

Online Appendix 1

<i>Abbreviation</i>	<i>Description</i>	<i>Unit</i>	<i>Availability</i>	<i>SOM input</i>
Adipon	adiponectin	mg/l	58%	yes
AHT	anti-hypertensive treatment	-	100%	-
ApoAI	serum apolipoprotein A-I	mg/dl	93%	yes
ApoAII	serum apolipoprotein A-II	mg/dl	89%	no
ApoB	serum apolipoprotein B	mg/dl	93%	yes
Chol	total cholesterol	mmol/l	97%	yes
CPept	C-peptide	pmol/l	76%	no
CRP	C-reactive protein	mg/l	82%	yes
DBP	diastolic blood pressure	mmHg	98%	-
DKD	diabetic kidney disease*	-	90%	-
DMDur	duration of type 1 diabetes	years	100%	-
DRP	diabetic retinal disease	-	99%	-
A1C	glycosylated hemoglobin A _{1c}	%	95%	yes
HDL ₂ C	high-density lipoprotein 2 cholesterol	mmol/l	92%	yes
HDL ₃ C	high-density lipoprotein 3 cholesterol	mmol/l	92%	no
LDLC	LDL cholesterol (Friedewald)	mmol/l	96%	no
MBL	mannan binding lectin	mg/l	59%	yes
MetS	metabolic syndrome	-	94%	-
SBP	systolic blood pressure	mmHg	98%	-
sCreat	serum creatinine	μmol/l	96%	yes
sK	serum potassium	mmol/l	55%	yes
sNa	serum sodium	mmol/l	55%	yes
TG	triglyceride concentration	mmol/l	97%	yes
24h-uAlb	24h-urine albumin excretion	mg	80%	yes
24h-uCreat	24h-urine creatinine excretion	mmol	66%	yes
24h-uK	24h-urine potassium excretion	mmol	62%	no
24h-uNa	24h-urine sodium excretion	mmol	62%	no
24h-uUrea	24h-urine urea excretion	mmol	61%	no

Only biochemical variables were included in the SOM input data. Furthermore, variables that were derived mathematically (Friedewald LDLC) or were strongly collinear with the other inputs (HDL₃C, ApoAII, 24h-uK, 24h-uNa, 24h-uUrea) were excluded (the low availability of the urine data caused additional artifacts in the analyses). C-peptide was close to the detection limit for almost all patients and was also excluded. SOM models that included all the biochemical variables were nevertheless tested to ensure that the model pruning did not reduce statistical sensitivity. Before analysis, the input data was rank-transformed for men and women separately.

The robustness of the analyses was tested by removing patients with low availability of data; the results for the more complete subsets were consistent with the full dataset (not shown). In theory, the SOM algorithm is able to automatically ignore any missing values, but due to the PCA initialization (illustrated in Online Appendix 2), the preprocessed data was imputed by least-squares linear regression; this also ensured that the missing values did not cause any non-linear artifacts.

*All patients were reviewed for DKD status, but 10% of the individuals could not be classified due to inconclusive clinical data.

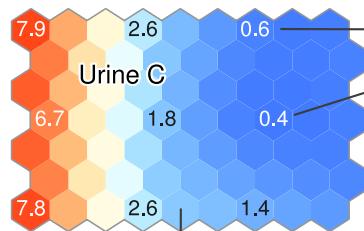
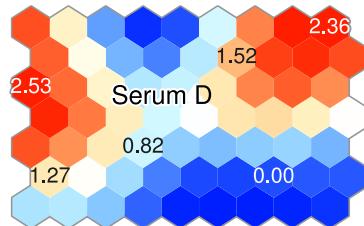
Self-organizing map

MeliKerion – free software for SOM analysis

The integration of numerous sources of biochemical and clinical materials sets a challenge to the researchers who are striving for a deeper insight into complex pathogenic mechanisms. Classical reductionist approaches often overlook non-linearity and rely on univariate diagnostic criteria as targets of inference, which may prevent the detection of complex multi-dimensional interactions. Unsupervised analysis of the regularities within a dataset is the first step in transforming the measurements into usable knowledge, but estimating the statistical significance of the observed patterns is often difficult.

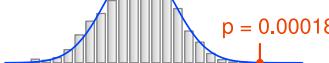
MeliKerion is an implementation of the Kohonen self-organizing neural network algorithm, and was designed for unsupervised analysis of clinical materials. A typical process flow is depicted on the right for a small dataset with diffuse phenotypical boundaries. In this simulated case, the emphasis is on the biochemical profiles, thus the first step is to preprocess the measurement data so that the variables become comparable in scale and mean value (Stage I). Next, linear decomposition (Stage II) is applied to create an initial map layout of patients (Stage III). The final layout is achieved after several iterations of a batch version of the Kohonen algorithm (Stage IV). Once the map is complete, the full dataset can be visualized based on the biochemical profiles, with significance estimates for clinical variables.

Internal component planes

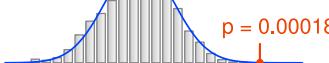


The dynamic range of the pseudo-colors is proportional to the expected random fluctuation; the colors are thus comparable between the component planes.

P-values for the observed regional variability can be estimated by simulating a situation where the biochemistry is not related to the clinical data (null hypothesis). If the observed statistic is very different from the null distribution, the observation is not likely to have occurred by chance. Here, 1000 such simulations were performed for each clinical group.

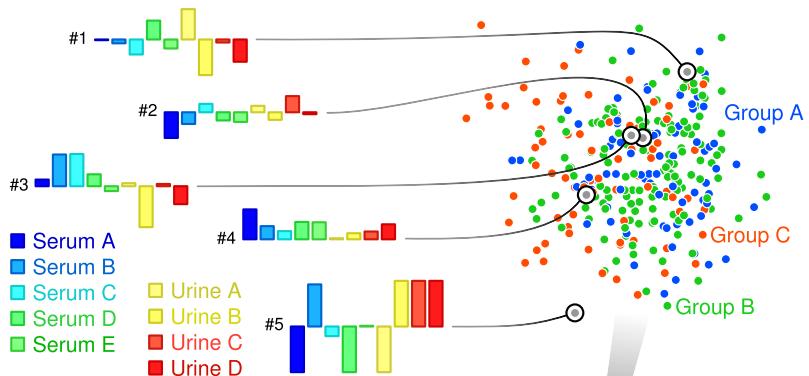


$p = 0.052$

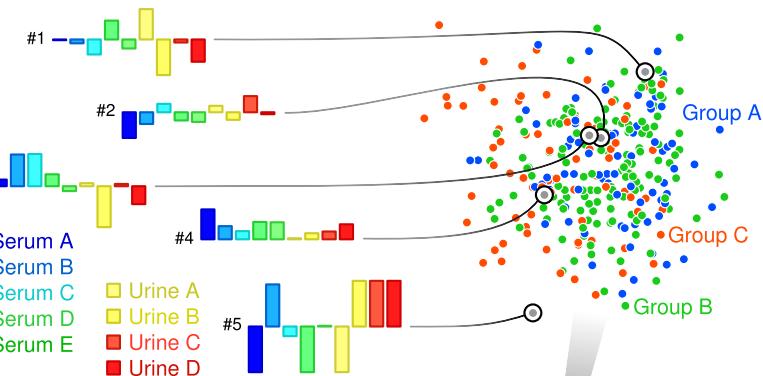


$p = 0.00018$

I Observed biochemical profiles



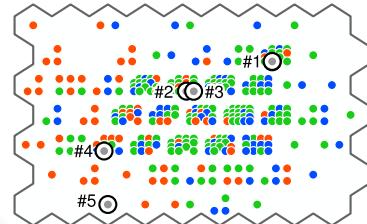
II Principal component analysis



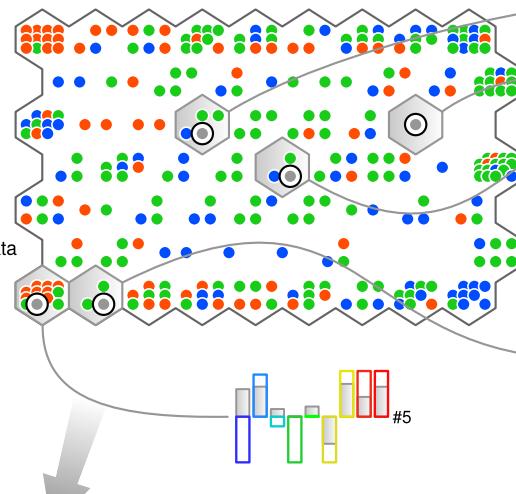
Hypothetical dataset

Patient records ($n = 300$) were created according to three independent latent traits. Nine biochemical variables were derived from the traits (some with non-linear formulae) with a 75% noise level. In addition, patients were divided into three groups based on two latent traits combined. The final dataset contained no distinguishable clusters and no direct algebraic links between the biochemistry and clinical categories. Data for sample #5 were altered to produce an outlier.

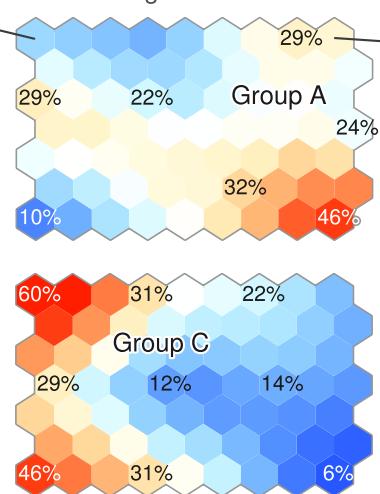
III Initial layout



IV Layout after self-organizing algorithm



External component planes with statistical significance estimates

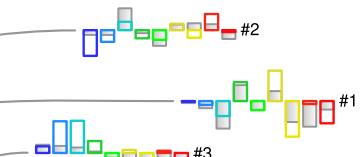


For categorical and binary variables, the map can be colored with respect to the demographic proportions on the hexagonal units.

Clusters in the data can be detected by estimating the rate of change of the model profiles between adjacent map units. Also, the areas between clusters would be void of individuals. In this case, the dataset was diffuse, hence the models change smoothly when moving across the map.

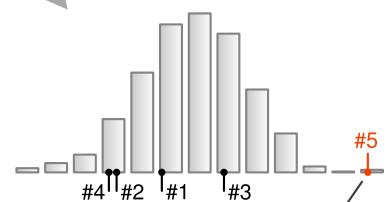
More details can be found in Supplement 1 of Mäkinen V-P et al.: *1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death*. *Molecular Systems Biology* 4:167, 2008.

Model profiles

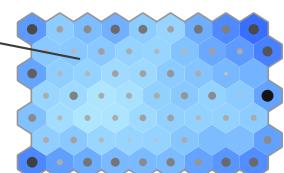


On each map unit, the neural network creates a model phenotype (grey) and together the models describe the recurring statistical features in the dataset. A patient's place on the map is set by the model that best matches the observed profile (colors).

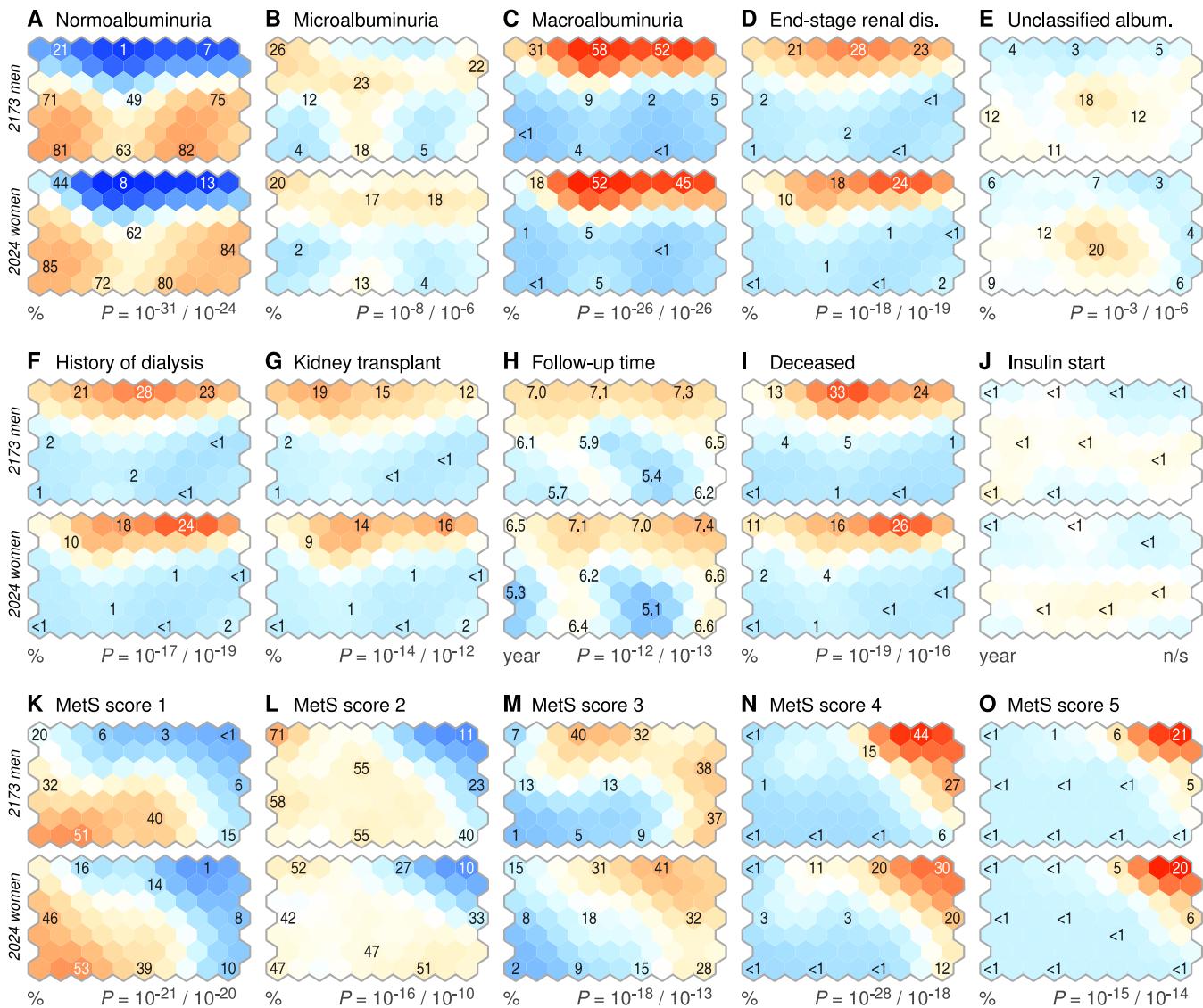
Map quality



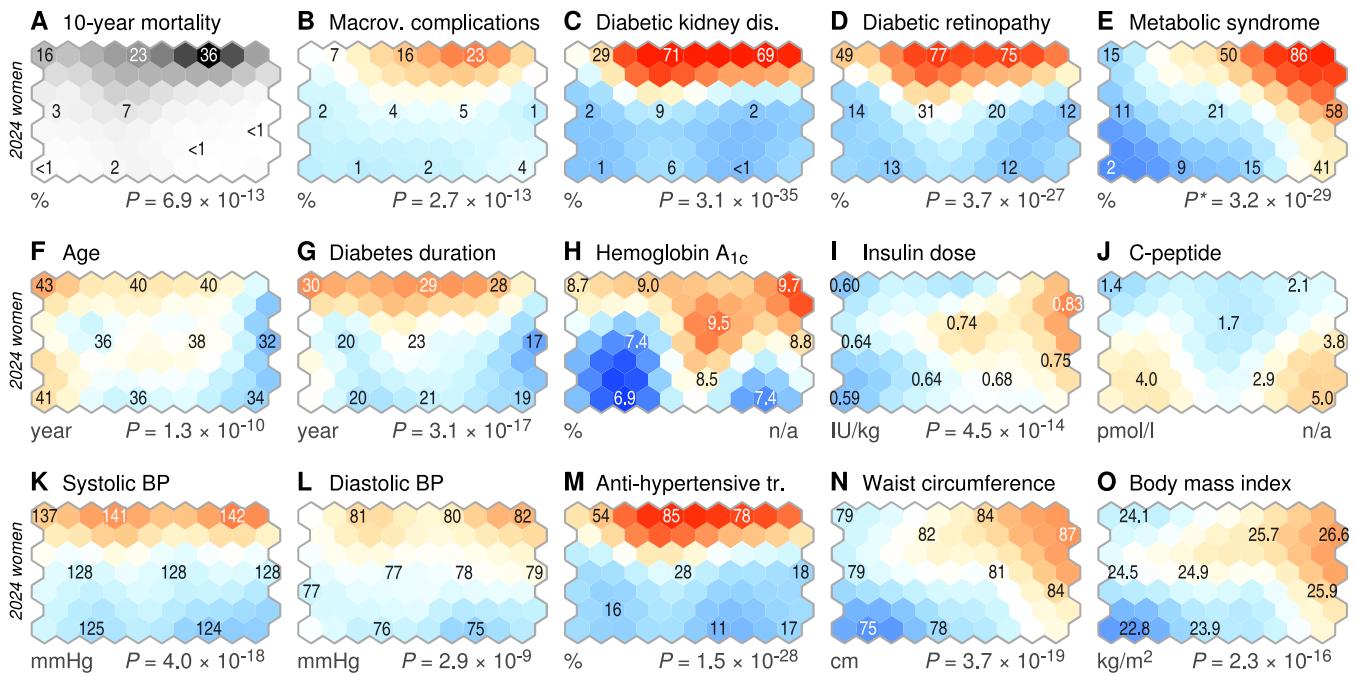
The error between an observed biochemical profile and the respective model reveals outliers.



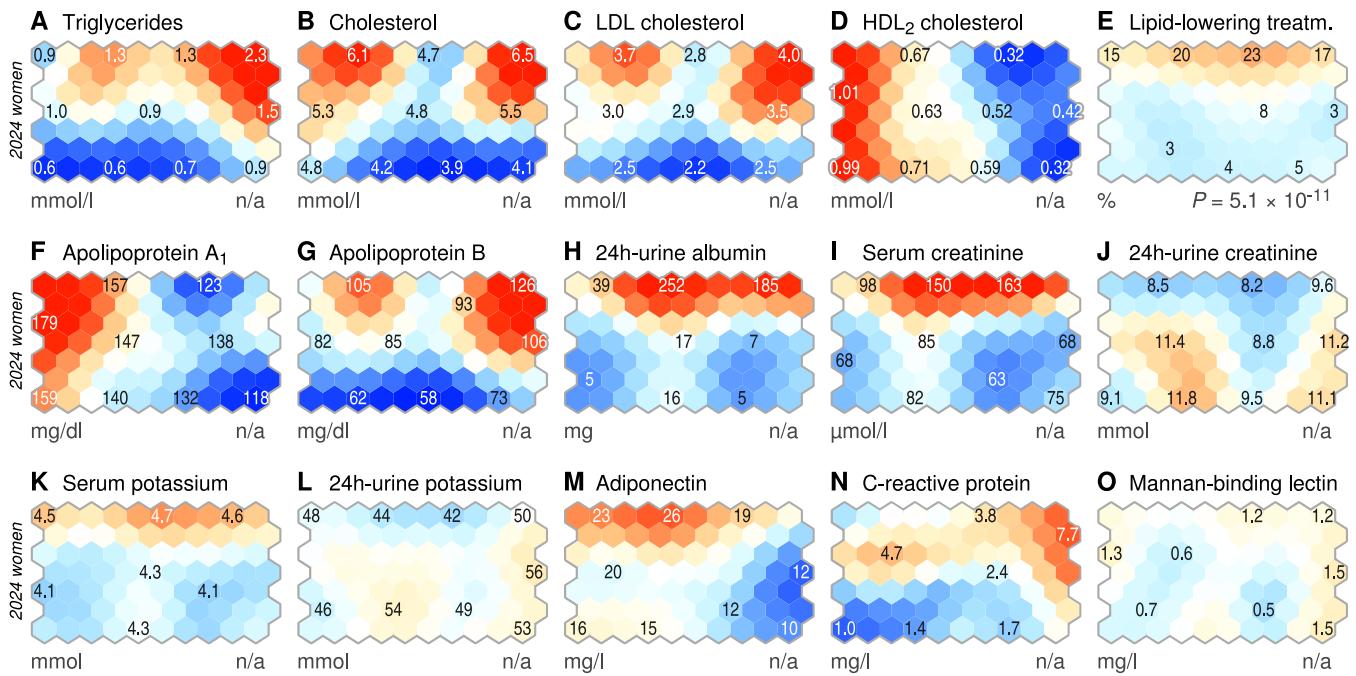
The MeliKerion is available at www.computationalmedicine.fi/software.



Online Appendix 3



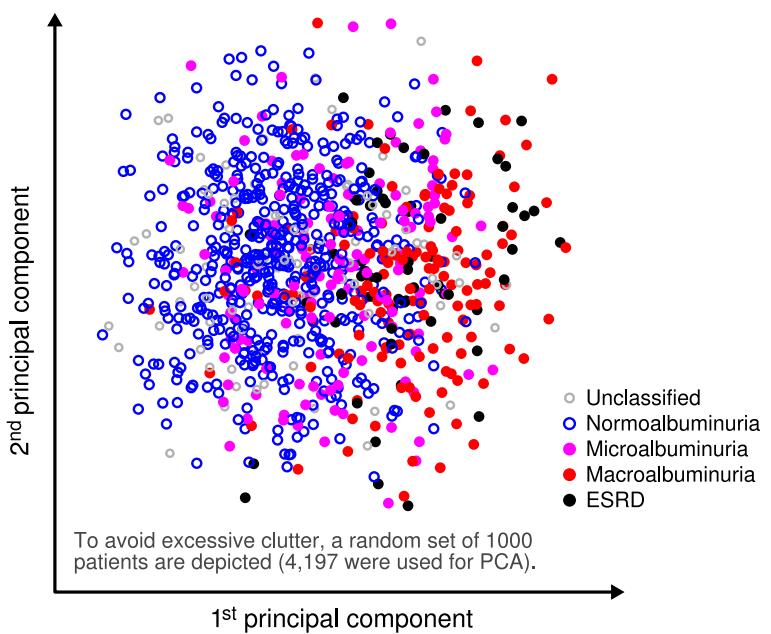
Online Appendix 4



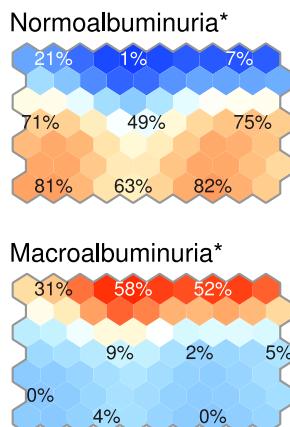
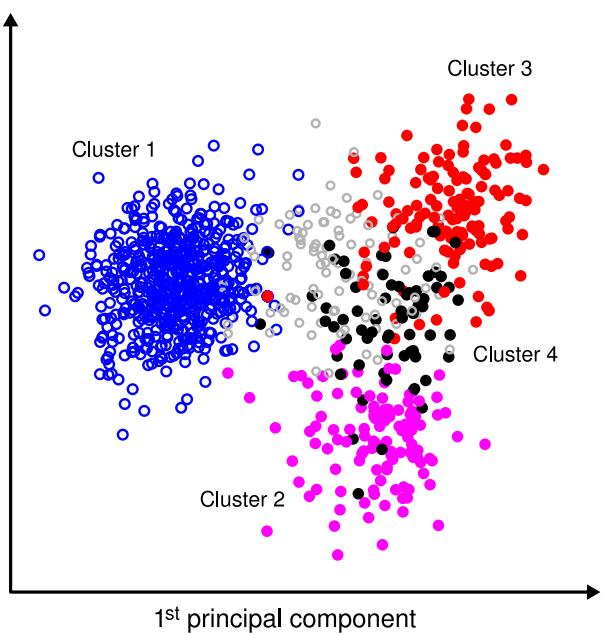
Online Appendix 5

Online Appendix 6

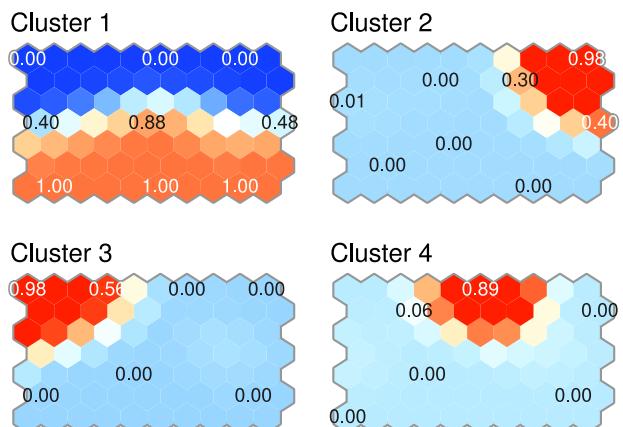
The FinnDiane dataset



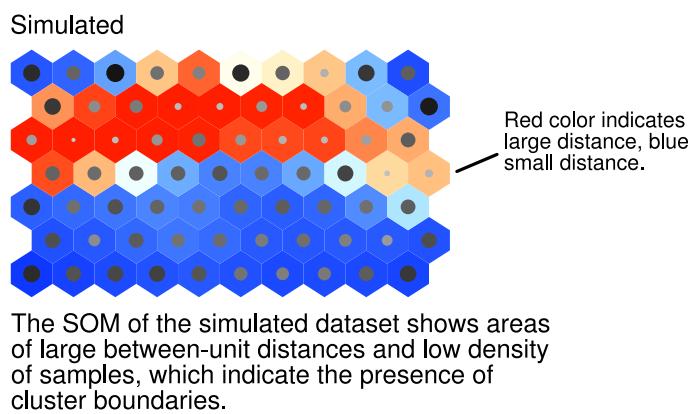
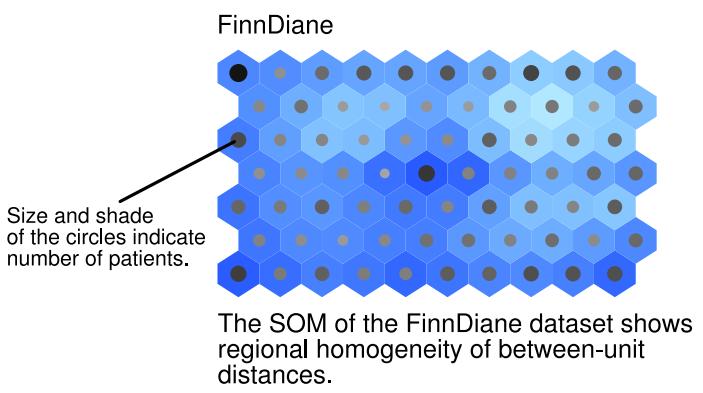
Simulated data with clusters



*Men only, women are shown in Online Appendix 2.



Euclidean distance in data space between map units



Online Appendix 7

The Finnish Diabetic Nephropathy Study Centers

Anjalankoski health care center	Sirpa Koivula Tarja Uggeldorf	Helsinki Herttoniemi Hospital	Markku Saraheimo Kerstin Wickström
Espoo health care center, Espoonlahti	Anita Nikkola Eila Ritola	Helsinki University Central Hospital	Paula Nikkanen Pirkko Salmi
Espoo health care center, Samaria	Eija Oukko-Ruponen	Hyvinkää health care center	Sisko Klemetti
Espoo health care center, Tapiola	Hanna Saarinen Merja Kataja-Tuomola		Teija Nyandoto
Espoo health care center, Viherlaakso	Marjatta Niska Anne Lytytinen Satu Raumavirta-Koivisto	Iisalmi Hospital	Eija Rontu Susanna Satuli-Autere
Etelä-Karjala Central Hospital	Marjo Kovanen Tuula Ensala Esko Hussi Jukka Toivonen Tuula Matikainen Raimo Vanamo Helga Hasu	Jokilaakso Hospital	Eija Toivanen Jaana Hyttiäinen
Heinola health care center	Päivi Hentunen Jukka Lagerstam	Jorvi Hospital	Tuula Laitinen Jukka Lagerstam
Helsinki health care center, Puistola	Kari Pohjola Heli Silvennoinen Tiina Simonen Hannele Kari	Jyväskylä health care center	Suvimarja Aranko Tarmo Leppälä Suvi Ervasti
Helsinki health care center, Suutarila	Anja Kaprio Birgitta Rantaeskola	Järvenpää health care center	Ritva Kauppinen-Mäkelin Arto Kuusisto
Helsinki health care center, Töölö	Jari Haaga Pirjo Kääriäinen Anna-Liisa Pietiläinen	Kainuu Central Hospital	Katriina Nikkilä Leena Pekkonen

Kanta-Häme Central Hospital	Tuula Myllykangas Pirjo Kinnunen Anja Orvola Tapani Salonen Annukka Vähänen Kerava health care center Keski-Pohjanmaa Central Hospital Keski-Suomi Central Hospital Kirkkonummi health care center Kivelä Hospital, Helsinki Koskela Hospital, Helsinki Kotka health care center Kouvola health care center	Helene Stuckey Paula Suominen Stefan Anderson Beatrice Asplund Ulrika Byskata Ingvor Byskata Toini Virkkala Päivi Liedes Terje Forslund Pirkko Koskiaho Marianne Laukkanen Juha Saltevo Marja Tiihonen Pirkko Koskiaho Anne Halonen Arvo Koistinen Esa Leppänen Anna-Mari Koski Anja Lappalainen Maisa Liimatainen Jaana Santaholma Ari Aimolahti Maarit Kärki Maija Lehtimäki Nina Uhlenius Eeva Huovinen Veikko Ilkka Eeva Pälkkö-Kontinen Annika Vanhanen Erkki Koskinen Anna Borgman Taimi Siitonen	Kuopio University Central Hospital Kuusamo health care center Kuusankoski Hospital Kymenlaakso Central Hospital Laakso Hospital, Helsinki Lahti City Hospital Länsi-Pohja Central Hospital Länsi-Uusimaa Hospital Lappeenranta health care center, Armila Hospital Lappi Central Hospital	Erja Huttunen Risto Ikäheimo Päivi Kiljander Markku Laakso Timo Lakka Leena Moilanen Leo Niskanen Ulla Tuovinen Ilkka Vauhkonen Erkki Voutilainen Pauli Karhapää Päivi Kekäläinen Erkki Lampainen Erja Isopoussu Tapio Kääriäinen Eero Vierimaa Esa Kilkki Liisa Riihelä Mare Riihelä Leena Ryysy Reino Paldanius Taina Meriläinen Paula Poukka Tuula Tikkanen Riitta Savolainen Aulikki Mäkelä Marjukka Tanner Eija Behm Pirkko Nyländen Hilkka Laukkanen Pirjo Myllykoski Inge-Maj Jousmaa Jukka Rinne Pirjo Linkola Irja Pulli Sirkka Severinkangas Timo Tulokas
-----------------------------	---	--	---	---

Lohja Hospital	Liisa Hyvärinen Tuula Granlund Mikko Saari Tuula Salonen Reija Tähtinen Pirjo Eloranta Arja Mäkelä Tarja Kalliomäki Juha Koskelainen Riitta Nikkanen Vesa Salonen Niilo Savolainen Hannu Sulonen Eija Valtonen Helena Lanki Seija Moilanen Marju Tilly-Kiesi Anne-Mirjamä Hänninen Ismo Pirttiniemi Anita Gynther Maarit Salminen Tuomo Vänttininen Mikko Honkasalo Marko Miettinen Päivi Nurminen Elina Jokelainen Pirkko-Liisa Jylkkä Elsi Kaarlela Jari Vuolaspuro Rauni Häkkinen Liisa Hiltunen Sirkka Keinänen-Kiukaanniemi Hannu Haapamäki Satu Hämäläinen Antero Helanterä Vesa Ilvesmäki Hilkka Miettinen	Palokka health care center, Palokka Palokka health care center, Vaajakoski Pieksämäki Hospital Pietarsaari Hospital Pohjois-Karjala Central Hospital Pori City Hospital Porvoo Hospital Raahe Hospital Rauma Hospital Riihimäki Hospital	Tarja Tick-Sinkkilä Leena Welling Kirsti Mäkinen Päivi Sopanen Mari Tamminen Markku Taskinen Maj-Len Holmbäck Bo Isomaa Leena Sarelin Laura Hyttinen Päivi Kekäläinen Maija Pietarinen Anneli Rissanen Risto Laitinen Hannu Turtola Ulla Maija Henttula Pirjo Merensalo Kirsi Sävelä Terttu Viitala Päivi Ahonen Heidi Johansson Sirkka Rämö Bjarne Rask Marjatta Kallion Anne Holma Matti Honkala Aino Tuomivaara Riitta Vainionpää Kaisa Laine Taru Salminen Paula Ketola Kaisa Saarinen Pirkko Aalto Marja Lanni Irmeli Laitaharju Eija Immonen Leena Juurinen
Loimaa health care center			
Lounais-Häme Hospital			
Malmi Hospital, Helsinki			
Mänttä Hospital			
Mikkeli Central Hospital			
Nurmijärvi health care center			
Oulaskangas Hospital			
Oulu health care center			
Päijät-Häme Central Hospital			

Salo Hospital	Jouni Lapinleimu Pirkko Rautio Merja Virtanen Anita Alanko Markku Asola Paavo Pääkkönen Marjo Pentti Marja Rautavirta Pertti Kunelius Sirkku Mannila Tuula Pulli Päivi Sallinen Eija Toivanen Helena Valtonen Aarne Vartia Eeva Korpi-Hyövälti Terhi Latvala Esko Leijala Tarja Nyrhilä Jukka Rinne Inge-Maj Jousma Pekka Alarotu Liisa Calonius Seppo Gummerus Paula Helin Tarja Kaitala Helena Kirkkopelto-Jokinen Eila Kujansuu Taru Niskanen Atte Vadén Timo Saaristo Marjo Kuortti Arja Kallio Ilpo Ala-Houhala Tuula Kuningas Paula Lampinen Marita Määttä	Tiirismaa health care center Turku health care center Turku University Central Hospital Vaajakoski health care center Vaasa Central Hospital Valkeakoski Hospital Vammala Hospital	Heikki Oksala Tuija Oksanen Kristiina Salonen Sirkku Tulokas Hilkka Tauriainen Leena Petlin Liisa Savolainen Tiina Kivelä Anne Artukka Irmeli Hämäläinen Päivi Kokkila Lea Sorvari Markku Vähätalo Hille Virtamo Kaija Breitholz Raija Eskola Kaj Metsärinne Riikka Neva Ulla Pietilä Pirjo Saarinen Riitta Tuominen Markku Asola Sari Äyräpää Päivi Sopanen Kirsti Mäkinen Stig Bergkulla Ulla Hautamäki Vivi-Ann Myllyniemi Irma Rusk Terhi Immonen Seppo Ojanen Maisa Rautiainen Eija Valtonen Hannu Ylönen Irma Isomäki Leena Mustaniemi Merja Tapiolinna-Mäkelä
Satakunta Central Hospital		Turku health care center	
Savonlinna Central Hospital			
Seinäjoki Central Hospital			
Tammisaari Hospital			
Tampere health care center			
Tampere University Central Hospital			

Vantaa health care center, Korso	Riitta Kroneld Jaana Vahtola Riitta Toivonen Raila Ahonen Marja Ivaska-Suomela Anu Jauhainen Sari Kultti Pirjo Kiviniemi Merja Laine Tiina Pellonpää Riitta Puranen Anne Airas Jukka Laakso Matti Eerola Eeva Jatkola Riitta Lönnblad Jari Mäkelä Anita Malm Elli Rautamo Hans Granlund Ann-Christin Johnsson Monica Forsen Satu Savolainen Tapani Marjanen
Vantaa health care center, Länsimäki	
Vantaa health care center, Martinlaakso	
Vantaa health care center, Myyrmäki	
Vantaa health care center, Rekola	
Vantaa health care center, Tikkurila	
Åland Central Hospital	
Äänekoski health care center	