

Data management
at NTNU

Introduction to data management resources

OPEN
SCIENCE
NTNU

Ane Møller Gabrielsen
NTNU University Library

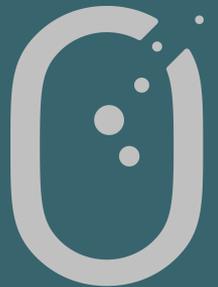


NORWEGIAN UNIVERSITY OF
SCIENCE AND TECHNOLOGY

Open Science at NTNU

Open Science at NTNU

- New policy for Open Science 2020
 - Previously: Policy for Open Access (2014) and policy for Open Research Data (2018)
- NTNU University Library point of contact and coordination for Open Science at NTNU
 - Support services, tools, and guidance
 - Collaboration with IT and departments /faculties



OPEN
SCIENCE
NTNU

Research data at NTNU:

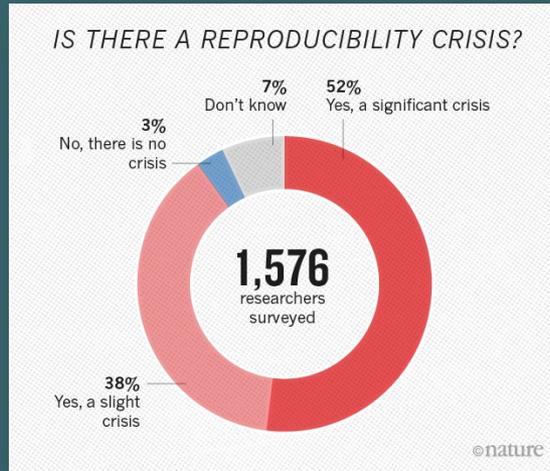
- As open as possible, as closed as necessary

- Data should be:



- NTNU guarantees for access to infrastructure and support services
- The researcher is responsible for the quality
- All projects should have a DMP

Transparency ...



1,500 scientists lift the lid on reproducibility
Survey sheds light on the 'crisis' rocking research.
Monya Baker

Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a

Miyakawa *Molecular Brain* (2020) 13:24
<https://doi.org/10.1186/s13041-020-0552-2>

Molecular Brain

EDITORIAL

Open Access

No raw data, no science: another possible source of the reproducibility crisis



Tsuyoshi Miyakawa

Abstract

A reproducibility crisis is a situation where many scientific studies cannot be reproduced. Inappropriate practices of science, such as HARKing, p-hacking, and selective reporting of positive results, have been suggested as causes of irreproducibility. In this editorial, I propose that a lack of raw data or data fabrication is another possible cause of irreproducibility.

As an Editor-in-Chief of *Molecular Brain*, I have handled 180 manuscripts since early 2017 and have made 41 editorial decisions categorized as "Revise before review," requesting that the authors provide raw data. Surprisingly, among those 41 manuscripts, 21 were withdrawn without providing raw data, indicating that requiring raw data drove away more than half of the manuscripts. I rejected 19 out of the remaining 20 manuscripts because of insufficient raw data. Thus, more than 97% of the 41 manuscripts did not present the raw data supporting their results when requested by an editor, suggesting a possibility that the raw data did not exist from the beginning, at least in some portions of these cases.

Considering that any scientific study should be based on raw data, and that data storage space should no longer be a challenge, journals, in principle, should try to have their authors publicize raw data in a public database or journal site upon the publication of the paper to increase reproducibility of the published results and to increase public trust in science.

Keywords: Raw data, Data fabrication, Open data, Open science, Misconduct, Reproducibility

... and reuse!

Where to find data?

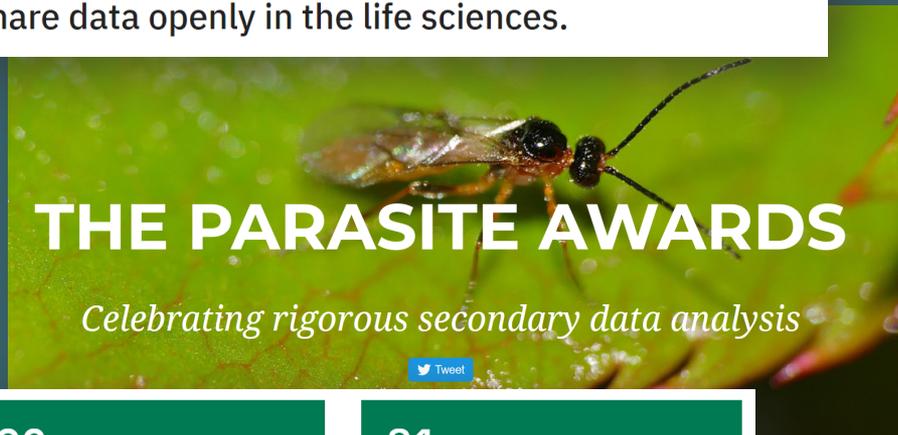
- BASE (Bielefeld Academic Search Engine)
 - www.base-search.net
- DataCite
 - search.datacite.org
- Elsevier Data Search
 - datasearch.elsevier.com
- Google Dataset Search
 - <https://datasetsearch.research.google.com/>

Registry of Research Data Repositories

- re3data.org

EMBL-EBI: Our impact

We collaborate with scientists and engineers all over the world, and provide the infrastructure needed to share data openly in the life sciences.



THE PARASITE AWARDS

Celebrating rigorous secondary data analysis

[Tweet](#)

**390
Petabytes**

of raw data storage



**81
million**

web requests made to our
websites on an average
day



NTNU Storage and Computing Resources

Storage: Information Security

- All information (including research data) should be classified, to help choose the correct storage
 - Open, Internal, Confidential, Highly confidential

Information classification

All information processed at NTNU must be classified according to confidentiality requirements, so that you know where and how the information can be processed. You can find more information about information classification on the page [Informasjonsklassifisering - informasjonssikkerhet](#). (Norwegian only)

The following classes are defined:



Confidentiality: How important is it that the information does not fall into the wrong hands?

Public

Internal

Confidential

Highly Confidential

Integrity: How important is it that the information is not modified by unauthorised persons or by accident?

Accessibility: How critical is it to lose access to the information for a period of time, or to lose it completely? (Example: If you store your data on M:, no one can access it if you quit.)

Data storage guide

This guide will help you select the correct solution for storage and processing of information, including research data. The storage guide gives an overview of the storage solutions assessed by NTNU, provides information about designated usage, and tells you how to access the different solutions. In addition, the [data collection guide](#) and the storage guide provides information that can be used when creating [data management plans](#).

Norsk versjon - [Lagringsguide](#)

[Topic page about information security](#) | [Pages labeled with information security](#)

Information security and classification

NTNU guidelines state that information must be classified in order to determine the value and identify the need for security and protection.

Public

Internal

Confidential

Highly Confidential

Classification of research data

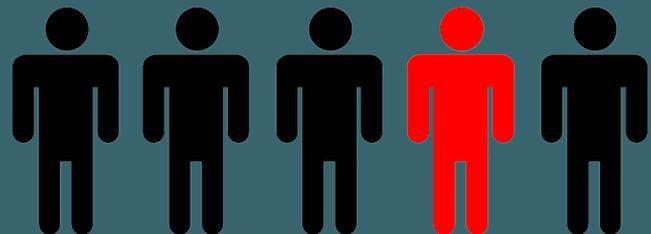
- Most research data will be classified as internal, but some types require more protection
- Examples:
 - Trade secrets
 - Data for commercial purposes, patents etc
 - *Personal data*

Personal data:

Any information that identifies a physical person, directly or indirectly

Examples:

- Name, ID-number, address, telephone number etc.
- IP-address, location information
- Images, sound, video, email, voice ...
- Background information that can be traced back to an individual, (e.g., institutional affiliations combined with information about age, sex, occupation, nationality, and so forth)



Special categories of personal data (sensitive):

- Race or ethnic origin
- Political opinions, religion, philosophical beliefs, union membership
- Sexual orientation and activity
- Criminal offence data
- Genetic and biometric data used for identification purposes
- Health data

What physical storage media can I use?

Physical storage media refers to local storage and processing of information, for instance your computer (Mac, PC or hard drive).

Information classification:	Public	Internal	Confidential	Highly confidential
Privately owned laptop	OK	NO	NO	NO
Privately owned desktop	OK	NO	NO	NO
NTNU-acquired desktop (self-administered)	OK	OK	NO	NO
NTNU-acquired laptop (self-administered)	OK	OK	NO	NO
NTNU-administered desktop – encrypted	OK	OK	OK	NO
NTNU-administered laptop – encrypted	OK	OK	OK	NO
USB drive/external hard drive	OK	OK	NO	NO
USB drive/external hard drive - encrypted	OK	OK	OK(1)	OK(2)

(1) The data must be stored in encrypted form on the storage media and the password kept in a separate location. [Read more about how to encrypt files.](#)

(2) The entire drive/disc must be encrypted with a strong password ([read more on how to make passwords](#)). The password must be kept in a separate location.

<https://innsida.ntnu.no/wiki/-/wiki/English/Data+storage+guide>



Storage services and collaboration platforms

Storage services and collaboration platforms refer to cloud services or servers at NTNU. Click on the different solutions for more information.

Information classification:	Public	Internal	Confidential	Highly confidential
Personal cloud storage (dropbox, google drive ++)	OK	NO	NO	NO
NTNU Home directory («M:-drive»)	OK	OK	OK	OK (1)
NTNU Shared directory (T-drive, group, project, etc.)	OK	OK	NO	NO
NTNU-administered Dropbox (contact Orakel)	OK	OK	NO	NO
NTNU-Box	OK	OK	NO	NO
Office 365 (SharePoint, Teams, Onedrive)	OK	OK	OK(1)	NO
NTNU NICE-1 - Storage solution with added security	OK	OK	OK	OK (1)
HUNT Cloud	OK	OK	OK	OK (2)
Uio TSD	OK	OK	OK	OK
NIRD (tidligere Norstore, driftes av Uninett Sigma2)	OK	OK	NO	NO

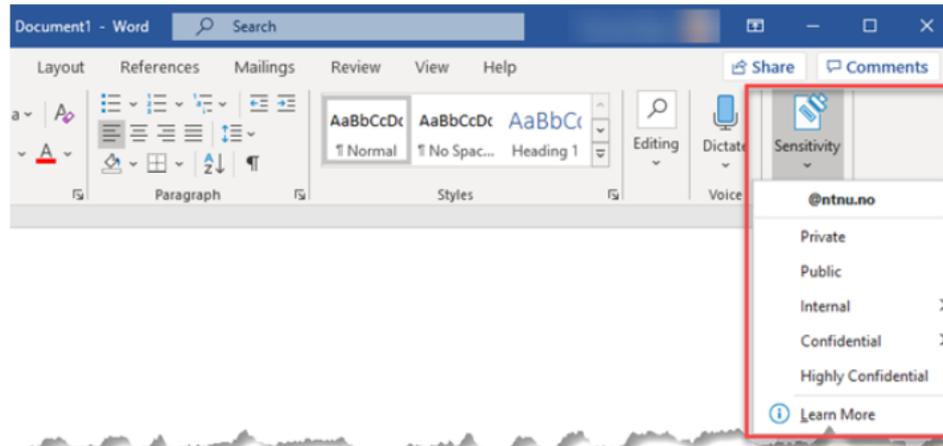
(1) Data must be encrypted. [Read more on how to encrypt O365 files using AIP here](#) or [how to encrypt other files with 7-Zip](#)

(2) Risk level is assessed on individual basis, see the [HUNT information page](#) for more information.

Encryption through AIP (Office 365)

Labels in Word, PowerPoint, Excel and Outlook

In Office 365, it is possible to determine the classification of files and documents, through the use of "labels" that mark the documents and which can trigger technical measures such as encryption. We use Azure Information Protection (AIP) to enable this feature. Once AIP is installed, you will find the classes as labels in the desktop version of Microsoft Office 365.



TSD: Secure environment for confidential and highly confidential data

TSD - Tjenester for sensitive data

På denne siden finner du grunnleggende informasjon om Tjenester for sensitive data (TSD), samt lenker til utfyllende informasjon om tjenesten og hvordan du kan ta den i bruk.

English version - TSD - Services for sensitive data

[Temaside om lagring](#) | [Sider merket med tsd](#)

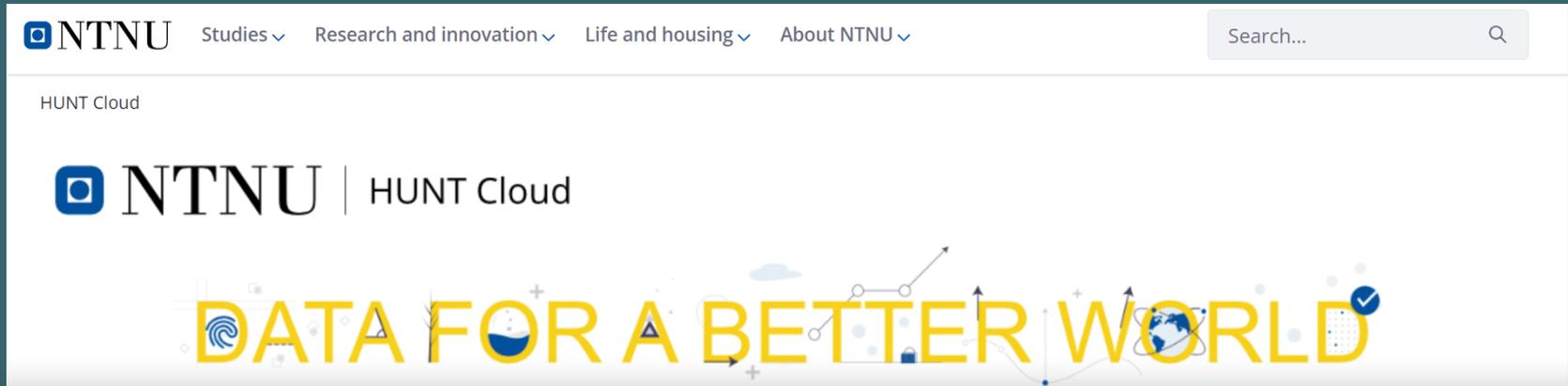
Om TSD

TSD er en trygg, privat sky med et fullt sett tjenester for innsamling, lagring og analyse av sensitive forskningsdata som krever et høyt sikkerhetsnivå. TSD utvikles og driftes av USIT ved UiO, og brukes av forskere ved flere nasjonale forskningsinstitusjoner, inkludert NTNU.

- sikkert prosjektområde
- integrert skjemaløsning for å samle inn sensitive data ([Nettskjema](#))
- tilgang fra hvor som helst i verden
- tilrettelagt for samarbeid med eksterne
- oppfyller lovens strenge krav til behandling og lagring av sensitive forskningsdata
- den eneste skybaserte løsningen som NTNU IT kan anbefale for strengt fortrolige data

- NTNU has an agreement with UiO for TSD usage
- Order project room from Innsida:
https://i.ntnu.no/wiki/-/wiki/No_rsk/TSD+-+registrere+nytt+prosjekt

HUNT Cloud: Secure services for confidential data (health data, genomics etc)



Services

HUNT Cloud provides services for digital lab management. This includes data value management services and scientific cloud services. The cloud services includes solutions for data storage, data computation and data transfers within a high-trust framework. Services are tailored towards large-scale sensitive data.

- [eBook with services overview](#)
- [Detailed services specifications](#)
- [HUNT Cloud user documentation](#)



RSE

Research
Software
Engineering

LAB

Laboratory
Instrumentation
and Support



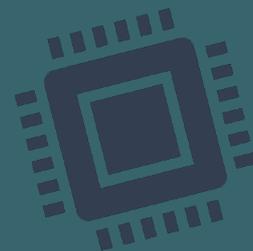
NTNU IT –
Research Support

Mime

IT support for
PhD students

HPC

High
Performance
Computing



hpc.ntnu.no/display/hpc/NTNU+HPC+GROUP

Confluence Områder Personer

HPC Wiki

- Sider
- Blogg

UNDERORDNEDE SIDER

- Sider
 - NTNU HPC GROUP
 - Vilje
 - Idun Cluster
 - Maur
 - Matlab for HPC
 - Python Numpy Scipy and Odespy
 - Training
 - Kongull (Shut down 2016)
 - HPC - Tre tiår med tungregning i ...
 - ansys fluent for hpc
 - Mimes Brønn
 - Parallel R for HPC system
 - Parallel Python

Dashbord

NTNU HPC GROUP

Seksjon for Vitenskapelig Databehandling

NEWS

Blogginnlegg

- Introduction courses to Parallel Programming and Jupyter Notebooks in January/February 2020 opprettet av John Floan
- Vilje /work unavailable due to disk error opprettet av John Floan
- Digital Lab Seminar will be held at NTNU opprettet av Tufan Arslan
- Idun is down for maintenance opprettet av Einar Næss Jensen
- Betzy is the name opprettet av Einar Næss Jensen

Support

System	Email
Fram/Saga/Vilje/NIRD	support@metacenter.no
Idun/Epic	help@hpc.ntnu.no
Maur	support-kongull@hpc.ntnu.no

twitter: <http://twitter.com/ntnuhpc>
facebook: <http://www.facebook.com/ntnuhpc>
Innsida-kanal: HPC at NTNU

Physical meetup:

- Sluppen: Sluppenveien 14, 4th floor, IT-drift.
- Gloschaugen: Mimes Brønn 2-105A (2.etasje)
Materialtekniske laboratorier, Trondheim. (Entrance from the west. Realifagsbygget in your back)

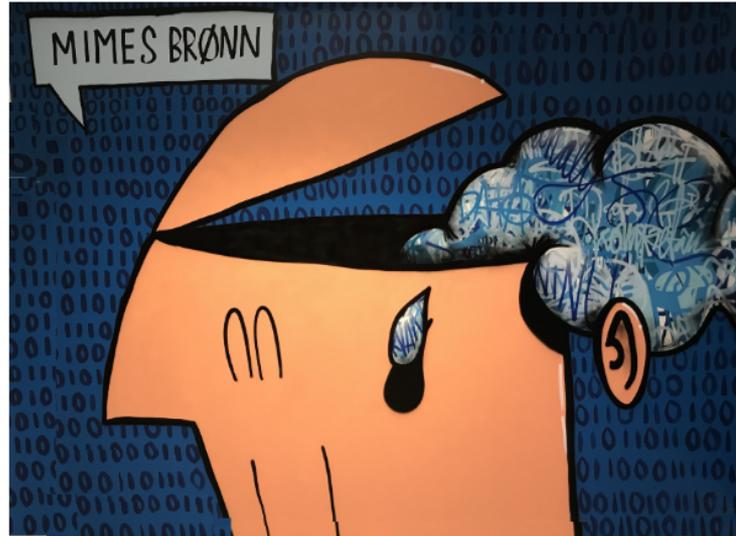


IT-support for PhD students - Mimes Brønn

«Mimes Brønn» (or in English *Mimir's well*) is a meeting place for **NTNU PhD students and IT professionals**.

The objective of Mimes Brønn is to **help PhD students** get started with their research as quickly as possible. By facilitating a **common physical location** for periods of one to four weeks, the IT department can help scientists **get started with their work faster**, reducing the time spent navigating the technological jungle.

In Norwegian: [Mimes Brønn - IT-støtte for ph.d.-er](#)



Mime

IT support for
PhD students

Get started

[Write to orakel@ntnu.no](mailto:orakel@ntnu.no) if you are interested in more information.

Key contacts: [Anders Christensen](#) and [Stein Stendahl](#)

NTNU Open Research Data

DataverseNO



NTNU Open Research Data

NTNU

DataverseNO > NTNU Open Research Data

 Find

[Advanced Search](#)

For data that does not fit into disciplinary or domain specific archives/repositories

Distributor Name: **NTNU Open Research Data** ✕

1 to 10 of 52 Results

Sort ▾

Replication Data for: Closure Law Model Uncertainty Quantification

Dec 5, 2021



Strand, Andreas; Kjelaas, Jørn; Bergström, Trond H.; Steinsland, Ingelin; Hellevik, Leif R., 2021, "Replication Data for: Closure Law Model Uncertainty Quantification", <https://doi.org/10.18710/3OJHDN>, DataverseNO, V1

The prediction uncertainty in simulators for industrial processes is due to uncertainties in the input variables and uncertainties in specification of the models, in particular the closure laws. In this work, the uncertainty in each closure law was modeled as a random variable an...

Replication data for: MultiPACK Project: Two years of data monitoring of all-CO2 retail stores within the MultiPACK project

Nov 24, 2021



Artuso, Paolo; Tosato, Giacomo; Rossetti, Antonio; Minetto, Silvia; Marinetti, Sergio, 2021, "Replication data for: MultiPACK Project: Two years of data monitoring of all-CO2 retail stores within the MultiPACK project", <https://doi.org/10.18710/Z33Q8X>, DataverseNO, V1

This dataset includes the data for two different integrated units installed in typical Italian neighbourhood supermarkets are considered. The two supermarkets are respectively located in the north-east area of Italy (Trento, Figure 1a), characterized by a relatively mild climate...

Replication Data for: MultiPACK Project: Field data of integrated CO2 heat pump systems for Italian hotels in the MultiPACK project

Nov 24, 2021



Artuso, Paolo; Tosato, Giacomo; Hafner, Armin; Rossetti, Antonio; Minetto, Silvia; Marinetti, Sergio, 2021, "Replication Data for: MultiPACK Project: Field data of integrated CO2 heat pump systems for Italian hotels in the MultiPACK project", <https://doi.org/10.18710/UN1YBZ>, DataverseNO, V1

The EU project MultiPACK supports the design, construction, installation and monitoring of reversible heat pump units using CO2 as the refrigerant, providing air conditioning, heating and domestic hot water. The design includes the use of two-phase ejectors and natural circulatio...

The effect of midbond functions on interaction energies computed using MP2 and CCSD(T)

Nov 1, 2021



Matveeva, Regina; Falck Erichsen, Merete; Koch, Henrik; Heyvik, Ida-Marie, 2021, "The effect of midbond functions on interaction energies computed using MP2 and CCSD(T)", <https://doi.org/10.18710/2FWECY>, DataverseNO, V1

This data set contains contains raw data (monomer and dimer energy contributions for the A24 data set; interaction energies for the A24 and S66 data sets; number of basis functions for 53 complexes of the S66 data set) as well as geometries of A24 and S66 complexes including midb...

Summary statistics for "Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis"

Version 1.0



Graham, Sarah E.; Willer, Cristen J., 2021, "Summary statistics for "Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis"", <https://doi.org/10.18710/J7ARK5>, DataVerseNO, V1

[Cite Dataset](#)
[Learn about Data Citation Standards.](#)
[Access Dataset](#)
[Contact Owner](#)
[Share](#)
[Dataset Metrics](#)

7 Downloads

Description

Chronic kidney disease (CKD) is a growing health burden currently affecting 10–15% of adults worldwide. Estimated glomerular filtration rate (eGFR) as a marker of kidney function is commonly used to diagnose CKD. We analyze eGFR data from the Nord-Trøndelag Health Study and Michigan Genomics Initiative and perform a GWAS meta-analysis with public summary statistics, more than doubling the sample size of previous meta-analyses. We identify 147 loci (53 novel) associated with eGFR, including genes involved in transcriptional regulation, kidney development, cellular signaling, metabolism, and solute transport. Additionally, sex-stratified analysis identifies one locus with more significant effects in women than men. Using genetic risk scores constructed from these eGFR meta-analysis results, we show that associated variants are generally predictive of CKD with only modest improvements in detection compared with other known clinical risk factors. Collectively, these results yield additional insight into the genetic factors underlying kidney function and progression to CKD. (2019-04-23)

Subject

Medicine, Health and Life Sciences

Keyword

Chronic kidney disease, CKD, Kidney function, GWAS, The Trøndelag Health Study (HUNT), eGFR, Michigan Genomics Initiative, Quantitative trait loci

Related Publication

Graham, S.E., Nielsen, J.B., Zawistowski, M. et al. Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nat Commun* 10, 1847 (2019). <https://doi.org/10.1038/s41467-019-09861-z> doi: 10.1038/s41467-019-09861-z

[Files](#)
[Metadata](#)
[Terms](#)
[Versions](#)
[Export Metadata](#)

DataverseNO Certified by  Search - Deposit Guide Support Log In

Citation Metadata

Dataset Persistent ID doi:10.18710/J7ARKS

Publication Date 2021-06-29

Title Summary statistics for "Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis"

Author Graham, Sarah E. (University of Michigan) - ORCID: 0000-0003-1271-2489
Willer, Cristen J. (University of Michigan) - ORCID: 0000-0001-5645-4966

Contact Use email button above to contact.
Gabrielsen, Malen E. (NTNU - Norwegian University of Science and Technology)

Description Chronic kidney disease (CKD) is a growing health burden currently affecting 10–15% of adults worldwide. Estimated glomerular filtration rate (eGFR) is a marker of kidney function is commonly used to diagnose CKD. We analyze eGFR data from the Nord-Trøndelag Health Study and Michigan Genomics Initiative and perform a GWAS meta-analysis with public summary statistics, more than doubling the sample size of previous meta-analyses. We identify 147 loci (53 novel) associated with eGFR, including genes involved in transcriptional regulation, kidney development, cellular signaling, metabolism, and solute transport. Additionally, sex-stratified analysis identifies one locus with more significant effects in women than men. Using genetic risk scores constructed from these eGFR meta-analysis results, we show that associated variants are generally predictive of CKD with only modest improvements in detection compared with other known clinical risk factors. Collectively, these results yield additional insight into the genetic factors underlying kidney function and progression to CKD. (2019-04-23)

Subject Medicine, Health and Life Sciences

Keyword Chronic kidney disease
CKD
Kidney function
GWAS
The Trøndelag Health Study (HUNT)
eGFR
Michigan Genomics Initiative
Quantitative trait loci

Related Publication Graham, S. E., Nielsen, J. B., Zawistowski, M. et al. Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nat Commun* 10, 1847 (2019). <https://doi.org/10.1038/s41467-019-09891-z> doi: 10.1038/s41467-019-09891-z <https://doi.org/10.1038/s41467-019-09891-z>

Producer Norwegian University of Science and Technology (NTNU) <https://www.ntnu.edu/>
University of Michigan <https://umich.edu/>

Distributor NTNU Open Research Data (Norwegian University of Science and Technology) (NTNU) <https://dataverse.no/dataverse/ntnu>

Distribution Date 2019-04-23

Depositor Tellefsen, Janne

Deposit Date 2021-06-22

Kind of Data Summary statistics (results) from GWAS meta- analysis.

Life Sciences Metadata

Design Type Select...

Other Design Type +

Factor Type Select...

Other Factor Type +

Organism Select...

Other Organism +

Measurement Type Select...

Other Measurement Type +

Technology Type Select...

Other Technology Type +

Technology Platform Select...

Other Technology Platform +

Cell Type +

Files Metadata Terms Versions

Terms of Use

Waiver

Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

CC0 - "Public Domain Dedication"



Ensures FAIRness

Log In

Log in or sign up with your institutional account — more [information about account creation](#). Leaving your institution? Please contact [DataverseNO](#) for assistance.

Your Institution

Feide - Norwegian educational ins...

[Allow me to type the name of my institution](#)

Other options

Log in with FEIDE user (NTNU user), see [wiki](#) for further information on how to publish data.



Nyheter Min profil For ansatte For studenter

Intranettet / Kunnskapsbasen / NTNU Open Data

NTNU Open Data

On this page you can find information about how you can archive your data set in the institutional repository NTNU Open Research Data.

[Topic page about research data](#) | [Pages labeled with Open Data](#)

[Norsk versjon - NTNU Open Data](#)

About NTNU Open Research Data

[NTNU Open Research Data](#) is an institutional repository for open data from all fields and disciplines. The archive is part of [DataverseNO](#), which is operated by UiT The Arctic University. NTNU Open Research Data adheres to the guidelines and policy of DataverseNO, and all data sets will be curated before publication. DataverseNO is a Core Trust Certified repository and assigns DOIs (Digital Object Identifiers) to data sets. The standard license is [CC0 \(Creative Commons Zero\)](#), but other open licenses can be considered if needed.

Open Source

Open source

This page contains information about where and how you can publish and share source code, things to consider before you do, and how to license source code.

[Topic page about research data](#) | [Pages labeled with Open Science](#)

[Norsk versjon - Åpen kildekode](#)

Why publish source code?

Source code and software developed in research projects and other academic activities, should be included as a part of published research results. Together with [publication of research data](#), this facilitates testing and validation of analyses and models, and enables reuse and further development of the code or software. Many aspects of best practice for research data management are applicable to source code. This includes versioning, documentation and metadata ensuring research output that is as FAIR, open and reproducible as possible. Nevertheless, there are some considerations and practicalities that are specific for source code, including licensing.

How to publish source code?

Several of the useful software development tools and services used for version control and collaboration, can also be used to make code and software openly available. Some commonly used examples include GitHub, GitLab and Bitbucket.

Help & Support

Support and guidance: Research Data @NTNU

- Central support service for research data
 - Coordinated by the library in close collaboration with IT
 - Support for data management throughout the data lifecycle



Research Data

For employees

What is Research Data @NTNU?

- a central support service for research data
- a service for researchers and students at NTNU
- a contact point for faculties and institutes at NTNU
- a collaboration between the University Library and NTNU IT

RESEARCH
DATA
@NTNU

NTNU requires good research data management

Research data at NTNU should be managed according to best practice and be as open as possible and as closed as necessary. Data containing personal information must be processed according to [GDPR and NTNU guidelines](#).



Data Management Plan (DMP)

All research projects at NTNU should develop a [Data Management Plan](#) describing how the research data will be managed.



NTNU Open Research Data

Research data at NTNU can be published and shared openly in our repository for research data, [NTNU Open Data](#).