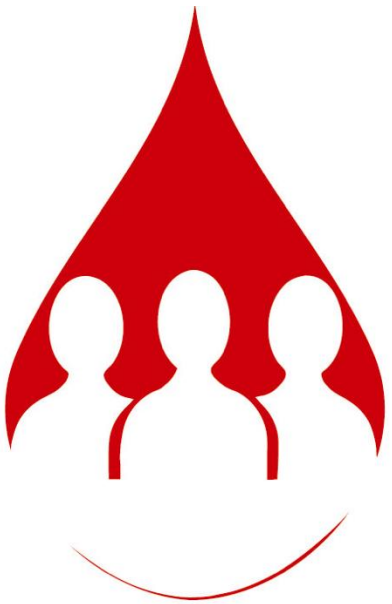
Big data, omics of donors/recipients, personalized medicine

Christian Erikstrup
Chief physician, associate professor
The Danish Blood Donor Study
Dept. of Clinical Immunology
Aarhus University Hospital



Definition of big data

Coined in: Cox and Ellsworth, Proceedings of the 8th conference on Visualization '97:

"Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data."

Mayer-Schönberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. 2013:

"The ability of society to harness information in novel ways to produce useful insights or goods and services of significant value"

Big data is not just large data sets

Data from several sources requiring interdisciplinary teams for analysis.

Integrative analysis – complex data, time trajectories.

Phenomics

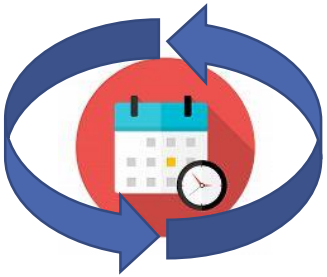
- Clinical data
- Questionnaires
- Register data

……

Genomics

- Genotypes (GWAS)
- Sequencing (WGS/WGS)
- EPIgenetics

……



Transcriptomics

- RNA Sequencing
- MicroRNA

……

Immunomics

- Response to pathogens
- Innate/adaptive immunity
- Cytokines

……

Proteomics, metabolomics, lipidomics, glycomics .....

Big data, omics in blood donors/recipients and personalized medicine

- Few studies in transfusion medicine qualify for following an approach using integrative analysis
- Several large studies have started data collection that will serve as platforms for "big data" studies in the years to come (INTERVAL Study, REDS-III, Donor InSight, Établissement Français du Sang)
- These studies may facilitate new discoveries for use in precision medicine
- Precision medicine fits nicely with transfusion medicine: Serologic crossmatch could qualify as one of the first examples of personalized medicine
- In the following, The Danish Blood Donor Study will be used as an example of a study which is gaining momentum to qualify for becoming a "big data" study in donors/recipients – focus on infectious diseases

Vision:

- To study the health impact of blood/plasma donation
- To identify risky donors
- To answer generic research questions not related to transfusion medicine
- To collaborate with other researchers

In a very cost efficient setup using the blood bank infrastructure

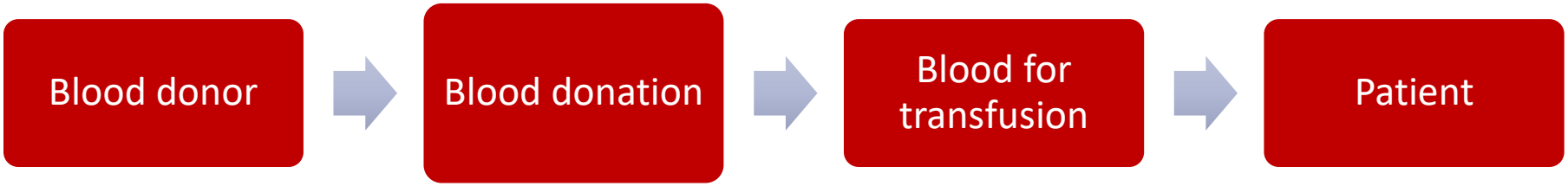
Blood bank infrastructure supports the implementation of large scale population studies

✓	Access to participants:	250,000 blood donations annually
✓	Collection sites:	200 sites nationally
✓	Educated staff:	Nurses, technicians, physicians, public relations experts
✓	Transport:	Look-back samples from each donation
✓	Laboratories:	Sample handling and analysis
✓	Storage Facilities:	-20C automated freezer, -80C freezers

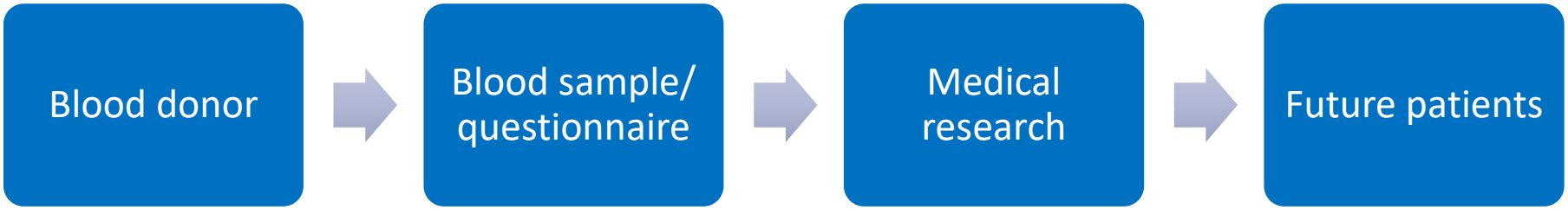


Donors want to participate in research projects: Participation rate >95%

Helping current patients



Helping future patients



Now part of the strategy for the Danish Blood Donor Association in the recruitment of new blood donors

Data collection and biobank law

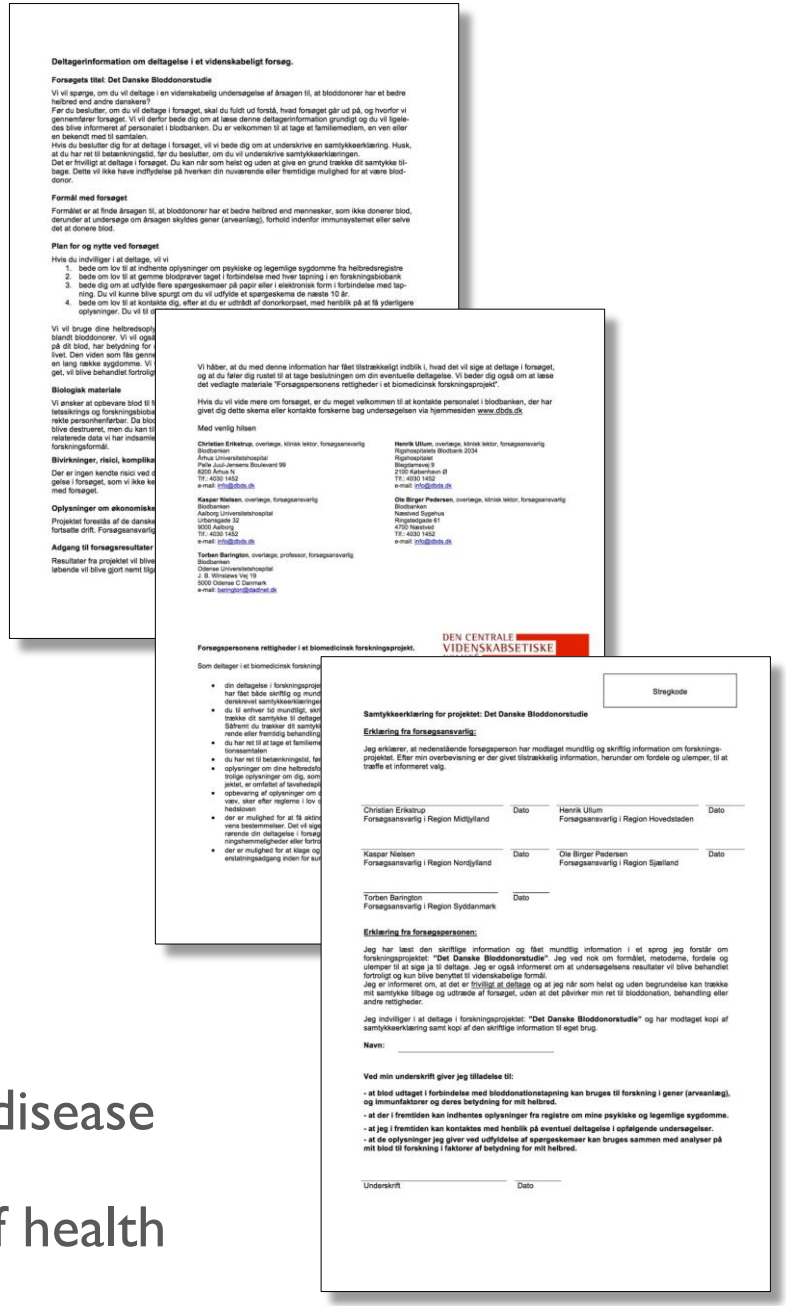
Facilitating factors of big data donor studies:

- Legal framework with support for:
  - the construction of large scale research bio banks
  - broad consent to facilitate the answering of research questions not formulated when donor was included
  - health register data accessible for researchers

Broad consent in the Danish Blood Donor Study

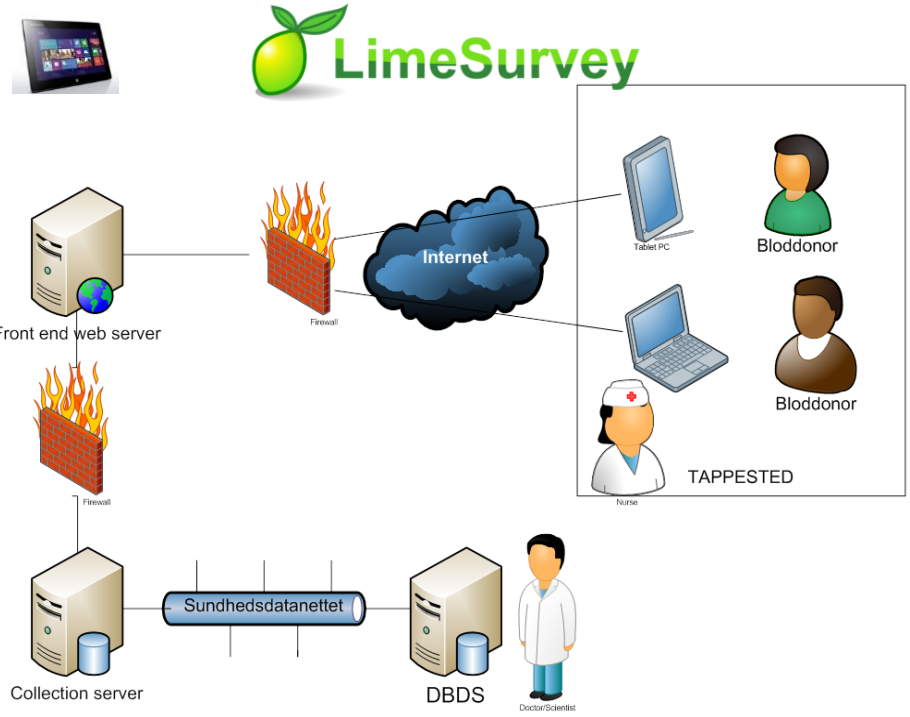
We have the permission to:

- ask donors to participate
- store samples in bio bank (plasma, DNA, nasal swabs)
- collect questionnaires – now and later
- contact donor again
- collect data from national registers
- study genetic predictors of health and disease
- study immunological plasma markers of health and disease



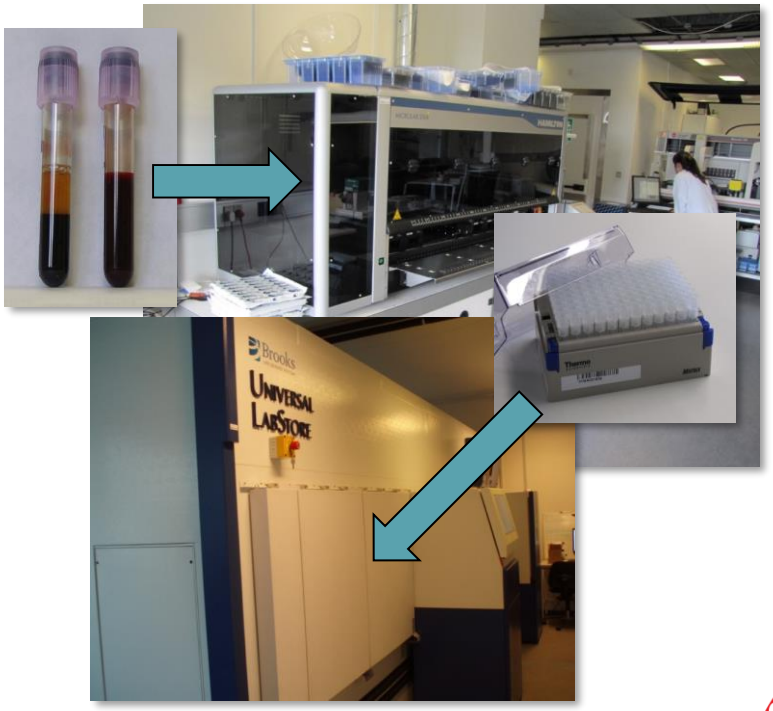
The collection of data

Tablet based questionnaire
Filled while waiting or
giving blood



Folding questions

Plasma and whole blood
biobank



Danish health registers:

National Patient Register

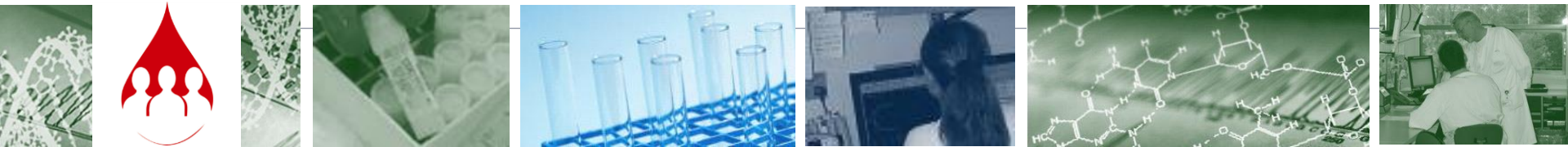
ICD-10 codes from all patient contacts with hospitals

National Prescription Register

ATX codes from all filled prescriptions

- Close to 100% coverage – reimbursement depends on registration

Socioeconomic data



Status

Inclusion initiated March 2010

115,000 blood donors have been included

Work dataset: approx. 110,000 participants; 475,000 person-years of follow-up by Dec 31 2016

All baseline samples transferred to automated sample management system

>600,000 plasma archive samples from <u>every donation</u> available for research



Blood donor compliance

- Can we identify risky donors through the donor health questionnaire?

Requirements for relying on the DHQ:

- Risk for relevant infections can be identified through questions regarding behaviour, symptoms
- The donors' answers are accurate

<u>Recent post-donation questionnaire studies among men who have sex with men:</u>

Blood donor deferral for men who have sex with men: the Blood Donation Rules Opinion Study (Blood DROPS), Custer et al (REDS-III), Transfusion 2015: *"Noncompliance with the MSM policy is evident."*

Views and experiences of men who have sex with men on the ban on blood donation: a cross sectional survey with qualitative interviews, Grenfell et al, BMJ, 2011: *"10.6% of men with experience of penetrative sex with a man reported having donated blood in Britain while ineligible under the exclusion criterion, and 2.5% had donated in the previous 12 months."*

Blood donor compliance

Criteria: defer donors who took systemic antibiotics within 28 days of donation

Filled prescriptions for systemic antibiotics before and after donation among all Danish donors, 1998-2012:

4.9 mill. donations from 463,129 donors

Possible reasons:

1. Some donors forget taking medicine
2. DHQ questions are not clear enough
3. Some donors perform self assessment of risk

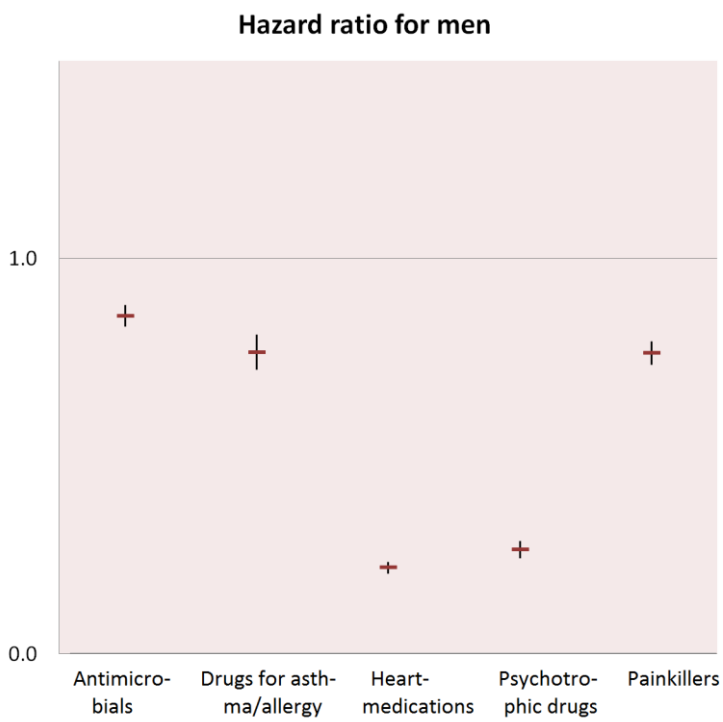
This study is different from the previously cited studies: this is real behaviour and not donors' response to questions regarding their behaviour

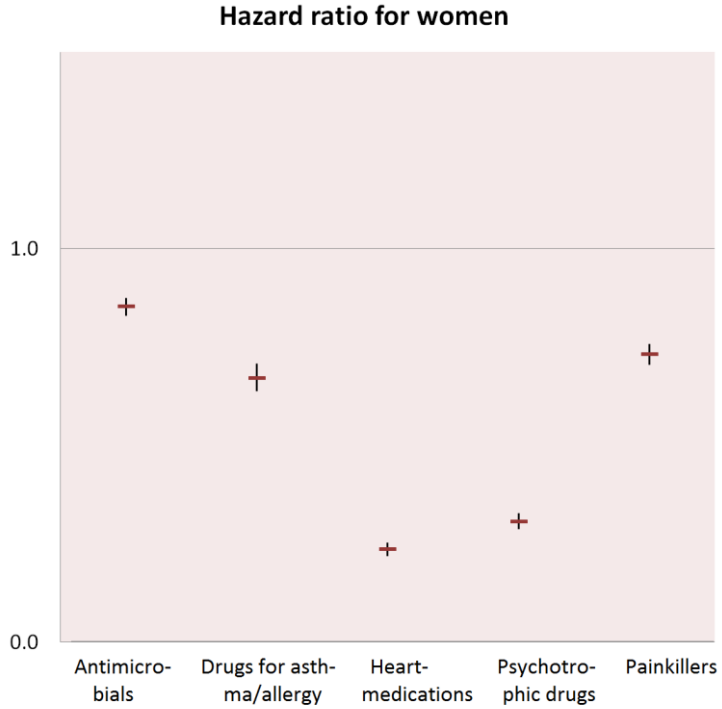
The healthy donor effect depends on the disease

37,808 participants and 44,917 randomly chosen controls from the general population matched for age, sex, and region of Denmark.

22,198 donors received at least one prescription during 48,492 person-years

26% of donors fill at least one prescription for an antibiotic every year!





(Hansen et al, unpublished)

InSight Study, Sanquin: several great studies on the healthy donor effect



Infectious agents in relation to psychiatric diseases and risk-taking

- *Toxoplasma gondii* and Cytomegalovirus have been associated with psychiatric diseases and cognitive deficits
- *T. gondii increases risk-taking in mice*

- Register data from 85,381 blood donors were analyzed to identify cases and controls
- Samples from cases and controls were collected from freezers; analysed for antibodies to *T. gondii*, CMV and others

Infectious agents in relation to psychiatric diseases

Cases		5,500 cases
Any psychiatric disorder	All included	2,900
Suicide or suicide attempts	Attempt age 15 years or older	400
Traffic accidents	Accident after 2008 included.	2,800
Controls	Matched on age at inclusion and gender	6,000 controls

Preliminary analysis:
T. gondii and CMV associated with schizophrenia and neurotic disease

Infectious agents in relation to psychiatric diseases

Pending analyses - do antibodies to *T. gondii*, CMV predict:

- Traffic accidents
- Risk behaviour

Fruitful collaboration with Copenhagen Business School: modelling risk behaviour

Are risk-taking donors risky for blood recipients?

If the speed limit is 80kph/50mph but you decide this does not apply to you what then about complying with criteria in the blood bank?

ORIGINAL RESEARCH

Annals of Internal Medicine

Transmission of Neurodegenerative Disorders Through Blood Transfusion

A Cohort Study

Gustaf Edgren, MD, PhD; Henrik Hjalgrim, MD, PhD, DrMedSci; Klaus Rostgaard, MSc; Paul Lambert, PhD; Agneta Wikman, MD, PhD; Rut Norda, MD, PhD; Kjell-Einar Titlestad, MD, PhD; Christian Erikstrup, MD, PhD; Henrik Ullum, MD, PhD; Mads Melbye, MD, DrMedSci; Michael P. Busch, MD, PhD; and Olof Nyrén, MD, PhD

2016

SCANDAT: The Scandinavian Transfusions and Donations database

1.5 mill. patients who received transfusions between 1968 and 2012

Patients with diagnosis of dementia of any type, Alzheimer disease or Parkinson disease after transfusion were identified i Swedish and Danish registers:

- Hazard ratios were calculated for donors with later disease and compared with healthy donors with regard to recipient risk.

No evidence for transmission of dementia was found – thus no evidence for a transmissible agent.

Control of the method: Transmission of chronic hepatitis before and after implementation of hepatitis C virus screening.

Good example how large studies with strong statistical power can give strong answers even though the findings are negative.

Interaction: Donation changing infection risk?



- 497,390 donors followed after 5,458,499 donations
- Hb below deferral guidelines was not associated with increased risk of infection during 3 months after donation.
- Conversely, a slightly reduced risk of filling a prescription for anti-microbials was observed:
  HR: 0.92, CI: 0.91-0.93.
- Does this indicate that iron loss may be beneficial?
- Nation-wide semi-cluster-randomised study of iron supplementation

Deferral for low hemoglobin is not associated with increased risk of infection in Danish blood donors

Sebastian R. Kotzé,[1] Ole B. Pedersen,[2] Mikkel S. Petersen,[1] Erik Sørensen,[3] Khoa Manh Dinh,[1] Kathrine Agergård Kaspersen,[1] Andreas S. Rigas,[3] Henrik Hjalgrim,[4] Klaus Rostgaard,[4] Gustaf Edgren,[5,6] Henrik Ullum,[3] and Christian Erikstrup[1]

Transfusion 2016

Increasing data volume and adding integrative analysis

- Collaboration between the Danish Blood Donor Study and deCODE, Iceland:
- Genome Wide Association study:
  110,000 donors to be typed for 650,000 SNPs using the Illumina Infinium Global Array
- Data to be linked with health registers, phenotypes and analysed on Computerome

COMPUTEROME

A Danish supercomputer dedicated solely to life science data



Computing Power: 16,000+ cores

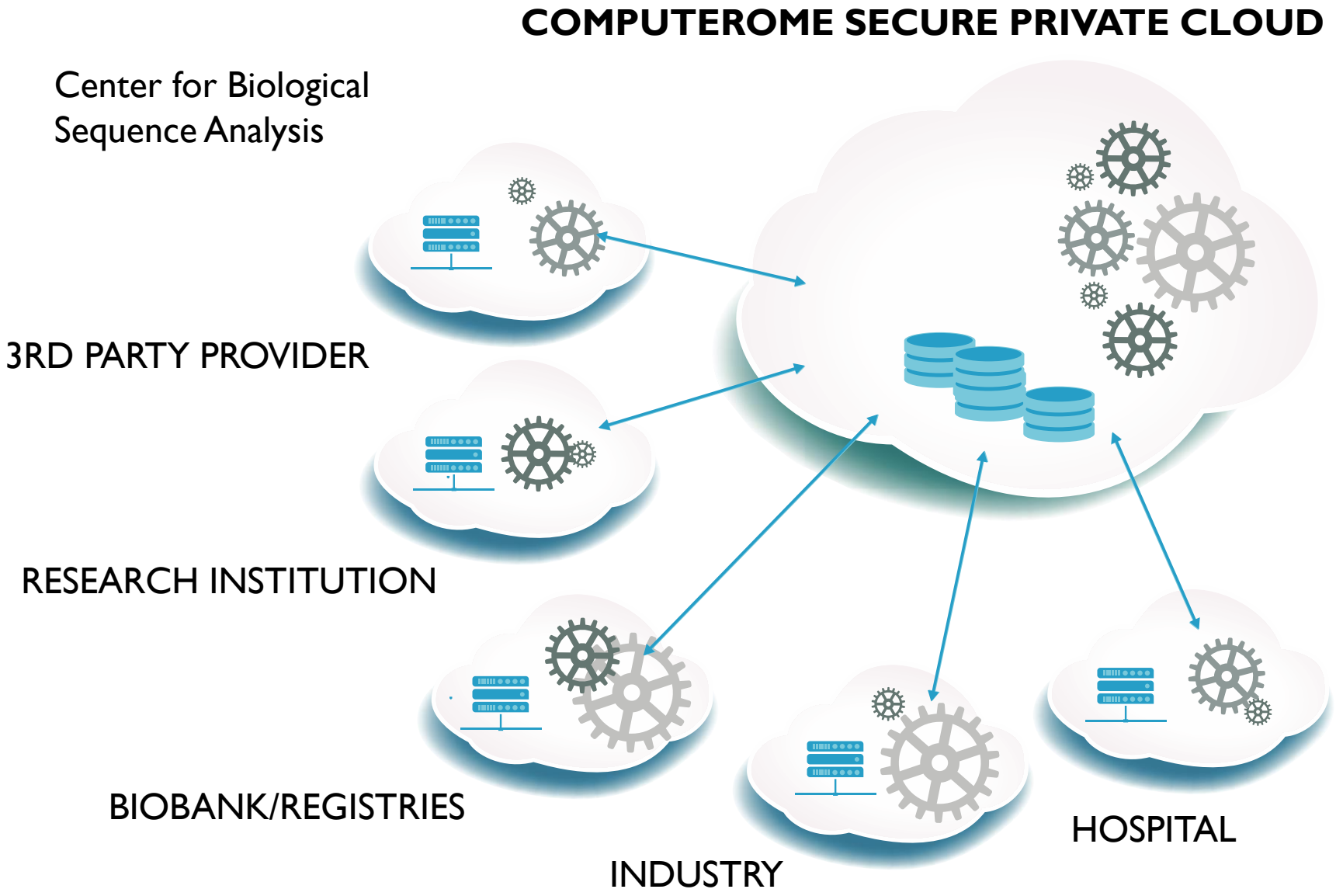
- 525 x Thin nodes
- 28 core Intel CPUs with 128 GB DDR4
- GPU Node – Nvidia K40
- 29 x FAT nodes
- 32 core Intel CPUs with 1TB DDR3 RAM

Storage: 8PB+

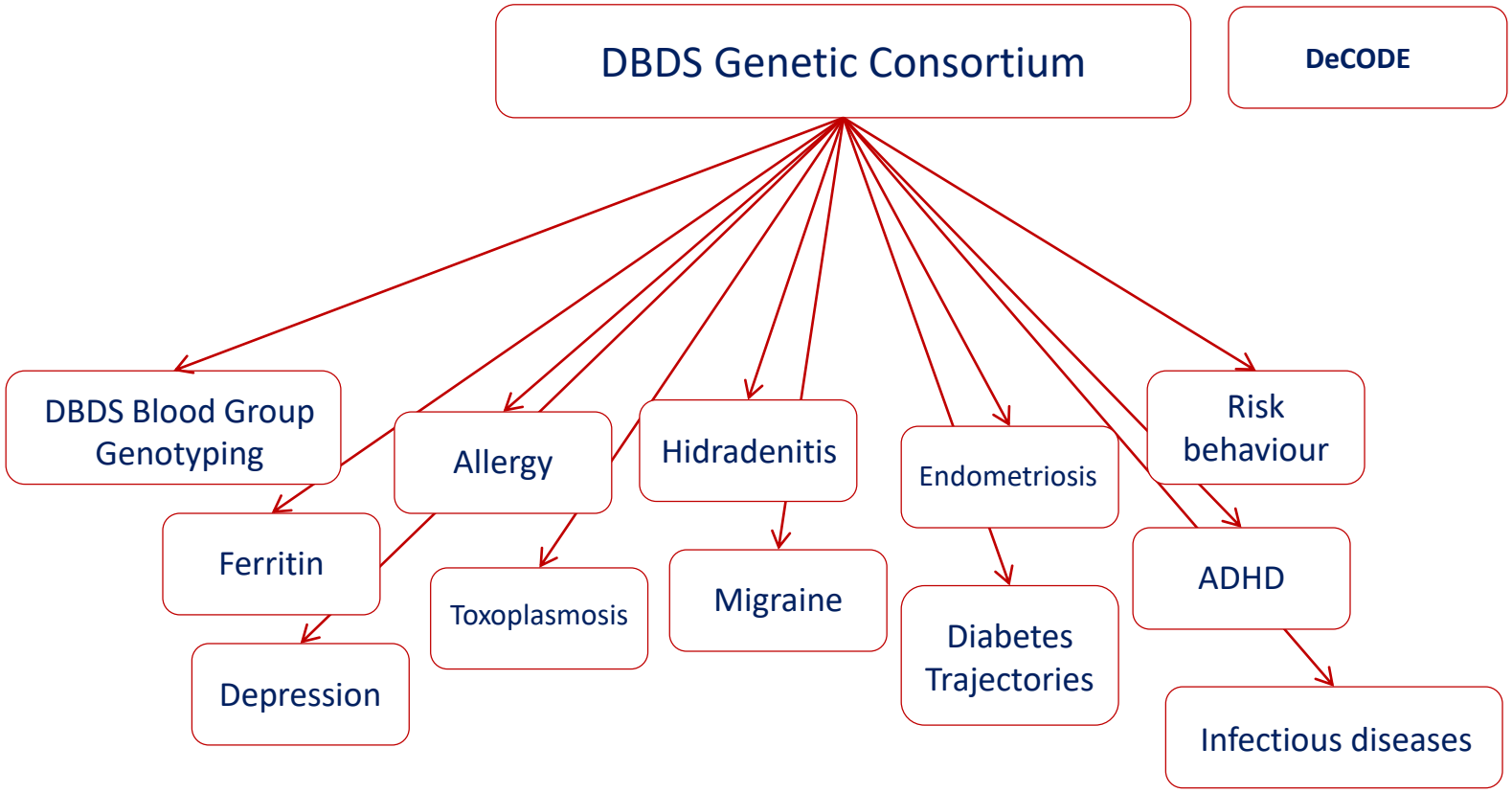
- 8 PB extremely high performance storage
- +50GB/s aggregated bandwidth

The Danish Blood Donor Study

Proprietary Danish secure cloud



DBDS Genetic Consortium
Collaborate to expand possibilities
One GWA array
– multiple outcomes, broad use of the data



Immediate use of big data in blood bank production

- Imputation from GWA array can be used to impute many more SNPs.
- We propose to investigate the possible use of GWA arrays in the imputation of:
  - HLA types for stem cell register donors
  - Blood group – especially rare blood groups – among blood donors

Proof of concept: 18,000 blood donors typed for 3 SNPs facilitated the identification of donors with rare types: Kpb neg (4 donors), Yta neg (35), HPA1a neg (478)

Concordance between imputed and sequenced HLA type

PLOS ONE | 2017

RESEARCH ARTICLE

Comparison of HLA allelic imputation programs

Jason H. Karnes[1], Christian M. Shaffer[2], Lisa Bastarache[3], Silvana Gaudieri[4,5,6], Andrew M. Glazer[2], Heidi E. Steiner[1], Jonathan D. Mosley[2], Simon Mallal[4,6,7], Joshua C. Denny[2,3], Elizabeth J. Phillips[2,4,6], Dan M. Roden[2,3,8]*

		European Americans		African Americans	
Allele	**Imputation Program**	**Concordance Rate**	**Call Rate**	**Concordance Rate**	**Call Rate**
HLA-A	SNP2HLA	0.983	0.999	0.969	0.995
	HLA*IMP:02	0.963	0.997	0.675	0.855
	HIBAG	0.986	0.996	0.960	0.796
HLA-B	SNP2HLA	0.969	1.00	0.884	1.00
	HLA*IMP:02	0.952	0.979	0.423	0.752
	HIBAG	0.978	0.967	0.953	0.403
HLA-C	SNP2HLA	0.987	1.00	0.884	1.00
	HLA*IMP:02	0.984	0.994	0.792	0.741
	HIBAG	0.987	0.992	0.957	0.619
HLA-DPB1	SNP2HLA	0.957	1.00	0.945	1.00
	HLA*IMP:02	0.829	0.987	0.567	0.708
	HIBAG	0.957	0.975	0.834	0.475
HLA-DQB1	SNP2HLA	0.988	1.00	0.907	1.00
	HLA*IMP:02	0.983	0.993	0.845	0.761
	HIBAG	0.988	0.990	0.904	0.654
HLA-DRB1	SNP2HLA	0.964	1.00	0.920	1.00
	HLA*IMP:02	0.924	0.961	0.414	0.791
	HIBAG	0.959	0.946	0.946	0.557



Prediction of chronic lymphatic leukemia

Proof of concept: can hematology parameters and register data be used to identify donors at risk for disease

- Automated hematology analyzers are used for hemoglobin measurements in most Danish bleeding facilities:
  Full white cell differential: 65,000 measurements /y
  Red cell parameters: approx. 150,000 measurements /y

- One small donation facility: 118,430 lymphocyte measurements from 15,448 donors during 13 years
- Only 32 donors had at least one measurements of lymphocytes >5 x 10^9 cells/L
- 7 donors were subsequently diagnosed with CLL – all were among the 32 donors with lymph. >5

Lymphocyte count and risk of CLL

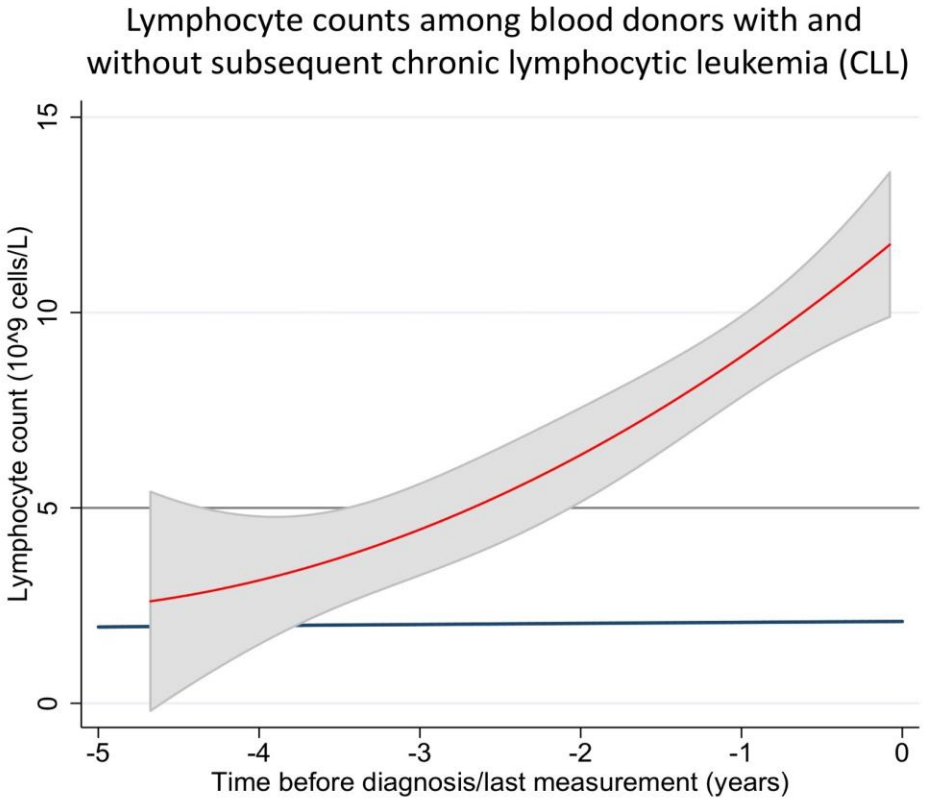
Criteria

- age>40 years
- Lymph. >5 x 10^9 cells/L

Sensitivity:
100% (59-100)

Specificity:
99,94% (99,89-99,97)

Positive predictive value:
41% (18-67)



Conclusions

- We bleed donors with CLL

No evidence of transmission of chronic lymphocytic leukemia through blood transfusion

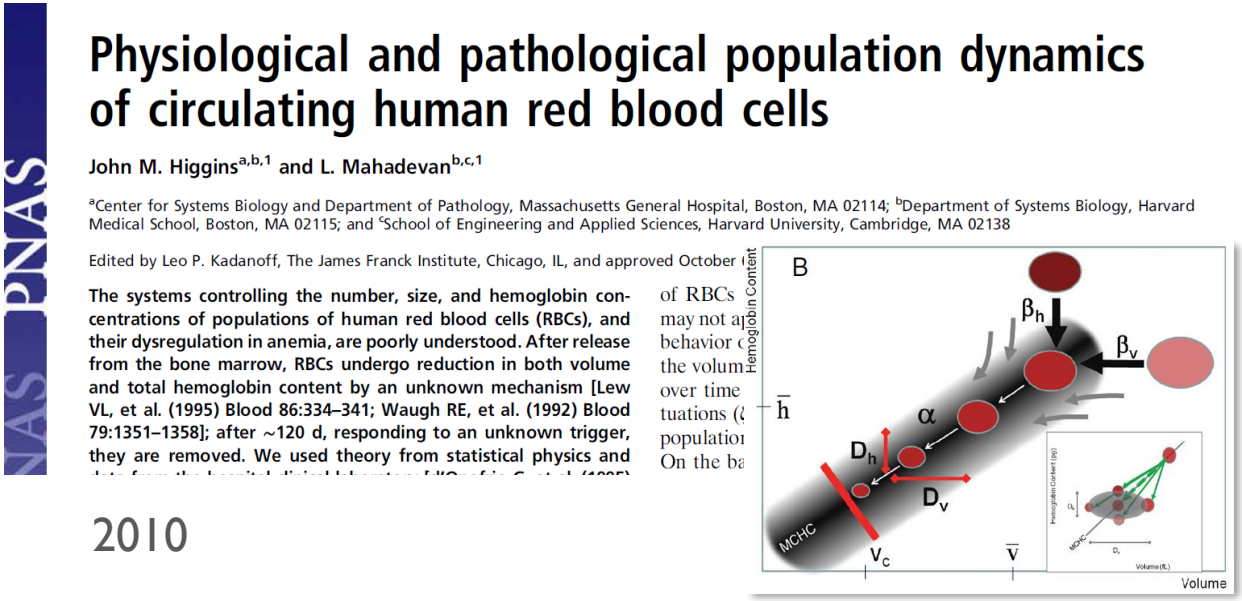
Henrik Hjalgrim,[1] Klaus Rostgaard,[1] Senthil K. Vasan,[2] Henrik Ullum,[3] Christian Erikstrup,[4] Ole B. V. Pedersen,[5] Kaspar R. Nielsen,[6] Kjell-Einar Titlestad,[7] Mads Melbye,[1] Olof Nyrén,[2] and Gustaf Edgren[2,8]

(Blood 2015)

- Donors should be informed and offered counselling
- New medicine – early intervention may be feasible for some mutations
- Data from hematology analyzers: immense possibilities

Bigger data?

Single cell data from hematology analyzers



PNAS

Physiological and pathological population dynamics of circulating human red blood cells

John M. Higgins[a,b,1] and L. Mahadevan[b,c,1]

[a]Center for Systems Biology and Department of Pathology, Massachusetts General Hospital, Boston, MA 02114; [b]Department of Systems Biology, Harvard Medical School, Boston, MA 02115; and [c]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Edited by Leo P. Kadanoff, The James Franck Institute, Chicago, IL, and approved October

The systems controlling the number, size, and hemoglobin concentrations of populations of human red blood cells (RBCs), and their dysregulation in anemia, are poorly understood. After release from the bone marrow, RBCs undergo reduction in both volume and total hemoglobin content by an unknown mechanism [Lew VL, et al. (1995) Blood 86:334–341; Waugh RE, et al. (1992) Blood 79:1351–1358]; after ~120 d, responding to an unknown trigger, they are removed. We used theory from statistical physics and

of RBCs
may not ap
behavior c
the volum
over time
tuations (
populatio
On the ba

2010

The INTERVAL Study, UK, has already started collecting (very) big data

Imagine the CLL study performed with an integrative approach using all single cell data over time with all diseases as outcome!



Allergy and asthma – influence of pollution

Upcoming:
Pollution pressure modelled for all participants at 1 minute intervals at the address level

More phenotypes!



Nasal microbiome

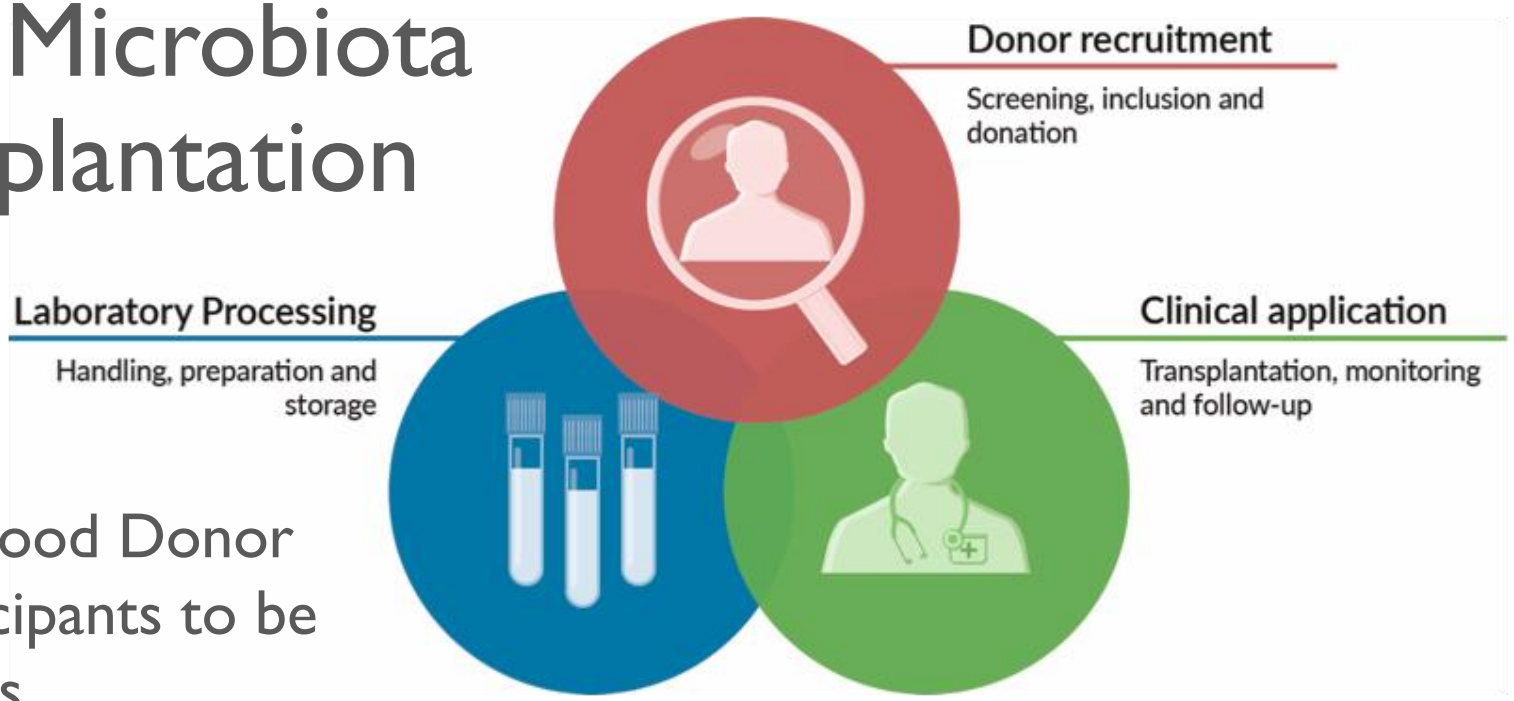
- A nasal swab has been obtained from 2,500 donors; aim 10,000
- Primary aim: to study the associations of *S. aureus* colonisation with morbidity (infections, metabolic disorders, autoimmune diseases)
- Secondary aims:
  Nasal microbiome and donor morbidity
  S. aureus/nasal microbiome and recipient outcome







Fecal Microbiota Transplantation



Including Blood Donor Study participants to be FMT donors

- Only 10% of blood donors declined to be fecal transplantation donors
- Only 4% declined because they found FMT unappealing
- 31 patients have been transplanted

Conclusion:

- Blood banks have the infrastructure for large scale population studies
- The donors want to participate
- We need to collect the phenotypes
- Full advantage of the data: interdisciplinary approach with strong collaborators to help with data analyses
- As a consequence: blood banks will be able to add significantly to new knowledge in the field of precision medicine
- Maybe, deferred donors can be research donors?
- Blood donor studies: a new business opportunity for blood centers

Dept. of Clinical Immunology, Naestved Hospital:

- Ole Birger Pedersen

Dept. of Clinical Immunology, Copenhagen University Hospital:

- Henrik Ullum
- Erik Sørensen
- Lise Wegner Thørner
- Kristoffer Sølvsten Burgdorf
- Andreas Striboldt Rigas

Dept. of Epidemiology Research, Statens Serum Institut:

- Henrik Hjalgrim
- Klaus Rostgaard

Dept. of Clinical Immunology, Aarhus University Hospital:

- Mikkel Steen Petersen
- Susan Mikkelsen
- Khoa Manh Dinh
- Kathrine Kaspersen

Dept. of Clinical Immunology, Aalborg University Hospital:

- Kaspar René Nielsen

The Blood Donors in Denmark

- Poul Erik Herner Petersen

Dept. of Clinical Immunology, Odense University Hospital:

- Helene Pårup

Dept. of Clinical Microbiology, Aarhus University Hospital:

- Lise Erikstrup

Collaborators:

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

- Gustaf Edgren

Blood Systems Research Institute

- Brian Custer



Give blood, save lives, create knowledge

