



# **Scientific output of 13 Swedish research projects**

with available blood samples and  
health data – A bibliometric analysis

# CONTENT

PREFACE	3
---------	---

ACKNOWLEDGEMENTS	4
------------------	---

ABBREVIATIONS	4
---------------	---

METHODS	5
---------	---

SUMMARY – BIBLIOMETRIC ANALYSIS OF ALL PUBLICATIONS	7
--	---

## INDIVIDUAL ANALYSES OF 11 PUBLICATION LISTS

1. Northern Sweden Health and Disease Study (NSHDS)	10
2. LifeGene	11
3. The Swedish Twin Registry (STR)	12
4. The Malmö cohorts: Malmö Preventive Project (MPP), Malmö Diet and Cancer Study (MDC), and Malmö Offspring Study (MOS)	13
5. Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research (SIMPLER)	14
6. Uppsala-Umeå Comprehensive Cancer Consortium (U-CAN)	15
7. Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA)	16
8. Human Glioblastoma Cell Cultures (HGCC)	17
9. The Importance of Migration and Ethnicity for Diabetes Development in Malmö (MEDIM)	18
10. The Environmental Determinants of Diabetes in the Young (TEDDY)	19
11. EIMS & IMSE	20

# APPENDIX

Read online



## SCIVAL REPORTS

- |   |   |
|---|---|
| i. Bibliometric analysis - All publications             | viii. Bibliometric analysis - U-CAN         |
| ii. Bibliometric analysis - NSHDS                       | ix. Bibliometric analysis - KARMA           |
| iii. Bibliometric analysis - NSHDS without EPIC studies | x. Bibliometric analysis - HGCC             |
| iv. Bibliometric analysis - LifeGene                    | xi. Bibliometric analysis - MEDIM           |
| v. Bibliometric analysis - STR                          | xii. Bibliometric analysis - TEDDY          |
| vi. Bibliometric analysis - The Malmö cohorts           | xiii. Bibliometric analysis - EIMS and IMSE |
| vii. Bibliometric analysis - SIMPLER                    |   |

# Preface

This report is a follow-up to the digital brochure *Fifteen Swedish research projects with blood samples and health data available for future research*.

In this report we aim to present a factual basis regarding the scientific output of selected Swedish research projects with sample collections during the period 2016–2020. We have chosen to use SciVal's standard model for bibliometric analysis in order to facilitate comparability with, for example, other national and international cohorts and research infrastructures.

We hope that these data will aid future analysis, discussions, and strategic decisions regarding how to

achieve a sustainable development of research that utilises Swedish cohorts with sample collections.

For questions regarding the report, please contact **Ulrika Morris** ([ulrika.morris@umu.se](mailto:ulrika.morris@umu.se)) or **Ingvar Bergdahl** ([ingvar.bergdahl@umu.se](mailto:ingvar.bergdahl@umu.se)).



<https://biobanksverige.se/digital-brochure-15-swedish-research-projects/>

## Acknowledgements

The bibliometric analyses have been conducted in collaboration with Umeå University Library. We would like to thank Cecilia Sandberg and Cristian Colliander for their assistance in this project.

## Abbreviations

Abbreviation	Explanation
EIMS	Epidemiological Investigation of Risk Factors for Multiple Sclerosis
EPIC	The European Prospective Investigation into Cancer and Nutrition
GWAS	Genome Wide Association Studies
HGCC	Human Glioblastoma Cell Cultures
IMSE	The Immunomodulation and Multiple Sclerosis Epidemiology study
KARMA	Karolinska Mammography Project for Risk Prediction of Breast Cancer
MDC	Malmö Diet Cancer
MEDIM	The Importance of Migration and Ethnicity for Diabetes Development in Malmö
MOS	Malmö Offspring Study
MPP	Malmö Preventive Project
MS	Multiple Sclerosis
NSHDS	Northern Sweden Health and Disease Study
SIMPLER	Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research
STR	Swedish Twin Registry
TEDDY	The Environmental Determinants of Diabetes in the Young
U-CAN	Uppsala-Umeå Comprehensive Cancer Consortium



# Methods

## Study inclusion

Fifteen cohorts included in the previous report *Fifteen Swedish research projects with blood samples and health data available for future research* (see QR code on page 3) were invited to submit publication lists for the five-year period 2016–2020, listing publications where the cohorts' health data and/or samples had been used.

The Malmö cohorts submitted a joint publication list for the Malmö Preventive Project, Malmö Diet Cancer, and the Malmö Offspring Study due to the large overlap in publications between the three cohorts. Two of the more recent studies included in the previous report, SCAPIS and BIG3, declined to take part in the bibliometric analysis since biological samples are not yet available to external researchers in these studies.

” the bibliometric analyses were based on 11 submitted publication lists

As a result, the bibliometric analyses were based on 11 submitted publication lists. Publications published during 2016–2020 that were available in Scopus at the time of data extraction (Date of data extraction: 2022-04-27) were included in the analysis. Published errata were excluded.

## Biobliometric reports in SciVal

A slightly customised version of SciVal's overview report, with the addition of Keyphrase analysis, Publications by Journal quartile, and Most cited publications, was used to generate bibliometric reports for the standard year range 2016–2020 for each publication list. A summary overview report was also generated for a joint list of all 1785 publications.

It is important to note that a few metrics in the overview report are based on Field-Weighted Citation Impact. Field-Weighted Citation Impact in SciVal indicates how the number of citations received by an entity's publications compares with the average number of citations received by all other similar publications (with the same publication year, publication type, and discipline) in the data universe.





The citations received in the year in which an item was published, and the following 3 years, are counted for in this metric.

In the SciVal overview reports the Field-Weighted Citation Impact is presented by year of publication. A number of these bar charts give an impression of a downward trend. This trend is likely to be an artefact produced by calculating the mean Field-Weighted Citation Impact in a relatively small data set that includes a few very highly cited publications. Highly cited publications may skew the Field-Weighted Citation Impact in data sets with a scholarly output less than 1000 publications.

In addition, some of the larger and older cohorts are part of research consortiums that generate a large number of publications. For example, the Northern Sweden Health and Disease Study (NSHDS) and Malmö Diet Cancer (MDC) are both part of The European Prospective Investigation into Cancer and Nutrition (EPIC) study. To assess how being a part of EPIC has impacted publication metrics, a second bibliometric analysis was conducted for NSHDS after publications generated through the EPIC cohort had been removed.

## Manual analysis of collaborations

In the SciVal reports the Scholarly Output by amount of international, national, and institutional collaboration is presented, where institutional collaboration represents any publication where all authors have the same affiliation (regardless of institution).

In order to assess institutional collaboration within and outside the cohort's host institution, raw data regarding

co-authors affiliations (country and institution) were exported from SciVal. The publications were re-assigned a single collaboration type based on the following categories:

1. Within host institution. All co-authors had affiliations to the cohort's host university or university hospital. For example, for cohorts managed by Karolinska Institutet, Karolinska Institutet and Karolinska University Hospital were both regarded as host institution.
2. Within Sweden. All authors had affiliations within Sweden only, but at least one co-author had an affiliation outside of the host institution.
3. Within EU/EES. At least one co-author had an affiliation outside of Sweden but within EU/EES. UK was also included in this category as the majority of publications were published before BREXIT.
4. At least one co-author had an affiliation outside the EU/EES or the UK.

## Manual analysis of last author affiliations

Similarly, raw data on last author affiliations were retrieved from Scopus for each publication list. In total, 38 publications where the last author affiliation was UNKNOWN (i.e., where the author was not linked to an address in Scopus) and two publications that did not contain any author information were excluded.

Each publication was manually assigned to one of the following categories:

1. Last author had an affiliation within the host institution (similarly as above).
2. Last author had no affiliation to the host institution but had an affiliation within Sweden.
3. Last author had no affiliation within Sweden but had an affiliation within EU/EES or UK.
4. Last author only had no affiliation within EU/EES or UK.

The last author affiliation was used as a proxy of the primary origin of the research. However, scoring "last author" as the author with overall responsibility has limitations. Author positions in medical science are increasingly moving away from the traditional first author/last author standard. For example, in large author lists, the first few authors are often responsible and remaining authors (including last author) may be published in alphabetical order. The affiliation of the corresponding author could perhaps have been a more representative proxy, however the data on corresponding author is often incomplete in SCOPUS making it difficult to use.

# Summary – bibliometric analysis of all publications

During the 5-year period 2016–2020, health data and/or samples from the 13 Swedish cohort studies included in this report have been used in 1785 publications. These publications include over 17 000 unique co-authors and have generated more than 78 000 citations in total (Appendix i). Over 25% (N = 458) of the publications are considered to be among the top 10% most cited publications in the world.

The publications are results of national and international studies of different sizes, often written in collaboration with some of the world's leading research institutes (Figure 1A, Table 1). Slightly more than half of the studies had a Swedish last author (Figure 1B).

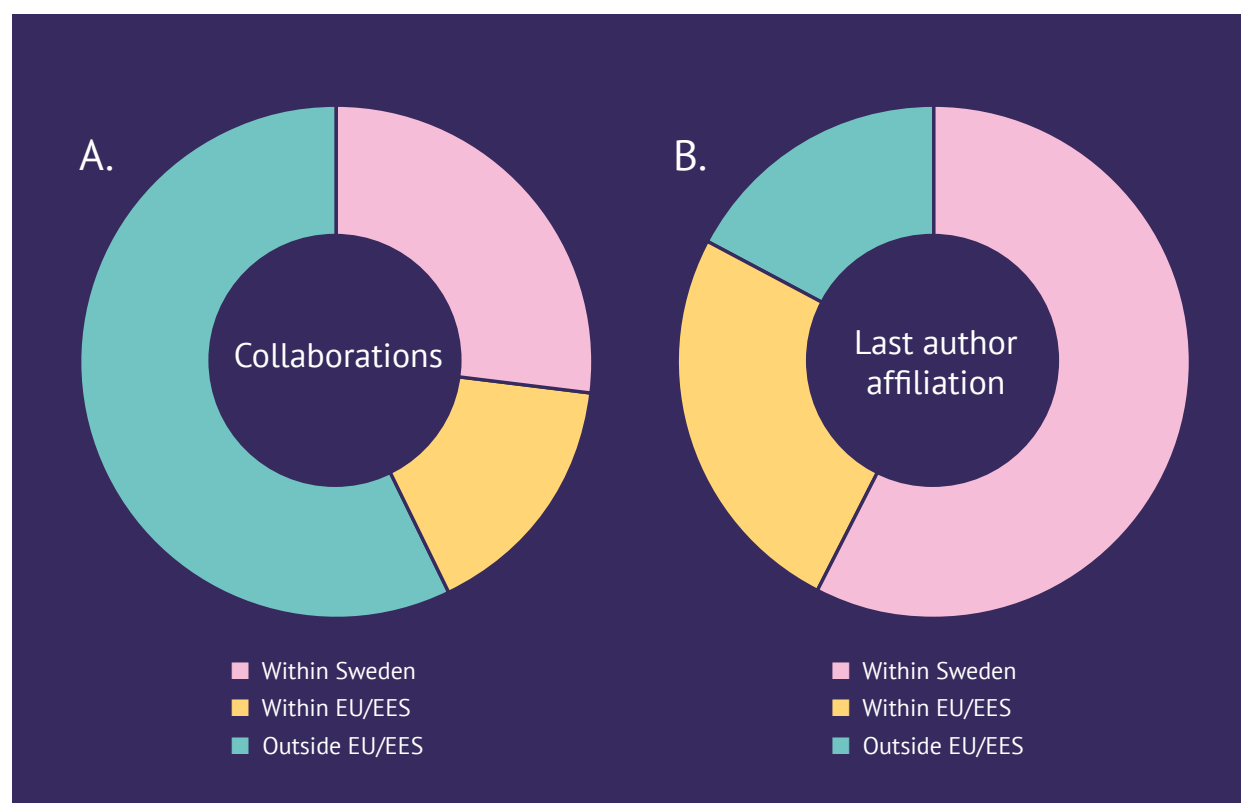


Figure 1. Pie charts showing collaborations and last author affiliation. A) Proportion of publications with co-authors from within Sweden only, from within EU/EES and UK, or including co-authors from outside of the EU/EES and UK. B) Proportion of publications where the last author had an affiliation within Sweden, within EU/EES and UK, or outside of the EU/EES and UK, as a proxy of the principle origin of the conducted research. Data source: Scopus. Year Range: 2016–2020.



Institution	Scholarly Output	Citations per Publication
Karolinska Institutet	1,052	56.1
Lund University	715	51.2
Umeå University	494	52.2
Harvard University	404	115.9
Uppsala University	352	72.3
Imperial College London	344	109.4
University of Oxford	337	119.7
National Institutes of Health	318	98.2
Inserm	308	88.0
University of Cambridge	308	117.5

Table 1. The institutions with the highest scholarly output in the set of 1785 publications, along with the mean number of citations per publication. Data source: Scopus. Year Range: 2016–2020.

The cohorts' scholarly output is correlated to the cohort size (Figure 2). The larger cohorts are typically older and have produced more publications during the five-year period.

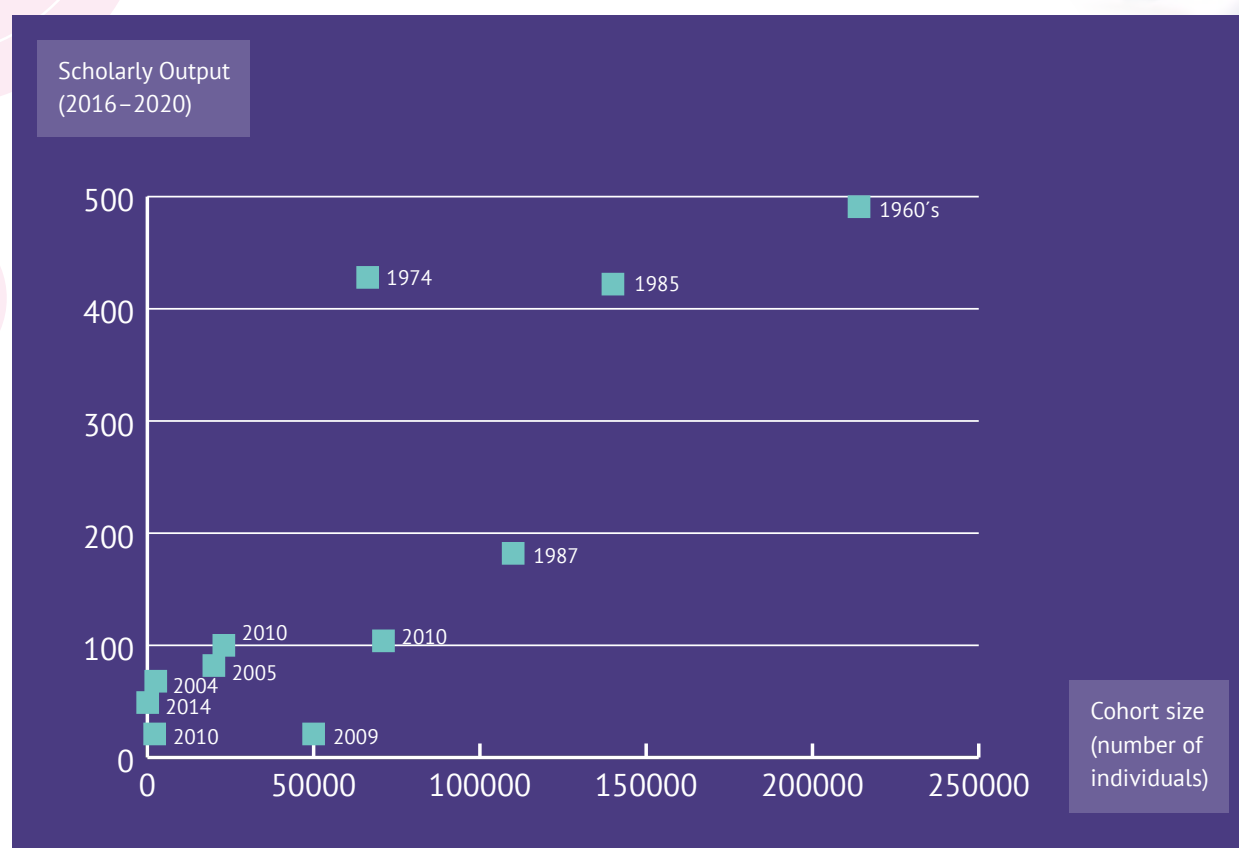


Figure 2. Scatter plot of the scholarly output (number of publications during 2016–2020) over cohort size. Data labels indicate the year the study cohort started.



The pairwise matrix (Table 2) shows co-occurrences of publications between different cohorts. The diagonal shows the number of publications in Scopus for each given cohort. The large number of co-occurring publications between NSHDS and the Malmö Cohorts is largely due to both being part of the EPIC cohort.

Few studies have combined data or samples from more than one cohort. Among the 1785 publications, 166 publications co-occurred in two publication lists (of which 72 were EPIC studies), seven publications co-occurred in three publication lists, and one publication co-occurred in four publication lists. The eight publications that occurred in more than two publication lists were predominantly Genome Wide Association Studies (GWAS) based on GWAS data from the larger cohorts (NSHDS, STR, the Malmö Cohorts, and SIMPLER).

To assess how being part of large consortiums such as EPIC has impacted publication metrics, a second bibliometric analysis for NSHDS was conducted after publications generated through the EPIC cohort had been removed (Appendix ii and Appendix iii). The main impact was on the number of publications. The impact on other SciVal metrics was small.

After noting a downward trend in the Field-Weighted Citation Impact in the summary report and some of the larger publication lists (NSHDS, STR, EIMS/IMSE), we looked at the top 20 publications with the highest Field-Weighted Citation Impact. After tabulating the publication year and to which cohort these publications belonged, it was clear that these highly cited outliers had skewed the trends over time through their individual impact on specific years (Table 3).

	NSHDS	LifeGene	STR	Malmö cohorts	SIMPLER	U-CAN	KARMA	HGCC	MEDIM	TEDDY	EIMS & IMSE
NSHDS	422										
LifeGene		21									
STR	16	3	491								
Malmö cohorts	93		9	428							
SIMPLER	26		2	2	182						
U-CAN	4					100					
KARMA	3	1	2		12		104				
HGCC						10		49			
MEDIM				2					21		
TEDDY										68	
EIMS & IMSE		2	5				1				82

Table 2. Co-publications between cohorts. A pairwise matrix showing the number of co-occurring publications between two cohorts. The diagonal shows the number of publications in Scopus for each given cohort.

Rank	Field-Weighted Citation Impact	Publication Year	Cohort	Rank	Field-Weighted Citation Impact	Publication Year	Cohort
1	281	2016	NSHDS	11	57	2020	U-CAN
2	261	2017	NSHDS + Malmö	12	47	2019	STR
3	248	2016	STR	13	46	2017	U-CAN
4	218	2016	Malmö	14	44	2019	STR
5	193	2020	Malmö	15	44	2019	Malmö
6	110	2017	Malmö	16	40	2018	STR
7	83	2017	Malmö	17	39	2018	EIMS/IMSE
8	64	2018	NSHDS + Malmö	18	37	2016	Malmö
9	59	2016	Malmö	19	36	2016	STR
10	59	2018	STR	20	33	2016	SIMPLER

Table 3. Top 20 publications with highest Field-Weighted Citation Impact.



NSHDS is a large cohort with over 250 000 sample collection occasions. Health data and/or samples have primarily been used in prospective cohort-, case-control-, and/or genetic studies in cancer and other major noncommunicable diseases such as diabetes. NSHDS is part of several large international consortia, including EPIC, through which NSHDS samples and dietary data have been used in a large number of publications (Appendix ii and iii) (see QR code on page 2). This may have contributed to the large number of last authors who have affiliations within the EU/EES but not in Sweden.

Study design:	Prospective cohort –general population
Recruitment:	1986-ongoing
Cohort size:	140 000
Host institution:	Umeå University
Scholarly Output (2016–2020):	422
Number of citations:	19 334
Number of co-authors:	6555
Citations per publication:	45.8
Field-Weighted Citation Impact:	3.74
Publications in top 10% most cited:	120 (28.4%)



Data source: Scopus, up to 27 Apr 2022

A A A relevance of keyphrase  
declining A A A growing

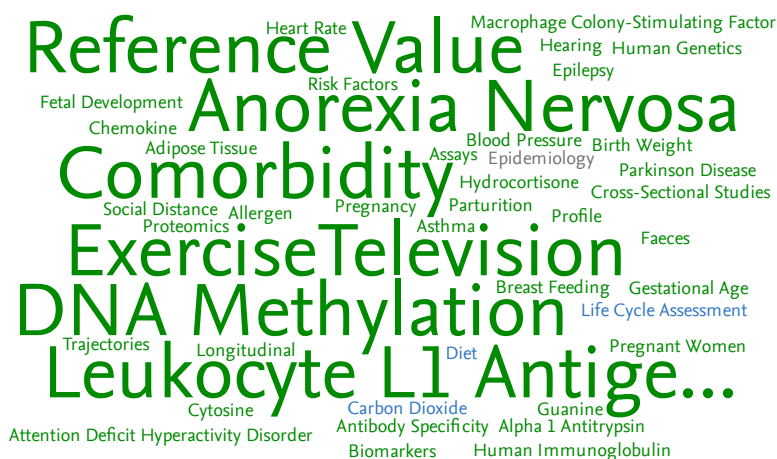


# LifeGene

LifeGene is a newer cohort with 30 000 sample collection occasions. The lower scholarly output may be explained by LifeGene being a more recent study. Health data and/or samples have primarily been used in epidemiological studies investigating associations between exposure and outcome of somatic and mental health. The last author affiliations show that a large proportion of the studies were led by Karolinska Institutet (host institution), and the other six Swedish universities with medical faculties that stand behind LifeGene.

## Summary metrics

Study design:	Prospective cohort –general population
Recruitment:	2009–2019
Cohort size:	50 000
Host institution:	Karolinska Institutet
Scholarly Output (2016–2020):	21
Number of citations:	467
Number of co-authors:	414
Citations per publication:	22.2
Field-Weighted Citation Impact:	2.12
Publications in top 10% most cited:	3 (14.3%)



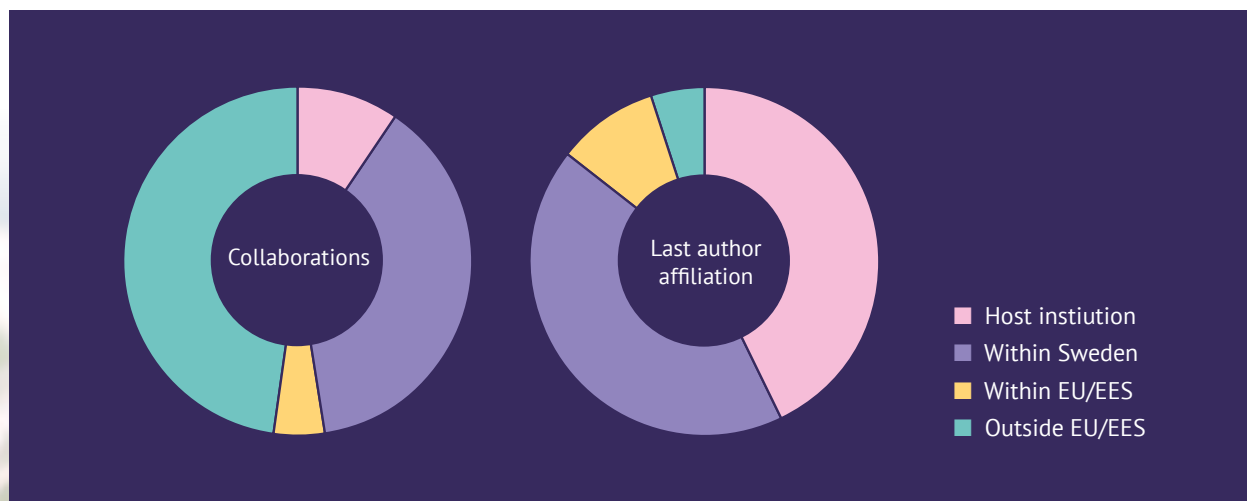
## Keyphrase analysis

Entity: LifeGene

Year range: 2016 to 2020

Data source: Scopus, up to 27 Apr 2022

AA relevance of keyphrase  
declining AA growing



STR is a large cohort with DNA from 60 000 twins and serum from 12 600 twins. Health data and/or samples have primarily been used in twin- and genome wide association studies. The twin design has, for example, been used to assess confounding genetics and the impact of shared environmental conditions when growing up. Two thirds of the studies have been conducted with collaborators outside of the EU/EES. Approximately half the publications have a last author affiliated to Karolinska Institutet.

Study design:	Prospective cohort – Twins
Recruitment:	1960's–ongoing
Cohort size:	107 000 sets of twins
Host institution:	Karolinska Institutet
Scholarly Output (2016–2020):	491
Number of citations:	26 957
Number of co-authors:	6203
Citations per publication:	54.9
Field-Weighted Citation Impact:	3.48
Publications in top 10% most cited:	137 (27.9%)



Data source: Scopus, up to 27 Apr 2022

A A A relevance of keyphrase  
declining A A A growing



Malmö Preventive Project (MPP), Malmö Diet and Cancer Study (MDC), and Malmö Offspring Study (MOS)

## Summary metrics

A word cloud visualization of research topics related to the Malmö study. The words are arranged in a circular pattern, with 'Malmö' being the largest and most central word. Other prominent words include 'Genome-Wide Association Study', 'Cancer Risk', 'Sweden', 'Diet Study Techniques', 'Ischaemic Heart Disease', 'Cardiovascular Risk Factors', 'Non Insulin Dependent Diabetes Mellitus', 'Abdominal Aortic Aneurysm', 'Case-Control Studies', 'Genetic Predisposition', 'Single Nucleotide Polymorphism', 'Lifestyle', 'Epic', 'Carotid Artery Diseases', 'Genetic Risk Score', 'Prospective', 'Blood Pressure', 'Pulse Wave Analysis', 'Arterial Stiffness', 'Nutrition', 'Coronary Events', 'Meta-Analysis', 'Human Life', 'Urinary Bladder Neoplasms', 'Chronic Kidney Failure', 'Obesity', 'Model Risk Cohort Studies', 'Hazard Ratio', 'Carotid Intima-Media Thickness', 'Coronary Artery Atherosclerosis', 'Body Weight Change', 'Pancreatic Neoplasms', 'Prostatic Neoplasms', 'Carotid Arteries', 'Brain Ischemia', 'Genetic Loci', 'Incidents', 'Atrial Fibrillation', 'Proteomics', 'Diabetes Mellitus', 'Vitamin D', 'Cadmium', 'Genetic Risk', 'Cardiovascular Disease', 'Diet', 'Genetic Predisposition', 'Case-Control Studies', 'Abdominal Aortic Aneurysm', and 'Non Insulin Dependent Diabetes Mellitus'.

Entity: Malmökohorterna  
Year range: 2016 to 2020  
Data source: Scopus, up to 27 Apr 2022

The figure consists of two donut charts. The left chart, titled 'Collaborations', shows a large teal segment (Outside EU/EES) and a large pink segment (Host institution). The right chart, titled 'Last author affiliation', shows a large pink segment (Host institution) and a large yellow segment (Within EU/EES). A legend on the right identifies the colors: pink for Host institution, purple for Within Sweden, yellow for Within EU/EES, and teal for Outside EU/EES.

Category	Host institution	Within Sweden	Within EU/EES	Outside EU/EES
Collaborations	~25%	~5%	~25%	~45%
Last author affiliation	~55%	~15%	~25%	~5%



SIMPLER comprises in total around 50 000 DNA samples with additional biological samples available from 12 400 individuals. Health data and/or samples have primarily been used in cohort- and genetic studies in cancer and other major noncommunicable diseases with a focus on dietary intake. Almost two thirds of the studies have been conducted with collaborators outside of the EU/ EES. Around half of the publications have a last author affiliated to Uppsala University or Karolinska Institutet.

Study design:	Prospective cohort –general population
Recruitment:	1987–1990 + 1997–1998
Cohort size:	110 000
Host institution:	Uppsala University, previously Karolinska Institutet
Scholarly Output (2016–2020):	182*
Number of citations:	6456
Number of co-authors:	3049
Citations per publication:	35.5
Field-Weighted Citation Impact:	2.61
Publications in top 10% most cited:	56 (30.8%)

[illegible]

Entity: SIMPLER  
Year range: 2016 to 2020  
Data source: Scopus, up to 27 Apr 2022

13

U-CAN is a biobank resource actively collecting samples and data from cancer patients in a longitudinal manner (i.e., repeated samples from the same patient during the course of the disease). As of 2022, over 24 000 individuals have consented to participate in U-CAN in Uppsala and Umeå. Health data and/or samples have been used in a wide range of scientific questions related to multiple forms of cancer including studies regarding tumour biomarkers, tumour biology, genetic variations, and epigenetics. Almost two thirds of the publications have a last author affiliated to one of the two host institutions, and almost half the studies have been conducted with international collaborators.

Study design:	Cancer patients
Recruitment:	2010–ongoing
Cohort size:	24 000
Host institutions:	Uppsala University and Umeå University
Scholarly Output (2016–2020):	100*
Number of citations:	3730
Number of co-authors:	807
Citations per publication:	37.3
Field-Weighted Citation Impact:	2.45
Publications in top 10% most cited:	20 (20.0%)

[illegible]

Entity: U-CAN  
Year range: 2016 to 2020  
Data source: Scopus, up to 27 Apr 2022

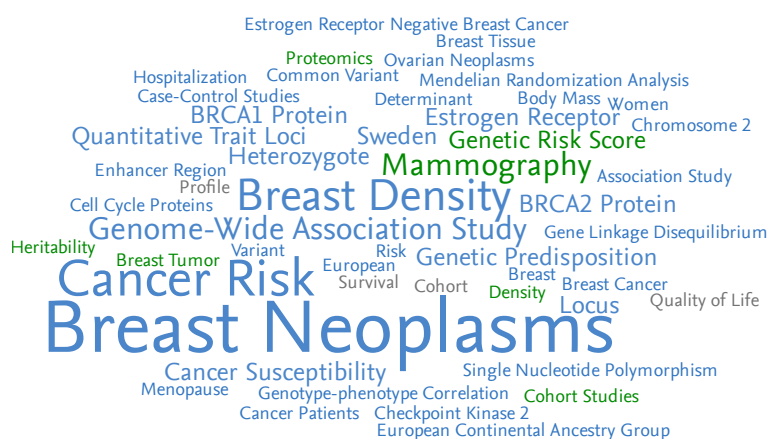
14

# Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA)

KARMA links a comprehensive information and image database with blood samples from all participants. Around 25% of participants have provided repeated samples. Samples and data have primarily been used in case control- and genetic studies for studying factors that influence the risk or prognosis for breast cancer. KARMA is part of the Breast Cancer Association Consortium, which may have contributed to the large proportion of international collaborations. Almost half the studies have a last author affiliated to Karolinska Institutet or Södersjukhuset Stockholm.

## Summary metrics

Study design:	Prospective cohort – females attending mammography screening
Recruitment:	primarily 2010–2013
Cohort size:	71 000
Host institutions:	Karolinska Institutet and Södersjukhuset Stockholm
Scholarly Output (2016-2020):	104
Number of citations:	3785
Number of co-authors:	2763
Citations per publication:	36.4
Field-Weighted Citation Impact:	2.19
Publications in top 10% most cited:	26 (25.0%)



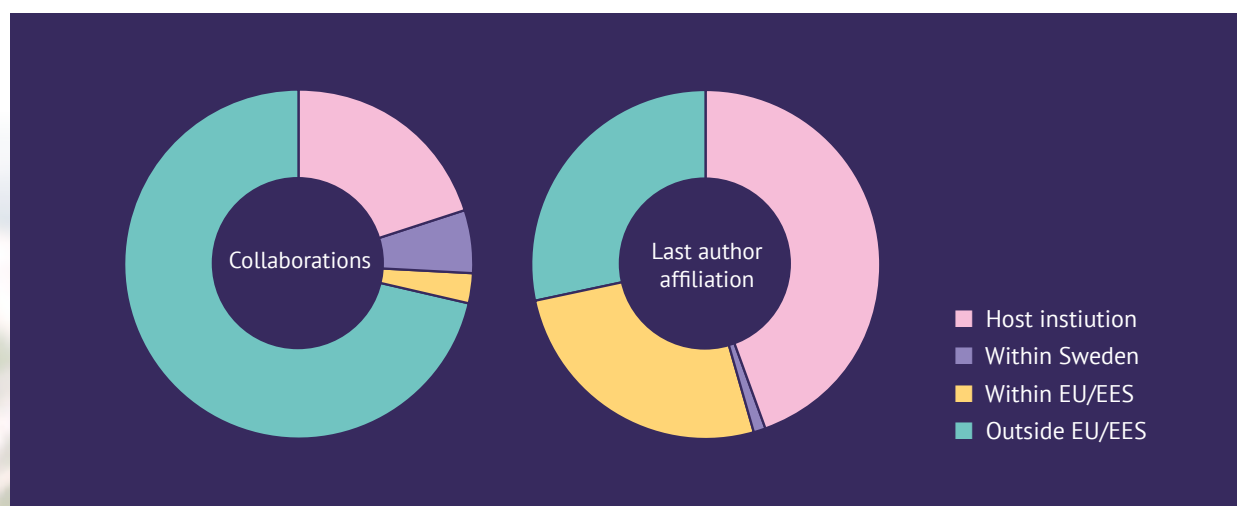
## Keyphrase analysis

Entity: KARMA

Year range: 2016 to 2020

Data source: Scopus, up to 27 Apr 2022

AA relevance of keyphrase  
declining AA growing



HGCC is a collection of 100 well-characterised cell lines derived from surgical samples from patients with glioblastoma. The cell lines have been used in research requiring Glioblastoma stem cell cultures to answer research questions related to basic biology and/or drug development. The cell lines have been used in collaboration with 50 laboratories in 17 countries, which may have contributed to an even distribution of national collaborations, international collaborations within EU/EES, and international collaborations outside of EU/EES. Around 40% of the publications had a last author affiliated to the host institution.

Study design:	Cell lines derived from surgical samples from patients with glioblastoma
Recruitment:	Not applicable
Cohort size:	100 cell lines
Host institution:	Uppsala University
Scholarly Output (2016-2020):	49
Number of citations:	896
Number of co-authors:	403
Citations per publication:	18.3
Field-Weighted Citation Impact:	1.31
Publications in top 10% most cited:	7 (14.3%)



Data source: Scopus, up to 27 Apr 2022

A A A relevance of keyphrase  
declining A A A growing



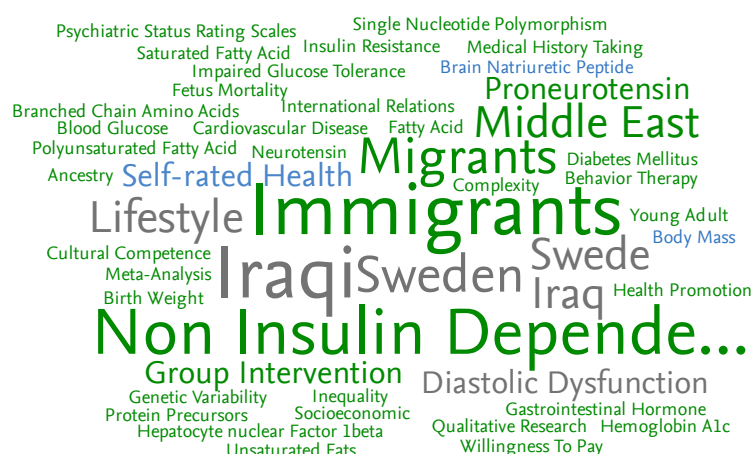


# The Importance of Migration and Ethnicity for Diabetes Development in Malmö (MEDIM)

MEDIM is a small cohort with blood samples and cross-sectional data collected from all 2220 participants. The samples and data have primarily been used in locally led research projects, as evidenced by the majority of last authors in the 21 publications being affiliated to Lund university or Skåne University Hospital. The studies have looked at country of birth and risk of type 2 diabetes, including analysis of metabolic biomarkers and mechanisms of glucose regulation.

## Summary metrics

Study design:	Prospective cohort – adults born in Iraq or Sweden and living in Malmö
Recruitment:	2010–2012
Cohort size:	2220
Host institution:	Lund University
Scholarly Output (2016-2020):	21
Number of citations:	150
Number of co-authors:	56
Citations per publication:	7.1
Field-Weighted Citation Impact:	0.73
Publications in top 10% most cited:	0 (0.0%)



## Keyphrase analysis

Entity: MEDIM

Year range: 2016 to 2020

Data source: Scopus, up to 27 Apr 2022

▲ ▲ ▲ relevance of keyphrase  
declining ▲ ▲ ▲ growing

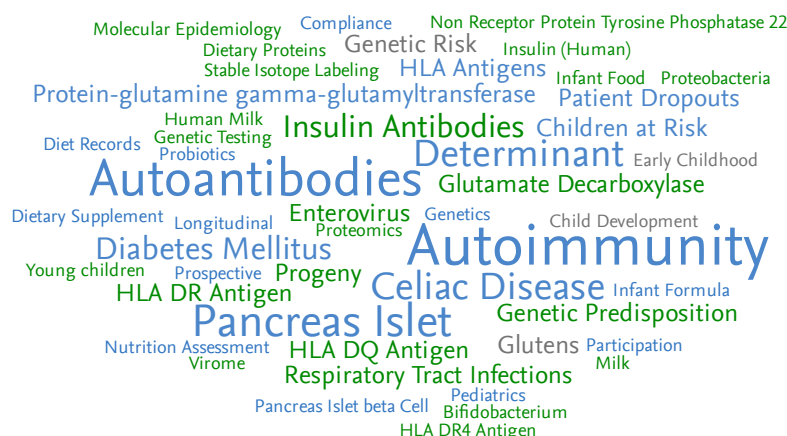


# The Environmental Determinants of Diabetes in the Young (TEDDY)

Teddy is an international cohort including children with increased hereditary risk of autoimmune diabetes in Skåne, Finland, Germany, and USA. Children are followed to the age of 15 with multiple sample collection occasions each year. Samples and data, including screening-, baseline-, study phase clinical data, and multiple analysis datasets, are made available through the National Institutes of Health in the United States. This may contribute to the lack of national collaborations within Sweden, and the majority of studies including collaborators in the United States. Studies have mainly focused on autoimmune disease in young children including the impact of dietary intake.

## Summary metrics

Study design:	Prospective cohort – children with increased hereditary risk of autoimmune diabetes
Recruitment:	2004–2010
Cohort size:	8667 of which 2528 in Sweden
Host institution:	Lund University
Scholarly Output (2016–2020):	68
Number of citations:	2485
Number of co-authors:	585
Citations per publication:	36.5
Field-Weighted Citation Impact:	2.78
Publications in top 10% most cited:	19 (27.9%)



## Keyphrase analysis

Entity: TEDDY study

Year range: 2016 to 2020

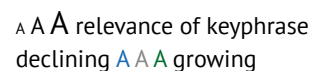
Data source: Scopus, up to 27 Apr 2022

A A A relevance of keyphrase  
declining A A A growing



Epidemiological Investigation of Risk Factors for Multiple Sclerosis (EIMS) and The Immunomodulation and Multiple Sclerosis Epidemiology study (IMSE) are a case-control study and a collection of clinical trials in Multiple Sclerosis (MS), respectively. Samples have primarily been used in genetic studies, for example in assessing genotype-environment interactions. Around two thirds of the studies have a last author affiliated to the host institute, with around 60% of the studies having been conducted with collaborators outside of the EU/EES.

Study design:	Case-control and clinical trials in MS
Recruitment:	2005–ongoing
Cohort size:	12 000 MS patients and 8000 controls
Host institution:	Karolinska Institutet
Scholarly Output (2016-2020):	82
Number of citations:	5278
Number of co-authors:	2272
Citations per publication:	64.4
Field-Weighted Citation Impact:	4.83
Publications in top 10% most cited:	35 (42.7%)





This document is part of our ambition to promote cohorts with sample collections in Sweden. We welcome all suggestions and feedback for future revisions of this document. Contact information: Ulrika Morris ([ulrika.morris@umu.se](mailto:ulrika.morris@umu.se)), Ingvar Bergdahl ([ingvar.bergdahl@umu.se](mailto:ingvar.bergdahl@umu.se)), and Alexander Hertzberg ([alexander.hertzberg@regionstockholm.se](mailto:alexander.hertzberg@regionstockholm.se)).