Laboratory. Adiponectin and mannan binding lectin were measured as previously described [Frystyk 2005, Thiel 2002]. Twenty-four–hour urinary urea (enzymatic), Na, and K (ion selective electrode) were measured on a Cobas Integra analyzer (Hoffmann-La Roche, Basel, Switzerland) by Medix Laboratories (Espoo, Finland). Glycated hemoglobin (A1c) was determined by standardized assays at local health care centers and hospitals. Serum concentration of the soluble receptor for advanced glycation end-products was measured by solid phase ELISA (Thomas *et al.* submitted).

Regression models of lipoprotein measures

Measurements of VLDL, IDL, LDL and HDL fractions were performed at the University of Oulu. The dataset included 863 individuals with no severe lipid abnormalities (total cholesterol <6.0 mmol/l, triglycerides <2.0 mmol/l) and individuals with familial cholesterol disorders or chronic renal failure, among others. The study subjects were recruited within several clinical studies [Savolainen 1991a, Savolainen 1991b, Hannuksela 1992, Hörkkö 1994]. In total, 1,775 plasma samples were available for the regression analyses.

Advanced lipoprotein measures were estimated *in silico* from total triglycerides, total cholesterol and HDL cholesterol by early-stopping committees of 20 neural networks (multi-layer perceptrons or MLPs for short). First, the original lipoprotein data from Oulu were randomly divided into two sets for each MLP (training and testing, respectively). Next, the MLPs were updated according to the respective training set and validated in the test set during each iteration. The process was halted if the test set indicated any deterioration in accuracy, that is, if the model was over-learning the training set. Finally, the outputs of all the committee members were averaged into a single estimate. The analysis software was implemented (by Mr Niemi) based on the MCMCStuff toolbox for Matlab

(http://www.lce.hut.fi/research/mm/mcmcstuff/).

Cox proportional hazard models of mortality were constructed to assess the biological significance of the lipoprotein estimates in n = 4,084 patients with type 1 diabetes from the FinnDiane Study. A separate age- and sex-adjusted model was fitted for each lipoprotein variable. The variables include the conventional risk markers (total triglycerides and total, LDL_F and HDL cholesterol), the estimates from the neural network modeling (apolipoprotein B-100 and A-I, IDL and LDL cholesterol, among others) and a number of derived indicators that have been previously linked to cardiovascular disease or mortality.

4.2 Proton NMR spectroscopy

Molecular windows

The NMR experiments were performed on a 500MHz Bruker Avance spectrometer located in the Instrument Centre in the Department of Chemistry at the University of Turku. A double-tube setup was adopted: the reference substance (sodium 3-trimethylsilyl[2,2,3,3-d_4]propionate 40 mmol/l, $MnSO_4$ 0.6 mmol/l in 99.8% D_2O) was placed in a separate small tube inside the sample container to prevent molecular mixing from disturbing the serum metabolite signals. Consequently, the results can be quantified with respect to the known volume and concentration of the reference substance. Before analysis, aliquots of 430 μl were extracted from the thawed serum samples and placed inside an automatic sample changer. All experiments were preformed at 37°C to mimic the physiological state of the serum macromolecules [Ala-Korpela 1995b].

The standard ^1H NMR spectrum of serum contains a complex envelope of signals from lipoprotein lipids, albumin-bound fatty acids and smaller features from a number of abundant molecules such as glucose and lactate (Figure 13). For this "LIPO" window, 128 transients were collected with a 90° flip angle, a 6.2 s acquisition time and a 0.1 s relaxation delay.

While the sample remains in the spectrometer, it is possible to change the relative visibility of metabolite signals by manipulating the electromagnetic pulses that excite magnetization in the target molecules. The low-molecular-weight or "LMWM" data were collected with a standard one-dimensional Carr-Purcell-Meiboom-Gill pulse sequence [Carr 1954] with a 325 ms T_2-filter and a fixed 400 ms echo delay to eliminate diffusion and J-modulation effects. Forty-eight transients were collected after 16 dummy scans with a 6.2 s acquisition time and an 8.7 s relaxation delay. The signals from macromolecules such as lipoproteins are proportionally more suppressed by the aforementioned sequence, so the spectral peaks from smaller molecules become visible.

Preprocessing

The spectra were obtained without water suppression at the instrument level. Water is by far the most abundant compound in blood serum, and thus produces a very strong background signal, which disturbs the nearby metabolite peaks from glucose (3.1-3.9 and ~5.2 ppm), lactate (1.23

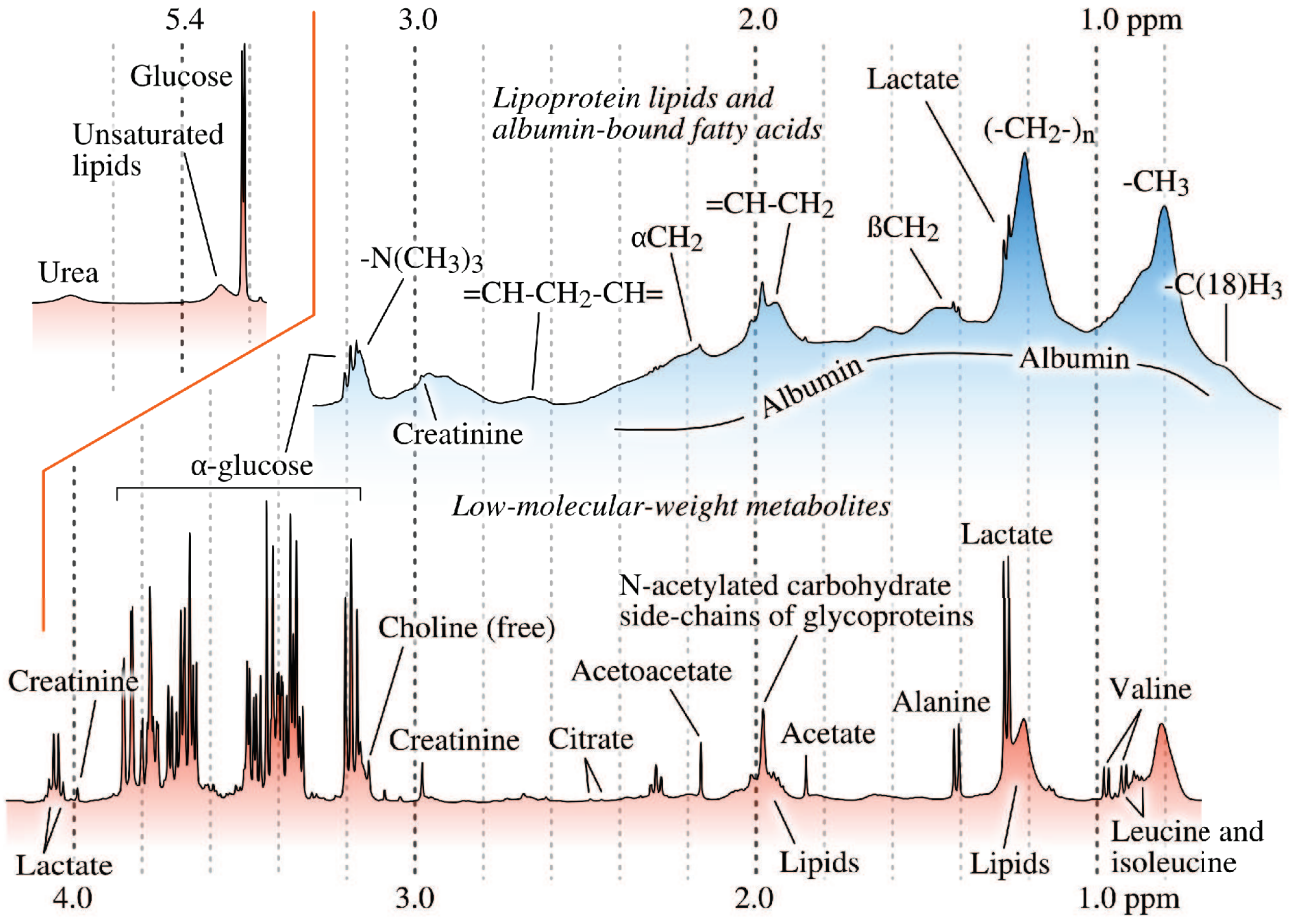

Figure 13: ^{1}H NMR spectra of serum obtained by a protocol with two molecular windows. The top-right spectrum represents the standard experiment with broad line-shapes from lipoprotein lipids and serum fatty acids. The bottom and left plots depict a T_2-filtered spectrum, where the broad resonances from macromolecules have been almost completely suppressed.

and ~4.1 ppm) and creatinine (~3.0 and ~4.0 ppm) among others. Before statistical analysis, the aliphatic tail of the water peak in the LMWM window was eliminated by fitting a Lorenzian curve and then subtracting it from the original spectrum. The aromatic tail was removed by piece-wise linear interpolation. A minor correction was also applied within 3.0-4.5 ppm in both molecular windows.

Selected metabolites (glucose, lactate and creatinine in particular) were quantified by estimating the total area under the corresponding peaks in the LMWM window. In theory, these areas can be normalized to absolute concentration units by knowing the suppression coefficients from the applied pulse sequence. Here, however, these coefficients were not readily available so the reference peak area (set to 0 ppm) was chosen as an arbitrary – but proportionally valid – concentration unit.

4.3 Self-organizing map

Dimension reduction

The self-organizing map is a two-dimen-

sional visualization algorithm for multi-dimensional datasets such as the biochemical data of the type 1 diabetic patients in this thesis. The algorithm is patient-centric by nature, that is, the mathematics are based on comparing the patients' metabolic profiles directly rather than relying on the statistical associations between the biochemical variables. The patients (or samples to be accurate) are spread on the map and their mutual positions indicate how similar they are to each other: adjacent patients on the map are relatively similar with respect to the metabolites, whereas patients that are far apart have differing metabolic profiles.

The finished SOM is analogous to an ordinary geographic map in the sense that the patients have a fixed "place of residence", just like the residents in a city have postal addresses that translate to two-dimensional coordinates. Consequently, it is possible to represent the regional characteristics of the patients on the SOM in the same way as one would characterize the different neighborhoods in the city. For instance, in a large city, one could find areas of high average income, education level and real estate prices in the center, and less affluent demographics in the under-developed outskirts. Similarly, old and obese patients are typically concentrated on different areas than lean young ones on the SOM due to the distinct metabolic profiles that reflect old age and excess body fat [Valkonen 2002].

Topographic statistic

Geographic maps can be colored according to the regional demographics and the same approach is directly applicable to the SOM. There is, however, a methodological caveat: for finite datasets there will always be some regional differences, even if there were no real statistical associations. A real-life analog would be to create a microscopically perfect planar surface on a pile of gravel. Hence it is difficult to determine a suitable color scale or dynamic range that would not give falsely significant patterns for purely random fluctuations. For instance, a simple adjustment by the variance of a variable would not accurately reflect whether the observed coloring were reliable or not, since statistical significance is also influenced by the number of data.

In this thesis, the problem of dynamic range was solved by random permutations according to a hypothesis-based, so-called frequentist, approach. The two challenges in any frequentist approach is to find i) a suitable test statistic that compresses the measured phenomenon into a univariate value and ii) the null distribution of the statistic in the random case. Here, the regional variability is the phenomenon to be measured.

A suitable statistic should be insensitive to orientation; whether the obese patients are clustered on the northern or southern SOM half is of no importance (independence of orientation). Furthermore, the mean value of the variable should not influence the results (independence of level). For instance, it would not matter whether body temperatures were given as degrees centigrade or Fahrenheit. It turns out that the regional variance, taken literally, is a robust and sensitive statistic that fulfills both requirements: if one takes the SOM unit averages, collects them in a column (independence of orientation)

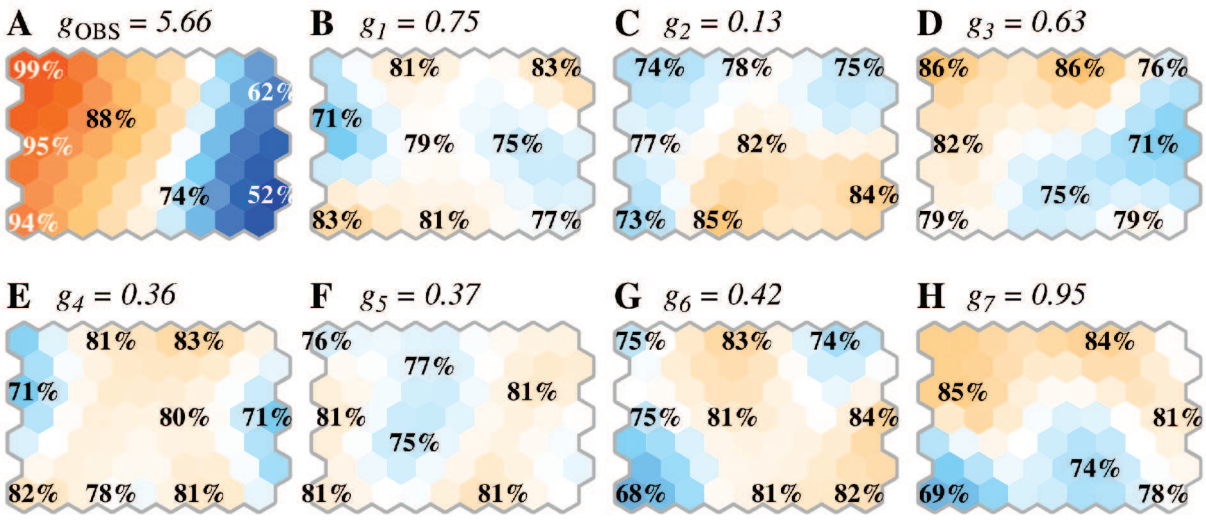

Figure 14: Estimation of empirical *P*-value for a variable that is not included in the SOM training set. **A** The map was created from the dataset of world countries: income level, literacy, infant mortality and population size were included in the training set (Figure 10). The SOM was then colored according to the availability of fresh drinking water (data from the United Nations). **B**-**H** During each permutation, the original water data is shuffled randomly and the map recolored (the positions of the countries remain the same). Also, the variance of the unit-specific values is estimated and represented by the normalized *g*-statistic. Here the observed statistic is larger than any of the seven permuted versions, so the upper limit for the estimate is 1 out of 7 or $P < 0.14$. A large number of permutations are typically required to attain accurate *P*-values.

and then calculates the variance over the values, one effectively estimates the volumetric space left between the mean elevation (independence of level) and the measured "gravel surface".

Statistical significance

If one assumes that a particular biochemical variable has no impact on the patients' positions on the map, then the regional averages (for the variable in question) on the SOM must come from purely random effects. Before the advent of computers, these random effects were estimated by mathematical formulas and rigorous constraints. On the other hand, non-association between a variable and the positions can be achieved artificially by randomly permuting either the positions or the variable. Therefore, it is possible to computationally simulate a situation of no statistical significance (null hypothesis) by randomly permuting the variable before calculating the regional averages on the SOM.

One simulated case of the null hypothesis is not enough, but a large number must be generated to accurately determine the null distribution, which then gives an estimate of the dynamic range of the random regional variation on the SOM, automatically adjusted by the map and population sizes. Furthermore, the procedure can be extended to data that were not included in the training set. In Publication I, for instance, the SOM was created based on the biochemical data, but the main results were obtained by color-

ing the map according to clinical endpoints and mortality – with statistical significance estimates directly available from the null distribution. An example of the permutation process is shown in Figure 14. The SOM statistics and visualizations are implemented in the Melikerion software package.

4.4 Association networks

Correlation matrix

The SOM is an excellent tool for investigating the diversity of metabolic profiles. However, it does not provide direct results on the correlation or association structure between variables. To complement the patient-centric analysis, a network based visualization approach was developed to characterize the inter-dependencies between clinical and biochemical risk factors for type 1 diabetes complications.

The size of a correlation matrix increases by the square of the number of variables. In clinical datasets it is typical that every variable is at least mildly correlated with every other variable. From a visualization perspective this is a problem since depicting an *n*-dimensional correlation structure in only two dimensions (printed media) leads to a complicated presentation that may be too difficult to interpret. Figure 15 depicts the correlation matrix of lipoproteins and their structural characteristics. The variables are grouped according to an existing taxonomy, which greatly helps in identifying functional blocks and biologically significant connections. In many other applications, however, such grouping is not known beforehand or is difficult to achieve by computational means. Also, the large size of the matrix may create typographical problems.

Network visualization

If presented with an excessively complex entity, the only choice is to simplify its presentation. One solution is to group certain variables together and present them as a single trait; these type of approaches are often based on hierarchical clustering algorithms [Ward 1963]. Reducing the number of correlation coefficients is another option. If the variables are regarded as nodes in a network and the connection strengths between the nodes are quantified by correlation coefficients, then the reduction of connections corresponds to pruning the link topology.

The pruning algorithm should preserve strong links, that is, strong correlation coefficients should not be removed first. On the other hand, heavily intra-connected cliques should not cause the fragmentation of the network into small dense islets of connections. Therefore, some weak links should also be preserved.

A spanning tree is an appealing construct to ensure that the network remains connected. The concept is also well suited for correlation networks since the tree can be chosen such that the sum of the link weights (i.e. correlation magnitudes) is maximized (denoted here by maximal spanning tree). By definition, a tree is an acyclic subgraph that connects all the nodes in the network, which means that with *n*-variables, a correlation tree contains exactly *n - 1* links. The tree contains

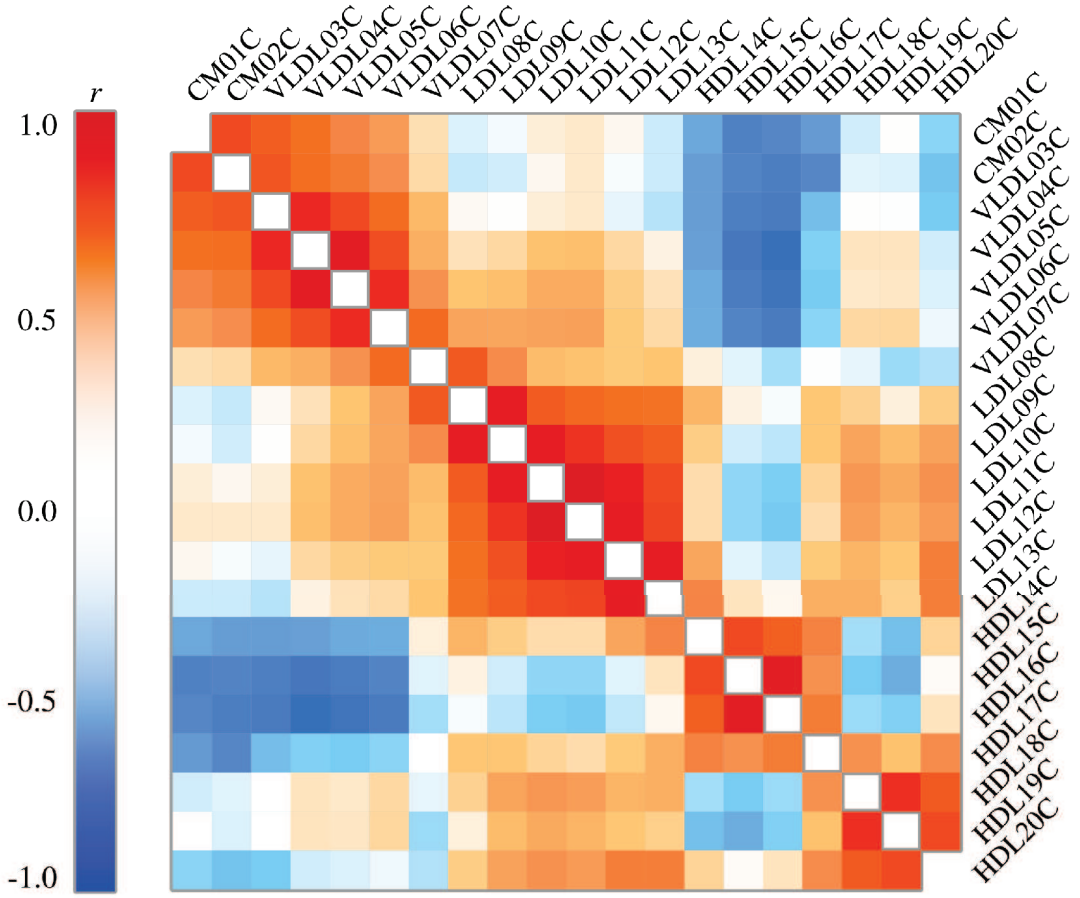

Figure 15: Spearman cross-correlations of lipoprotein cholesterol content in a set of 100 human serum samples. Each pixel corresponds to a pair of variables, determined by the row and column. The color of the pixel depicts the correlation coefficient between the two variables. The data were measured by the LipoSearch method (Skylight Biotech Inc). The subclasses from 1-20 follow the LipoSearch taxonomy (based on particle size) and are not directly comparable with the commonly used notation.

only branches so it is difficult to ascertain if multiple variables have a high degree of inter-connectedness. In some sense, this is the opposite case to the small islets with dense internal connections.

Augmenting the first spanning tree with a second is a compromise between the two extremes. Assuming that the network is fully connected (correlations are non-zero), then removing the first maximal spanning tree from the network does not disconnect the variables. Another spanning graph can now be calculated for the remainder and so on until the desired level of connectedness is achieved. In most situations, two spanning trees combined is enough for a visually pleasing layout (Figure 16).

Pruning the network is not enough to draw the picture. Positioning the nodes in an optimal configuration with as few intersecting links as possible is the final step. Here, a hybrid algorithm based on force-directed spring kinetics and simulated annealing was used (the Himmeli

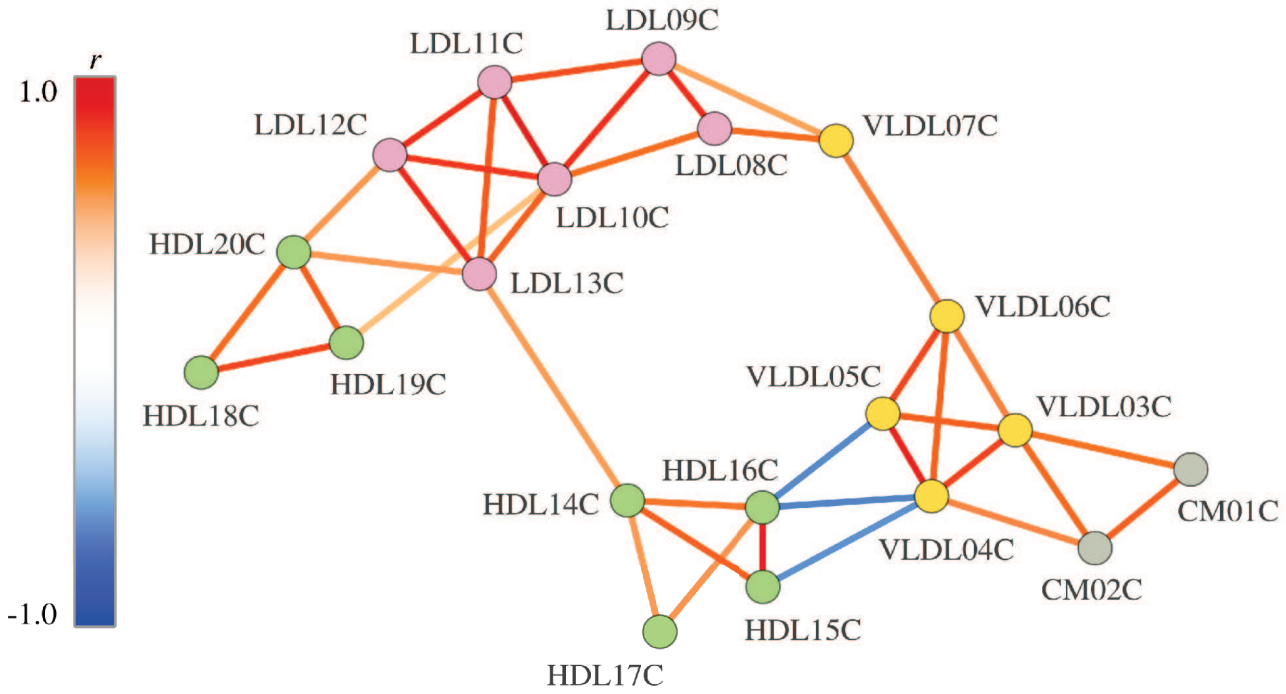

Figure 16: Pruned correlation network from the correlation matrix in Figure 15. The nodes represent lipoprotein cholesterol and the links depict the topologically important correlations. The colors of the links indicate the respective correlation coefficient. Compared with Figure 15, the network visualization is more compact and highlights the connections between the blocks more clearly. For instance, it is easy to see that the cholesterol in larger HDL-particles is inversely correlated with VLDL cholesterol whereas the smaller particles are associated with the LDL particles. The numbers in the abbreviations indicate ordering by particle size.

graph visualization software).

Statistical robustness

The networks of continuous variables were based on pair-wise Spearman's correlation coefficients. It is a challenge to isolate the influence of methodological, physiological, pathological and data collection from the final network topology. Hence it does not make sense to investigate whether the observed structures are non-random – they inevitably are but mostly for the wrong reasons. Instead, by dividing the material into subsets and then comparing the differences between the subset correlation networks it is possible to eliminate the irrelevant phenomena and focus on those that are related to the subset division.

Comparisons of subset networks (formed according to the kidney disease status) were validated by permutation analysis. First, two subsets were selected from the dataset, designated as cases and controls. Next, the difference network was calculated (observation). Then the null distribution of differences was simulated by randomly shuffling the case-control labels 10,000 times (effectively creating a large number of random subsets), and each time recording the differences. Finally, the observation was compared with

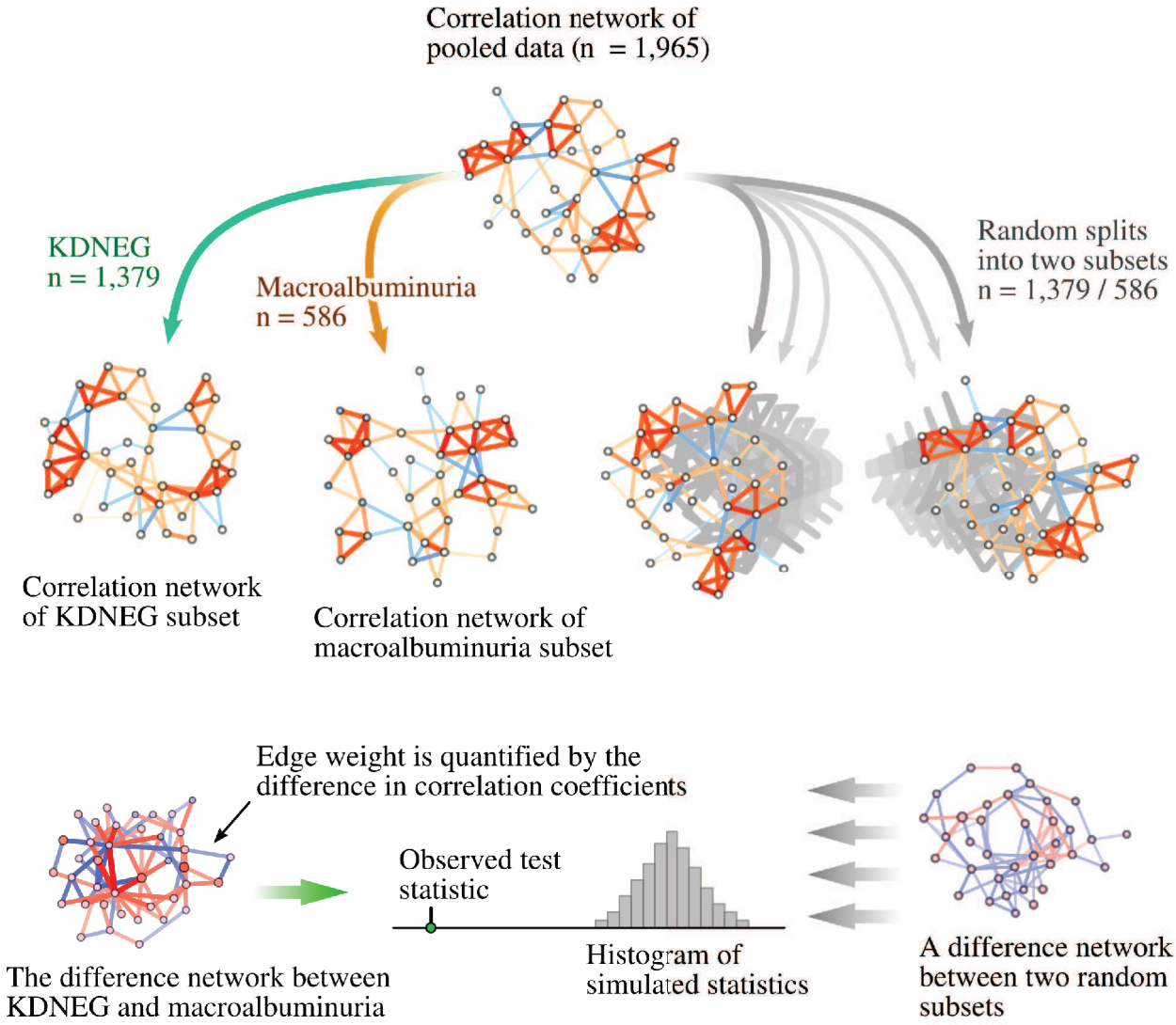

Figure 17: Permutation process for subset graph comparison. Type 1 diabetic patients without kidney disease (KDNEG) and patients with macroalbuminuria where chosen for analysis, as described in Publication V. First, the correlation networks of the two subsets are calculated and then the observed difference network is created by subtracting the KDNEG correlation coefficients from the corresponding macroalbuminuria coefficients (left side of the figure). The variance of the differences was chosen as the test statistic. Next, the process is repeated for a large number of random divisions of the pooled dataset to determine the null distribution of the test statistic (right side). Finally, the observed statistic is compared against the histogram: a clear separation can be seen and thus one can conclude that the division by kidney disease makes a statistically significant difference network.

the simulated distribution to obtain an estimate for *P*-value (Figure 17).

Non-continuous data

The correlation coefficient is suitable for measuring the linear dependence between two continuous variables, but medical datasets usually contain both quantitative and qualitative traits. A real-valued variable has a higher "information density" than a nominal or binary trait, so one can convert categorical data into continuous variables without the loss of precision – at least from a computer programming perspective.

A Bayesian probit model was separately created for each binary variable [Johnson 1999]. Standard logistic regression was not numerically stable due to the high number of dimensions. The binary target variable was estimated based on the real-valued variables, and then the linear predictor of the model was stored as a surrogate continuous trait. The procedure was repeated for every variable (least-squares ridge-regression for the continuous) to ensure that the binary-binary links were comparable to binary-continuous and continuous-continuous links. To reduce artificial inflation of correlation, the dataset was divided into two halves, one of which was then used to predict the other.

After the binary and continuous data were converted to linear predictors, the network was constructed by calculating the correlations between the predictors instead of the observed values (regression-correlation). Permutation analysis of statistical significance was not feasible due to the excessive computational load.

4.5 Software for clinical data analysis

Melikerion

The integration of numerous sources of biochemical and clinical information sets a challenge to the medical scientists who are striving for a deeper understanding of human disease mechanisms. Melikerion is an implementation of the Kohonen self-organizing neural network algorithm, and was designed for unsupervised analysis of clinical materials, in particular. The goal was to make an easily accessible web-based system that can read tabulated data and cope with missing data without the need of excessive manual preprocessing of the files (visit www.finndiane.fi/software/ for more details).

A typical process flow - with emphasis on the biochemical profiles - begins with the preparation of the measurement data so that the variables become comparable in scale and mean value. Next, linear decomposition is applied to create an initial map layout of patients. The final layout is achieved after several iterations of a batch version of the Kohonen algorithm. Once the map is complete, the full dataset is visualized based on the biochemical profiles, with statistical significance estimates for the clinical variables.

The first version of the software was based on the SOM Toolbox [Vesanto 2000] on the Matlab programming environment (Mathworks Inc., Natick MA, USA). However, it was necessary to create a stand-alone package for maximum usability in the academic field and to enable the web-based service. Octave by Eaton JW *et al.* is an open-source clone of Matlab so it was chosen as the soft-

ware platform for the current version.

The Matlab/Octave environment is designed for matrix algebra and the current implementation is programmatically inefficient when it comes to certain types of tasks that involve element-by-element operations of multiple data arrays. The next generation of Melikerion, currently under development, is programmed in C/C++, which is better suited for the most time-consuming components of the software.

Himmeli and Katiska

A general graph constitutes a multi-dimensional object so to draw it on a flat surface involves a coarse simplification of its nature. The Himmeli software is designed for the most difficult cases of weighted and densely connected graphs were no regularities can be exploited. Consequently, large graphs take a long time to process and the textbook examples of exact combinatorial constructs might not be visualized optimally.

The core algorithm is a mix of simulated annealing and molecular dynamics and employs a cell grid memory structure to reduce unnecessary computing [Fruchtermann 1991, Davidson 1996]. The nodes of the network are regarded as particles that repel each other according to a non-linear function of distance. The links correspond to strings that work as opposite forces and try to pull the nodes together.

In a closed system, the nodes and links would oscillate indefinitely. Hence an annealing process is imposed upon the network: first the nodes are free to move larger distances, but gradually the maximum move per iteration is limited until a near-zero value is reached. One can think of this as putting the network in a viscous fluid that ultimately drains the kinetic energy and the structure is stabilized to an optimal configuration were all the forces cancel each other out. The starting configuration is critical for reaching a good node layout. Himmeli calculates the spanning tree of the graph to create a planar connected subgraph and then employs the Walker's algorithm to create the initial layout [Walker 1990b, Mäkinen 2005].

Katiska is a web-based interface to make the visualization of correlation networks easier. It works according to the same principles as the Melikerion software and accepts incomplete datasets. An Octave-based prototype is available on the web site (www.finndiane.fi/software), but the final implementation will be in C/C++ and will also feature the statistical significance estimates for subset networks.

Lipido

The neural networks in Publication IV were trained with the MCMCStuff package [Vanhatalo 2006] for Matlab. A full open-source version for the web server was not feasible due to dependencies on proprietary toolboxes, but the neural models were portable from Matlab to Octave. The online system provides estimates for VLDL triglycerides, IDL, LDL and HDL_2 cholesterol, and apolipoproteins A-I and B-100 from the measured total cholesterol, HDL cholesterol and triglycerides. The service is available at www.finndiane.fi/software and www.computationalmedicine.fi.

5 Results and discussion

5.1 Metabolic phenotypes and mortality

Gender effects

The observed metabolic profiles were gender-specific. For instance, women are shorter and have larger waist-hip ratio on average. Their lower body mass also produces less urine (i.e. smaller amounts of 24h-urine metabolites), and women tend to have higher HDL cholesterol and serum adiponectin concentrations than men. The initial SOM analyses revealed a clear division into male- and female-dominated map halves, which clouded the more biologically interesting relationships between metabolism and clinical phenotypes.

To eliminate the differences in univariate statistics between the sexes, the dataset was split into two and then a rank transformation was applied to both the male and the female subsets. The ranks were scaled to within [-1, 1] and the final values were calculated with the formula $x = (z^3 + z) / 2$ to make the rank-based uniform distributions closer to a bell-shaped curve. The latter step mimics a Gaussian probability density, and is therefore compatible with the use of Euclidean distances between the patient profiles in the standard SOM algorithm. This makes the spread of the patients more balanced on the map. The two subsets were then pooled together to train the SOM. The modified rank transform successfully removed the spatial division between men and women (Figure 18).

Model profiles and map colorings

Each SOM unit represents a model metabolic profile that summarizes the characteristics of nearby patients. In Figure 18, the bar profiles depict the relative deviations of selected biochemical markers from the study set average. For example, the patients that are located in the north-east corner of the map (row 2, column 10) have relatively high concentrations of serum triglycerides, cholesterol apolipoprotein B-100, and C-reactive protein. This type of profile is consistent with the metabolic syndrome phenotype [Laaksonen 2004, Bloomgarden 2005].

In the center-north area (1,6), the patients have relatively high serum creatinine concentrations, and potassium and 24h urinary albumin excretion, which indicates a phenotype dominated by kidney disease [Sircar 2008]. Moving to the west (1,1) improves the lipid profile with a decrease in triglycerides and increase in apolipoprotein A-I, while having some signs of kidney disease (minor increase in urinary albumin). Lastly, the patients in the south-west have a profile of high HDL_2 cholesterol, low C-reactive protein and low serum creatinine, which suggests that these individuals have a low-risk metabolic phenotype (7,1).

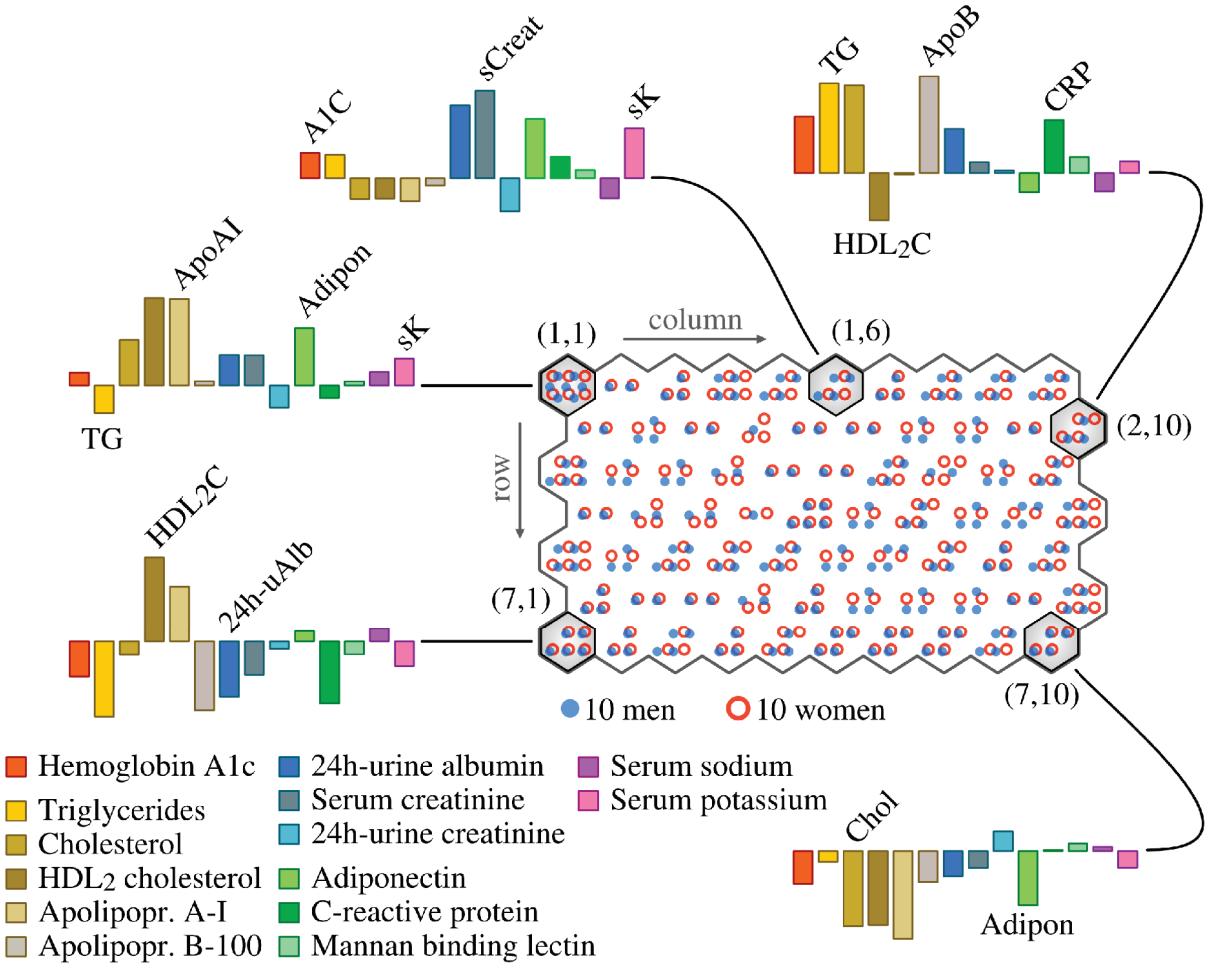

Figure 18: Self-organizing map of the 4,197 FinnDiane patients with type 1 diabetes. The colored bars indicate the typical increase or decrease for a given variable in the highlighted regions, in reference to the dataset mean. The bar heights are adjusted by standard deviations and thus they indicate proportional change. The circular markers represent the positions of the patients on the map (one marker corresponds to a group of 10 local residents).

Map colorings reveal the previous findings from another perspective, one variable at a time (Figure 19). The spatial patterns are similar between men and women, hence only the men (i.e. the larger group) are depicted to avoid redundancy. Nevertheless, the mean values are different between the sexes so the regional averages are listed also for women in the text.

The highest triglyceride values in the north-east exceed the NCEP ATP III guideline of 1.7 mmol/l for the presence of the metabolic syndrome lipid component. Furthermore, the patients in the eastern half have lower HDL_2 cholesterol than those in the western half (0.23 vs. 0.69-0.80 mmol/l for men and 0.32-0.42 vs. 0.99-1.01 mmol/l for women). High concentrations of serum creatinine – indicative of reduced kidney function – prevail on the northern regions (134-194 µmol/l for men and 150-163 µmol/l for women). Interestingly, adiponectin is elevated in the center north (up to 21 mg/l for men and 26 mg/l for women). C-reactive protein is also increased, but it peaks in the north-east (4.3 mg/l for men and 7.7 mg/l for women). The signifi-

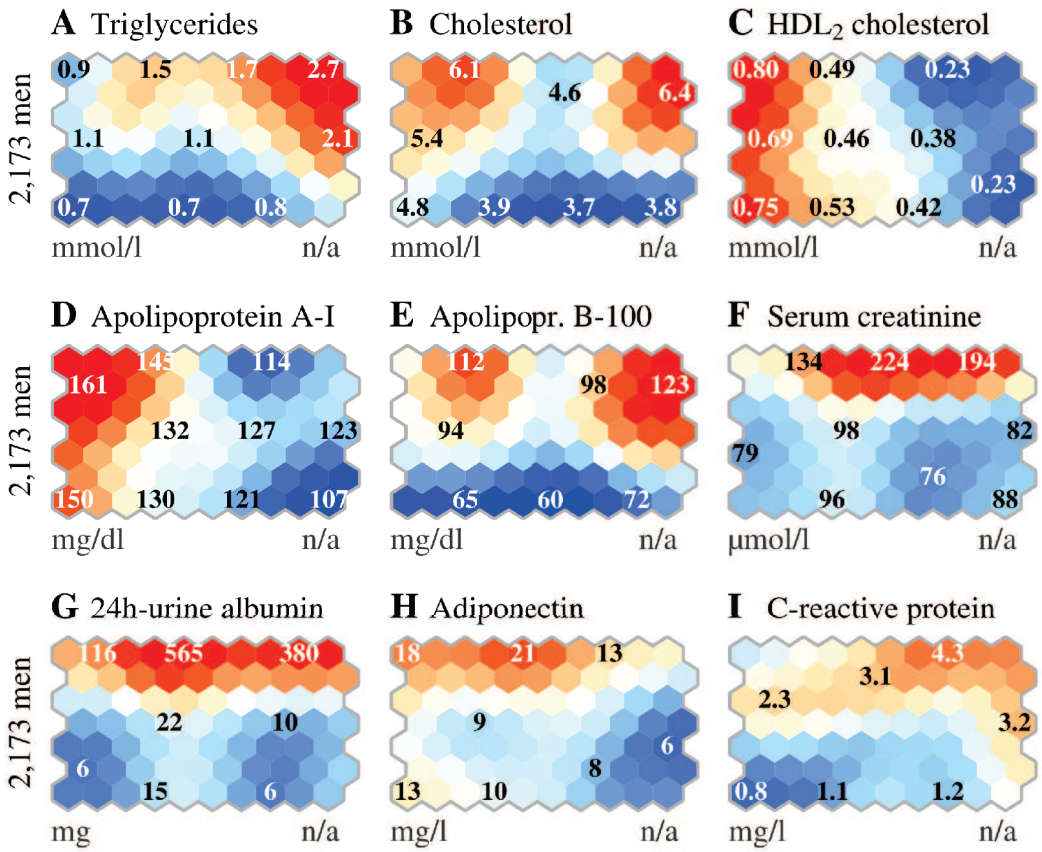

Figure 19: Map colorings for biochemical variables. The plots were qualitatively the same for women, except that women had generally lower triglycerides and 24h urinary albumin and higher HDL_2 cholesterol, adiponectin and C-reactive protein (data not shown). These data were part of the training set for the self-organizing map, hence the *P*-values for the regional variation could not be estimated.

cance of adiponectin is discussed in the next section.

Cholesterol exhibits a split pattern of high values on the map (6.1 & 6.4 mmol for men and 6.1 & 6.5 mmol/l for women). This is not, however, conclusive evidence that there would be two classes of patients with high cholesterol. The pattern probably represents the folding of a multi-dimensional structure into two dimensions; the continuum of cholesterol values is split in order to "fit" the data onto the canvas. On the other hand, the pattern shows that high total cholesterol does not necessarily accompany high triglycerides or low HDL_2 cholesterol, a clue that could be missed by conventional linear modeling.

Clinical traits

The SOM was trained with biochemical data only. It is therefore possible to estimate the statistical significance between the SOM layout and the clinical variables by permutation analysis, as explained earlier. Figure 20 depicts the 10-year all-cause mortality, diabetic kidney disease, retinopathy, the prevalence of the metabolic syndrome and other clinical characteristics on the map. As expected, the highest metabolic syndrome scores are located in the north-east (69-88%) and

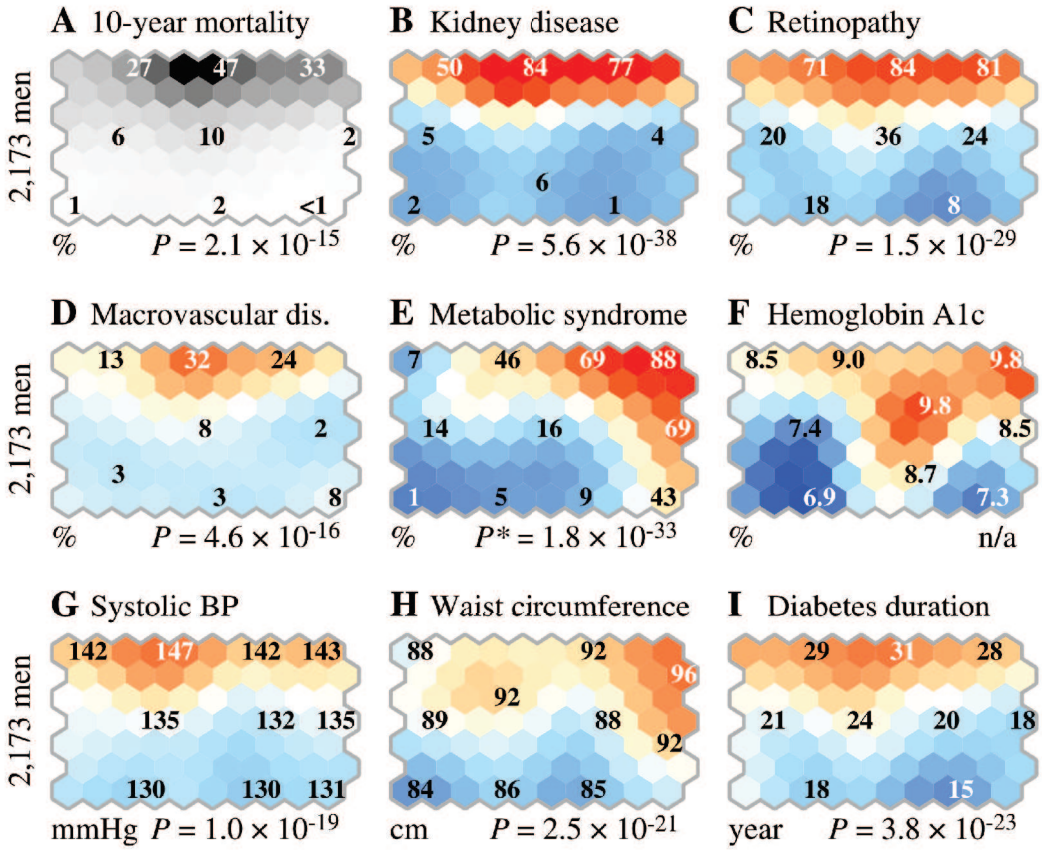

Figure 20: Clinical characteristics of 2,173 men with type 1 diabetes. Similar color patterns were observed for women (data not shown) although the group means were different: women had lower average mortality and prevalence of complications by approximately 10 percentage points. *The metabolic syndrome definition includes some variables that were part of the self-organizing map training set, and therefore the *P*-value may not be accurate.

the center-north area has the highest rates of ESRD or macroalbuminuria (50-84%). The south-west region is characterized by good glycemic control (hemoglobin A1c < 7.5%), which was reflected also in the favorable metabolic profile in Figure 18. The highest macrovascular disease prevalence (32%) coincides with the highest mortality (47% per decade). These patients have also the longest diabetes duration of 31 years, on average.

Figure 21 depicts the main findings that were reported in Publication I. There was a ten percentage-point difference between male and female 10-year mortality, however, this difference could be explained by the sex difference in the entire Finnish population (Figure 5). Population-adjusted mortality ratio was highest for patients that had features from both the obesity-related metabolic syndrome and advanced kidney disease. Both men and women with the adverse phenotype were more than 10-times likely to die before their average Finnish peers. No statistically significant excess risk was detected for patients with a favorable metabolic phenotype, although this last observation may be explainable by the younger age of the low-risk patients.

Overall, the results obtained by the SOM

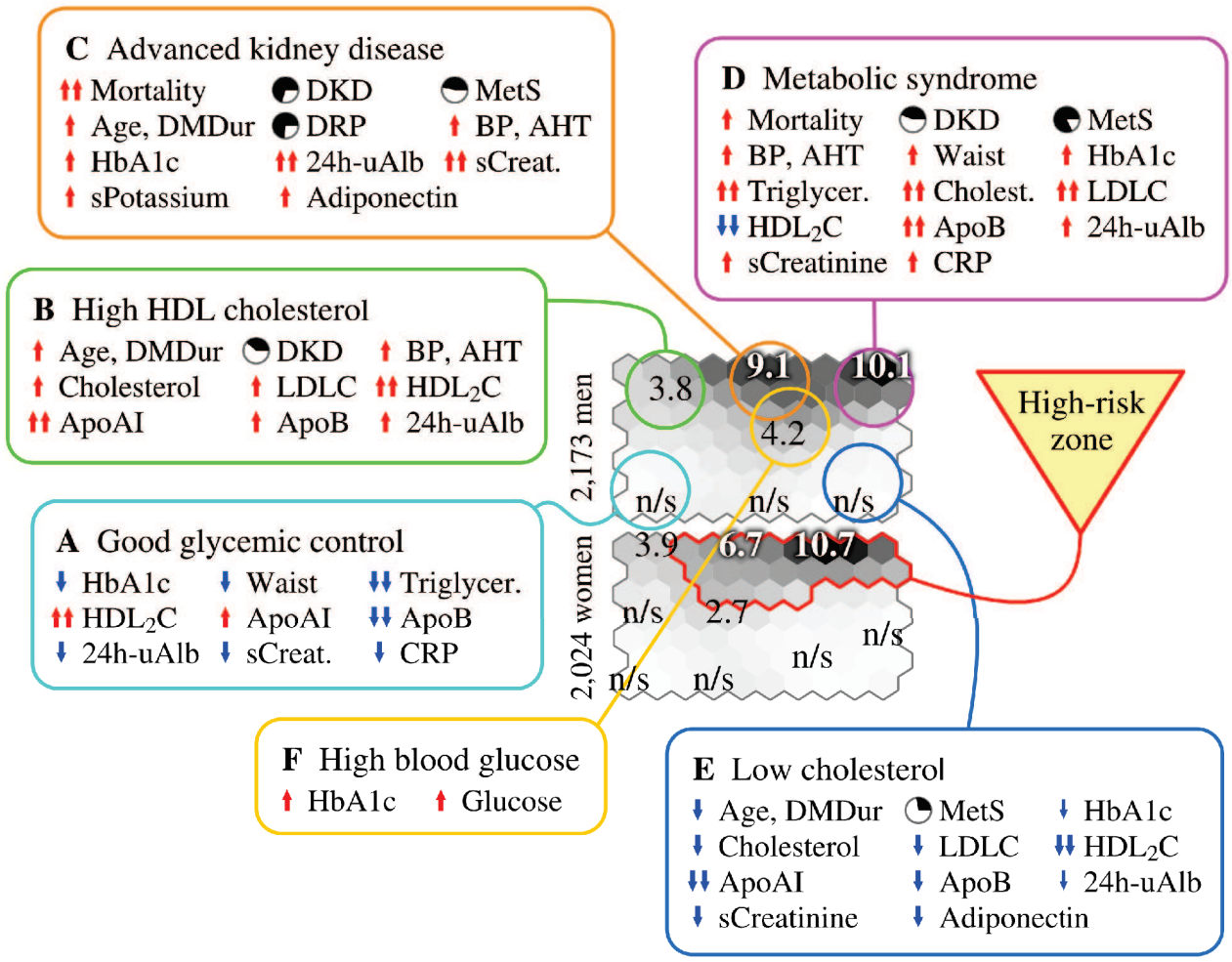

Figure 21: Metabolic phenotypes and the risk of premature death. The relative risk of death for men and women was estimated against the expected mortality (n/s stands for non-significant). **A-E** Five model phenotypes were constructed based on observations from the self-organizing map of 4,197 FinnDiane patients. **F** The glucose concentration was obtained from a subset of patients that had ^{1}H NMR spectra of serum available. The models do not represent distinct clusters in the dataset, but they summarize the characteristics of those patients that are located close to the highlighted area. Serum biomarkers are denoted by 's' to distinguish from urinary measurements (when necessary).

software were consistent with previous knowledge and thus validate the somewhat unorthodox approach. Risk ratios up to 37-fold for death or macrovascular events have been reported in other studies [Borch-Johnsen 1987, Soedamah-Muthu 2004, Stadler 2006], although direct comparisons are difficult. Nevertheless, a study of the FinnDiane material that focused on the clinical albuminuria categories and employed the classical statistical tools produced similar ratios despite the difference in statistical methodology [Groop 2009].

Strengths and weaknesses of the self-organizing map

If conventional methods can produce the same results, why should one adopt the more complicated SOM approach? First, the notion of complicated analysis is, in

fact, misleading since the method can process large numbers of data automatically. In a traditional setting, the researcher would spend considerable time in designing qualitative criteria that best revealed significant aspects from the material. A typical example is the division of the patients into age groups and then investigating the group means and prevalences of clinical end-points. In the worst-case scenario, these divisions have to be made for a number of variables one-by-one, which can be time-consuming. Also, the statistical power to detect associations is diminished due to the categorization step. As to the SOM, no categorization of the data is necessary, and most of the exploration is performed by the computer: the task left to the researcher is to interpret the colorings and profiles which, admittedly, requires some training to master.

Non-linearity is another benefit, or a down-side, of the SOM depending on the situation. There is evidence that many biological phenomena have non-linear components. Alcohol consumption is a famous example: a modest increase from absolutism improves cardiovascular health but too high doses are detrimental [Athyros 2007, O'Keefe 2007]. Combining multiple factors can also lead to unexpected results. A study of *Caenorhabditis Elegans* found that prolonging the worm's life by glucose restriction or antioxidants required a specific genetic background, and thus a simple analysis of high and low glucose could not accurately capture the whole situation [Schulz 2007]. The downside of non-linearity is the increasing danger of overfitting and unstable models, but this can be avoided by increasing the level of smoothing (decrease in spatial resolution), which prevents the SOM from adapting too much.

The SOM offers an automated exploratory tool that can resolve apparent controversies from simpler analyses. The somewhat mirroring relationship between adiponectin and C-reactive protein was suggested by the earlier analysis. Adiponectin is usually considered to be inversely correlated with insulin resistance and obesity [Lara-Castro 2007], which in turn are thought to be heavily involved in diabetic small vessel complications [Groop 2005]. However, studies have shown that patients with type 1 diabetes complications have higher serum adiponectin concentrations [Saraheimo 2005b, Jorsal 2008]. The profiles in Figure 18 and the colorings in Figures 19 and 20 point to one possible explanation: the presence or increased risk of kidney disease and/or retinopathy is associated with elevated adiponectin (Figures 19H and 20B), but within these patients the inverse relationship with obesity (Figures 19H and 20E,H) is still preserved, albeit masked by the complications. On the other hand, C-reactive protein, a marker of systemic inflammation, is positively associated with obesity and lipid toxicity [Cave 2008, Mathieu 2009], so the higher concentrations are located more to the triglyceride-rich north-east areas of the SOM (Figures 19A,I and 20E,H).

The interpretation of the SOM figures directly influences the final conclusions. A non-linear model with a large number of parameters is prone to overfitting the data, that is, to produce patterns where none exist in reality. The relationship between adiponectin and C-reactive protein was weak on the map, and the discussion

Table 3: Clinical characteristics of the case-control subsets in Publication III. Median values ± standard deviation are listed for continuous variables. BP denotes blood pressure.

	Controls	*Unclassified*	*Cases*
	Normal AER	Microalbuminuria	Macroalbuminuria
Number	251	137	225
Male	47%	54%	53%
Age [year]	40 ± 12	36 ± 12	41 ± 10
Diabetes dur. [year]	24 ± 10	25 ± 11	28 ± 8
Waist-hip ratio	0.84 ± 0.08	0.86 ± 0.08	0.89 ± 0.09
Systolic BP [mmHg]	130 ± 15	135 ± 16	143 ± 20
Diastolic BP [mmHg]	78 ± 9	80 ± 10	82 ± 10
Retinopathy	26%	48%	79%
Macrovasc. disease	3%	5%	16%
Metabolic syndrome	29%	40%	59%
Follow-up time [year]	8.4 ± 0.7	8.5 ± 0.4	8.5 ± 0.6
Died within follow-up	2%	7%	19%
Urinary alb. [mg/24h]	7.7 ± 5.1	47 ± 101	532 ± 1186

in the preceding paragraph should not be regarded as final evidence on the matter. It is unlikely that the main observation is false, but the finer details such as the adiponectin-obesity gradient within the kidney disease group should be verified in an independent study. The SOM offers ample opportunity to jump to erroneous conclusions – unless one accepts that it offers a human-friendly view to complex datasets, but may exaggerate statistically insignificant features. This applies especially to the biochemical (or other) variables that are included in the training set.

5.2 Classification of the metabolic state by NMR

Biochemical information

Publication II is, to the author's knowledge, the first human study of diabetic kidney disease by ^{1}H NMR spectroscopy of serum. Originally, a subset of 182 patients with type 1 diabetes were selected from the FinnDiane Study for a case-control analysis of normal AER vs. macroalbuminuria. The sample selection was later expanded in Publication III (Table 3).

The two-window NMR protocol (discussed previously) provides information on lipoprotein lipids, albumin, creatinine and other abundant metabolites, some of which had already been measured by specific biochemical assays by the FinnDiane Study. It was therefore of interest to investigate the pattern recognition ability of the ^{1}H NMR metabonomics framework, and to compare it with the corresponding conventional measurements.

Figures 22 and 23 depict the univariate associations between the measured spectral intensities and the non-NMR data, calculated for the expanded dataset from Publication III (missing data excluded). In the LIPO window (Figure 22), a negative correlation can be observed with the background signal and 24h urinary albu-

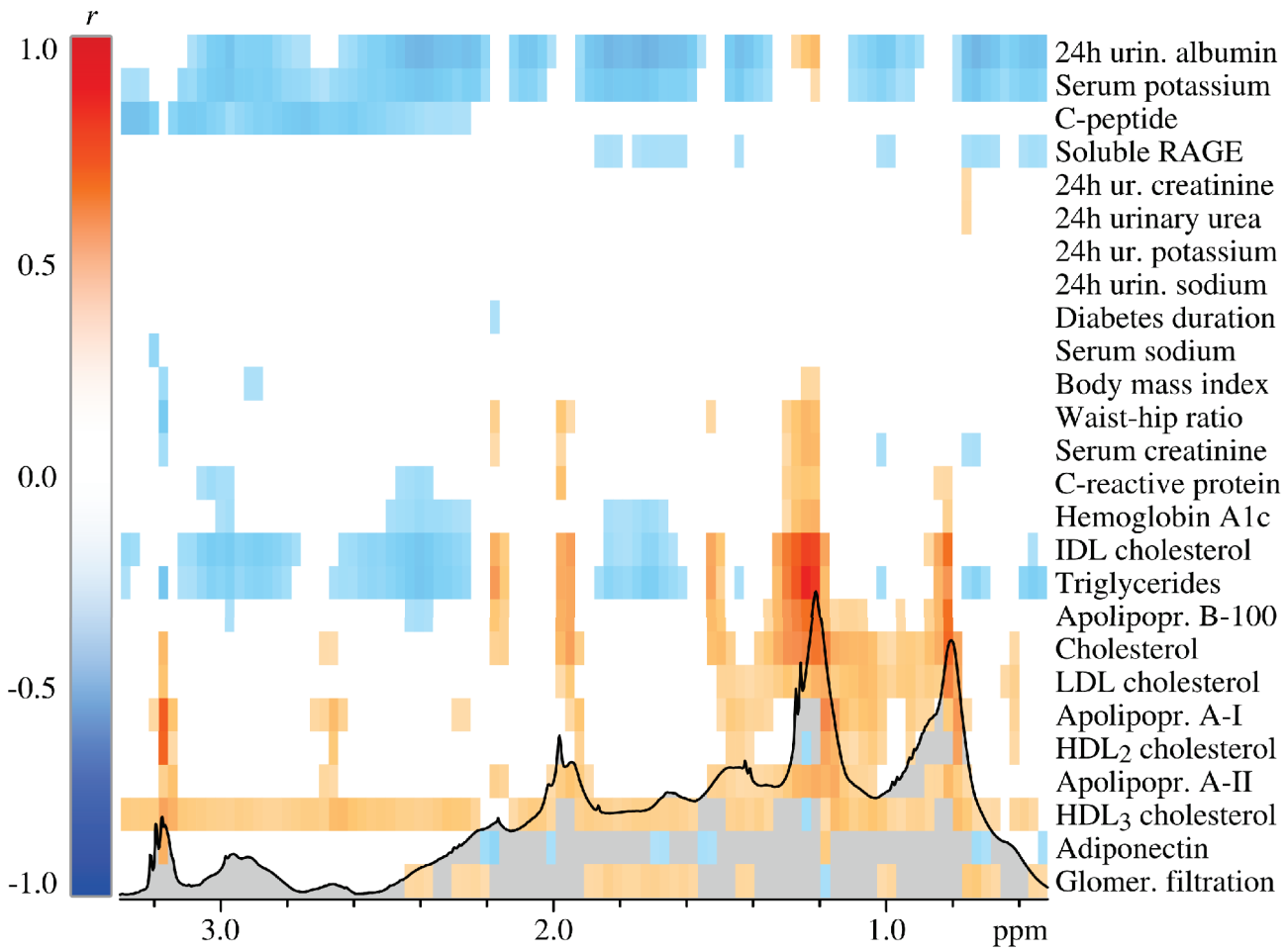

Figure 22: Correlations between ^{1}H NMR spectral intensities and biochemical and clinical data of 540 patients with type 1 diabetes. The NMR data was measured accrding to the LIPO protocol. A pixel depicts the Spearman correlation coefficient between a spectral region and an independently measured variable (the color indicates magnitude). Glomerular filtration was estimated according to the Cockcroft-Gault formula [Cockcroft 1976]. Only pixels with a single test $P < 0.001$ are included.

min, which suggests that the concentration (or NMR signal intensity) of serum albumin and albumin-bound fatty acids decreases as urinary albumin excretion increases.

Patients with significantly impaired kidney function have lower serum albumin concentrations [Kaysen 1998]; here none of the subjects had ESRD. On the other hand, the coupling between serum potassium and urinary and serum albumin suggests a reduced ability to maintain electrolyte balance, which is likely to be a sign of decreased kidney function.

The lipoprotein lipids produce positive correlations at the expected locations [Ala-Korpela 2007]. The CH_2-groups mostly from triglycerides produce a signal at 1.25 ppm, and the CH_3 signals from lipoprotein lipids including cholesterol, surface phospholipids and triglycerides are visible at 0.85 ppm. The phospholipids from HDL particles can be seen at 3.18 ppm. There are also background correlations: triglycerides are correlated with urine albumin and thus

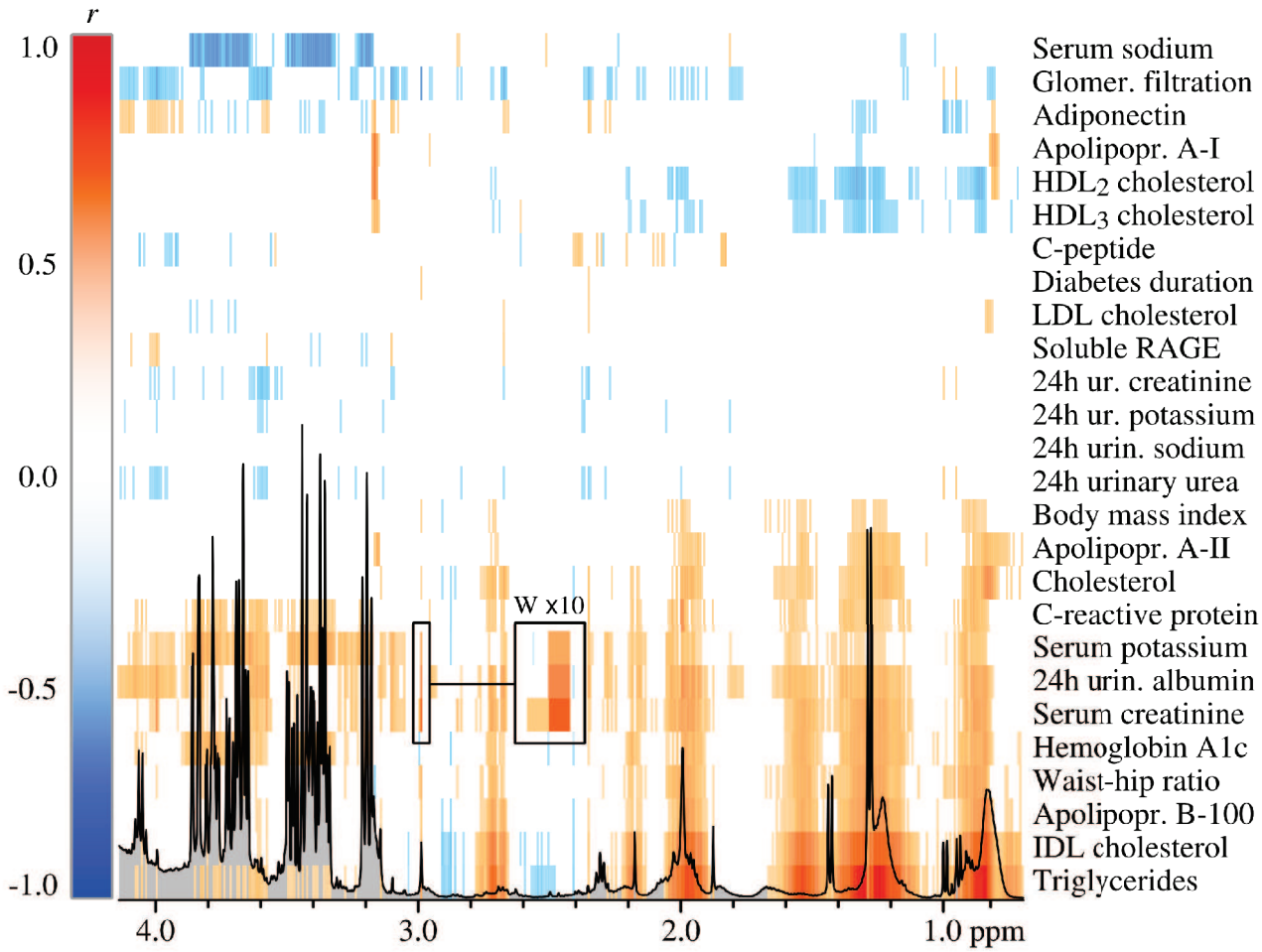

Figure 23: Correlations between ^{1}H NMR spectral intensities in the LMWM window and biochemical and clinical data of 540 patients with type 1 diabetes. A pixel depicts the Spearman correlation coefficient between a spectral region and an independently measured variable (the color indicates magnitude). Only pixels with a single test $P < 0.001$ are included.

are related to the general background phenomenon. Surprisingly, HDL_3 cholesterol shows a pattern of positive correlations.

In the LMWM window (Figure 23), the most mobile lipid species are still detectable at 0.85, 1.25 and 1.92 ppm with traces left also around 3.18 ppm. The negative HDL signal comes from the biological interaction with triglycerides. The creatinine singlet peaks at 2.98 and 3.99 ppm are correlated with creatinine itself, as expected, and with 24h urinary albumin via the biological link with kidney function. Glucose peaks between 3 and 4 ppm have negative correlations with serum sodium.

Figures 22 and 23 contain a number of unexpected correlations with the spectral background and the biochemical variables. It is obvious that any negative correlation must be the result of a biological or methodological dependence, and not from a direct NMR signal. Furthermore, any ions, large proteins, urinary metabolites, or molecules with nanomolar concentrations cannot be detected by the serum NMR protocol.

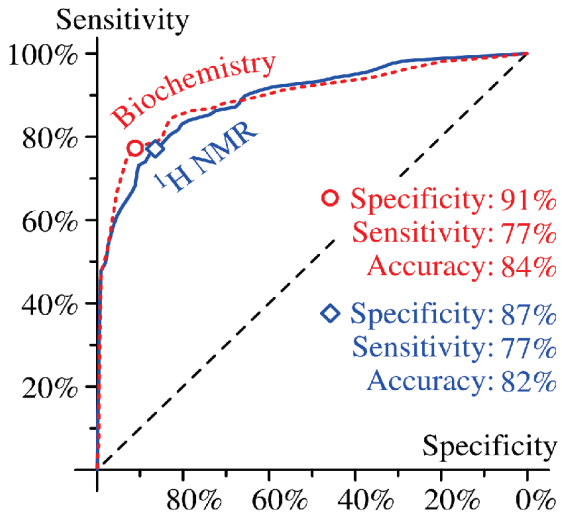

Figure 24: Receiver operator characteristic curves for diabetic kidney disease classification of 203 controls and 212 cases. Patients with any missing biochemical data were excluded. Two logistic regression models were constructed: the red model (dashed line) includes 14 serum biomarkers from the standard FinnDiane laboratory data and the blue model (solid line) is based on the first 14 principal components of the two-window ^{1}H NMR spectra of serum. Sensitivity and specificity were estimated by leave-one-out cross-validation. The circle and the diamond indicate optimal cutoffs.

C-peptide concentrations (not visible by NMR) are low in patients with type 1 diabetes and the values for most patients are at the assay detection limit. Some correlations were observed, but it is possible that they are produced by changes in laboratory protocols over the years, or by a few atypical cases of diabetes. On the other hand, the links between glucose, sodium, potassium and albumin could indicate a change in fluid and electrolyte balance. It can also be speculated that the observations may be caused by altered NMR visibility of the albumin and fatty acid background due to glycation or other systemic effects.

Biomarkers and disease diagnostics

Next, the non-NMR variables were quantified from the NMR data by linear regression to investigate the biochemical information yield from ^{1}H NMR experiments. Triglycerides (88% of total variance explained by the regression model) and creatinine (61% explained) could be estimated with reasonable accuracy, as was expected from the correlation analysis. Apolipoprotein A-I (64% explained) and B-100 (53% explained), HDL_2 cholesterol (58% explained) and even 24h urinary albumin (41% explained) were associated with the spectral data.

The third part of Publication II discussed the classification accuracy of selected spectral features and a set of non-NMR serum biomarkers. Figure 24 shows the receiver operator characteristic curves for 212 cases and 203 controls. The dataset includes also patients from Publication III (samples with missing data were excluded), but the results were similar to those obtained for the set of 182 patients in Publication II. In the figure, the NMR model was built on PCA components. In the original article, a kernel-based NMR model with eight spectral features that represented HDL cholesterol, triglycerides, creatinine and albumin (two features per metabolic marker) was also used successfully.

The non-NMR model includes total cholesterol, triglycerides, HDL_2 and HDL_3 cholesterol, apolipoproteins A-I and B-100, creatinine, sodium, potassium and hemoglobin A1c as independent serum (blood) biomarkers. All the NMR and non-NMR models were successful at

separating most of the patients with macroalbuminuria from those with normal AER (81-86% correct cross-validated classification).

It is noteworthy that the results in Figure 24 are similar to those obtained in Publication II, despite the addition of patients and a number of new biomarkers. One can therefore conclude that improving the biochemical quantification or coverage would not be likely to improve the classification performance. This outcome could, perhaps, be anticipated for two reasons: i) the clinical diagnosis is essentially a cutoff of a single biomarker (urinary albumin), which means the target variable is noisy and ii) there was only one time point per patient so the biological information at an individual level is limited.

The classification results are probably over-optimistic, since the dataset was purposefully chosen to reflect the clinically different sets of cases and controls. In a population-based setting, one would have more patients with microalbuminuria and those with a short diabetes duration. Therefore, Publication II has only minor clinical relevance. Nevertheless, NMR was shown to be a viable screening technique and in the next phase the method was employed in large-scale metabolic characterization.

5.3 Metabolic characterization by NMR and the self-organizing map

Unsupervised framework

In the second NMR study (Publication III), the number of patients was tripled and the unsupervised SOM approach was developed. The emphasis was on the inherent information content rather than on the specific classification of albuminuria. Before analysis, the water signal was removed from the spectra as described earlier to make the biological signals clearer. The multitude of glucose peaks between 3.22 and 3.88 ppm were suppressed to 0.1% of original intensity to prevent the undesired influence from the erratic glucose concentrations that are characteristic of type 1 diabetes.

Specifically quantified lipoprotein lipids or low-molecular-weight metabolites (except a select few) were not available from the LIPO or LMWM spectra. The SOM was therefore constructed directly from the preprocessed spectral curves. The upside of this choice was that all available information was accessible to the SOM algorithm but, on the other hand, also less relevant information was included and the emphasis on clinically significant features was absent. Estimating the classification performance for any particular clinical category was therefore not a priority since the statistical model was not optimized for the task in the first place. Also, examples of supervised analysis were already reported in Publication II.

Spectral profiles

Only the spectra were used for the training of the SOM. Nevertheless, diabetic kidney disease emerged as the defining clinical characteristic on the map (Figure 25-27): the lowest prevalence of macroalbuminuria was 16% in the north-east, compared with a maximum of 70% in the west. Unlike in Publication I, the centrally measured 24h urinary albumin was not included in the training set, nor were

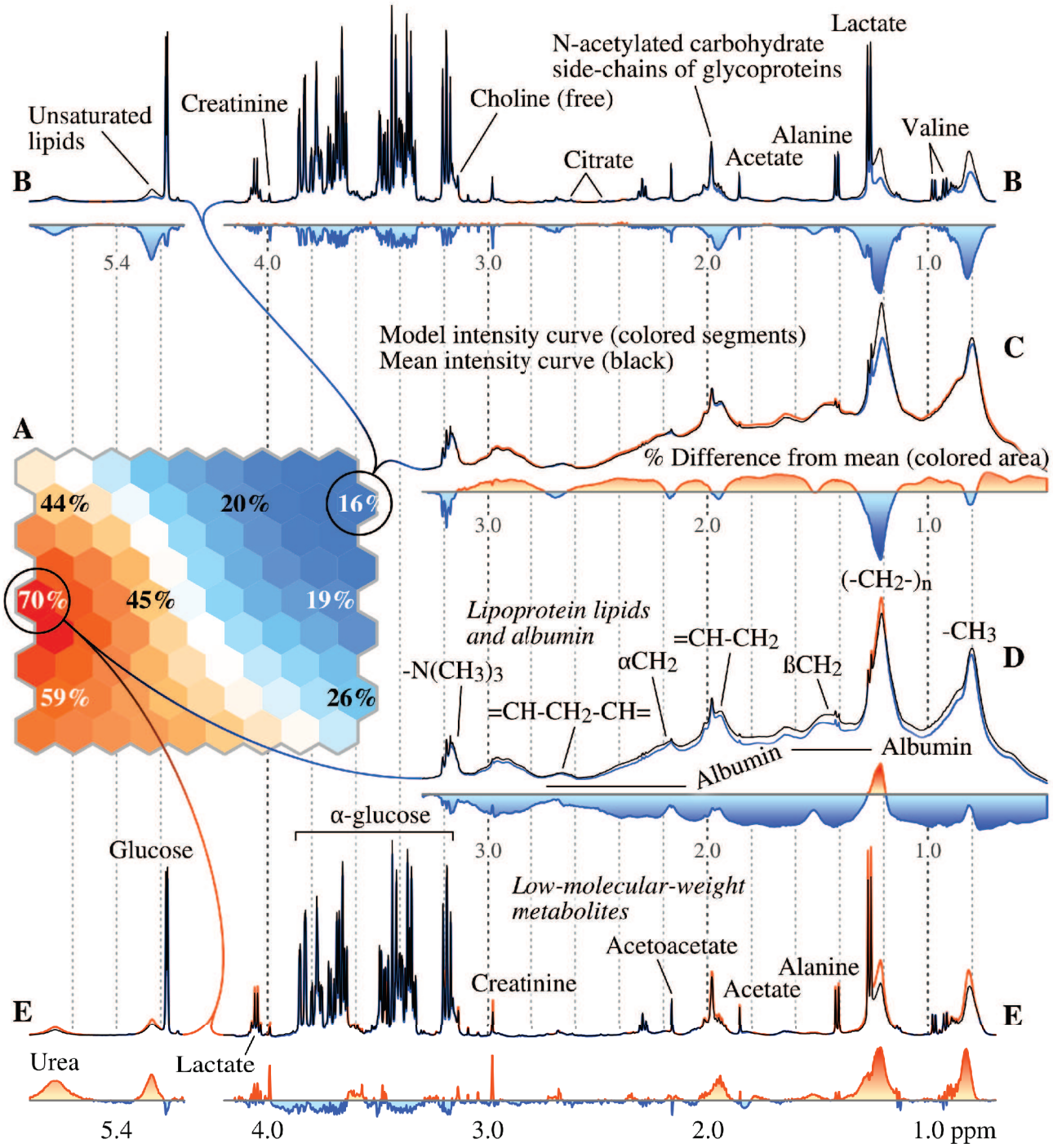

Figure 25: NMR spectral profiles of type 1 diabetes. **A** The self-organizing map was constructed from 613 × 2 ^{1}H NMR spectra of serum and colored according to the prevalence estimates within the map regions. Each hexagonal map unit defines a specific model spectrum and a corresponding subset of patients, the spectra of which best match the model in question. **B** The low molecular weight metabolites (LMWM) model spectrum and **C** the lipoprotein lipid and albumin (LIPO) model spectrum for the patient subset within the map unit with the lowest kidney disease prevalence. The colored curve segments indicate the fitted model, whereas the solid black curve indicates the mean spectrum over all data, thus serving as a constant reference. The colored areas below the model spectra represent the proportional differences of the unit-specific model and the mean model. **D** The LIPO model and **E** the LMWM model for the map unit with the highest diabetic kidney disease prevalence.

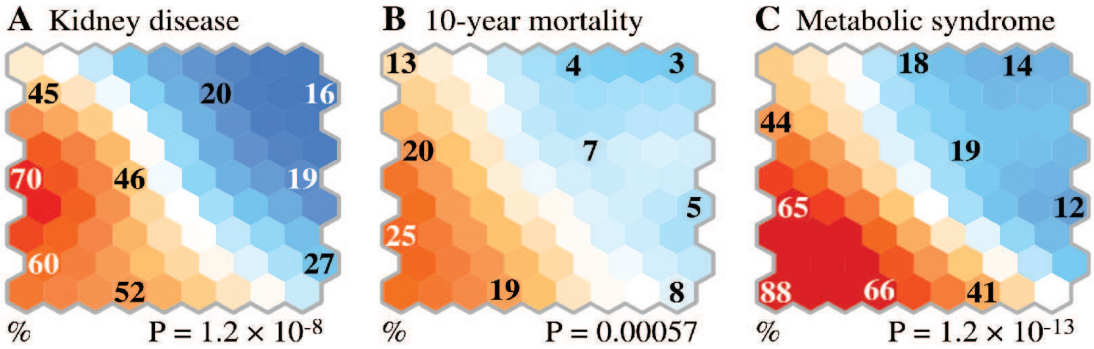

Figure 26: Prevalence pattern of kidney disease (macroalbuminuria), all-cause mortality rate and the metabolic syndrome on the self-organizing map of 613 × 2 ^{1}H NMR spectra of serum. The color of each hexagonal map unit indicates the estimated proportion of cases with respect to the total number of patients who reside on the unit in question. For mortality, the estimates were normalized by follow-up time.

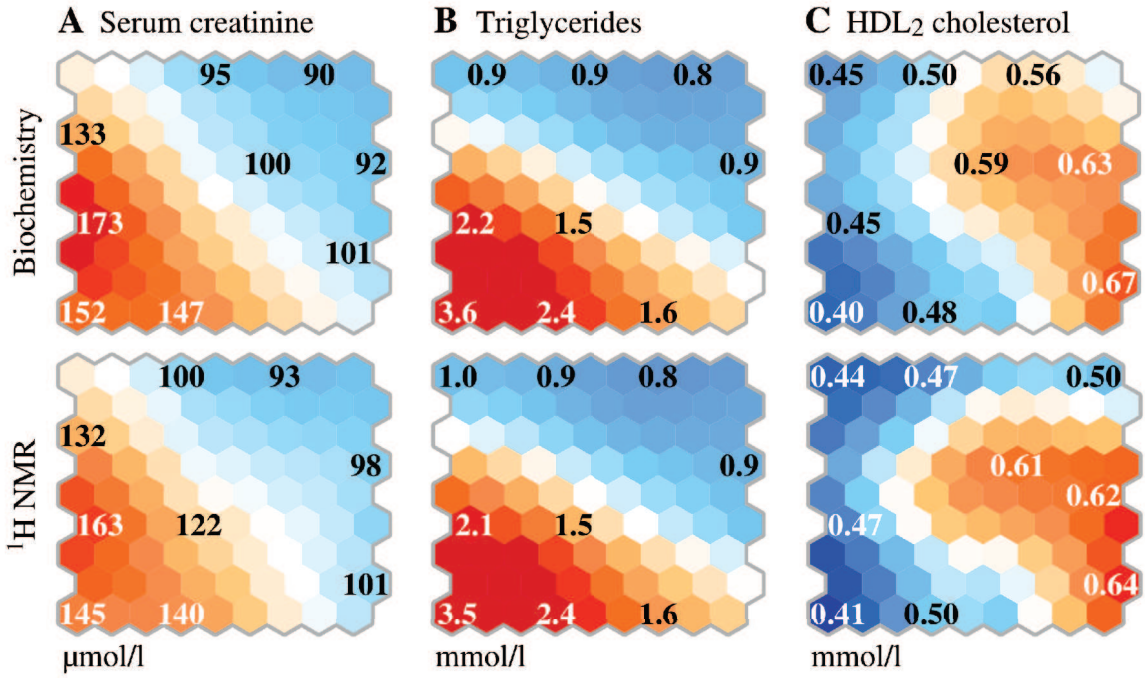

Figure 27: Comparison of biomarker concentrations from direct assays or regression modeling of ^{1}H NMR spectra of serum. The results are depicted via the SOM in Figure 25 that was constructed from the NMR spectra. The colorings on the bottom row were produced by first fitting a regression model to estimate the direct assays, and then coloring the map based on the estimated concentrations in different regions.

any of the urine metabolites, which explains the less clear result. Please note that the orientation of the map (i.e. whether high prevalence is in the south or in the north) has no significance in either analyses and that the patterns in the two SOMs (Publication I vs. III) are not spatially comparable.

The SOM model profiles confirmed earlier observations from Publication II. Relatively high albumin background, low triglycerides (0.85 and 1.25 ppm) and low creatinine (2.98 and 3.99 ppm) are visible from the model spectra in the north-east on the unit with the lowest kidney disease prevalence (Figure 25A-C). Moreover, the interactive version of the figure revealed that the highest triglyceride signals were concentrated on the south-west corner (data not shown), whereas the strongest creatinine signals were located in the west close to the unit with the highest kidney disease prevalence (Figure 25A,E).

Mortality was the highest in the areas of high serum creatinine and triglycerides in the south-western quadrant (Figures 26B and 27A,B). The pattern is consistent with Publication I, where advanced kidney disease and the metabolic syndrome (with its high-triglyceride phenotype) were only partially overlapping on the map, and the risk of death peaked where there was overlap (Figures 20A,B,E and 21).

Metabolite yield

The quantification results from Publication II were replicated indirectly. New regression models were trained for the larger sample set and then the map was colored according to the estimates (Figure 27). The colorings were nearly identical to the colorings from the non-NMR counterparts, which suggests that most of the biologically relevant information can be captured by the NMR experiments, even though the quantitative performance is not perfect.

The NMR spectra contain also information about metabolites that are not among the standard FinnDiane laboratory data. The albumin background was higher on the eastern half of the SOM, where also the prevalence of diabetic kidney disease was lower (Figure 25A,C). Urea is a waste product of protein metabolism and is used by the body to transport excess nitrogen in water-soluble form to the kidneys and out with urine. The nitrogen-containing amine groups produce signals in the aromatic LMWM region and a wide peak can be observed between 5.6 and 5.8 ppm in the areas with a high proportion of kidney patients (Figure 25A,E). The signal is strongly correlated with creatinine and the highest intensities can be observed in the middle of the western edge (data not shown).

Lactate, acetate and glucose are three low-molecular-weight metabolites that are all involved in carbohydrate metabolism (Figure 28). Increased concentration of lactate or lactic acid can be a sign of insufficient oxygenation of tissues. In healthy individuals, vigorous exercise often leads to the accumulation of lactate beyond the normal clearance capacity [Sircar 2008]. In this dataset, high concentrations of lactate and acetate were associated with the triglyceride-rich phenotype (Figures 27B and 28A) and could indicate an impaired transport of oxygen

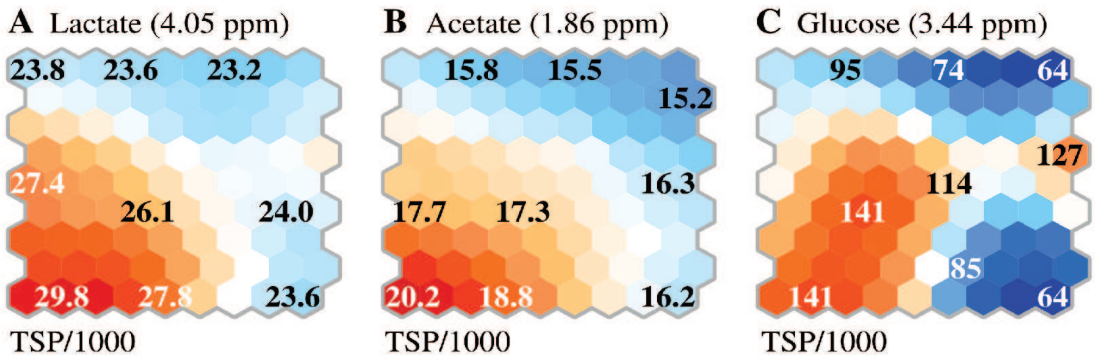

Figure 28: Direct concentration estimates from ^{1}H NMR spectra of serum by peak area integration. The unit is proportional to the intensity of the reference signal of sodium 3-trimethylsilyl[2,2,3,3-d_4]propionate (TSP).

into peripheral tissues or reduced clearance by the kidneys [Chatham 1999, Sugden 2003].

5.4 Advanced lipoprotein estimates from standard measurements

Low-density lipoproteins

Lipoprotein particle size and composition is highly heterogeneous and depends on the type of function the particle fulfills (Figure 6). Epidemiological studies and animal models have shown that the small and dense LDL particles are associated with the development of atherosclerotic lesions inside arteries, although the LDL phenotype is not likely to be the only significant factor [Gardner 1996, Lyons 2006, Stampfer 1996, Davidson 2008]. However, the exact number or size of the LDL particles is difficult to measure in clinical settings.

The work by Friedewald *et al.* has been the basis for LDL estimation in the clinical practice [Friedewald 1972]. At the time, lipoproteins were divided into four classes (chylomicrons, VLDL, LDL_F and HDL) based on density cutoffs in ultracentrifugation experiments. Most blood samples are taken within the fasting state, so chylomicrons can usually be ignored. Each VLDL and LDL_F particle contains a single apolipoprotein B-100 molecule and they can be removed by chemical precipitation, with the HDL left in the remaining fluid. Friedewald *et al.* used the total triglyceride and cholesterol concentration of serum and the cholesterol in the HDL remainder to estimate the amount of cholesterol in the LDL_F lipoproteins.

More sensitive measurements have revealed additional details of lipoprotein metabolism. The conventional LDL_F is now divided into subclasses and an intermediate-density fraction (IDL) and researchers began to apply direct measurements by NMR or other methods to investigate lipoprotein metabolism, instead of the Friedewald LDL formula [Groop 1996]. Nevertheless, many prospective clinical studies are running out of baseline samples, so modernizing the LDL formula is the only way to obtain up-to-date measures for these cohorts.

Quantification accuracy

Simple surrogate equations of unobtain-

Table 4: Cross-validated quantification accuracy of the artificial neural network models for advanced lipoprotein estimates. The explained proportion of the observed variance is denoted by r^2, and $|\Delta|$ indicates the mean absolute bias according to the Bland-Altman method [Bland 1986]. The 95% error interval around the model estimate is denoted by ∂. LDL_F was estimated by the Friedewald formula [Friedewald 1972].

	r^2	$\|\Delta\|$ ∂
		mmol/l
HDL_2 cholesterol	85%	0.00 ± 0.45
LDL cholesterol	83%	0.00 ± 0.99
LDL_F cholesterol	85%	0.42 ± 1.04
IDL cholesterol	61%	0.00 ± 0.28
LDL+IDL cholesterol	86%	0.01 ± 0.93
VLDL triglycerides	96%	0.00 ± 0.36
		mg/dl
Apolipoprotein A-I	85%	0.67 ± 23
Apolipoprotein B-100	90%	0.89 ± 15

able measures are preferred in clinical research since they can be applied without complicated arithmetics. On the other hand, the computer has become a ubiquitous resource and any new estimation formulas need not be human readable, especially in the research setting. In this respect, artificial neural networks (Figure 9) are appealing black-box models for estimation problems [Bishop 1996].

In the past, great emphasis was put on the topology of the neural net to ensure reliable estimates. Over-training can also be avoided by committees or by estimating the parameters in a Bayesian framework [Vanhatalo 2006], both of which are numerically intensive approaches. Here, an early-stopping committee was employed to find robust estimates for the more advanced lipoprotein quantities that are not routinely available.

A summary of the cross-validated results from Publication IV is listed in Table 4. VLDL triglycerides (95% error interval: -0.35, 0.36 mmol/l), IDL cholesterol (-0.28, 0.28 mmol/l), LDL cholesterol (-0.99, 0.98 mmol/l), HDL_2 cholesterol (-0.45, 0.45 mmol/l), and apolipoproteins A-I (-22, 23 mg/dl) and B-100 (-14, 16 mg/dl) were estimated from total triglycerides, cholesterol and HDL cholesterol. The IDL estimate was negatively biased in proportion of the concentration level with a Pearson correlation coefficient r = -0.23 between the model residual and absolute value.

The performance of the neural model for LDL+IDL cholesterol and the Friedewald equation were similar. However, the LDL_F systematically underestimated the LDL+IDL cholesterol in the training data (95% error interval: -1.46, 0.62 mmol/l). The most likely cause is the ultracentrifugation wastage that was not defined in the study by Friedewald *et al*.

From a practical point of view, the LDL_F estimate is close to the pure LDL cholesterol, since the negative bias balances the extra contribution from IDL particles. Consequently, the average levels of LDL_F that have been reported in numerous studies are likely to reflect the correct concentrations. Nevertheless, it is difficult to exclude the effects of ultracentrifugation protocols in different laboratories. The comparison between a 35-year old setup and more recent measurements cannot, after all, be considered a rigorous test.

The LDL_F cholesterol should not be calculated for those individuals with triglycerides above 4.52 mmol/l according to

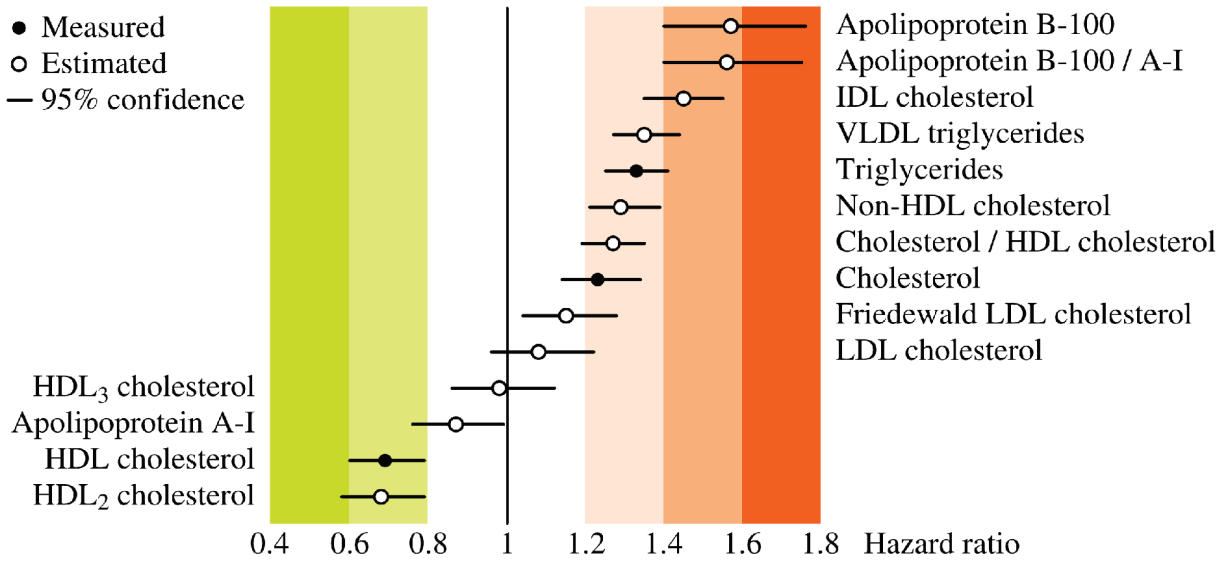

Figure 29: Results from Cox hazard regression of lipoprotein markers and all-cause mortality in 4,048 patients with type 1 diabetes. A separate age- and sex-adjusted Cox model was fitted for each marker. The hazard ratio indicates the relative risk of death after having an increase of one standard deviation in the target variable.

the original article [Friedewald 1972]. In Publication IV, the triglyceride limits were extended to 9.60 mmol/l for the lipid measures and to 6.60 mmol/l for the apolipoprotein measures. These limits are based on the range of values that were available in the training set.

The addition of the HDL_2 subclass and apolipoprotein estimates should give clinicians more information on the patient [Superko 2009]. The cost of the quantification *in silico* is negligible and can therefore be easily incorporated into standard hospital protocols.

Clinical significance in type 1 diabetes

The FinnDiane protocol already includes the basic apolipoproteins and the HDL_2 (non-HDL_3) cholesterol is obtained by direct measurements of total HDL and the HDL_3 subclass. However, the LDL and IDL cholesterol are not available directly and they were estimated in Publication IV.

Separate Cox hazard models [Cox 1972] of all-cause mortality were constructed for each of the estimated lipoprotein measures (Figure 29). The highest hazard ratios were observed for the estimated apolipoprotein B-100 (1.57, $P < 0.0001$) and for the ratio B-100/A-I (1.56, $P < 0.0001$). IDL cholesterol was the next strongest predictor of death (1.45, $P < 0.0001$). Both the estimated VLDL and the measured total triglycerides were significantly predictive (1.35 and 1.33, $P < 0.0001$). High estimated concentration of HDL_2 cholesterol was protective against premature death (0.68, $P < 0.0001$), as was also the measured total HDL cholesterol (0.69, $P < 0.0001$).

Lipoprotein metabolism has been associated with diabetic kidney disease and other complications. Lipid-rich lipopro-

teins from the liver (i.e. those containing apolipoprotein B-100) are increased in type 1 diabetic patients with kidney disease, including the IDL fraction [Groop 1996] and the ratio of the apolipoprotein A-I containing HDL particles and the A-I & A-II containing HDL particles is negatively correlated with the risk of cardiovascular disease [Groop 2007]. Furthermore, the clearance of the atherogenic IDL particles is impaired in patients with end-stage renal disease [Ikewaki 2005, Shoji 1998].

Apolipoprotein B-100 and IDL cholesterol are the most important lipoprotein predictors of death in Publication IV, which is consistent with cardiovascular events in non-diabetic populations [Kastelein 2008]. Interestingly, the estimated LDL cholesterol was not significantly associated with mortality, but the Friedewald LDL_F was, probably via the contribution from IDL. Based on Publication I, apolipoprotein B-100 was highest for those patients with a metabolic syndrome profile, but high concentrations occur also in some individuals with the advanced kidney disease phenotype (Figures 19E and 20B,E). Nevertheless, the combination of high B-100 and low A-I is a highly significant risk marker in both publications and also other studies have reached similar conclusions [McQueen 2008, Walldius 2006].

5.5 Connecting risk factors and clinical end-points

Network of continuous factors

It is a challenge to determine the most significant variables in a clinical dataset, especially in a cross-sectional setting where the individuals are of varying age or have other potentially confounding characteristics. Linear regression is the classical approach, and the model coefficients indicate the most significant variables, given the current model. However, the relative contribution from biochemical and clinical traits may be different depending on what variables are included. Also, the conventional regression models cannot show complex patterns of interaction, that is, they are not well suited to study the correlation structure from a multi-variate perspective despite being multi-variate methods.

The SOM does not give direct information on the associations between risk factors, although it is, in principle, possible to quantify the similarity of the observed spatial patterns between any two variables. In Publication V, the correlation structure of the FinnDiane dataset was investigated directly with a network approach to augment the information that was obtained from the patient-centric analysis in Publication I.

Figure 30 illustrates the correlation network of the biochemical and clinical characteristics for the set of 4,197 patients with type 1 diabetes. There are strong links between methodologically and biochemically dependent variables: markers of body mass (weight, BMI and WHR), 24h urinary excretion (potassium, urea, sodium and creatinine), HDL-related biochemistry (HDL cholesterol, apolipoprotein A-I and A-II) and other lipoprotein quantities (triglycerides, total cholesterol, apolipoprotein B) form positively intra-correlated cliques, that is, subnetworks with densely connected nodes.

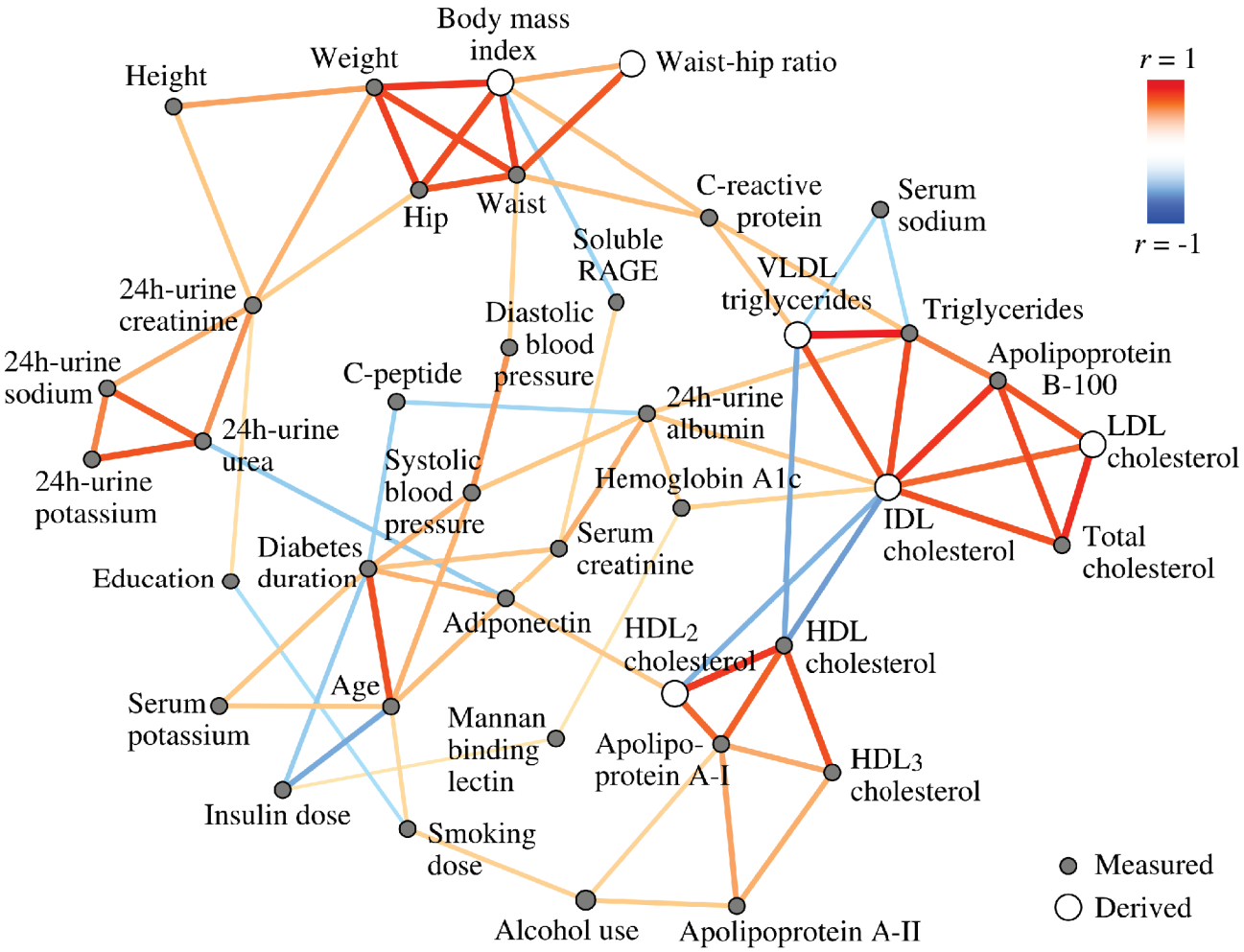

Figure 30: A pruned visualization of the correlation structure within the set of 4,197 patients with type 1 diabetes. Prior to the analysis, the data were adjusted for gender. Each variable is presented with a symbol; those quantities that were measured directly are filled with ink and the open circles denote derived variables. The width and color of the links indicate the correlation magnitude and type, as shown in the legend. The r denotes Spearman correlation.

The HDL-clique and the estimated IDL cholesterol are inversely correlated. Furthermore, triglycerides and body mass are connected via C-reactive protein, and 24h urinary albumin acts as the connector between triglycerides, IDL cholesterol, hemoglobin A1c, blood pressure and serum creatinine. Adiponectin links HDL metabolism with urinary metabolites and soluble RAGE is located between serum creatinine and body mass. Smoking and alcohol intake are correlated; alcohol consumption is also reflected in apolipoprotein A-I and A-II concentrations, and smoking dose is linked to lower education and – by definition – to higher age.

Topological features of albuminuria

The diagnosis of diabetic kidney disease depends on persistent urinary albumin excretion; albuminuria alone is a quantitative risk factor for premature death and large vessel diseases [Gross 2005]. Simultaneously, the biological

significance of the correlation network cannot be measured directly from the graph topology due to the heavy confounding effect of variable selection and methodological dependencies. It was therefore sensible to divide the material according to the clinical kidney disease classification and then apply the subset comparisons (see Methods) to identify the specific changes in the network topology that are statistically significant.

Table 5 lists the significant changes in link weights with respect to the KDNEG subset. Age and blood pressure show a mixed trend: diastolic blood pressure has a negligible age-dependence in the KDNEG subset, but an inverse correlation in the macroalbuminuria subset (r = 0.02 vs. -0.20, $P = 2.5 \times 10^{-5}$), whereas systolic blood pressure shows stronger dependence in the KDNEG subset (r = 0.43 vs. 0.28, $P = 4.0 \times 10^{-4}$). Adiponectin is also age-dependent in the KDNEG subset, but uncorrelated in the macroalbuminuria subset (r = 0.32 vs. 0.08, $P = 8.5 \times 10^{-7}$).

Serum creatinine is connected to adiponectin (r = 0.05 vs. 0.29, $P = 6.3 \times 10^{-8}$) and soluble RAGE (r = 0.03 vs. 0.33, $P = 1.5 \times 10^{-10}$) in the macroalbuminuria, but not in the KDNEG subset. The associations between albumin excretion and other variables are also negligible in the KDNEG subset. On the other hand, urinary albumin is significantly correlated with total cholesterol (r = 0.02 vs. 0.23, $P = 3.0 \times 10^{-6}$), IDL cholesterol (r = 0.06 vs. 0.30, $P = 3.0 \times 10^{-8}$) and triglycerides (r = 0.08 vs. 0.23, $P = 3.0 \times 10^{-4}$) in the macroalbuminuria subset.

Glucose balance, as indicated by hemoglobin A1c, is not among the most significant factors. The result is in contrast with the strong evidence from the Diabetes Control and Complications Trial that reported the lasting beneficial effects of intensive blood glucose management [DCCT 1993, DCCT 2000]. The discrepancy may be explained by the biological inaccuracy of a single hemoglobin A1c measurement and the observational nature of the FinnDiane Study.

Network of clinical traits

Figure 31 collects the clinical and biochemical data of the FinnDiane dataset into a single simplified picture. Diabetic kidney disease, high-blood pressure and retinopathy comprise the devastating small vessel complications triangle on the top part of the figure, with strong links to death and to reduced working ability. Metabolically, the clique connects to high serum potassium, creatinine, soluble RAGE and albuminuria, which can be attributed to impaired kidney function. Genetics and life style may also be involved: patients with complications have more diabetic siblings and have smoked more.

Adiponectin is negatively correlated via the urine metabolites (not adjusted for body surface area) to the body mass indicators. Women are smaller and tend to have higher concentrations, whereas the male gender and obesity (bigger adipose mass) is known to reduce the level of circulating adiponectin [Lara-Castro 2007]. Importantly, though, there is a positive correlation with the microvascular clique, which was already seen to mask the expected behavior in the SOM analysis (Figures 19 and 20).

Table 5: Comparison of the correlations in the KDNEG subset network against the micro-, macroalbuminuria and ESRD networks. The links were selected according to statistical significance ($P < 0.01$ for at least one comparison) and their topological status (see Methods). Urine samples were not available from most patients with ESRD (72% missing); the r values presented were obtained from the imputed dataset. The links are sorted alphabetically.

	KDNEG r	Microalb. r	Macroalb. r	ESRD r
Age – Diastolic blood pressure	0.02	-0.15*	-0.20**	-0.29**
Age – Systolic blood pressure	0.43	0.35	0.28*	0.06**
Adiponectin – Age	0.32	0.30	0.08**	-0.01**
Adiponectin – HDL cholesterol	0.45	0.36	0.21**	0.10**
ApoA-II – HDL_2 cholesterol	0.13	0.17	0.34**	0.33*
ApoA-II – Waist	0.16	-0.02*	0.02*	0.05
Total cholesterol – Education	-0.03	-0.18*	-0.08	-0.01
Serum creatinine – Adiponectin	0.05	0.03	0.29**	0.18
Serum creatinine – Diabetes duration	0.07	0.22*	0.17	-0.06
Serum creatinine – Insulin dose	-0.01	-0.17*	-0.15*	-0.13
Serum creatinine – Soluble RAGE	0.03	0.05	0.33**	0.40**
Serum creatinine – 24h-uAlb	0.06	0.07	0.15	0.44†
CRP – Age	-0.10	0.09*	0.05*	-0.01
CRP – Serum potassium	-0.05	0.12*	-0.01	-0.02
CRP – Waist-hip ratio	0.18	0.34*	0.23	0.22
IDL cholesterol – LDL cholesterol	0.72	0.63*	0.53**	0.53**
LDL cholesterol – Education	-0.01	-0.17*	-0.07	0.01
MBL – 24h-urine urea	0.08	-0.10*	-0.05	-0.02†
Serum potassium – Diabetes duration	0.27	0.26	-0.02**	-0.02**
VLDL triglycerides – 24h-uAlb	0.07	0.12	0.22*	0.51†
24h-uAlb – ApoB	0.07	0.19	0.27**	0.31†
24h-uAlb – Total cholesterol	0.02	0.17*	0.23**	0.16†
24h-uAlb – HDL cholesterol	-0.06	-0.04	-0.12	-0.46†
24h-uAlb – IDL cholesterol	0.06	0.16	0.30**	0.50†
24h-uAlb – Triglycerides	0.08	0.13	0.23*	0.50†
24h-uAlb – 24h-urine creatinine	0.11	-0.05*	0.02	-0.36†
24h-uAlb – 24h-urine urea	0.04	-0.06	-0.06	-0.41†

*$P < 0.01$, **$P < 0.0001$, comparison with normoalbuminuria; †imputed

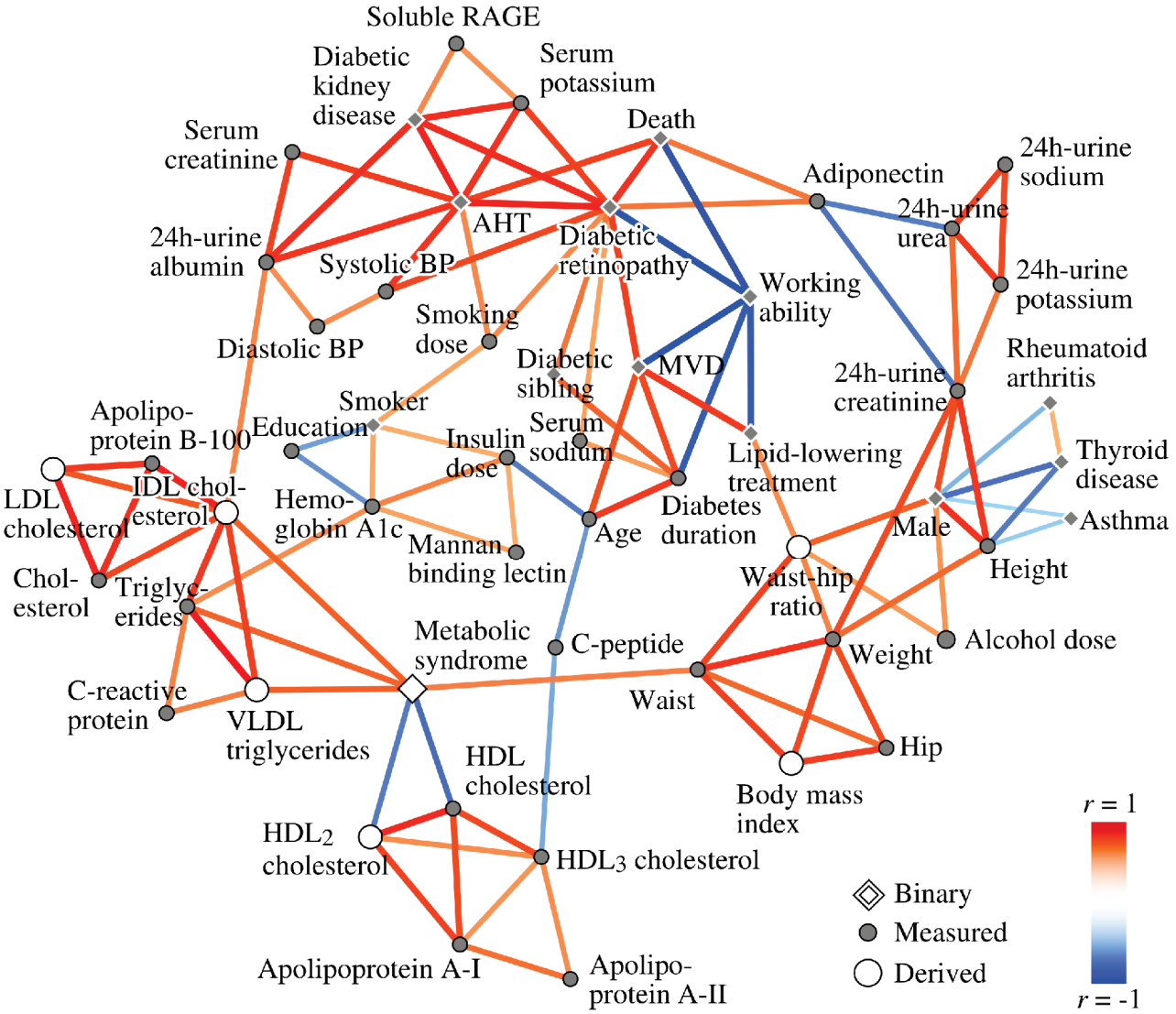

Figure 31: A pruned visualization of the correlation network from regression modeling. The data were not adjusted for gender prior to the analysis. Each variable was converted to a surrogate linear predictor before computations. The symbols in the figure correspond to the source of information: directly observed variables are filled, whereas derived variables are denoted by open symbols. A circle is used for continuous quantities, and a diamond for binary traits. The width and color of the links indicate the association magnitude and type, as shown in the legend. The r denotes the correlation of the linear predictors and is not comparable with Figure 30.

The available selection of variables poses a critical bias to the network structure, and should be taken into consideration when interpreting the results. For example, the metabolic syndrome is a derived clinical construct the combines HDL-metabolism, triglycerides, obesity, blood pressure and glucose homeostasis (criteria from Publication I). Accordingly, the syndrome definition creates a connector between most of the aforementioned components. The absence of the blood pressure link can be explained by medication: if the original NCEP ATP III criteria [NCEP 2002] were used, then those patients with anti-hypertensive medication would have been "awarded" one point even without exceeding the diastolic or systolic limit, which would have strengthened the link to blood pressure.

The previous example shows also the dangers of applying clinically practical categorizations when trying to understand complex phenomena. A small change in guidelines can cause significant alterations in the results and, to some extent, also in the conclusions.

5.6 Concluding remarks

Paradigm shift in biomedical research

The standard way to study biomedicine is i) to create a taxonomy of the observed phenomena, ii) to isolate the potential agents that are responsible for the observations and iii) to reproduce the conditions of each of the taxonomic classes in a controlled environment to verify the findings. Put differently, the biological behavior of an organism is dissected into distinct mechanisms that each produce a specific symptom. This reductionist approach is highly successful in uncovering infectious agents, severe genetic defects, trauma to specific tissues or the effects of a toxic compound. Reductionism has also direct practical applicability in the form of differential diagnosis: a set of simple rules is an efficient means to isolate the correct branch in the taxonomic tree.

At present, chronic metabolic diseases and cancers are the greatest threats to human health and longevity in the developed world. The "easy" problems have been solved; thanks to the discovery of insulin, type 1 diabetes is no longer the prelude to quick death. Reductionism worked well here since the phenotype is qualitatively clear (the severe depletion of beta-cell capacity within a relatively short time period). Furthermore, the diagnosis can be confirmed by a few straightforward tests.

Dig a little deeper and things get difficult. The primary causes for the beta-cell destruction remain obscure and the mechanisms of the long-term complications are still hidden. Instead of a single causative agent, the two aforementioned phenomena may depend on multiple genetic defects that interact with the environment. In the extreme, this can mean that every individual has a unique disease at the molecular level, even though the perceived symptoms at the physical level are compatible.

The uniqueness can be viewed also from another perspective: there may be a great number of small defects – which will have a deterministic effect on an individual in a reductionist fashion – but every patient has a unique combination of them. The disease classification should

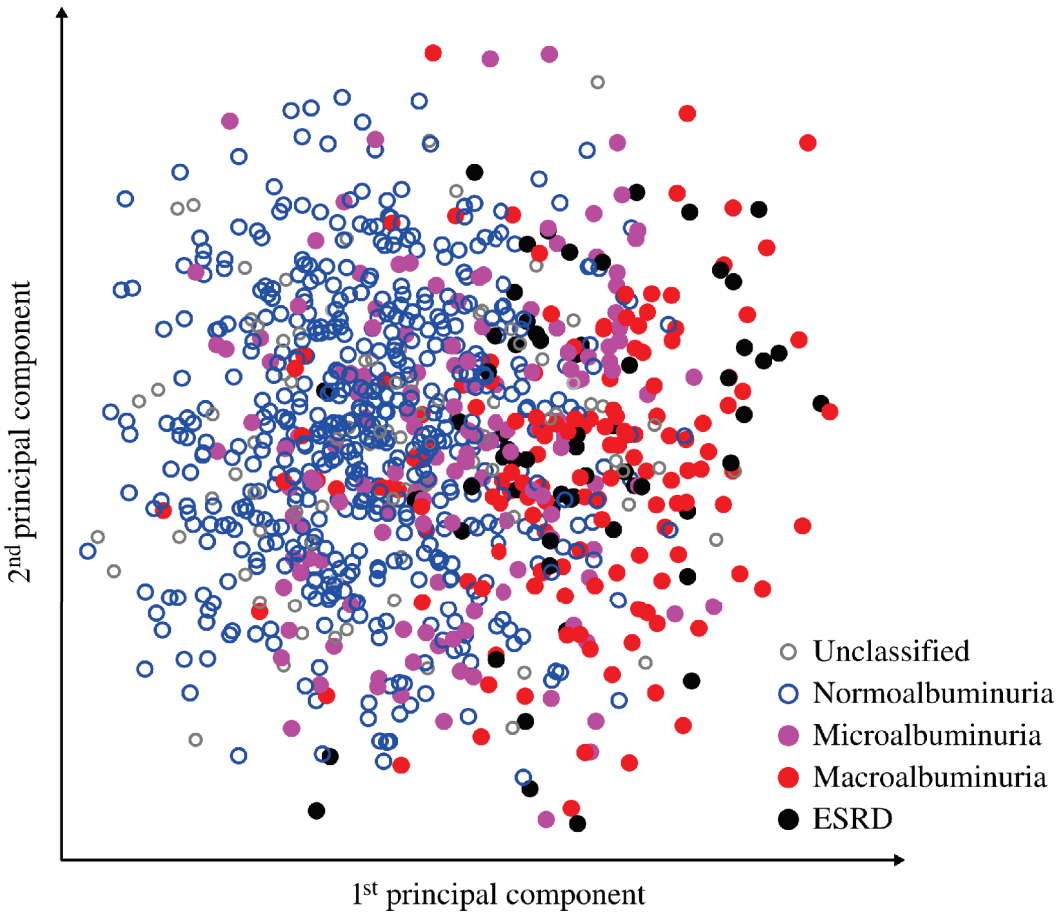

Figure 32: Principal component analysis of the biochemical data from the FinnDiane Study. The training set (from Publication I) includes 4,197 patients, of which only 1,000 randomly selected individuals are depicted to avoid clutter.

therefore cover all the naturally occurring combinations. However, the combinatorial explosion will quickly make it impossible to create an accurate taxonomy of such a phenomenon.

The explorative analyses produced tangible evidence on the magnitude of the phenotyping challenge. The clinical categories did not form isolated clusters in the data space. If anything, the opposite was the typical situation: the patients comprise a point cloud where the prevalence of kidney disease increased from one side to the other in a continuous fashion (Figure 32). Based on the statistics, there does not seem to be an inherent metabolic threshold for the presence of diabetic kidney disease. By contrast, a clear separation could be observed between healthy non-diabetic controls and patients with type 1 diabetes, as expected (data not shown).

Clinical significance

Figures 18-21, 30 and 31 are among the first multi-variate visualization of a large set of patients with type 1 diabetes. They reveal a bleak picture for those individuals that are affected by chronic complications but, on the other hand, the majority of the Finnish patients seem to cope well with their disease. The material did not

contain any direct evidence on the acute complications such as severe episodes of low blood glucose. Nevertheless, younger patients and those without serious complications had comparable mortality to the general population, which suggests that the life-preserving insulin treatment is highly successful in the short term.

The long-term end-organ damage is the true challenge: it causes most of the human suffering, but the early stages cannot be detected with the current clinical means nor has the precise set of disease mechanisms been identified. The networks in Figures 30 and 31 show no strong links between the complications clique, lipid clique or the miscellaneous structures. In particular, there are few biomarkers that preferentially connect to the clinical classifications, except serum creatinine and urinary albumin which, by definition, are connected to kidney disease.

It is mathematically impossible to find the perfect biomarker if the target diagnosis is less than perfect. This is probably the current situation in diabetic kidney disease and thus explains the (lack of) observed connections. Nevertheless, the networks do show that obesity, high blood pressure, abnormal lipoproteins, aging and unhealthy life-style all contribute to the disease process.

The results in this thesis indicate a large biological variance in the cross-sectional FinnDiane dataset and suggest that increasing the accuracy or coverage of a measurement at a single time point may not help to improve the diagnostic or prognostic accuracy. Neither cells, organs nor people live in isolation, but are constantly adapting to environmental pressures. Yet clinical epidemiology is almost exclusively focused on fasting blood samples in order to exclude environmental effects such as food ingestion. It is questionable whether these static investigations are sufficient for uncovering the early causative agents that may be detectable only during systemic perturbations.

Practical implications

Although multiple time points were not available in this thesis, the FinnDiane dataset provides a unique opportunity to develop new ways to integrate metabolic and clinical data. There is no shortage of available methods in the literature, but the algorithms and models are often designed for the statisticians themselves. Here, the emphasis was on visualization and explorative analysis, and computational methods that are accessible for people from a wide range of disciplines.

Methods such as the self-organizing map are needed to improve the knowledge discovery process from vast biological datasets. As discussed earlier, the lack of clear disease categories and simultaneous interaction of multiple small defects may prevent the application of the conventional paradigm with controlled isolated effects. Instead of exhaustively going through a great number of combinations, the SOM and the networks can quickly pinpoint interesting multi-variate phenomena that can then be investigated in more detail.

An engineer's job is to come up with technological solutions to difficult problems. However, there are two obstacles

for an efficient use of computational resources in biomedical research. First, the biologist may not be familiar with the capabilities of modern statistical methods and does not therefore pose difficult enough questions for the engineer. On the other hand, the engineer has little knowledge of the biological significance of the phenomenon and thus tends to focus on irrelevant technical details instead of the scientific objectives.

The multi-variate description of the type 1 diabetic condition is the main contribution to science in this thesis. For the clinician, the results themselves may not be entirely new, but the way the data were acquired and presented should inspire a new type of thinking and possibly invoke more interest in the use of modern computational and analytical resources. For the engineer, the self-organizing map and the network figures offer a fast data-driven portal to the metabolic features of type 1 diabetes and they explicitly show the complexity of diabetic complications.

For the general public, the message is mixed. Type 1 diabetes is a serious but manageable disease, and most patients can expect a healthy and long life but for some, the gradual damage to the body causes considerable suffering. On a more positive note, genetics is not likely to be the only determinant; patients themselves can affect the course of the disease. This thesis shows how a certain metabolic profile is associated with the absence of complications. Metabolism, in turn, is influenced by diet, exercise and mental health. Improving the quality of life on all fronts simultaneously is as important for the patients with diabetes, as it is for everyone else.

6 Summary and conclusions

1. A quarter of patients with type 1 diabetes exhibit an obesity-related phenotype with the characteristic lipids (high triglycerides, cholesterol, apolipoprotein B-100 and low HDL cholesterol) and chronic inflammation (high C-reactive protein). A third of individuals have a diabetic kidney disease phenotype (high urinary albumin and serum creatinine). The combination of the two was associated with the highest population-adjusted mortality.

2. Patients who did not have an adverse metabolic phenotype were not at an increased risk of premature death within the follow-up period. Therefore, the prevention and early treatment of the high-risk phenotypes is the key to improved survival in long-standing type 1 diabetes.

3. ^{1}H NMR spectroscopy of serum is a robust alternative to numerous conventional biomarkers and is suitable for the metabolic screening of diabetic complications.

4. The combination of the self-organizing map and the correlation network are mutually complementary unsupervised methods that can visualize the main characteristics of complex heterogeneous datasets.

5. Advanced lipoprotein estimates are available from the three standard lipid measures (triglycerides, total and HDL cholesterol). In particular, the IDL fraction may be the most significant lipoprotein covariate in the pathogenesis of diabetic complications.

6. The variables in this study were methodologically, physically and biochemically linked. Separating these components of associations may not be possible, thus the reductionist approach that relies on controlled environment and isolated effects may not capture all the relevant features of the phenomenon.

7 Acknowledgments

This thesis is the culmination of a collaborative effort between the Department of Biomedical Engineering and Computational Science at Helsinki University of Technology, and FinnDiane Study Group in Folkhälsan Research Center in Biomedicum Helsinki. Two central figures, Prof Kimmo Kaski and Prof Per-Henrik Groop have been able to provide a stimulating, well resourced environment for this cross-disciplinary research.

The friendly and supportive attitude to younger scientists has been even more important. The two have never dismissed my ideas (even when they have been a little too much out there) and given me the trust and freedom to pursue unexplored areas of science. It is also fun hanging out with these characters, although after a while you do learn most of their anecdotes by heart. Nevertheless, I give my warmest thanks to both Kimmo and Perra for all of the friendship and support over and beyond the past years.

Getting started with the thesis was slow due to the steep learning curve in medicine. However, when Prof Mika Ala-Korpela came along, things started to move quickly. We had converging ideas on the study of metabolism, and he gave me excellent opportunities to advance my work into new areas. Things were also lively on the entertainment front: our small computational medicine group became (in)famous for its enthusiastic approach to social events in and out of the professional sphere. My thanks come from the introduction to the wonderful world of lipids and also on the unforgettable shared experiences we had ingesting them.

Dr Carol Forsblom has had an immeasurable influence on this dissertation. In fact, I used his thesis as a template for my own, and he has provided an inexhaustible source of medical knowledge, not to mention all those interesting stories of music and of the everyday existence (funny things seem to happen to him on a daily basis). My gratitude glows warm for Carol for all the work he has done for me and for the entire FinnDiane Study Group.

Savonia is an exotic region of Finland with its own language and culture, and even the Savonian mafia. I had the good fortune of collaborating with Dr Pasi Soininen from Kuopio, who performed most of the “real” analysis of the serum samples. His work was instrumental in the metabonomics studies. I also thank Prof Reino Laatikainen for his jolly demeanor, fried fish and excellent sauna in the heartland of Savonia.

Dr Soininen also made measurements in Turku, in collaboration with Dr Petri Ingman. I have to say that the Savonians may have their quirks, but even they

could not come up with a stranger conference venue than the fishing village that was the location of the Finnish NMR meeting in 2007. Thank you for the experience.

In order to cover the inhabited regions of Finland, we also had collaboration with the University of Oulu in the north. I thank Sanna Mäkelä, Dr Minna Hannuksela and Prof Markku Savolainen for their contributions to the lipoprotein studies in this thesis.

I have shared office space with several people over the years. Dr Jukka-Pekka Onnela was my fellow student and a trusted collegue for several years and we had many invigorating discussions on various topics from work to social sciences and music. I thank JP for the depth of vision he gave me. I had the also the priviledge of working with Dr Sebastian "Sebu" von Alfthan, Dr Marko Sysi-Aho and last – but not least – with Jörkki Hyvönen.

My thanks go to Dr Jari Saramäki, Dr Riku Linna, Dr Jukka Heikkonen, Dr Riitta Toivonen and Dr Aki Vehtari for the discussions and education they have given me. I am also grateful to Dr Margareta Segerståhl for the challenge of keeping up with her speeding mind.

Maikki Parkkonen and I continue to maintain the sprawling database of the FinnDiane Study and I thank her for the patience and efforts that made my studies possible. I thank Dr Sara Fröjdö and Lisa Sjölind for the few early years we shared offices in the Biomedicum building. I also wish to thank Milla Rosengård-Bärlund, Dr Maija Wessman and Dr Riitta Sallinen for their good humor and joyful attitudes.

Other FinnDiane members include Dr Kim "Peffen" Petterson-Fernholm, Dr Johan Fagerudd, Dr Markku Saraheimo and the more recent additions Dr Markku Lehto and Dr Valma Harjutsalo. Dr Sanna Lehtonen and Mervi Hyvönen have also been recently recruited into the group, along with Kustaa Hietala, Janne Kytö and Aila Ahola. I thank you all for the friendly atmosphere and stimulating exchange of ideas.

The nurses and laboratory technicians do the real work in our group. Anna Sandelin, Anna-Reetta Salonen, Tuula Soppela, Jaana Tuomikangas, Sinikka Lindh and Tarja Vesisenaho, you have all provided the necessary resources for this thesis to be completed. And, thruth be told, life would have been nearly imbossible in the office without the managerial skills of Jaana Welin-Haapamäki and her counterparts in Otaniemi (Eeva Lampinen and Dr Kaija Virolainen).

I have had lots of fun with the younger members of the FinnDiane Group. Our latest expedition took us to the Mediterranean beaches, although the success of that trip may be put to question. We have also been to the movies together and shared strange lunch conversations and I hope that this subculture continues. My thanks go to Jenny Söderlund, Johan Wadén, Dr Daniel Gordin, Dr Lena Thorn, Nina Tolonen, Mariann Nymark, Outi Heikkilä, Anna Hoverfält, Laura Salovaara, Emma Fagerholm, Dr Pekka Ihalmo, Anna Tiitu and all the other people from FinnDiane I forget to mention.

The computational medicine gang used to gather on he third floor of Innopoli 2, to treat the caffeine and sparkling wine habits (I myself, of course, was free of the former). My thanks go to Jaakko Niemi, Linda Kumpula, Taru Tukiainen, Pasi Jylänki, Antti Kangas, Johanna Hokkanen, Janne Ojanen and Tomi Peltola for the fun and games at various instances.

Tomi has continued the collaboration between FinnDiane and BECS by working part-time in Biomedicum. Niina Sandholm has also crossed over – from engineering to genetics – and I am thankful to both for their enthusiasm for discovery and perpetual good spirits.

Dr Teemu Leppänen and I grew up in the same neighborhoods and went to the same schools as children. Later, we ended up in the then Laboratory of Computational Engineering. In fact, it was Teemu who hinted me about the summer work in LCE, if my memory serves me right. For that, I think that we can all be grateful. With his wife Satu Matinlauri we have had many in-depth debates about science, humanity and the meaning of life. Your opinions have been an enourmous help in forming my own and I thank you for all the friendship and support you have provided for me over the years.

This thesis would not have been possible without funding from various organisations. I give thanks to the Graduate Schools of Electical and Communications Engineering and of Computing Science at Helsinki University of Technology, to the Jenny and Antti Wihuri Foundation, to the Finnish Cultural Foundation, to Samfundet Folkhälsan, to the Academy of Finland and to the European Union.

Finally, I give my utmost gratitude to my family, my grandparents, parents and brother for giving me their love and affection. I especially wish to thank both my grandmothers for their unconditional attention to my welfare. My grandfathers have been the perfect role models and I owe a great deal of respect for both of them. My parents are also important people, I would like to especially thank my mother for all the work and warm feelings she has invested in me. As to my little brother, he is a unique package and I would not have it any other way.

December 2009

Ville-Petteri Mäkinen

8 References

ADA (2007) **Definition and description of diabetes mellitus**. *Diabetes Care* 27:S5-S10.

Abrahamson M, Olafsson I, Palsdottir A, Ulvsbäck M, Lundwall A, Jensson O & Grubb A (1990) **Structure and expression of the human cystatin C gene**. *Biochem J* 268:287-294.

Adosraku R, Choi G, Constantinou-Kokotos V, Anderson M & Gibbons W (1994) **NMR lipid profiles of cells, tissues, and body fluids: proton NMR analysis of human erythrocyte lipids**. *J Lipid Res* 35:1925-1931.

Agardh CD, Agardh E & Torffvit O (1997) **The association between retinopathy, nephropathy, cardiovascular disease and long-term metabolic control in type 1 diabetes mellitus: a 5 year follow-up study of 442 adult patients in routine care**. *Diabetes Res Clin Pract* 35:113-121.

Ahlgren AR, Sundkvist G, Sandgren T & Länne T (2002) **Female gender increases stiffness of elastic but not of muscular arteries in type I diabetic patients**. *Clin Physiol Funct Imaging* 22:409-415.

Ahmed AT, Karter AJ & Liu J (2006) **Alcohol consumption is inversely associated with adherence to diabetes self-care behaviours**. *Diabet Med* 23:795-802.

Al-Kateb H, Mirea L, Xie X, Sun L, Liu M, Chen H, Bull SB, Boright AP, Paterson AD (2007) **Multiple variants in vascular endothelial growth factor (VEGFA) are risk factors for time to severe retinopathy in type 1 diabetes: the DCCT/EDIC genetics study**. *Diabetes* 56:2161-2168.

Al-Kateb H, Boright AP, Mirea L, Xie X, Sutradhar R, Mowjoodi A, Bharaj B, Liu M, Bucksa JM, Arends VL, Steffes MW, Cleary PA, Sun W, Lachin JM, Thorner PS, Ho M, McKnight AJ, Maxwell AP, Savage DA, Kidd KK, Kidd JR, Speed WC, Orchard TJ, Miller RG, Sun L, Bull SB, Paterson AD (2008) **Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Genetics study**. *Diabetes* 57:218-228.

Ala-Korpela M, Korhonen A, Keisala J, Hörkkö S, Korpi P, Ingman LP, Jokisaari J, Savolainen MJ & Kesäniemi YA (1994) **^{1}H NMR-based absolute quantitation of human lipoproteins and their lipid contents directly from plasma**. *J Lipid Res* 35:2292-2304.

Ala-Korpela M, Hiltunen Y & Bell JD (1995a) **Quantification of biomedical NMR data using artificial neural network analysis: lipoprotein lipid profiles from ^{1}H NMR data of human plasma**. *NMR Biomed* 8:235-244.

Ala-Korpela M (1995b) **^{1}H NMR spectroscopy of human blood plasma**. *Progr Nucl Magn Reson Spectr* 27:475–554.

Ala-Korpela M, Lankinen N, Salminen A, Suna T, Soininen P, Laatikainen R, Ingman P, Jauhiainen M, Taskinen M, Héberger K & Kaski K (2007) **The inherent accuracy of ^{1}H NMR spectroscopy to quantify plasma lipoproteins is subclass dependent**. *Atherosclerosis* 190:352-358.

Alberti KG & Zimmet PZ (1998) **Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation**. *Diabet Med* 15:539-553.

Alberti KG, Zimmet P & Shaw J (2006) **Metabolic syndrome – a new world-wide defi-**

nition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* 23:469-480.

Alibrahim E, Donaghue KC, Rogers S, Hing S, Jenkins AJ, Chan A & Wong TY (2006) **Retinal vascular caliber and risk of retinopathy in young patients with type 1 diabetes**. *Ophthalmology* 113:1499-1503.

Almaas E (2007) **Biological impacts and context of network theory**. *J Exp Biol* 210:1548-1558.

Alpaydin E & Gurgen F (1998) **Comparison of statistical and neural classifiers and their applications to optical character recognition and speech classification**. Neural network systems techniques and applications (Ed. Leondes CT), Academic Press 61-88.

Amari S, Murata N, Muller K, Finke M & Yang H (1997) **Asymptotic Statistical Theory of Overtraining and Cross-Validation**. *IEEE Transactions on Neural Networks* 8:985-996.

Anderson RJ, Freedland KE, Clouse RE & Lustman PJ (2001) **The prevalence of comorbid depression in adults with diabetes: a meta-analysis**. *Diabetes Care* 24:1069-1078.

Antoniucci D, Valenti R, Migliorini A, Parodi G, Moschi G, Memisha G, Santoro GM & Cerisano G (2004) **Impact of insulin-requiring diabetes mellitus on effectiveness of reperfusion and outcome of patients undergoing primary percutaneous coronary intervention for acute myocardial infarction**. *Am J Cardiol* 93:1170-1172.

Argacha JF, Fontaine D, Adamopoulos D, Ajose A, van de Borne P, Fontaine J & Berkenboom G (2008) **Acute effect of sidestream cigarette smoke extract on vascular endothelial function**. *J Cardiovasc Pharmacol* 52:262-267.

Ashby D (2006) **Bayesian statistics in medicine: a 25 year review**. *Stat Med* 25:3589-3631.

Astrup AS, Tarnow L, Rossing P, Pietraszek L, Riis Hansen P & Parving H (2005) **Improved prognosis in type 1 diabetic patients with nephropathy: a prospective follow-up study**. *Kidney Int* 68:1250-1257.

Atherton HJ, Bailey NJ, Zhang W, Taylor J, Major H, Shockcor J, Clarke K & Griffin JL (2006) **A combined ^{1}H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome**. *Physiol Genomics* 27:178-186.

Athyros VG, Liberopoulos EN, Mikhailidis DP, Papageorgiou AA, Ganotakis ES, Tziomalos K, Kakafika AI, Karagiannis A, Lambropoulos S & Elisaf M (2007) **Association of drinking pattern and alcohol beverage type with the prevalence of metabolic syndrome, diabetes, coronary heart disease, stroke, and peripheral arterial disease in a Mediterranean cohort**. *Angiology* 58:689-697.

Auryan S & Itamar R (2008) **Gender-specific care of diabetes mellitus: particular considerations in the management of diabetic women**. *Diabetes Obes Metab* 10:1135-1156.

Ayodele OE, Alebiosu CO & Salako BL (2004) **Diabetic nephropathy – a review of the natural history, burden, risk factors and treatment**. *J Natl Med Assoc* 96:1445-1454.

Badr GA, Tang J, Ismail-Beigi F & Kern TS (2000) **Diabetes downregulates GLUT1 expression in the retina and its microvessels but not in the cerebral cortex or its microvessels**. *Diabetes* 49:1016-1021.

Bain SC, Gill GV, Dyer PH, Jones AF, Murphy M, Jones KE, Smyth C & Barnett AH (2003) **Characteristics of type 1 diabetes of over 50 years duration (the Golden Years Cohort)**. *Diabet Med* 20:808-811.

Balamuthusamy S, Srinivasan L, Verma M, Adigopula S, Jalandara N, Hathiwala S & Smith E (2008) **Renin angiotensin system blockade and cardiovascular outcomes in patients with chronic kidney disease and proteinuria: a meta-analysis**. *Am Heart J* 155:791-805.

Banday AA & Lokhandwala MF (2008) **Oxidative stress-induced renal angiotensin AT1 receptor upregulation causes increased stimulation of sodium transporters and hyperten-**

sion. *Am J Physiol Renal Physiol* 295:F698-706.

Banting F, Best C, Collip J, Campbell W & Fletcher A (1922) **Pancreatic extracts in the treatment of diabetes mellitus, preliminary report**. *Can Med Assoc J* 12:141-146.

Barabasi A & Albert R (1999) **Emergence of scaling in random networks**. *Science* 286:509-512.

Barnoya J & Glantz SA (2005) **Cardiovascular effects of secondhand smoke: nearly as large as smoking**. *Circulation* 111:2684-2698.

Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I & Wrana JL (2005) **High-throughput mapping of a dynamic signaling network in mammalian cells**. *Science* 307:1621-1625.

Basta G, Lazzerini G, Massaro M, Simoncini T, Tanganelli P, Fu C, Kislinger T, Stern DM, Schmidt AM & De Caterina R (2002) **Advanced glycation end products activate endothelium through signal-transduction receptor RAGE: a mechanism for amplification of inflammatory responses**. *Circulation* 105:816-822.

Baynes JW (2003) **Chemical modification of proteins by lipids in diabetes**. *Clin Chem Lab Med* 41:1159-1165.

Beckman JA, Creager MA & Libby P (2002) **Diabetes and atherosclerosis: epidemiology, pathophysiology, and management**. *JAMA* 287:2570-2581.

Beckonert O, Monnerjahn J, Bonk U & Leibfritz D (2003) **Visualizing metabolic changes in breast-cancer tissue using ^{1}H-NMR spectroscopy and self-organizing maps**. *NMR Biomed* 16:1-11.

Beltramo E, Berrone E, Giunti S, Gruden G, Perin PC & Porta M (2006) **Effects of mechanical stress and high glucose on pericyte proliferation, apoptosis and contractile phenotype**. *Exp Eye Res* 83:989-994.

Bengio Y (1995) **Convergence properties of the k-means algorithms**. Advances in Neural Information Processing Systems 7, MIT Press 585-592.

Bensinger SJ & Tontonoz P (2008) **Integration of metabolism and inflammation by lipid-activated nuclear receptors**. *Nature* 454:470-477.

Benson M & Breitling R (2006) **Network theory to understand microarray studies of complex diseases**. *Curr Mol Med* 6:695-701.

Berg K (1963) **A new serum type system in man: the Lp system**. *Acta Pathol Microbiol Scand* 59:362-382.

Bergholdt R, Størling ZM, Lage K, Karlberg EO, Olason PI, Aalund M, Nerup J, Brunak S, Workman CT & Pociot F (2007) **Integrative analysis for finding genes and networks involved in diabetes and other complex diseases**. *Genome Biol* 8:R253.

Beulens JWJ, Kruidhof JS, Grobbee DE, Chaturvedi N, Fuller JH & Soedamah-Muthu SS (2008) **Alcohol consumption and risk of microvascular complications in type 1 diabetes patients: the EURODIAB Prospective Complications Study**. *Diabetologia* 51:1631-1638.

Bishop C (1996) **Neural networks for pattern recognition**. Oxford University Press.

Bishop C, Parker GC & Coscina DV (2004) **Systemic nicotine alters whole-body fat utilization in female rats**. *Physiol Behav* 80:563-567.

Bland JM & Altman DG (1986) **Statistical methods for assessing agreement between two methods of clinical measurement**. *Lancet* 1:307-310.

Bloomgarden ZT (2005) **2nd International Symposium on triglycerides and HDL: metabolic syndrome**. *Diabetes Care* 28:2577-2584.

Bloomgarden ZT (2008) **Diabetic nephropathy**. *Diabetes Care* 31:823-827.

Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, Rodriguez MA & Balakireva

L (2009) **Clickstream data yields high-resolution maps of science**. *PLoS ONE* 4:e4803.

Bonnet F, Cooper ME, Kawachi H, Allen TJ, Boner G & Cao Z (2001) **Irbesartan normalises the deficiency in glomerular nephrin expression in a model of diabetes and hypertension**. *Diabetologia* 44:874-877.

Borch-Johnsen K & Kreiner S (1987) **Proteinuria: value as predictor of cardiovascular mortality in insulin dependent diabetes mellitus**. *BMJ (Clin Res Ed)* 294:1651-1654.

Borch-Johnsen K, Nørgaard K, Hommel E, Mathiesen ER, Jensen JS, Deckert T & Parving HH (1992) **Is diabetic nephropathy an inherited complication?** *Kidney Int* 41:719-722.

Boren SA, Gunlock TL, Schaefer J & Albright A (2007) **Reducing risks in diabetes self-management: a systematic review of the literature**. *Diabetes Educ* 33:1053-77; discussion 1078-9.

Boright AP, Paterson AD, Mirea L, Bull SB, Mowjoodi A, Scherer SW, Zinman B (2005) **Genetic variation at the ACE gene is associated with persistent microalbuminuria and severe nephropathy in type 1 diabetes: the DCCT/EDIC Genetics Study**. *Diabetes* 54:1238-1244.

Bostom AG, Bausserman L, Jacques PF, Liaugaudas G, Selhub J & Rosenberg IH (1999) **Cystatin C as a determinant of fasting plasma total homocysteine levels in coronary artery disease patients with normal serum creatinine**. *Arterioscler Thromb Vasc Biol* 19:2241-2244.

Brange J (1997) **The new era of biotech insulin analogues**. *Diabetologia* 40 Suppl 2:S48-53.

Breed LM (1918) **Metabolism in diabetes, nephritis and cholecystitis**. *Cal State J Med* 16:327-330.

Bretzel RG (1997) **Prevention and slowing down the progression of the diabetic nephropathy through antihypertensive therapy**. *J Diabetes Complications* 11:112-122.

Brewster UC & Perazella MA (2004) **The renin-angiotensin-aldosterone system and the kidney: effects on kidney disease**. *Am J Med* 116:263-272.

Brooks B, Molyneaux L & Yue DK (1999) **Augmentation of central arterial pressure in type 1 diabetes**. *Diabetes Care* 22:1722-1727.

Brownlee M (2001) **Biochemistry and molecular cell biology of diabetic complications**. *Nature* 414:813-820.

Bunn HF, Gabbay KH & Gallop PM (1978) **The glycosylation of hemoglobin: relevance to diabetes mellitus**. *Science* 200:21-27.

Burr L & Fry H (2005) **Biased regression: the case for cautious application**. *Technometrics* 47:284-296.

Burton I, Quilliam M & Walter J (2005) **Quantitative ^{1}H NMR with external standards: use in preparation of calibration solutions for algal toxins and other natural products**. *Analytical Chemistry* 77:3123-3131.

Busik JV, Mohr S & Grant MB (2008) **Hyperglycemia-induced reactive oxygen species toxicity to endothelial cells is dependent on paracrine mediators**. *Diabetes* 57:1952-1965.

Buysschaert M, Joudi I, Wallemacq P & Hermans MP (2003) **Performance of serum cystatin-C versus serum creatinine in subjects with type 1 diabetes**. *Diabetes Care* 26:1320.

Bylesjö M, Rantalainen M, Nicholson JK, Holmes E & Trygg J (2008) **K-OPLS package: kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space**. *BMC Bioinformatics* 9:106.

Caramori ML, Fioretto P & Mauer M (2006) **Enhancing the predictive value of urinary albumin for diabetic nephropathy**. *J Am Soc Nephrol* 17:339-352.

Carr H & Purcell E (1954) **Effects of diffusion on free precession in nuclear magnetic resonance experiments**. *Phys Rev* 94:9630–9638.

Carrero JJ, Qureshi AR, Parini P, Arver S, Lindholm B, Bárány P, Heimbürger O & Stenvinkel P (2009) **Low serum testosterone increases**

mortality risk among male dialysis patients. *J Am Soc Nephrol* 20:613-620.

Carter PJ, Cutfield WS, Hofman PL, Gunn AJ, Wilson DA, Reed PW & Jefferies C (2008) **Ethnicity and social deprivation independently influence metabolic control in children with type 1 diabetes**. *Diabetologia* 51:1835-1842.

Cave MC, Hurt RT, Frazier TH, Matheson PJ, Garrison RN, McClain CJ & McClave SA (2008) **Obesity, inflammation, and the potential application of pharmaconutrition**. *Nutr Clin Pract* 23:16-34.

Kearney PM, Blackwell L, Collins R, Keech A, Simes J, Peto R, Armitage J & Baigent C (2008) **Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis**. *Lancet* 371:117-125.

Chan ECY, Koh PK, Mal M, Cheah PY, Eu KW, Backshall A, Cavill R, Nicholson JK & Keun HC (2009) **Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS)**. *J Proteome Res* 8:352-361.

Chatham JC, Gao ZP, Bonen A & Forder JR (1999) **Preferential inhibition of lactate oxidation relative to glucose oxidation in the rat heart following diabetes**. *Cardiovasc Res* 43:96-106.

Chaturvedi N, Fuller JH, Taskinen MR (2001) **Differing associations of lipid and lipoprotein disturbances with the macrovascular and microvascular complications of type 1 diabetes**. *Diabetes Care* 24:2071-2077.

Chen H, Ho C & Li C (2006) **Age and sex may significantly interact with diabetes on the risks of lower-extremity amputation and peripheral revascularization procedures: evidence from a cohort of a half-million diabetic patients**. *Diabetes Care* 29:2409-2414.

Chiolero A, Faeh D, Paccaud F & Cornuz J (2008) **Consequences of smoking for body weight, body fat distribution, and insulin resistance**. *Am J Clin Nutr* 87:801-809.

Chiou WL & Hsu FH (1975) **Pharmacokinetics of creatinine in man and its implications in the monitoring of renal function and in dosage regimen modifications in patients with renal insufficiency**. *J Clin Pharmacol* 15:427-434.

Christodoulou CI, Pattichis CS, Pantziaris M & Nicolaides A (2003) **Texture-based classification of atherosclerotic carotid plaques**. *IEEE Trans Med Imaging* 22:902-912.

Chuang PY, Yu Q, Fang W, Uribarri J & He JC (2007) **Advanced glycation endproducts induce podocyte apoptosis by activation of the FOXO4 transcription factor**. *Kidney Int* 72:965-976.

Clayton TA, Lindon JC, Cloarec O, Antti H, Charuel C, Hanton G, Provost J, Le Net J, Baker D, Walley RJ, Everett JR & Nicholson JK (2006) **Pharmaco-metabonomic phenotyping and personalized drug treatment**. *Nature* 440:1073-1077.

Cleary PA, Orchard TJ, Genuth S, Wong ND, Detrano R, Backlund JC, Zinman B, Jacobson A, Sun W, Lachin JM, Nathan DM (2006) **The effect of intensive glycemic treatment on coronary artery calcification in type 1 diabetic participants of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study**. *Diabetes* 55:3556-3565.

Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, Nicholson JK & Holmes E (2005) **Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^{1}H NMR spectroscopic metabonomic studies.** *Anal Chem* 77:517-526.

Cockcroft D & Gault M (1976) **Prediction of creatinine clearance from serum creatinine**. *Nephron* 16:31-41.

Collins P, Rosano G, Casey C, Daly C, Gambacciani M, Hadji P, Kaaja R, Mikkola T, Palacios S, Preston R, Simon T, Stevenson J & Stramba-Badiale M (2007) **Management of cardiovascular risk in the peri-menopausal woman: a consensus statement of European cardiolo-**

gists and gynaecologists. *Eur Heart J* 28:2028-2040.

Comper WD, Osicka TM, Clark M, MacIsaac RJ & Jerums G (2004) **Earlier detection of microalbuminuria in diabetic patients using a new urinary albumin assay**. *Kidney Int* 65:1850-1855.

Constantinou M, Papakonstantinou E, Spraul M, Sevastiadou S, Costalos C, Koupparis M, Shulpis K, Tsantili-Kakoulidou A & Emmanuel M (2005) **^{1}H NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine**. *Analytica Chimica Acta* 542:169–177.

Cooper RG (2006) **Effect of tobacco smoking on renal function**. *Indian J Med Res* 124:261-268.

Cordova CM, Schneider CR, Juttel ID & Cordova M (2004) **Comparison of LDL-cholesterol direct measurement with the estimate using the Friedewald formula in a sample of 10,664 patients**. *Arq Bras Cardiol* 83:482-7; 476-81.

Cossrow N & Falkner B (2004) **Race/ethnic issues in obesity and obesity-related comorbidities**. *J Clin Endocrinol Metab* 89:2590-2594.

Cotman CW & Engesser-Cesar C (2002) **Exercise enhances and protects brain function**. *Exerc Sport Sci Rev* 30:75-79.

Coughlan MT, Thorburn DR, Penfold SA, Laskowski A, Harcourt BE, Sourris KC, Tan ALY, Fukami K, Thallas-Bonke V, Nawroth PP, Brownlee M, Bierhaus A, Cooper ME & Forbes JM (2009) **RAGE-induced cytosolic ROS promote mitochondrial superoxide generation in diabetes**. *J Am Soc Nephrol* 20:742-752

Cover T & Hart P (1967) **Nearest neighbor pattern classification**. *IEEE Transactions on Information Theory*, 13:21-27.

Cox D (1972) **Regression models and life tables**. *J Royal Stat Soc* 34:187-220.

Craig A, Cloarec O, Holmes E, Nicholson J & Lindon J (2006) **Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets**. *Anal Chem* 78:2262-2267.

Dahl-Jørgensen K, Larsen JR & Hanssen KF (2005) **Atherosclerosis in childhood and adolescent type 1 diabetes: early disease, early treatment?** *Diabetologia* 48:1445-1453.

Dahlquist G, Möllsten A, Källén B (2008) **Hospitalization for vascular complications in childhood onset type 1 diabetes – effects of gender and age at onset**. *Acta Paediatr* 97:483-488.

Daneman D (2006) **Type 1 diabetes**. *Lancet* 367:847-858.

Davidson R & Harel D (1996) **Drawing graphs nicely using simulated annealing**. *ACM Transactions on Graphics* 15:301 - 331 .

Davidson MH & Toth PP (2007a) **High-density lipoprotein metabolism: potential therapeutic targets**. *Am J Cardiol* 100:n32-40.

Davidson JA, Ciulla TA, McGill JB, Kles KA & Anderson PW (2007b) **How the diabetic eye loses vision**. *Endocrine* 32:107-116.

Davidson MH (2008) **Is LDL-C passed its prime? The emerging role of non-HDL, LDL-P, and ApoB in CHD risk assessment**. *Arterioscler Thromb Vasc Biol* 28:1582-1583.

Day C (2007) **Metabolic syndrome, or what you will: definitions and epidemiology**. *Diab Vasc Dis Res* 4:32-38.

DCCT (1993) **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus**. *N Engl J Med* 329:977–986.

DCCT (1995a) **The effect of intensive diabetes therapy on the development and progression of neuropathy**. *Ann Int Med* 122:561-568.

DCCT (1995b) **Progression of retinopathy with intensive versus conventional treatment in the Diabetes Control and Complications Trial**. *Ophthalmology* 102:647-661.

DCCT (1998) **Early worsening of diabetic retinopathy in the Diabetes Control and**

Complications Trial. *Arch Ophthalmol* 116:874-886.

DCCT (2000) **Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy**. *N Engl J Med* 342:381-389.

DCCT (2005) **Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes**. *N Engl J Med* 353:2643-2653.

De Block CEM, De Leeuw IH & Van Gaal LF (2005) **Impact of overweight on chronic microvascular complications in type 1 diabetic patients**. *Diabetes Care* 28:1649-1655.

De Boer IH, Sibley SD, Kestenbaum B, Sampson JN, Young B, Cleary PA, Steffes MW, Weiss NS, Brunzell JD (2007) **Central obesity, incident microalbuminuria, and change in creatinine clearance in the Epidemiology of Diabetes Interventions and Complications Study**. *J Am Soc Nephrol* 18:235-243.

De Graaf R & Behar K (2003) **Quantitative ^{1}H NMR spectroscopy of blood plasma metabolites**. *Anal Chem* 75:2100-2104.

De Groot M, Anderson R, Freedland KE, Clouse RE & Lustman PJ (2001) **Association of depression and diabetes complications: a meta-analysis**. *Psychosom Med* 63:619-630.

De Groot J (2004) **The AGE of the matrix: chemistry, consequence and cure**. *Curr Opin Pharmacol* 4:301-305.

De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC & Van Criekinge W (2008) **NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm**. *Anal Chem* 80:3783-3790.

Demiriz A, Bennett KP, Breneman CM & Embrechts MJ (2001) **Support vector machine regression in chemometrics**. Computing Science and Statistics: Proceedings of the 33rd Symposium on the Interface.

Desai ND, Cohen EA, Naylor CD, Fremes SE (2004) **A randomized comparison of radial-artery and saphenous-vein coronary bypass grafts**. *N Engl J Med* 351:2302-2309.

Després J & Lemieux I (2006) **Abdominal obesity and metabolic syndrome**. *Nature* 444:881-887.

Dewar BJ, Bradford BU & Thurman RG (2002) **Nicotine increases hepatic oxygen uptake in the isolated perfused rat liver by inhibiting glycolysis**. *J Pharmacol Exp Ther* 301:930-937.

Domon B & Aebersold R (2006) **Mass spectrometry and protein analysis**. *Science* 312:212-217.

Doublier S, Salvidio G, Lupia E, Ruotsalainen V, Verzola D, Deferrari G & Camussi G (2003) **Nephrin expression is reduced in human diabetic nephropathy: evidence for a distinct role for glycated albumin and angiotensin II**. *Diabetes* 52:1023-1030.

Drummond K, Mauer M (2002) **The early natural history of nephropathy in type 1 diabetes: II. Early renal structural changes in type 1 diabetes**. *Diabetes* 51:1580-1587.

Dubé M, Valois P, Prud'homme D, Weisnagel SJ & Lavoie C (2006) **Physical activity barriers in diabetes: development and validation of a new scale**. *Diabetes Res Clin Pract* 72:20-27.

Dunn W, Bailey N & Johnson H (2005) **Measuring the metabolome: current analytical technologies**. *Analyst* 130:606–625.

Durvasula RV & Shankland SJ (2008) **Activation of a local renin angiotensin system in podocytes by glucose**. *Am J Physiol Renal Physiol* 294:F830-9.

Duvillard L, Florentin E, Lalanne-Mistrich ML, Petit JM, Baillot-Rudoni S, Brun-Pacaud A, Brun JM, Gambert P & Vergès B (2005) **Normal metabolism of apolipoprotein B100-containing lipoproteins despite qualitative abnormalities in type 1 diabetic men**. *Diabetologia* 48:1366-1372.

Eckel RH, Grundy SM & Zimmet PZ (2005) **The metabolic syndrome**. *Lancet* 365:1415-1428.

Edwards JL, Vincent AM, Cheng HT & Feldman EL (2008) **Diabetic neuropathy: mechanisms to management**. *Pharmacol Ther* 120:1-34.

Efron B & Tibshirani R (1995) **Cross-validation and the bootstrap: estimating the error rate of a prediction rule**. Natural Science and Engineering Research Council of Canada.

Eisenreich W & Adelbert Bacher A (2007) **Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry**. *Phytochemistry* 68:2799–2815.

Eliasson B, Mero N, Taskinen MR & Smith U (1997) **The insulin resistance syndrome and postprandial lipid intolerance in smokers**. *Atherosclerosis* 129:79-88.

Ellis DI, Dunn WB, Griffin JL, Allwood JW & Goodacre R (2007) **Metabolic fingerprinting as a diagnostic tool**. *Pharmacogenomics* 8:1243-1266.

Enzlin P, Mathieu C, Van Den Bruel A, Vanderschueren D & Demyttenaere K (2003) **Prevalence and predictors of sexual dysfunction in patients with type 1 diabetes**. *Diabetes Care* 26:409-414.

Ergon R (2004) **Informative PLS score-loading plots for process understanding and monitoring**. *J Process Contr* 14:889-897.

Ergün UU, Serhatlioğlu S, Hardalaç F & Güler I (2004) **Classification of carotid artery stenosis of patients with diabetes by neural network and logistic regression**. *Comput Biol Med* 34:389-405.

Evers IM, de Valk HW & Visser GHA (2004) **Risk of complications of pregnancy in women with type 1 diabetes: nationwide prospective study in the Netherlands**. *BMJ* 328:915.

Ewens KG, George RA, Sharma K, Ziyadeh FN & Spielman RS (2005) **Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test**. *Diabetes* 54:3305-3318.

Fagerudd JA, Tarnow L, Jacobsen P, Stenman S, Nielsen FS, Pettersson-Fernholm KJ, Grönhagen-Riska C, Parving HH & Groop PH (1998) **Predisposition to essential hypertension and development of diabetic nephropathy in IDDM patients**. *Diabetes* 47:439-444.

Failla M, Grappiolo A, Carugo S, Calchera I, Giannattasio C & Mancia G (1997) **Effects of cigarette smoking on carotid and radial artery distensibility**. *J Hypertens* 15:1659-1664.

Fajans SS, Bell GI & Polonsky KS (2001) **Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young**. *N Engl J Med* 345:971-980.

Farkas O & Héberger K (2005) **Comparison of ridge regression, partial least-Squares, pairwise correlation, forward- and best subset selection methods for prediction of retention indices for aliphatic alcohols**. *J Chem Inf Model* 45:339-346.

Ferrara CT, Wang P, Neto EC, Stevens RD, Bain JR, Wenner BR, Ilkayeva OR, Keller MP, Blasiole DA, Kendziorski C, Yandell BS, Newgard CB & Attie AD (2008) **Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling**. *PLoS Genet* 4:e1000034.

Ferris FL3, Davis MD & Aiello LM (1999) **Treatment of diabetic retinopathy**. *N Engl J Med* 341:667-678.

Filippi CM & von Herrath MG (2008) **Viral trigger for type 1 diabetes: pros and cons**. *Diabetes* 57:2863-2871.

Finne P, Reunanen A, Stenman S, Groop P & Grönhagen-Riska C (2005) **Incidence of end-stage renal disease in patients with type 1 diabetes**. *JAMA* 294:1782-1787.

Fioretto P, Sutherland DER, Najafian B & Mauer M (2006) **Remodeling of renal interstitial and tubular lesions in pancreas transplant recipients**. *Kidney Int* 69:907-912.

Fisher R (1936) **The use of multiple measurements in taxonomic problems**. *Ann Eugen* 7:179-188.

Fisher BM (1997) **Heart abnormalities in**

IDDM. *Diabetologia* 40 Suppl 2:S127-9.

Forbes JM, Coughlan MT & Cooper ME (2008) **Oxidative stress as a major culprit in kidney disease in diabetes**. *Diabetes* 57:1446-1454.

Forsén A, Kangro M, Sterner G, Norrgren K, Thorsson O, Wollmer P & Sundkvist G (2004) **A 14-year prospective study of autonomic nerve function in Type 1 diabetic patients: association with nephropathy**. *Diabet Med* 21:852-858.

Frank LL, Sorensen BE, Yasui Y, Tworoger SS, Schwartz RS, Ulrich CM, Irwin ML, Rudolph RE, Rajan KB, Stanczyk F, Bowen D, Weigle DS, Potter JD & McTiernan A (2005) **Effects of exercise on metabolic risk variables in overweight postmenopausal women: a randomized clinical trial**. *Obes Res* 13:615-625.

Friedewald W, Levy R & Fredrickson D (1972) E**stimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge**. *Clin Chem* 18:499-502.

Fröjdö S, Sjölind L, Parkkonen M, Mäkinen V, Kilpikari R, Pettersson-Fernholm K, Forsblom C, Fagerudd J, Tikellis C, Cooper ME, Wessman M, Groop P (2005) **Polymorphisms in the gene encoding angiotensin I converting enzyme 2 and diabetic nephropathy**. *Diabetologia* 48:2278-2281.

Fruchtermann T & Reingold E (1991) **Graph drawing by force-directed placement**. *Software Pract Exp* 21:1129–1164.

Frystyk J, Tarnow L, Hansen TK, Parving H & Flyvbjerg A (2005) **Increased serum adiponectin levels in type 1 diabetic patients with microvascular complications**. *Diabetologia* 48:1911-1918.

Fukami K, Ueda S, Yamagishi S, Kato S, Inagaki Y, Takeuchi M, Motomiya Y, Bucala R, Iida S, Tamaki K, Imaizumi T, Cooper ME & Okuda S (2004) **AGEs activate mesangial TGF-beta-Smad signaling via an angiotensin II type I receptor interaction**. *Kidney Int* 66:2137-2147.

Gardner CD, Fortmann SP & Krauss RM (1996) **Association of small low-density lipoprotein particles with the incidence of coronary artery disease in men and women**. *JAMA* 276:875-881.

Gershater MA, Löndahl M, Nyberg P, Larsson J, Thörne J, Eneroth M & Apelqvist J (2009) **Complexity of factors related to outcome of neuropathic and neuroischaemic/ischaemic diabetic foot ulcers: a cohort study**. *Diabetologia* 52:398-407.

Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, Mewes H, Wichmann H, Weinberger KM, Adamski J, Illig T & Suhre K (2008) **Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum**. *PLoS Genet* 4:e1000282.

Giorgino F, Laviola L, Cavallo Perin P, Solnica B, Fuller J & Chaturvedi N (2004) **Factors associated with progression to macroalbuminuria in microalbuminuric Type 1 diabetic patients: the EURODIAB Prospective Complications Study**. *Diabetologia* 47:1020-1028.

Girvan M & Newman MEJ (2002) **Community structure in social and biological networks**. *Proc Natl Acad Sci USA* 99:7821-7826.

Gnudi L, Thomas SM & Viberti G (2007) **Mechanical forces in diabetic kidney disease: a trigger for impaired glucose metabolism**. *J Am Soc Nephrol* 18:2226-2232.

Godsland IF, Leyva F, Walton C, Worthington M & Stevenson JC (1998) **Associations of smoking, alcohol and physical activity with risk factors for coronary heart disease and diabetes in the first follow-up cohort of the Heart Disease and Diabetes Risk Indicators in a Screened Cohort study (HDDRISC-1)**. *J Intern Med* 244:33-41.

Goebel-Fabbri AE, Fikkan J, Franko DL, Pearson K, Anderson BJ & Weinger K (2008) **Insulin restriction and associated morbidity and mortality in women with type 1 diabetes**. *Diabetes Care* 31:415-419.

Goh K, Cusick ME, Valle D, Childs B, Vidal M & Barabási A (2007) **The human disease network**. *Proc Natl Acad Sci USA* 104:8685-8690.

Goldin A, Beckman JA, Schmidt AM & Creager MA (2006) **Advanced glycation end products: sparking the development of diabetic vascular injury**. *Circulation* 114:597-605.

Goldstein L & Brown S (1977) **The low-density lipoprotein pathway and its relation to atherosclerosis**. *Ann Rev Biochem* 46:897-930.

Gomes MB & Gonçalves MF (2001) **Is there a physiological variability for albumin excretion rate? Study in patients with diabetes type 1 and non-diabetic individuals**. *Clin Chim Acta* 304:117-123.

Gordin D, Rönnback M, Forsblom C, Heikkilä O, Saraheimo M & Groop PH (2007a) **Acute hyperglycaemia rapidly increases arterial stiffness in young patients with type 1 diabetes**. *Diabetologia* 50:1808-1814.

Gordin D, Hiilesmaa V, Fagerudd J, Rönnback M, Forsblom C, Kaaja R, Teramo K, Groop PH (2007b) **Pre-eclampsia but not pregnancy-induced hypertension is a risk factor for diabetic nephropathy in type 1 diabetic women**. *Diabetologia* 50:516-522.

Gordin D, Forsblom C, Rönnback M & Groop PH (2008a) **Acute hyperglycaemia disturbs cardiac repolarization in Type 1 diabetes**. *Diabet Med* 25:101-105.

Gordin D, Rönnback M, Forsblom C, Mäkinen VP, Saraheimo M & Groop PH (2008b) **Glucose variability, blood pressure and arterial stiffness in type 1 diabetes**. *Diabetes Res Clin Pract* 80:e4-7.

Gordois A, Scuffham P, Shearer A, Oglesby A & Tobian JA (2003) **The health care costs of diabetic peripheral neuropathy in the US**. *Diabetes Care* 26:1790-1795.

Griffin JL & Nicholls AW (2006) **Metabolomics as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders**. *Pharmacogenomics* 7:1095-1107.

Groop PH, Elliott T, Ekstrand A, Franssila-Kallunki A, Friedman R, Viberti GC & Taskinen MR (1996) **Multiple lipoprotein abnormalities in type I diabetic patients with renal disease**. *Diabetes* 45:974-979.

Groop PH, Forsblom C & Thomas MC (2005) **Mechanisms of disease: pathway-selective insulin resistance and microvascular complications of diabetes**. *Nat Clin Pract Endocrinol Metab* 1:100-110.

Groop PH, Thomas MC, Rosengård-Bärlund M, Mills V, Rönnback M, Thomas S, Forsblom C, Taskinen M & Viberti G (2007) **HDL composition predicts new-onset cardiovascular disease in patients with type 1 diabetes**. *Diabetes Care* 30:2706-2707.

Groop PH, Thomas MC, Moran JL, Wadèn J, Thorn LM, Mäkinen VP, Rosengård-Bärlund M, Saraheimo M, Hietala K, Heikkilä O, Forsblom C (2009) **The presence and severity of chronic kidney disease predicts all-cause mortality in type 1 diabetes**. *Diabetes* 58:1651-1658.

Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML & Zelmanovitz T (2005) **Diabetic nephropathy: diagnosis, prevention, and treatment**. *Diabetes Care* 28:164-176.

Guimerà R, Mossa S, Turtschi A & Amaral LAN (2005) **The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles**. *Proc Natl Acad Sci USA* 102:7794-7799.

Ha H, Hwang I, Park JH & Lee HB (2008) **Role of reactive oxygen species in the pathogenesis of diabetic nephropathy**. *Diabetes Res Clin Pract* 82 Suppl 1:S42-5.

Hadjadj S, Duly-Bouhanick B, Bekherraz A, BrIdoux F, Gallois Y, Mauco G, Ebran J & Marre M (2004) **Serum triglycerides are a predictive factor for the development and the progression of renal and retinal complications in patients with type 1 diabetes**. *Diabetes Metab* 30:43-51.

Hammes H, Lin J, Renner O, Shani M, Lundqvist A, Betsholtz C, Brownlee M & Deutsch U (2002) **Pericytes and the pathogenesis of diabetic retinopathy**. *Diabetes* 51:3107-3112.

Hannuksela M, Marcel YL, Kesäniemi YA & Savolainen MJ (1992) **Reduction in the con-**

centration and activity of plasma cholesteryl ester transfer protein by alcohol. *J Lipid Res* 33:737-744.

Hansen TK, Tarnow L, Thiel S, Steffensen R, Stehouwer CD, Schalkwijk CG, Parving H & Flyvbjerg A (2004) **Association between mannose-binding lectin and vascular complications in type 1 diabetes**. *Diabetes* 53:1570-1576.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JT, Galver LM, Fan J, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier J, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok P, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui L, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R & Stewart J (2007) **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 449:851-861.

Harjutsalo V, Katoh S, Sarti C, Tajima N & Tuomilehto J (2004) **Population-based assessment of familial clustering of diabetic nephropathy in type 1 diabetes**. *Diabetes* 53:2449-2454.

Harjutsalo V, Podar T & Tuomilehto J (2005) **Cumulative incidence of type 1 diabetes in 10,168 siblings of Finnish young-onset type 1 diabetic patients**. *Diabetes* 54:563-569.

Harjutsalo V, Reunanen A & Tuomilehto J (2006) **Differential transmission of type 1 diabetes from diabetic fathers and mothers to their offspring**. *Diabetes* 55:1517-1524.

Harjutsalo V, Sjöberg L & Tuomilehto J (2008) **Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study**. *Lancet* 371:1777-1782.

Harrison RF & Kennedy RL (2005) **Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation**. *Ann Emerg Med* 46:431-439.

Hasin T, Hochadel M, Gitt AK, Behar S, Bueno H & Hasin Y (2009) **Comparison of treatment and outcome of acute coronary syndrome in patients with versus patients without diabetes mellitus**. *Am J Cardiol* 103:772-778.

Hassan K, Loar R, Anderson BJ & Heptulla RA

(2006) **The role of socioeconomic status, depression, quality of life, and glycemic control in type 1 diabetes mellitus**. *J Pediatr* 149:526-531.

Haverkos HW, Battula N, Drotman DP & Rennert OM (2003) **Enteroviruses and type 1 diabetes mellitus**. *Biomed Pharmacother* 57:379-385.

Haykin S (1994) **Neural networks. A comprehensive foundation**. Maxmillan College.

He B, Osterholm A, Hoverfält A, Forsblom C, Hjörleifsdóttir EE, Nilsson A, Parkkonen M, Pitkäniemi J, Hreidarsson A, Sarti C, McKnight AJ, Maxwell AP, Tuomilehto J, Groop P & Tryggvason K (2009) **Association of genetic variants at 3q22 with nephropathy in patients with type 1 diabetes mellitus**. *Am J Hum Genet* 84:5-13.

Heidarianpour A, Hajizadeh S, Khoshbaten A, Niaki AG, Bigdili MR & Pourkhalili K (2007) **Effects of chronic exercise on endothelial dysfunction and insulin signaling of cutaneous microvascular in streptozotocin-induced diabetic rats**. *Eur J Cardiovasc Prev Rehabil* 14:746-752.

Heikkilä O, Lundbom N, Timonen M, Groop P, Heikkinen S & Mäkimattila S (2008) **Hyperglycaemia is associated with changes in the regional concentrations of glucose and myo-inositol within the brain**. *Diabetologia* 52:534-540.

Hempe JM, Gomez R, McCarter RJJ & Chalew SA (2002) **High and low hemoglobin glycation phenotypes in type 1 diabetes: a challenge for interpretation of glycemic control**. *J Diabetes Complications* 16:313-320.

Herbst A, Kordonouri O, Schwab KO, Schmidt F, Holl RW (2007) **Impact of physical activity on cardiovascular risk factors in children with type 1 diabetes: a multicenter study of 23,251 patients**. *Diabetes Care* 30:2098-2100.

Hietala K, Forsblom C, Summanen P, Groop P (2008) **Heritability of proliferative diabetic retinopathy**. *Diabetes* 57:2176-2180.

Hillier TA, Rousseau A, Lange C, Lépinay P, Cailleau M, Novak M, Calliez E, Ducimetière P, Balkau B (2006) **Practical way to assess metabolic syndrome using a continuous score obtained from principal components analysis**. *Diabetologia* 49:1528-1535.

Hoek FJ, Kemperman FAW & Krediet RT (2003) **A comparison between cystatin C, plasma creatinine and the Cockcroft and Gault formula for the estimation of glomerular filtration rate**. *Nephrol Dial Transplant* 18:2024-2031.

Hofbauer LC, Brueck CC, Singh SK & Dobnig H (2007) **Osteoporosis in patients with diabetes mellitus**. *J Bone Miner Res* 22:1317-1328.

Hoffman PC, Mauer AM & Vokes EE (2000) **Lung cancer**. *Lancet* 355:479-485.

Hogg RJ, Furth S, Lemley KV, Portman R, Schwartz GJ, Coresh J, Balk E, Lau J, Levin A, Kausz AT, Eknoyan G, Levey AS (2003) **National Kidney Foundation's Kidney Disease Outcomes Quality Initiative clinical practice guidelines for chronic kidney disease in children and adolescents: evaluation, classification, and stratification**. *Pediatrics* 111:1416-1421.

Holl RW, Grabert M, Thon A & Heinze E (1999) **Urinary excretion of albumin in adolescents with type 1 diabetes: persistent versus intermittent microalbuminuria and relationship to duration of diabetes, sex, and metabolic control**. *Diabetes Care* 22:1555-1560.

Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, Chan Q, Ebbels T, De Iorio M, Brown IJ, Veselkov KA, Daviglus ML, Kesteloot H, Ueshima H, Zhao L, Nicholson JK & Elliott P (2008a) **Human metabolic phenotype diversity and its association with diet and blood pressure**. *Nature* 453:396-400.

Holmes E, Wilson ID & Nicholson JK (2008b) **Metabolic phenotyping in health and disease**. *Cell* 134:714-717.

Holzgrabe U, Deubner R, Schollmayer C & Waibel B (2005) **Quantitative NMR spectroscopy – applications in drug analysis**. *J Pharmac Biomed Anal* 38:806–812.

Honeyman MC, Brusic V, Stone NL & Harrison LC (1998) **Neural network-based prediction of candidate T-cell epitopes**. *Nat Biotechnol* 16:966-969.

Hörkkö S, Huttunen K, Korhonen T & Kesäniemi YA (1994) **Decreased clearance of low-density lipoprotein in patients with chronic renal failure**. *Kidney Int* 45:561-570.

Horvat S & Jakas A (2004) **Peptide and amino acid glycation: new insights into the Maillard reaction**. *J Pept Sci* 10:119-137.

Howard BV & Howard WJ (1990) **The compelling case for smoking cessation in diabetics**. *Circulation* 82:299-301.

Hsu H, Hoffmann S, Endlich N, Velic A, Schwab A, Weide T, Schlatter E & Pavenstädt H (2008) **Mechanisms of angiotensin II signaling on cytoskeleton of podocytes**. *J Mol Med* 86:1379-1394.

Huang J, Brennan D, Sattler L, Alderman J, Lane B & O'Mathuna C (2002) **A comparison of calibration methods based on calibration data size and robustness**. *Chemom Intell Lab Syst* 62:25-35.

Hudson BI, Bucciarelli LG, Wendt T, Sakaguchi T, Lalla E, Qu W, Lu Y, Lee L, Stern DM, Naka Y, Ramasamy R, Yan SD, Yan SF, D'Agati V & Schmidt AM (2003) **Blockade of receptor for advanced glycation endproducts: a new target for therapeutic intervention in diabetic complications and inflammatory disorders**. *Arch Biochem Biophys* 419:80-88.

Huttunen NP (1976) **Congenital nephrotic syndrome of Finnish type. Study of 75 patients**. *Arch Dis Child* 51:344-348.

Iafusco D (2006) **Diet and physical activity in patients with type 1 diabetes**. *Acta Biomed* 77 Suppl 1:41-46.

Ikewaki K, Schaefer JR, Frischmann ME, Okubo K, Hosoya T, Mochizuki S, Dieplinger B, Trenkwalder E, Schweer H, Kronenberg F, Koenig P & Dieplinger H (2005) **Delayed in vivo catabolism of intermediate-density lipoprotein and low-density lipoprotein in hemodialysis patients as potential cause of premature atherosclerosis**. *Arterioscler Thromb Vasc Biol* 25:2615-2622.

Ishak KG, Zimmerman HJ & Ray MB (1991) **Alcoholic liver disease: pathologic, pathogenetic and clinical aspects**. *Alcohol Clin Exp Res* 15:45-66.

Ishida T, Haneda M, Maeda S, Koya D & Kikkawa R (1999) **Stretch-induced overproduction of fibronectin in mesangial cells is mediated by the activation of mitogen-activated protein kinase**. *Diabetes* 48:595-602.

Jahromi MM & Eisenbarth GS (2007) **Cellular and molecular pathogenesis of type 1A diabetes**. *Cell Mol Life Sci* 64:865-872.

Jain A, Duin R & Mao J (2000) **Statistical pattern recognition: a review**. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22:4-37.

Jeffcoate SL (2004) **Diabetes control and complications: the role of glycated haemoglobin, 25 years on**. *Diabet Med* 21:657-665.

Jenkins AJ, Lyons TJ, Zheng D, Otvos JD, Lackland DT, McGee D, Garvey WT, Klein RL (2003) **Lipoproteins in the DCCT/EDIC cohort: associations with diabetic nephropathy**. *Kidney Int* 64:817-828.

Jenkins AJ, Best JD, Klein RL & Lyons TJ (2004) **Lipoproteins, glycoxidation and diabetic angiopathy**. *Diabetes Metab Res Rev* 20:349-368.

Jenkins AJ, Rothen M, Klein RL, Moller K, Eldridge L, Zheng D, Durazo-Arvizu R, McGee D, Lackland D, Thorpe SR, Garvey WT, Lyons TJ (2008) **Cross-sectional associations of C-reactive protein with vascular risk factors and vascular complications in the DCCT/EDIC cohort**. *J Diabetes Complic* 22:153-163.

Jensen EX, Fusch C, Jaeger P, Peheim E & Horber FF (1995) **Impact of chronic cigarette smoking on body composition and fuel metabolism**. *J Clin Endocrinol Metab* 80:2181-2185.

Jeon CY, Lokken RP, Hu FB & van Dam RM

(2007) **Physical activity of moderate intensity and risk of type 2 diabetes: a systematic review**. *Diabetes Care* 30:744-752.

Jessen AB, Toubro S & Astrup A (2003) **Effect of chewing gum containing nicotine and caffeine on energy expenditure and substrate utilization in men**. *Am J Clin Nutr* 77:1442-1447.

Jia J, Ding G, Zhu J, Chen C, Liang W, Franki N & Singhal PC (2008) **Angiotensin II infusion induces nephrin expression changes and podocyte apoptosis**. *Am J Nephrol* 28:500-507.

Jo Y, Talmage DA & Role LW (2002) **Nicotinic receptor-mediated effects on appetite and food intake**. *J Neurobiol* 53:618-632.

Johnson V & Albert J (1999) **Ordinal data modeling**. Springer-Verlag New York.

Joles JA, van Goor H & Koomans HA (1998) **Estrogen induces glomerulosclerosis in analbuminemic rats**. *Kidney Int* 53:862-868.

Jones KL, Russo A, Berry MK, Stevens JE, Wishart JM & Horowitz M (2002) **A longitudinal study of gastric emptying and upper gastrointestinal symptoms in patients with diabetes mellitus**. *Am J Med* 113:449-455.

Joosten MM, Beulens JWJ, Kersten S & Hendriks HFJ (2008) **Moderate alcohol consumption increases insulin sensitivity and ADIPOQ expression in postmenopausal women: a randomised, crossover trial**. *Diabetologia* 51:1375-1381.

Jorsal A, Tarnow L, Frystyk J, Lajer M, Flyvbjerg A, Parving H, Vionnet N & Rossing P (2008) **Serum adiponectin predicts all-cause mortality and end stage renal disease in patients with type I diabetes and diabetic nephropathy**. *Kidney Int* 74:649-654.

Jukarainen N, Korhonen S, Laakso M, Korolainen M, Niemitz M, Soininen P, Tuppurainen K, Vepsalainen J, Pirttilä T & Laatikainen R (2008) **Quantification of ^{1}H NMR spectra of human cerebrospinal fluid: a protocol based on constrained total-line-shape analysis**. *Metabolomics* 4:150–160.

Juutilainen A, Lehto S, Rönnemaa T, Pyörälä K & Laakso M (2005) **Type 2 diabetes as a "coronary heart disease equivalent": an 18-year prospective population-based study in Finnish subjects**. *Diabetes Care* 28:2901-2907.

Juutilainen A, Lehto S, Rönnemaa T, Pyörälä K & Laakso M (2008) **Similarity of the impact of type 1 and type 2 diabetes on cardiovascular mortality in middle-aged subjects**. *Diabetes Care* 31:714-719.

Kahn SE, Hull RL & Utzschneider KM (2006) **Mechanisms linking obesity to insulin resistance and type 2 diabetes**. *Nature* 444:840-846.

Karim R, Buchanan T, Hodis H, Li Y & Mack W (2004) **The association of smoking and subclinical atherosclerosis in type 2 diabetes: modification by duration of diabetes**. *Diab Med* 22:81-87.

Kaseta JR, Skafar DF, Ram JL, Jacober SJ & Sowers JR (1999) **Cardiovascular disease in the diabetic woman**. *J Clin Endocrinol Metab* 84:1835-1838.

Kastelein JJP, van der Steeg WA, Holme I, Gaffney M, Cater NB, Barter P, Deedwania P, Olsson AG, Boekholdt SM, Demicco DA, Szarek M, LaRosa JC, Pedersen TR, Grundy SM, (2008) **Lipids, apolipoproteins, and their ratios in relation to cardiovascular events with statin treatment**. *Circulation* 117:3002-3009.

Kaysen GA (1998) **Biological basis of hypoalbuminemia in ESRD**. *J Am Soc Nephrol* 9:2368-2376.

Keeler J (2002) **Understanding NMR spectroscopy**. John Wiley & Sons UK.

Keenan HA, Costacou T, Sun JK, Doria A, Cavellerano J, Coney J, Orchard TJ, Aiello LP & King GL (2007) **Clinical factors associated with resistance to microvascular complications in diabetic patients of extreme disease duration: the 50-year medalist study**. *Diabetes Care* 30:1995-1997.

Kell DB (2006) **Systems biology, metabolic modelling and metabolomics in drug discov-**

ery and development. *Drug Discov Today* 11:1085-1092.

Kemperman FA, Silberbusch J, Slaats EH, van Zanten AP, Weber JA, Krediet RT & Arisz L (1999) **Glomerular filtration rate estimation from plasma creatinine after inhibition of tubular secretion: relevance of the creatinine assay**. *Nephrol Dial Transplant* 14:1247-1251.

Kestilä M, Lenkkeri U, Männikkö M, Lamerdin J, McCready P, Putaala H, Ruotsalainen V, Morita T, Nissinen M, Herva R, Kashtan CE, Peltonen L, Holmberg C, Olsen A & Tryggvason K (1998) **Positionally cloned gene for a novel glomerular protein – nephrin – is mutated in congenital nephrotic syndrome**. *Mol Cell* 1:575-582.

Khaw K, Dowsett M, Folkerd E, Bingham S, Wareham N, Luben R, Welch A & Day N (2007) **Endogenous testosterone and mortality due to all causes, cardiovascular disease, and cancer in men: European prospective investigation into cancer in Norfolk (EPIC-Norfolk) Prospective Population Study**. *Circulation* 116:2694-2701.

Kiers H & Smilde A (2007) **A comparison of various methods for multivariate regression with highly collinear variables**. *Stat Meth Appl* 16:193-228.

Kilpatrick ES, Rigby AS & Atkin SL (2007a) **Variability in the relationship between mean plasma glucose and HbA1c: implications for the assessment of glycemic control**. *Clin Chem* 53:897-901.

Kilpatrick ES, Rigby AS & Atkin SL (2007b) **Insulin resistance, the metabolic syndrome, and complication risk in type 1 diabetes: "double diabetes" in the Diabetes Control and Complications Trial**. *Diabetes Care* 30:707-712.

Kilpatrick ES, Rigby AS, Goode K & Atkin SL (2007c) **Relating mean blood glucose and glucose variability to the risk of multiple episodes of hypoglycaemia in type 1 diabetes**. *Diabetologia* 50:2553-2561.

Kim WY, Astrup AS, Stuber M, Tarnow L, Falk E, Botnar RM, Simonsen C, Pietraszek L, Hansen PR, Manning WJ, Andersen NT & Parving H (2007) **Subclinical coronary and aortic atherosclerosis detected by magnetic resonance imaging in type 1 diabetes with and without diabetic nephropathy**. *Circulation* 115:228-235.

Kim SH, Abbasi F, Lamendola C & Reaven GM (2009) **Effect of moderate alcoholic beverage consumption on insulin sensitivity in insulin-resistant, nondiabetic individuals**. *Metabolism* 58:387-392.

Kimmelstiel P & Wilson C (1936) **Benign and malignant hypertension and nephrosclerosis: a clinical and pathological study**. *Am J Pathol* 12:45-48.

King JY, Ferrara R, Tabibiazar R, Spin JM, Chen MM, Kuchinsky A, Vailaya A, Kincaid R, Tsalenko A, Deng DX, Connolly A, Zhang P, Yang E, Watt C, Yakhini Z, Ben-Dor A, Adler A, Bruhn L, Tsao P, Quertermous T & Ashley EA (2005) **Pathway analysis of coronary atherosclerosis**. *Physiol Genomics* 23:103-118.

Kirschenlohr HL, Griffin JL, Clarke SC, Rhydwen R, Grace AA, Schofield PM, Brindle KM & Metcalfe JC (2006) **Proton NMR analysis of plasma is a weak predictor of coronary artery disease**. *Nat Med* 12:705-710.

Klein RL, McHenry MB, Lok KH, Hunter SJ, Le N, Jenkins AJ, Zheng D, Semler AJ, Brown WV, Lyons TJ, Garvey WT (2004) **Apolipoprotein C-III protein concentrations and gene polymorphisms in type 1 diabetes: associations with lipoprotein subclasses**. *Metabolism* 53:1296-1304.

Klein R, Klein BEK & Moss SE (2005) **Ten-year incidence of self-reported erectile dysfunction in people with long-term type 1 diabetes**. *J Diabetes Complic* 19:35-41.

Klein R, Knudtson MD, Lee KE, Gangnon R & Klein BEK (2008) **The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XXII the twenty-five-year progression of retinopathy in persons with type 1 diabetes**. *Ophthalmology* 115:1859-1868.

Knip M & Siljander H (2008) **Autoimmune mechanisms in type 1 diabetes**. *Autoimmun*

Rev 7:550-557.

Koschinsky ML & Marcovina SM (2003) **The relationship between lipoprotein(a) and the complications of diabetes mellitu**. *Acta Diabetol* 6:695-701.

Kohonen T (2001) **Self-organizing maps**. Springer Series in Information Sciences.

Komoroski R (2000) **Applications of ^{7}Li NMR in biomedicine**. *Magnetic Resonance Imaging* 18:103–116.

Koolman J & Roehm K (2005) **Color Atlas of Biochemistry**. Georg Thieme Verlag.

Korpelainen R, Korpelainen J, Heikkinen J, Väänänen K & Keinänen-Kiukaanniemi S (2006) **Lifelong risk factors for osteoporosis and fractures in elderly women with low body mass index – a population-based study**. *Bone* 39:385-391.

Kraut A, Walld R, Tate R & Mustard C (2001) **Impact of diabetes on employment and income in Manitoba, Canada**. *Diabetes Care* 24:64-68.

Krolewski AS (1999) **Genetics of diabetic nephropathy: evidence for major and minor gene effects**. *Kidney Int* 55:1582-1596.

Kumpula LS, Kumpula JM, Taskinen M, Jauhiainen M, Kaski K & Ala-Korpela M (2008) **Reconsideration of hydrophobic lipid distributions in lipoprotein particles**. *Chem Phys Lipids* 155:57-62.

Kwiterovich POJ (2000) **The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: a current review**. *Am J Cardiol* 86:5L-10L.

Laaksonen DE, Niskanen L, Nyyssönen K, Punnonen K, Tuomainen T, Valkonen V, Salonen R & Salonen JT (2004) **C-reactive protein and the development of the metabolic syndrome and diabetes in middle-aged men**. *Diabetologia* 47:1403-1410.

Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN (2008) **Effect of glycemic exposure on the risk of microvascular complications in the diabetes control and complications trial – revisited**. *Diabetes* 57:995-1001.

Laing SP, Swerdlow AJ, Slater SD, Burden AC, Morris A, Waugh NR, Gatling W, Bingley PJ & Patterson CC (2003a) **Mortality from heart disease in a cohort of 23,000 patients with insulin-treated diabetes**. *Diabetologia* 46:760-765.

Laing SP, Swerdlow AJ, Carpenter LM, Slater SD, Burden AC, Botha JL, Morris AD, Waugh NR, Gatling W, Gale EAM, Patterson CC, Qiao Z & Keen H (2003b) **Mortality from cerebrovascular disease in a cohort of 23 000 patients with insulin-treated diabetes**. *Stroke* 34:418-421.

Lampinen J & Vehtari A (2001) **Bayesian approach for neural networks – review and case studies.** *Neural Networks* 14:14-13.

Lara-Castro C, Fu Y, Chung BH & Garvey WT (2007) **Adiponectin and the metabolic syndrome: mechanisms mediating risk for metabolic and cardiovascular disease**. *Curr Opin Lipidol* 18:263-270.

Larsen J & Hansen LK (1995) **Empirical generalization assessment of neural network models**. Proceedings of the IEEE Workshop on Neural Networks for Signal Processing V, Piscataway 30-39.

Laugesen E, Hansen KW, Knudsen ST, Erlandsen M, Ebbehøj E, Mogensen CE & Poulsen PL (2009) **Increased ambulatory arterial stiffness index and pulse pressure in microalbuminuric patients with type 1 diabetes**. *Am J Hypertens* 22:513-519.

Lee SJ, Lee DW, Kim KS & Lee IK (2001) **Effect of estrogen on endothelial dysfunction in postmenopausal women with diabetes**. *Diabetes Res Clin Pract* 54 Suppl 2:S81-92.

Lee HB, Yu M, Yang Y, Jiang Z & Ha H (2003) **Reactive oxygen species-regulated signaling pathways in diabetic nephropathy**. *J Am Soc Nephrol* 14:S241-5.

Lee D, Park J, Kay KA, Christakis NA, Oltvai ZN & Barabási A (2008) **The implications of human metabolic network topology for dis-**

ease comorbidity. *Proc Natl Acad Sci USA* 105:9880-9885.

Lefrandt JD, Hoogenberg K, van Roon AM, Dullaart RP, Gans RO & Smit AJ (1999) **Baroreflex sensitivity is depressed in microalbuminuric Type I diabetic patients at rest and during sympathetic manoeuvres**. *Diabetologia* 42:1345-1349.

Lefrandt JD, Bosma E, Oomen PHN, Hoeven JH, Roon AM, Smit AJ & Hoogenberg K (2003) **Sympathetic mediated vasomotion and skin capillary permeability in diabetic patients with peripheral neuropathy**. *Diabetologia* 46:40-47.

Legato MJ, Gelzer A, Goland R, Ebner SA, Rajan S, Villagra V, Kosowski M (2006) **Gender-specific care of the patient with diabetes: review and recommendations**. *Gend Med* 3:131-158.

Lehto S, Rönnemaa T, Pyörälä K & Laakso M (2000) **Cardiovascular risk factors clustering with endogenous hyperinsulinaemia predict death from coronary heart disease in patients with type II diabetes**. *Diabetologia* 43:148-155.

Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G (2003) **National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification**. *Ann Intern Med* 139:137-147.

Lewis GF (2006) **Determinants of plasma HDL concentrations and reverse cholesterol transport**. *Curr Opin Cardiol* 21:345-352.

Lewko B, Bryl E, Witkowski JM, Latawiec E, Angielski S & Stepinski J (2005) **Mechanical stress and glucose concentration modulate glucose transport in cultured rat podocytes**. *Nephrol Dial Transplant* 20:306-311.

Li C, Ford ES, Strine TW & Mokdad AH (2008) **Prevalence of depression among U.S. adults with diabetes: findings from the 2006 behavioral risk factor surveillance system**. *Diabetes Care* 31:105-107.

Linden J & Nicholson J (1997) **Recent advances in high-resolution NMR spectroscopic methods in bioanalytical chemistry**. *Trends Anal Chem* 16:190-200.

Linden E, Cai W, He JC, Xue C, Li Z, Winston J, Vlassara H & Uribarri J (2008) **Endothelial dysfunction in patients with chronic kidney disease results from advanced glycation end products (AGE)-mediated inhibition of endothelial nitric oxide synthase through RAGE activation**. *Clin J Am Soc Nephrol* 3:691-698.

Lindström J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemiö K, Hämäläinen H, Härkönen P, Keinänen-Kiukaanniemi S, Laakso M, Louheranta A, Mannelin M, Paturi M, Sundvall J, Valle TT, Uusitupa M, Tuomilehto J (2006) **Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study**. *Lancet* 368:1673-1679.

LipoScience (2009) LipoScience Inc. Raleigh US, www.lipoprofile.com.

Liu Y (2007) **Dimensionality Reduction for Mass Spectrometry Data**. Advanced Data Mining and Applications in Lecture Notes in Computer Science, Springer 203-213.

Lockhart CJ, Hamilton PK, McVeigh KA & McVeigh GE (2008) **A cardiologist view of vascular disease in diabetes**. *Diabetes Obes Metab* 10:279-292.

Lopes-Virella MF, McHenry MB, Lipsitz S, Yim E, Wilson PF, Lackland DT, Lyons T, Jenkins AJ, Virella G (2007) **Immune complexes containing modified lipoproteins are related to the progression of internal carotid intima-media thickness in patients with type 1 diabetes**. *Atherosclerosis* 190:359-369.

Lorenzo C, Serrano-Ríos M, Martínez-Larrad MT, Gabriel R, Williams K, Gómez-Gerique JA, Stern MP & Haffner SM (2003) **Central adiposity determines prevalence differences of the metabolic syndrome**. *Obes Res* 11:1480-1487.

Loscalzo J, Kohane I & Barabasi A (2007a) **Human disease classification in the postgenomic era: a complex systems approach to**

human pathobiology. *Mol Syst Biol* 3:124.

Loscalzo J (2007b) **Association studies in an era of too much information: clinical analysis of new biomarker and genetic data**. *Circulation* 116:1866-1870.

Lucas P (2004) **Bayesian analysis, pattern analysis, and data mining in health care**. *Curr Opin Crit Care* 10:399-403.

Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, Wang S, Drake TA, Lusis AJ & Schadt EE (2006) **Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes**. *J Neurochem* 97:S50-S62.

Lyons TJ, Jenkins AJ, Zheng D, Klein RL, Otvos JD, Yu Y, Lackland DT, McGee D, McHenry MB, Lopes-Virella M, Garvey WT (2006) **Nuclear magnetic resonance-determined lipoprotein subclass profile in the DCCT/EDIC cohort: associations with carotid intima-media thickness**. *Diabet Med* 23:955-966.

Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P & Groop L (2008) **Clinical risk factors, DNA variants, and the development of type 2 diabetes**. *N Engl J Med* 359:2220-2232.

Lyssenko V, Nagorny CLF, Erdos MR, Wierup N, Jonsson A, Spégel P, Bugliani M, Saxena R, Fex M, Pulizzi N, Isomaa B, Tuomi T, Nilsson P, Kuusisto J, Tuomilehto J, Boehnke M, Altshuler D, Sundler F, Eriksson JG, Jackson AU, Laakso M, Marchetti P, Watanabe RM, Mulder H & Groop L (2009) **Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion**. *Nat Genet* 41:82-88.

Macor C, Ruggeri A, Mazzonetto P, Federspil G, Cobelli C & Vettor R (1997) **Visceral adipose tissue impairs insulin secretion and insulin sensitivity but not energy expenditure in obesity**. *Metabolism* 46:123-129.

Magee GM, Bilous RW, Cardwell CR, Hunter SJ, Kee F & Fogarty DG (2009) **Is hyperfiltration associated with the future risk of developing diabetic nephropathy? A meta-analysis**. *Diabetologia* 52:691-697.

Maguire AM, Craig ME, Craighead A, Chan AKF, Cusumano JM, Hing SJ, Silink M, Howard NJ & Donaghue KC (2007) **Autonomic nerve testing predicts the development of complications: a 12-year follow-up study**. *Diabetes Care* 30:77-82.

Mäkinen VP, Parkkonen M, Wessman M, Groop P, Kanninen T & Kaski K (2005) **High-throughput pedigree drawing**. *Eur J Hum Genet* 13:987-989.

Malone JI, Morrison AD, Pavan PR, Cuthbertson DD (2001) **Prevalence and significance of retinopathy in subjects with type 1 diabetes of less than 5 years' duration screened for the Diabetes Control and Complications Trial**. *Diabetes Care* 24:522-526.

Malthouse E, Tamhane A & Mah R (1997) **Nonlinear partial least squares**. *Computers Chem Eng* 21:875-890.

Maric C & Sullivan S (2008) **Estrogens and the diabetic kidney**. *Gend Med* 5:S103-S113.

Marques RG, Fontaine MJ & Rogers J (2004) **C-peptide: much more than a byproduct of insulin biosynthesis**. *Pancreas* 29:231-238.

Marra G, Cotroneo P, Pitocco D, Manto A, Di Leo MAS, Ruotolo V, Caputo S, Giardina B, Ghirlanda G & Santini SA (2002) **Early increase of oxidative stress and reduced antioxidant defenses in patients with uncomplicated type 1 diabetes: a case for gender difference**. *Diabetes Care* 25:370-375.

Marshall SM & Flyvbjerg A (2006) **Prevention and early detection of vascular complications of diabetes**. *BMJ* 333:475-480.

Martin CL, Albers J, Herman WH, Cleary P, Waberski B, Greene DA, Stevens MJ, Feldman EL (2006) **Neuropathy among the diabetes control and complications trial cohort 8 years after trial completion**. *Diabetes Care* 29:340-344.

Martínez AM & Kak AC (2001) **PCA versus LDA**. *IEEE Transactions on Pattern Analysis*

and Machine Intelligence 23:228-233.

Masley S, Roetzheim R & Gualtieri T (2009) **Aerobic Exercise Enhances Cognitive Flexibility**. *J Clin Psychol Med Settings* 16:186-193.

Mason NJ, Jenkins AJ, Best JD & Rowley KG (2006) **Exercise frequency and arterial compliance in non-diabetic and type 1 diabetic individuals**. *Eur J Cardiovasc Prev Rehabil* 13:598-603.

Mathieu P, Pibarot P, Larose E, Poirier P, Marette A & Després J (2008) **Visceral obesity and the heart**. *Int J Biochem Cell Biol* 40:821-836.

Mathieu P, Poirier P, Pibarot P, Lemieux I & Després J (2009) **Visceral obesity: the link among inflammation, hypertension, and cardiovascular disease**. *Hypertension* 53:577-584.

May O, Arildsen H, Damsgaard EM & Mickley H (2000) **Cardiovascular autonomic neuropathy in insulin-dependent diabetes mellitus: prevalence and estimated risk of coronary heart disease in the general population**. *J Intern Med* 248:483-491.

McKellar JD, Humphreys K & Piette JD (2004) **Depression increases diabetes symptoms by complicating patients' self-care adherence**. *Diabetes Educ* 30:485-492.

McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, Steyn K, Sanderson JE, Hasani M, Volkova E, Kazmi K, Yusuf S (2008) **Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study**. *Lancet* 372:224-233.

Mercer N, Ahmed H, Etcheverry SB, Vasta GR & Cortizo AM (2007) **Regulation of advanced glycation end product (AGE) receptors and apoptosis by AGEs in osteoblast-like cells**. *Mol Cell Biochem* 306:87-94.

Mielke S & Khrisnan V (2005) **Estimation of protein secondary structure content directly from NMR spectra using an improved empirical correlation with averaged chemical shift**. *J Struct Funct Genomics* 6:281–285.

Miljkovic-Gacic I, Bunker CH, Ferrell RE, Kammerer CM, Evans RW, Patrick AL & Kuller LH (2006) **Lipoprotein subclass and particle size differences in Afro-Caribbeans, African Americans, and white Americans: associations with hepatic lipase gene variation**. *Metabolism* 55:96-102.

Miller RG, Prince CT, Klein R & Orchard TJ (2009) **Retinal vessel diameter and the incidence of coronary artery disease in type 1 diabetes**. *Am J Ophthalmol* 147:653-660.

Milton J & Arnold J (1995) **Introduction to probability and statistics**. McGraw-Hill.

Misra A, Wasir JS & Vikram NK (2005) **Waist circumference criteria for the diagnosis of abdominal obesity are not applicable uniformly to all populations and ethnic groups**. *Nutrition* 21:969-976.

Mogensen CE (1995) **Microalbuminuria in prediction and prevention of diabetic nephropathy in insulin-dependent diabetes mellitus patients**. *J Diabetes Complications* 9:337-349.

Mohamed Q, Gillies MC & Wong TY (2007) **Management of diabetic retinopathy: a systematic review**. *JAMA* 298:902-916.

Molarius A, Seidell JC, Sans S, Tuomilehto J & Kuulasmaa K (1999) **Varying sensitivity of waist action levels to identify subjects with overweight or obesity in 19 populations of the WHO MONICA Project**. *J Clin Epidemiol* 52:1213-1224.

Molitch ME (2006) **Management of dyslipidemias in patients with diabetes and chronic kidney disease**. *Clin J Am Soc Nephrol* 1:1090-1099.

Monami M, Marchionni N & Mannucci E (2009) **Long-acting insulin analogues vs. NPH human insulin in type 1 diabetes. A meta-analysis**. *Diabetes Obes Metab* 11:372-378.

Monti MC, Lonsdale JT, Montomoli C, Montross R, Schlag E & Greenberg DA (2007) **Familial risk factors for microvascular complications and differential male-female risk in a**

large cohort of American families with type 1 diabetes. *J Clin Endocrinol Metab* 92:4650-4655.

Moore G (1965) **Cramming more components onto integrated circuits**. *Electronics* 38(8).

Mora S, Szklo M, Otvos JD, Greenland P, Psaty BM, Goff DCJ, O'Leary DH, Saad MF, Tsai MY & Sharrett AR (2007) **LDL particle subclasses, LDL particle size, and carotid atherosclerosis in the Multi-Ethnic Study of Atherosclerosis (MESA)**. *Atherosclerosis* 192:211-217.

Mora S (2009) **Advanced lipoprotein testing and subfractionation are not (yet) ready for routine clinical use**. *Circulation* 119:2396-2404.

Morrish NJ, Wang SL, Stevens LK, Fuller JH & Keen H (2001) **Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes**. *Diabetologia* 44 Suppl 2:S14-21.

Mueller MJ, Zou D & Lott DJ (2005) **"Pressure gradient" as an indicator of plantar skin injury**. *Diabetes Care* 28:2908-2912.

Mueller PW, Rogus JJ, 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**. *J Am Soc Nephrol* 17:1782-1790.

Mukamal KJ, Jensen MK, Grønbaek M, Stampfer MJ, Manson JE, Pischon T & Rimm EB (2005) **Drinking frequency, mediating biomarkers, and risk of myocardial infarction in women and men**. *Circulation* 112:1406-1413.

Mussap M, Dalla Vestra M, Fioretto P, Saller A, Varagnolo M, Nosadini R & Plebani M (2002) **Cystatin C is a more sensitive marker than creatinine for the estimation of GFR in type 2 diabetic patients**. *Kidney Int* 61:1453-1461.

Nádas J, Putz Z, Fövényi J, Gaál Z, Gyimesi A, Hídvégi T, Hosszúfalusi N, Neuwirth G, Oroszlán T, Pánczél P, Vándorfi G, Winkler G, Wittmann I & Jermendy G (2009) **Cardiometabolic risk and educational level in adult patients with type 1 diabetes**. *Acta Diabetol* 46:159-162.

Najafian B, Crosson JT, Kim Y & Mauer M (2006) **Glomerulotubular junction abnormalities are associated with proteinuria in type 1 diabetes**. *J Am Soc Nephrol* 17:S53-60.

Natarajan R & Nadler JL (2004) **Lipid inflammatory mediators in diabetic vascular disease**. *Arterioscler Thromb Vasc Biol* 24:1542-1548.

NCEP (2002) **Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)**. *Circulation* 106:3143-3421.

Nejentsev S, Howson JMM, Walker NM, Szeszko J, Field SF, Stevens HE, Reynolds P, Hardy M, King E, Masters J, Hulme J, Maier LM, Smyth D, Bailey R, Cooper JD, Ribas G, Campbell RD, Clayton DG, Todd JA (2007) **Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A**. *Nature* 450:887-892.

Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, Hitman GA & Hawke C (2005) **Systematic review on urine albumin testing for early detection of diabetic complications**. *Health Technol Assess* 9:iii-vi, xiii-163.

Nicholson J, O'Flynn M & Sadler P (1984) **Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects**. *Biochem J* 217:365-375.

Nicholson JK, Connelly J, Lindon JC & Holmes E (2002) **Metabonomics: a platform for studying drug toxicity and gene function**. *Nat Rev Drug Discov* 1:153-161.

Nishikawa T, Edelstein D, Du XL, Yamagishi S, Matsumura T, Kaneda Y, Yorek MA, Beebe D, Oates PJ, Hammes H, Giardino I & Brownlee M (2000) **Normalizing mitochondrial superoxide**

production blocks three pathways of hyperglycaemic damage. *Nature* 404:787-790.

Nordwall M, Bojestig M, Arnqvist HJ, Ludvigsson J (2004) **Declining incidence of severe retinopathy and persisting decrease of nephropathy in an unselected population of Type 1 diabetes-the Linköping Diabetes Complications Study**. *Diabetologia* 47:1266-1272.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M & Bustamante CD (2008) **Genes mirror geography within Europe**. *Nature* 456:98-101.

O'Donnell MP, Kasiske BL & Keane WF (1988) **Glomerular hemodynamic and structural alterations in experimental diabetes mellitus**. *FASEB J* 2:2339-2347.

O'Keefe JH, Bybee KA & Lavie CJ (2007) **Alcohol and cardiovascular health: the razor-sharp double-edged sword**. *J Am Coll Cardiol* 50:1009-1014.

Olson RE (1998) **Discovery of the lipoproteins, their role in fat transport and their significance as risk factors**. *J Nutr* 128:439S-443S.

Omenn GS (2006) **Strategies for plasma proteomic profiling of cancers**. *Proteomics* 6:5662-5673.

Onnela J, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J & Barabási A (2007) **Structure and tie strengths in mobile communication networks**. *Proc Natl Acad Sci USA* 104:7332-7336.

Opstad KS, Ladroue C, Bell BA, Griffiths JR & Howe FA (2007) **Linear discriminant analysis of brain tumour (1)H MR spectra: a comparison of classification using whole spectra versus metabolite quantification**. *NMR Biomed* 20:763-770.

Orchard TJ, Costacou T, Kretowski A & Nesto RW (2006) **Type 1 diabetes and coronary artery disease**. *Diabetes Care* 29:2528-2538.

Ordovas JM (2007) **Gender, a significant factor in the cross talk between genes, environment, and health**. *Gend Med* 4 Suppl B:S111-22.

Oresic M, Simell S, Sysi-Aho M, Näntö-Salonen K, Seppänen-Laakso T, Parikka V, Katajamaa M, Hekkala A, Mattila I, Keskinen P, Yetukuri L, Reinikainen A, Lähde J, Suortti T, Hakalax J, Simell T, Hyöty H, Veijola R, Ilonen J, Lahesmaa R, Knip M & Simell O (2008) **Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes**. *J Exp Med* 205:2975-2984.

Österby R (1995) **Microalbuminuria in diabetes mellitus – is there a structural basis?** *Nephrol Dial Transplant* 10:12-14.

Østergaard J, Hansen TK, Thiel S & Flyvbjerg A (2005) **Complement activation and diabetic vascular complications**. *Clin Chim Acta* 361:10-19.

Österholm A, He B, Pitkäniemi J, Albinsson L, Berg T, Sarti C, Tuomilehto J & Tryggvason K (2007) **Genome-wide scan for type 1 diabetic nephropathy in the Finnish population reveals suggestive linkage to a single locus on chromosome 3q**. *Kidney Int* 71:140-145.

Otvos J, Jeyarajah E & Bennett D (1991) **Quantification of plasma lipoproteins by proton nuclear magnetic resonance spectroscopy**. *Clinical Chemistry* 37:377-386.

Ozdemir AM, Hopfer U, Erhard P, Monnier VM & Weiss MF (2005) **Processing advanced glycation end product-modified albumin by the renal proximal tubule and the early pathogenesis of diabetic nephropathy**. *Ann NY Acad Sci* 1043:625-636.

Pajunen P, Taskinen MR, Nieminen MS & Syvänne M (2000) **Angiographic severity and extent of coronary artery disease in patients with type 1 diabetes mellitus**. *Am J Cardiol* 86:1080-1085.

Pambianco G, Costacou T & Orchard TJ (2007) **The prediction of major outcomes of type 1 diabetes: a 12-year prospective evaluation of three separate definitions of the metabolic syndrome and their components and esti-**

mated glucose disposal rate: the Pittsburgh Epidemiology of Diabetes Complications Study experience. *Diabetes Care* 30:1248-1254.

Paraoan L & Grierson I (2007) **Focus on molecules: cystatin C**. *Exp Eye Res* 84:1019-1020.

Parhofer K & Barrett P (2006) **What we have learned about VLDL and LDL metabolism from human kinetics studies**. *J Lipid Res* 47:1620-1630.

Park J & Edington DW (2001) **A sequential neural network model for diabetes prediction**. *Artif Intell Med* 23:277-293.

Parving HH, Tarnow L & Rossing P (1996) **Genetics of diabetic nephropathy**. *J Am Soc Nephrol* 7:2509-2517.

Parving HH, Hommel E, Jensen BR & Hansen HP (2001) **Long-term beneficial effect of ACE inhibition on diabetic nephropathy in normotensive type 1 diabetic patients**. *Kidney Int* 60:228-234.

Pätäri A, Forsblom C, Havana M, Taipale H, Groop P & Holthöfer H (2003) **Nephrinuria in diabetic nephropathy of type 1 diabetes**. *Diabetes* 52:2969-2974.

Patrakka J, Kestilä M, Wartiovaara J, Ruotsalainen V, Tissari P, Lenkkeri U, Männikkö M, Visapää I, Holmberg C, Rapola J, Tryggvason K & Jalanko H (2000) **Congenital nephrotic syndrome (NPHS1): features resulting from different mutations in Finnish patients**. *Kidney Int* 58:972-980.

Pawson T & Linding R (2008) **Network medicine**. *FEBS Lett* 582:1266-1270.

Pearson K (1901) **On lines and planes of closest fit to systems of points in space**. *Philosoph Magazine* 2:559-572.

Perrin NESS, Torbjörnsdotter TB, Jaremko GA & Berg UB (2006) **The course of diabetic glomerulopathy in patients with type I diabetes: a 6-year follow-up with serial biopsies**. *Kidney Int* 69:699-705.

Petermann AT, Hiromura K, Blonski M, Pippin J, Monkawa T, Durvasula R, Couser WG & Shankland SJ (2002) **Mechanical stress reduces podocyte proliferation in vitro**. *Kidney Int* 61:40-50.

Petersen M, Dyrby M, Toubro S, Engelsen S, Nørgaard L, Pedersen H & Dyerberg J (2005) **Quantification of lipoprotein subclasses by proton nuclear magnetic resonance–based partial least-Squares regression models**. *Clin Chem* 51:1457–1461.

Petrovsky N & Silva D (2004) **The power of an integrated informatic and molecular approach to type 1 diabetes research**. *Ann NY Acad Sci* 1037:216-224.

Pettersson-Fernholm K, Fröjdö S, Fagerudd J, Thomas MC, Forsblom C, Wessman M, Groop P (2006) **The AT2 gene may have a gender-specific effect on kidney function and pulse pressure in type I diabetic patients**. *Kidney Int* 69:1880-1884.

Phillips C, Murugasu G, Owens D, Collins P, Johnson A & Tomkin GH (2000) **Improved metabolic control reduces the number of postprandial apolipoprotein B-48-containing particles in type 2 diabetes**. *Atherosclerosis* 148:283-291.

Pieroni E, de la Fuente van Bentem S, Mancosu G, Capobianco E, Hirt H & de la Fuente A (2008) **Protein networking: insights into global functional organization of proteomes**. *Proteomics* 8:799-816.

Pitkäniemi J, Moltchanova E, Haapala L, Harjutsalo V, Tuomilehto J & Hakulinen T (2007) **Genetic random effects model for family data with long-term survivors: analysis of diabetic nephropathy in type 1 diabetes**. *Genet Epidemiol* 31:697-708.

Pittenger GL, Liu D & Vinik AI (1997) **The apoptotic death of neuroblastoma cells caused by serum from patients with insulin-dependent diabetes and neuropathy may be Fas-mediated**. *J Neuroimmunol* 76:153-160.

Prigent A (2008) **Monitoring renal function and limitations of renal function tests**. *Semin Nucl Med* 38:32-46.

Proctor SD, Vine DF & Mamo JCL (2004) **Arterial permeability and efflux of apolipoprotein B-containing lipoproteins assessed by in situ perfusion and three-dimensional quantitative confocal microscopy**. *Arterioscler Thromb Vasc Biol* 24:2162-2167.

Pundziute G, Schuijf JD, Jukema JW, Boersma E, Scholte AJHA, Kroft LJM, van der Wall EE & Bax JJ (2007) **Noninvasive assessment of plaque characteristics with multislice computed tomography coronary angiography in symptomatic diabetic patients**. *Diabetes Care* 30:1113-1119.

Raile K, Galler A, Hofer S, Herbst A, Dunstheimer D, Busch P & Holl RW (2007) **Diabetic nephropathy in 27,805 children, adolescents, and adults with type 1 diabetes: effect of diabetes duration, A1C, hypertension, dyslipidemia, diabetes onset, and sex**. *Diabetes Care* 30:2523-2528.

Rebsomen L, Khammar A, Raccah D & Tsimaratos M (2008) **C-peptide effects on renal physiology and diabetes**. *Exp Diabetes Res* 2008:281536.

Reckelhoff JF, Yanes LL, Iliescu R, Fortepiani LA & Granger JP (2005) **Testosterone supplementation in aging men and women: possible impact on cardiovascular-renal disease**. *Am J Physiol Renal Physiol* 289:F941-8.

Rehm J, Room R, Graham K, Monteiro M, Gmel G & Sempos CT (2003) **The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview**. *Addiction* 98:1209-1228.

Retnakaran R & Zinman B (2008) **Type 1 diabetes, hyperglycaemia, and the heart**. *Lancet* 371:1790-1799.

Reunanen A (2006) **Diabetes yleistyy Suomessa entistä kiivaammin, mutta diabeetikoiden ennuste paranee ja lisäsairauksien ilmaantuminen vähenee**. *Diabetes ja lääkäri* 35:7-10 (in Finnish).

Rimm EB, Manson JE, Stampfer MJ, Colditz GA, Willett WC, Rosner B, Hennekens CH & Speizer FE (1993) **Cigarette smoking and the risk of diabetes in women**. *Am J Public Health* 83:211-214.

Rimm EB, Klatsky A, Grobbee D & Stampfer MJ (1996) **Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits**. *BMJ* 312:731-736.

Riser BL, Ladson-Wofford S, Sharba A, Cortes P, Drake K, Guerin CJ, Yee J, Choi ME, Segarini PR & Narins RG (1999) **TGF-beta receptor expression and binding in rat mesangial cells: modulation by glucose and cyclic mechanical strain**. *Kidney Int* 56:428-439.

Robertson MD, Parkes M, Warren BF, Ferguson DJP, Jackson KG, Jewell DP & Frayn KN (2003) **Mobilisation of enterocyte fat stores by oral glucose in humans**. *Gut* 52:834-839.

Roglic G, Colhoun HM, Stevens LK, Lemkes HH, Manes C & Fuller JH (1998) **Parental history of hypertension and parental history of diabetes and microvascular complications in insulin-dependent diabetes mellitus: the EURODIAB IDDM Complications Study**. *Diabet Med* 15:418-426.

Rönnback M, Fagerudd J, Forsblom C, Pettersson-Fernholm K, Reunanen A, Groop P (2004) **Altered age-related blood pressure pattern in type 1 diabetes**. *Circulation* 110:1076-1082.

Roos JF, Doust J, Tett SE & Kirkpatrick CMJ (2007) **Diagnostic accuracy of cystatin C compared to serum creatinine for the estimation of renal dysfunction in adults and children – a meta-analysis**. *Clin Biochem* 40:383-391.

Rose M, Fliege H, Hildebrandt M, Schirop T & Klapp BF (2002) **The network of psychological variables in patients with diabetes and their importance for quality of life and metabolic control**. *Diabetes Care* 25:35-42.

Rosipal R & Krämer N (2006) **Overview and recent advances in partial least squares**. Subspace, Latent Structure and Feature Selection Techniques in Lecture Notes in Computer Science, Springer 34-51.

Rosolowsky ET, Niewczas MA, Ficociello LH, Perkins BA, Warram JH & Krolewski AS

(2008) **Between hyperfiltration and impairment: demystifying early renal functional changes in diabetic nephropathy**. *Diabetes Res Clin Pract* 82:S46-S53.

Rossing P (2005) **The changing epidemiology of diabetic microangiopathy in type 1 diabetes**. *Diabetologia* 48:1439-1444.

Ruggeri ZM (2002) **Platelets in atherothrombosis**. *Nat Med* 8:1227-1234.

Ruotsalainen V, Ljungberg P, Wartiovaara J, Lenkkeri U, Kestilä M, Jalanko H, Holmberg C & Tryggvason K (1999) **Nephrin is specifically located at the slit diaphragm of glomerular podocytes**. *Proc Natl Acad Sci USA* 96:7962-7967.

Safar ME, London GM & Plante GE (2004) **Arterial stiffness and kidney function**. *Hypertension* 43:163-168.

Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, Nunez D, Sweatman BC, Haselden JN, Cox RD, Connor SC & Griffin JL (2007) **A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human**. *Physiol Genomics* 29:99-108.

Samnegård B, Jacobson SH, Jaremko G, Johansson B, Ekberg K, Isaksson B, Eriksson L, Wahren J & Sjöquist M (2005) **C-peptide prevents glomerular hypertrophy and mesangial matrix expansion in diabetic rats**. *Nephrol Dial Transplant* 20:532-538.

Saraheimo M, Teppo A, Forsblom C, Fagerudd J & Groop P (2003) **Diabetic nephropathy is associated with low-grade inflammation in Type 1 diabetic patients**. *Diabetologia* 46:1402-1407.

Saraheimo M, Forsblom C, Hansen TK, Teppo A, Fagerudd J, Pettersson-Fernholm K, Thiel S, Tarnow L, Ebeling P, Flyvbjerg A, Groop P (2005a) **Increased levels of mannan-binding lectin in type 1 diabetic patients with incipient and overt nephropathy**. *Diabetologia* 48:198-202.

Saraheimo M, Forsblom C, Fagerudd J, Teppo A, Pettersson-Fernholm K, Frystyk J, Flyvbjerg A, Groop P (2005b) **Serum adiponectin is increased in type 1 diabetic patients with nephropathy**. *Diabetes Care* 28:1410-1414.

Sattar N, McConnachie A, Shaper AG, Blauw GJ, Buckley BM, de Craen AJ, Ford I, Forouhi NG, Freeman DJ, Jukema JW, Lennon L, Macfarlane PW, Murphy MB, Packard CJ, Stott DJ, Westendorp RG, Whincup PH, Shepherd J & Wannamethee SG (2008) **Can metabolic syndrome usefully predict cardiovascular disease and diabetes? Outcome data from two prospective studies**. *Lancet* 371:1927-1935.

Savolainen MJ, Rantala M, Kervinen K, Järvi L, Suvanto K, Rantala T & Kesäniemi YA (1991a) **Magnitude of dietary effects on plasma cholesterol concentration: role of sex and apolipoprotein E phenotype**. *Atherosclerosis* 86:145-152.

Savolainen MJ, Korhonen T, Aalto-Setälä K, Kontula K & Kesäniemi YA (1991b) **Screening for a prevalent LDL receptor mutation in patients with severe hypercholesterolaemia**. *Hum Genet* 87:125-128.

Schiffrin E (2004) **Vascular Stiffening and Arterial Compliance: Implications for Systolic Blood Pressure**. *Am J Hypertens* 17:39S–48S.

Schram MT, Schalkwijk CG, Bootsma AH, Fuller JH, Chaturvedi N, Stehouwer CDA (2005) **Advanced glycation end products are associated with pulse pressure in type 1 diabetes: the EURODIAB Prospective Complications Study**. *Hypertension* 46:232-237.

Schramm TK, Gislason GH, Køber L, Rasmussen S, Rasmussen JN, Abildstrøm SZ, Hansen ML, Folke F, Buch P, Madsen M, Vaag A & Torp-Pedersen C (2008) **Diabetes patients requiring glucose-lowering therapy and nondiabetics with a prior myocardial infarction carry the same cardiovascular risk: a population study of 3.3 million people**. *Circulation* 117:1945-1954.

Schulz TJ, Zarse K, Voigt A, Urban N, Birringer M & Ristow M (2007) **Glucose restriction extends Caenorhabditis elegans life span by inducing mitochondrial respiration and increasing oxidative stress**. *Cell Metab* 6:280-293.

Scott LJ, Warram JH, Hanna LS, Laffel LM, Ryan L & Krolewski AS (2001) **A nonlinear effect of hyperglycemia and current cigarette smoking are major determinants of the onset of microalbuminuria in type 1 diabetes**. *Diabetes* 50:2842-2849.

Scott AR, Dhindsa P, Forsyth J, Mansell P (2004) **Effect of hormone replacement therapy on cardiovascular risk factors in postmenopausal women with diabetes**. *Diabetes Obes Metab* 6:16-22.

Scragg R & Metcalf P (2004) **Do triglycerides explain the U-shaped relation between alcohol and diabetes risk? Results from a cross-sectional survey of alcohol and plasma glucose**. *Diabetes Res Clin Pract* 66:147-156.

Seaquist ER, Goetz FC, Rich S & Barbosa J (1989) **Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy**. *N Engl J Med* 320:1161-1165.

Seissler J & Scherbaum WA (2006) **Autoimmune diagnostics in diabetes mellitus**. *Clin Chem Lab Med* 44:133-137.

Selhub J, Jacques PF, Bostom AG, D'Agostino RB, Wilson PW, Belanger AJ, O'Leary DH, Wolf PA, Schaefer EJ & Rosenberg IH (1995) **Association between plasma homocysteine concentrations and extracranial carotid-artery stenosis**. *N Engl J Med* 332:286-291.

Serkova N, Fuller TF, Klawitter J, Freise CE & Niemann CU (2005) **^{1}H-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants**. *Kidney Int* 67:1142-1151.

Seshiah PN, Weber DS, Rocic P, Valppu L, Taniyama Y & Griendling KK (2002) **Angiotensin II stimulation of NAD(P)H oxidase activity: upstream mediators**. *Circ Res* 91:406-413.

Shahid S & Tabassum Mahboob T (2007) **Cigarette smoking: an environmental risk factor for progression of nephropathy in diabetes**. *Int J Diab Dev Ctries* 27:104-107.

Sharrett AR, Ding J, Criqui MH, Saad MF, Liu K, Polak JF, Folsom AR, Tsai MY, Burke GL & Szklo M (2006) **Smoking, diabetes, and blood cholesterol differ in their associations with subclinical atherosclerosis: the Multiethnic Study of Atherosclerosis (MESA)**. *Atherosclerosis* 186:441-447.

Shelness GS & Ledford AS (2005) **Evolution and mechanism of apolipoprotein B-containing lipoprotein assembly**. *Curr Opin Lipidol* 16:325-332.

Shlens J (2005) **A tutorial on principal component analysis**. University of California, US.

Shoji T, Nishizawa Y, Kawagishi T, Kawasaki K, Taniwaki H, Tabata T, Inoue T & Morii H (1998) **Intermediate-density lipoprotein as an independent risk factor for aortic atherosclerosis in hemodialysis patients**. *J Am Soc Nephrol* 9:1277-1284.

Sibley SD, Thomas W, de Boer I, Brunzell JD & Steffes MW (2006) **Gender and elevated albumin excretion in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) cohort: role of central obesity**. *Am J Kidney Dis* 47:223-232.

Siebenhofer A, Plank J, Berghold A, Horvath K, Sawicki PT, Beck P & Pieber TR (2004) **Meta-analysis of short-acting insulin analogues in adult patients with type 1 diabetes: continuous subcutaneous insulin infusion versus injection therapy**. *Diabetologia* 47:1895-1905.

Silbiger S & Neugarten J (2008) **Gender and human chronic renal disease**. *Gend Med* 5:S3-S10.

Sircar S (2008) **Principles of Medical Physiology**. Georg Thieme Verlag, New York.

Sittiwet C, Gylling H, Hallikainen M, Pihlajamäki J, Moilanen L, Laaksonen DE, Niskanen L, Agren JJ, Laakso M & Miettinen TA (2007) **Cholesterol metabolism and non-cholesterol sterol distribution in lipoproteins of type 1 diabetes: the effect of improved glycemic control**. *Atherosclerosis* 194:465-472.

Skrivarhaug T, Fosmark DS, Stene LC, Bangstad H, Sandvik L, Hanssen KF & Joner G

(2006) **Low cumulative incidence of proliferative retinopathy in childhood-onset type 1 diabetes: a 24-year follow-up study**. *Diabetologia* 49:2281-2290.

Small M, Walker DM & Tse CK (2007) **Scale-free distribution of avian influenza outbreaks**. *Phys Rev Lett* 99:188702.

Snell-Bergeon JK, Hokanson JE, Jensen L, MacKenzie T, Kinney G, Dabelea D, Eckel RH, Ehrlich J, Garg S & Rewers M (2003) **Progression of coronary artery calcification in type 1 diabetes: the importance of glycemic control**. *Diabetes Care* 26:2923-2928.

Sobolewski BA, Zgibor JC & Orchard TJ (2004) **ACE inhibitors and calcium channel blockers: patterns of use and associations with mortality in type 1 diabetes**. *Diabetes Res Clin Pract* 65:37-43.

Soedamah-Muthu SS, Chang Y, Otvos J, Evans RW, Orchard TJ (2003) **Lipoprotein subclass measurements by nuclear magnetic resonance spectroscopy improve the prediction of coronary artery disease in Type 1 diabetes. A prospective report from the Pittsburgh Epidemiology of Diabetes Complications Study**. *Diabetologia* 46:674-682.

Soedamah-Muthu SS, Chaturvedi N, Toeller M, Ferriss B, Reboldi P, Michel G, Manes C, Fuller JH (2004) **Risk factors for coronary heart disease in type 1 diabetic patients in Europe: the EURODIAB Prospective Complications Study**. *Diabetes Care* 27:530-537.

Soedamah-Muthu SS, Chaturvedi N, Teerlink T, Idzior-Walus B, Fuller JH, Stehouwer CDA (2005) **Plasma homocysteine and microvascular and macrovascular complications in type 1 diabetes: a cross-sectional nested case-control study**. *J Intern Med* 258:450-459.

Sogabe M, Okahisa T, Tsujigami K, Okita Y, Hayashi H, Taniki T, Hukuno H, Nakasono M, Muguruma N, Okamura S & Ito S (2005) **Ultrasonographic assessment of gastric motility in diabetic gastroparesis before and after attaining glycemic control**. *J Gastroenterol* 40:583-590.

Vesanto J, Himberg J, Alhoniemi E & Parhankangas J (2000) SOM Toolbox for Matlab 5. Helsinki University of Technology, Finland.

Stadler M, Auinger M, Anderwald C, Kästenbauer T, Kramar R, Feinböck C, Irsigler K, Kronenberg F & Prager R (2006) **Long-term mortality and incidence of renal dialysis and transplantation in type 1 diabetes mellitus**. *J Clin Endocrinol Metab* 91:3814-3820.

Stampfer MJ, Krauss RM, Ma J, Blanche PJ, Holl LG, Sacks FM & Hennekens CH (1996) **A prospective study of triglyceride level, low-density lipoprotein particle diameter, and risk of myocardial infarction**. *JAMA* 276:882-888.

Stanimirova I, Walczak B, Massart D & Simeonov V (2004) **A comparison between two robust PCA algorithms**. *Chemom Intell Lab Syst* 71:83 – 95.

Stern DM, Yan SD, Yan SF & Schmidt AM (2002) **Receptor for advanced glycation endproducts (RAGE) and the complications of diabetes**. *Ageing Res Rev* 1:1-15.

Stone ML, Craig ME, Chan AK, Lee JW, Verge CF & Donaghue KC (2006) **Natural history and risk factors for microalbuminuria in adolescents with type 1 diabetes: a longitudinal study**. *Diabetes Care* 29:2072-2077.

Sugden MC & Holness MJ (2003) **Recent advances in mechanisms regulating glucose oxidation at the level of the pyruvate dehydrogenase complex by PDKs**. *Am J Physiol Endocrinol Metab* 284:E855-62.

Suna T, Salminen A, Soininen P, Laatikainen R, Ingman P, Mäkelä S, Savolainen MJ, Hannuksela ML, Jauhiainen M, Taskinen M, Kaski K & Ala-Korpela M (2007) **^{1}H NMR metabonomics of plasma lipoprotein subclasses: elucidation of metabolic clustering by self-organising maps**. *NMR Biomed* 20:658-672.

Superko HR (2009) **Advanced lipoprotein testing and subfractionation are clinically useful**. *Circulation* 119:2383-2395.

Suzuma I, Hata Y, Clermont A, Pokras F, Rook SL, Suzuma K, Feener EP & Aiello LP (2001)

Cyclic stretch and hypertension induce retinal expression of vascular endothelial growth factor and vascular endothelial growth factor receptor-2: potential mechanisms for exacerbation of diabetic retinopathy by hypertension. *Diabetes* 50:444-454.

Tarnow L, Groop P, Hadjadj S, Kazeem G, Cambien F, Marre M, Forsblom C, Parving H, Trégouët D, Thévard A, Farrall M, Gut I, Gauguier D, Cox R, Matsuda F, Lathrop M, Vionnet N (2008) **European rational approach for the genetics of diabetic complications – EURAGEDIC: patient populations and strategy**. *Nephrol Dial Transplant* 23:161-168.

Taskinen M (2003) **Diabetic dyslipidaemia: from basic research to clinical practice**. *Diabetologia* 46:733-749.

Teede HJ (2007) **Sex hormones and the cardiovascular system: effects on arterial function in women**. *Clin Exp Pharmacol Physiol* 34:672-676.

Teo KK, Ounpuu S, Hawken S, Pandey MR, Valentin V, Hunt D, Diaz R, Rashed W, Freeman R, Jiang L, Zhang X, Yusuf S (2006) **Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study**. *Lancet* 368:647-658.

Theodoridis S (2003) Pattern Recognition, Academic Press Elsevier.

Thiel S, Møller-Kristensen M, Jensen L & Jensenius JC (2002) **Assays for the functional activity of the mannan-binding lectin pathway of complement activation**. *Immunobiology* 205:446-454.

Thomas MC, Forbes JM & Cooper ME (2005) **Advanced glycation end products and diabetic nephropathy**. *Am J Ther* 12:562-572.

Thomas MC, Rosengård-Bärlund M, Mills V, Rönnback M, Thomas S, Forsblom C, Cooper ME, Taskinen M, Viberti G & Groop P (2006a) **Serum lipids and the progression of nephropathy in type 1 diabetes**. *Diabetes Care* 29:317-322.

Thomas MC & Atkins RC (2006b) **Blood pressure lowering for the prevention and treatment of diabetic kidney disease**. *Drugs* 66:2213-2234.

Thorn LM, Forsblom C, Fagerudd J, Thomas MC, Pettersson-Fernholm K, Saraheimo M, Wadén J, Rönnback M, Rosengård-Bärlund M, Björkesten CA, Taskinen M, Groop P (2005) **Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycemic control (the FinnDiane study)**. *Diabetes Care* 28:2019-2024.

Thorn LM, Forsblom C, Fagerudd J, Pettersson-Fernholm K, Kilpikari R, Groop P (2007) **Clustering of risk factors in parents of patients with type 1 diabetes and nephropathy**. *Diabetes Care* 30:1162-1167.

Thorn LM, Forsblom C, Wadén J, Saraheimo M, Tolonen N, Hietala K, Groop P (2009a) **The metabolic syndrome as a risk factor for cardiovascular disease, mortality, and progression of diabetic nephropathy in type 1 diabetes**. *Diabetes Care* 32:950-952.

Thorn LM, Forsblom C, Wadén J, Söderlund J, Rosengård-Bärlund M, Saraheimo M, Heikkilä O, Hietala K, Pettersson-Fernholm K, Ilonen J, Groop P (2009b) **Effect of parental type 2 diabetes on offspring with type 1 diabetes**. *Diabetes Care* 32:63-68.

Tidman M, Sjöström P & Jones I (2008) **A Comparison of GFR estimating formulae based upon s-cystatin C and s-creatinine and a combination of the two**. *Nephrol Dial Transplant* 23:154-160.

Tikhonov A (1963) **Resolution of ill-posed problems and the regularization method**. *Dokl Akad Nauk SSSR* 151:501–504 (in Russian).

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JHM, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AAC, Ovington NR, Allen J, Adlem E, Leung H, Wallace C, Howson JMM, Guja C, Ionescu-Tîrgovişte C, , Simmonds MJ, Heward JM, Gough SCL, , Dunger DB, Wicker LS & Clayton DG (2007) **Robust associations**

of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39:857-864.

Tolonen N, Forsblom C, Thorn L, Wadén J, Rosengård-Bärlund M, Saraheimo M, Heikkilä O, Pettersson-Fernholm K, Taskinen M, Groop P (2008) **Relationship between lipid profiles and kidney function in patients with type 1 diabetes**. *Diabetologia* 51:12-20.

Tomaszewski M, Charchar FJ, Maric C, Kuzniewicz R, Gola M, Grzeszczak W, Samani NJ & Zukowska-Szczechowska E (2009a) **Association between lipid profile and circulating concentrations of estrogens in young men**. *Atherosclerosis* 203:257-262.

Tomaszewski M, Charchar FJ, Maric C, Kuzniewicz R, Gola M, Grzeszczak W, Samani NJ & Zukowska-Szczechowska E (2009b) **Inverse associations between androgens and renal function: the Young Men Cardiovascular Association (YMCA) study**. *Am J Hypertens* 22:100-105.

Tomlinson JW, Owen KR & Close CF (2003) **Treating hypertension in diabetic nephropathy**. *Diabetes Care* 26:1802-1805.

Trygg J (2002) **O2-PLS for qualitative and quantitative analysis in multivariate calibration**. *J Chemom* 16:283 - 293.

Trygg J, Holmes E & Lundstedt T (2007) **Chemometrics in metabonomics**. *J Proteome Res* 6:469-479.

Tukiainen T, Tynkkynen T, Mäkinen V, Jylänki P, Kangas A, Hokkanen J, Vehtari A, Gröhn O, Hallikainen M, Soininen H, Kivipelto M, Groop P, Kaski K, Laatikainen R, Soininen P, Pirttilä T & Ala-Korpela M (2008) **A multi-metabolite analysis of serum by ^{1}H NMR spectroscopy: early systemic signs of Alzheimer's disease**. *Biochem Biophys Res Commun* 375:356-361.

Twitchell EW (1907) **The prognosis of diabetes**. *Cal State J Med* 5:264-265.

Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S & Boyce J (2004) **Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening**. *Diabet Med* 21:84-90.

Vaarala O (2006) **Is it dietary insulin?** *Ann NY Acad Sci* 1079:350-359.

Vaidya D, Szklo M, Liu K, Schreiner PJ, Bertoni AG & Ouyang P (2007) **Defining the metabolic syndrome construct: Multi-Ethnic Study of Atherosclerosis (MESA) cross-sectional analysis**. *Diabetes Care* 30:2086-2090.

Valkonen V, Kolehmainen M, Lakka H & Salonen JT (2002) **Insulin resistance syndrome revisited: application of self-organizing maps**. *Int J Epidemiol* 31:864-871.

Vallon V, Blantz RC & Thomson S (2003) **Glomerular hyperfiltration and the salt paradox in early [corrected] type 1 diabetes mellitus: a tubulo-centric view**. *J Am Soc Nephrol* 14:530-537.

Van Harten B, de Leeuw F, Weinstein HC, Scheltens P & Biessels GJ (2006) **Brain imaging in patients with diabetes: a systematic review**. *Diabetes Care* 29:2539-2548.

Vanhatalo J & Vehtari A (2006) MCMC Methods for MLP-network and Gaussian process and stuff: a documentation for Matlab Toolbox MCMCstuff. Helsinki University of Technology, Finland.

Veglio M, Borra M, Stevens LK, Fuller JH & Perin PC (1999) **The relation between QTc interval prolongation and diabetic complications. The EURODIAB IDDM Complication Study Group**. *Diabetologia* 42:68-75.

Vehtari A, Mäkinen VP, Soininen P, Ingman P, Mäkelä SM, Savolainen MJ, Hannuksela ML, Kaski K & Ala-Korpela M (2007) **A novel Bayesian approach to quantify clinical variables and to determine their spectroscopic counterparts in ^{1}H NMR metabonomic data**. *BMC Bioinformatics* 8:S8.

Venna J & Kaski S (2001) Neighborhood preservations in non-linear projection methods: An experimental study. Helsinki University of Technology, Finland.

Vérier-Mine O, Chaturvedi N, Webb D & Fuller JH (2005) **Is pregnancy a risk factor for mi-**

crovascular complications? The EURODIAB Prospective Complications Study. *Diabet Med* 22:1503-1509.

Veselkov KA, Lindon JC, Ebbels TMD, Crockford D, Volynkin VV, Holmes E, Davies DB & Nicholson JK (2009) **Recursive segment-wise peak alignment of biological ^{1}H NMR spectra for improved metabolic biomarker recovery**. *Anal Chem* 81:56-66.

Vigneau E, Devaux M, Qannari E & Robert P (1997) **Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration**. *J Chemom* 11:239–249.

Villar E, Chang SH & McDonald SP (2007) **Incidences, treatments, outcomes, and sex effect on survival in patients with end-stage renal disease by diabetes status in Australia and New Zealand (1991-2005)**. *Diabetes Care* 30:3070-3076.

Vincent AM, Brownlee M & Russell JW (2002) **Oxidative stress and programmed cell death in diabetic neuropathy**. *Ann NY Acad Sci* 959:368-383.

Vionnet N, Tregouët D, Kazeem G, Gut I, Groop P, Tarnow L, Parving H, Hadjadj S, Forsblom C, Farrall M, Gauguier D, Cox R, Matsuda F, Heath S, Thévard A, Rousseau R, Cambien F, Marre M & Lathrop M (2006) **Analysis of 14 candidate genes for diabetic nephropathy on chromosome 3q in European populations: strongest evidence for association with a variant in the promoter region of the adiponectin gene**. *Diabetes* 55:3166-3174.

Virtanen SM & Knip M (2003) **Nutritional risk predictors of beta cell autoimmunity and type 1 diabetes at a young age**. *Am J Clin Nutr* 78:1053-1067.

Von Hertzen L, Forsblom C, Stumpf K, Pettersson-Fernholm K, Adlercreutz H, Groop P (2004) **Highly elevated serum phyto-oestrogen concentrations in patients with diabetic nephropathy**. *J Intern Med* 255:602-609.

Von Sengbusch S, Müller-Godeffroy E, Häger S, Reintjes R, Hiort O & Wagner V (2005) **Mobile diabetes education and care: intervention for children and young people with type 1 diabetes in rural areas of northern Germany**. *Diabet Med* 23:122–127.

Wadén J, Forsblom C, Thorn LM, Saraheimo M, Rosengård-Bärlund M, Heikkilä O, Lakka TA, Tikkanen H, Groop P (2008) **Physical activity and diabetes complications in patients with type 1 diabetes: the Finnish Diabetic Nephropathy (FinnDiane) Study**. *Diabetes Care* 31:230-232.

Walker H, Hall W & Hurst J (1990a) **Clinical methods: the history, physical, and laboratory examinations**. Butterworth Publishers, Stoneham MA.

Walker J (1990b) **A node-positioning algorithm for general trees**. *Software Pract Exp* 20:685 – 705.

Walldius G, Aastveit AH & Jungner I (2006) **Stroke mortality and the apoB/apoA-I ratio: results of the AMORIS prospective study**. *J Intern Med* 259:259-266.

Wang C, Kong H, Guan Y, Yang J, Gu J, Yang S & Xu G (2005) **Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis**. *Anal Chem* 77:4108-4116.

Ward J (1963) **Hierarchical grouping to optimize an objective function**. *JASA* 58:236-244.

Weinger K, Jacobson AM, Musen G, Lyoo IK, Ryan CM, Jimerson DC & Renshaw PF (2008) **The effects of type 1 diabetes on cerebral white matter**. *Diabetologia* 51:417-425.

Wessels AM, Scheltens P, Barkhof F & Heine RJ (2008) **Hyperglycaemia as a determinant of cognitive decline in patients with type 1 diabetes**. *Eur J Pharmacol* 585:88-96.

Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J & van Dorsten F (2008) **Assessment of PLSDA cross validation**. *Metabolomics* 4:81-89.

White KE, Bilous RW, Marshall SM, El Nahas M, Remuzzi G, Piras G, De Cosmo S & Viberti

G (2002) **Podocyte number in normotensive type 1 diabetic patients with albuminuria**. *Diabetes* 51:3083-3089.

Wider G & Dreier L (2006) **Measuring protein concentrations by NMR Spectroscopy**. *J Am Chem Soc* 128:2571-2576.

Wijekoon EP, Brosnan ME & Brosnan JT (2007) **Homocysteine metabolism in diabetes**. *Biochem Soc Trans* 35:1175-1179.

Wijndaele K, Duvigneaud N, Matton L, Duquet W, Thomis M, Beunen G, Lefevre J & Philippaerts RM (2007) **Muscular strength, aerobic fitness, and metabolic syndrome risk in Flemish adults**. *Med Sci Sports Exerc* 39:233-240.

Williams KJ (2008) **Molecular processes that handle – and mishandle – dietary lipids**. *J Clin Invest* 118:3247-3259.

WMA (2008) World Medical Association Declaration of Helsinki, ethical principles for medical research involving human subjects.

Wold H (1973) Nonlinear iterative partial least squares (NIPALS) modeling: some current developments. multivariate analysis (Ed. Krishnaiah, PR), Academic Press New York 383-487.

Wolf G, Chen S & Ziyadeh FN (2005) **From the periphery of the glomerular capillary wall toward the center of disease: podocyte injury comes of age in diabetic nephropathy**. *Diabetes* 54:1626-1634.

Yang J, Xu G, Hong Q, Liebich HM, Lutz K, Schmülling R & Wahl HG (2004) **Discrimination of type 2 diabetic patients from healthy controls by using metabonomics method based on their serum fatty acid profiles**. *J Chromatogr B Analyt Technol Biomed Life Sci* 813:53-58.

Ye J (2007) Least squares linear discriminant analysis. New York, NY, USA 1087-1093.

Yeap BB (2009) **Testosterone and ill-health in aging men**. *Nat Clin Pract Endocrinol Metab* 5:113-121.

Young RA (2000) **Biomedical discovery with DNA arrays**. *Cell* 102:9-15.

Yudkin JS (2007) **Insulin resistance and the metabolic syndrome – or the pitfalls of epidemiology**. *Diabetologia* 50:1576-1586.

Zgibor JC, Piatt GA, Ruppert K, Orchard TJ & Roberts MS (2006) **Deficiencies of cardiovascular risk prediction models for type 1 diabetes**. *Diabetes Care* 29:1860-1865.

Zhang SL, Filep JG, Hohman TC, Tang SS, Ingelfinger JR & Chan JS (1999) **Molecular mechanisms of glucose action on angiotensinogen gene expression in rat proximal tubular cells**. *Kidney Int* 55:454-464.

Zhang S, Day I & Ye S (2001) **Nicotine induced changes in gene expression by human coronary artery endothelial cells**. *Atherosclerosis* 154:277-283.

Zhang L, Krzentowski G, Albert A & Lefèbvre PJ (2003) **Factors predictive of nephropathy in DCCT type 1 diabetic patients with good or poor metabolic control**. *Diabet Med* 20:580-585.

Zhao G, Ford ES, Li C & Mokdad AH (2008) **Compliance with physical activity recommendations in US adults with diabetes**. *Diabet Med* 25:221-227.

Zhu M & Ghodsi A (2006) **Automatic dimensionality selection from the scree plot via the use of profile likelihood**. *Comp Stat Data Anal* 51:918 – 930.

Zoungas S & Asmar RP (2007) **Arterial stiffness and cardiovascular outcome**. *Clin Exp Pharmacol Physiol* 34:647-651.