Constructing a whole genome reference dataset for the Swedish population

Adam Ameur

National Genomics Infrastructure, SciLifeLab, Uppsala, Sweden

How it started - in 2015

- A national SciLifeLab project for whole genome sequencing
- Aim: to build a reference dataset for the Swedish population
  - Based on 1000 individuals
  - With whole genome sequencing
  - Make available for research
  - Make available for health care

Ulf Gyllensten

Why 1000 Swedish whole genomes?

- The dataset has many possible uses:
  - Look up genetic variant frequencies
  - Use as matched controls
  - Study population genetics
  - Study human evolutionary history

High demand for the data from many different groups:

➔ Make the data available as **quickly** and **openly** as possible!

Deciding on a cohort to use for the project

The Swedish Twin Registry:

- Inclusion based on twinning
- Distribution like population density
- General population-prevalence of disease
- 10,000 individuals have been analysed with SNP arrays

1,000 individuals selected based on genetic structure across Sweden

Whole Genome Sequencing

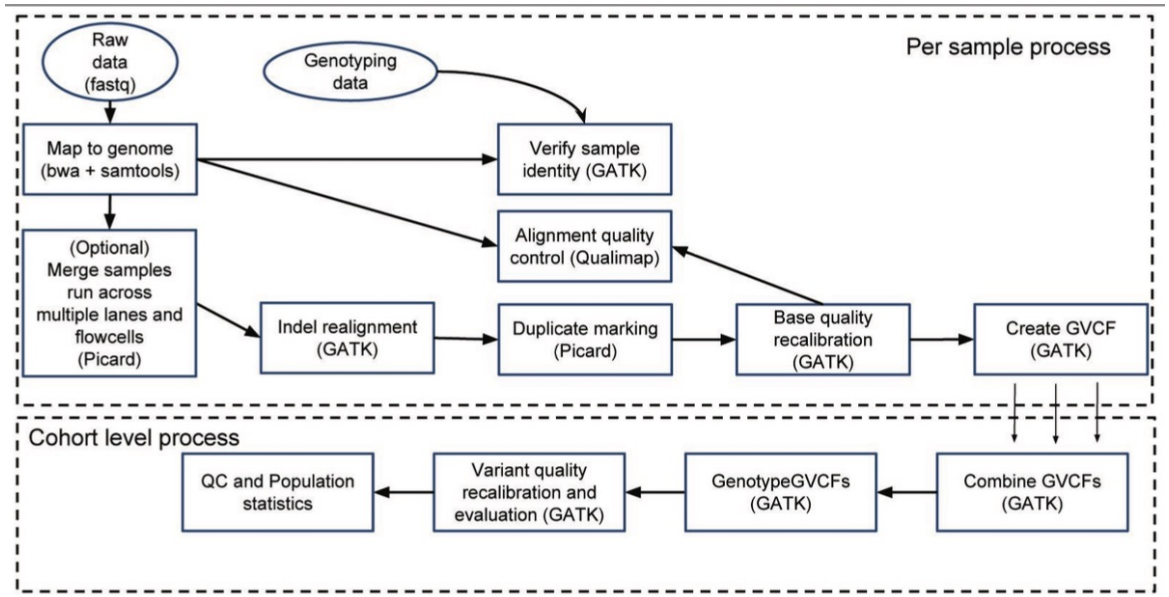
30X Illumina WGS data generated for all 1,000 individuals

- 509 samples sequenced at NGI Stockholm
- 491 samples sequenced at NGI Uppsala

Sequencing of all 1000 samples completed in September 2016

Data management and analysis

Analysis pipeline developed for mapping and variant calling



- About 100,000 Gb data generated within the project
- Over 2 million CPU hours used for the first round of analyses

All computations were done on Bianca: **sens2016003**

A Swedish 1000 Genomes Project

From SweGen release party on Oct 19th 2016!

European Journal of Human Genetics

Explore content · About the journal · Publish with us

nature › european journal of human genetics › articles › article

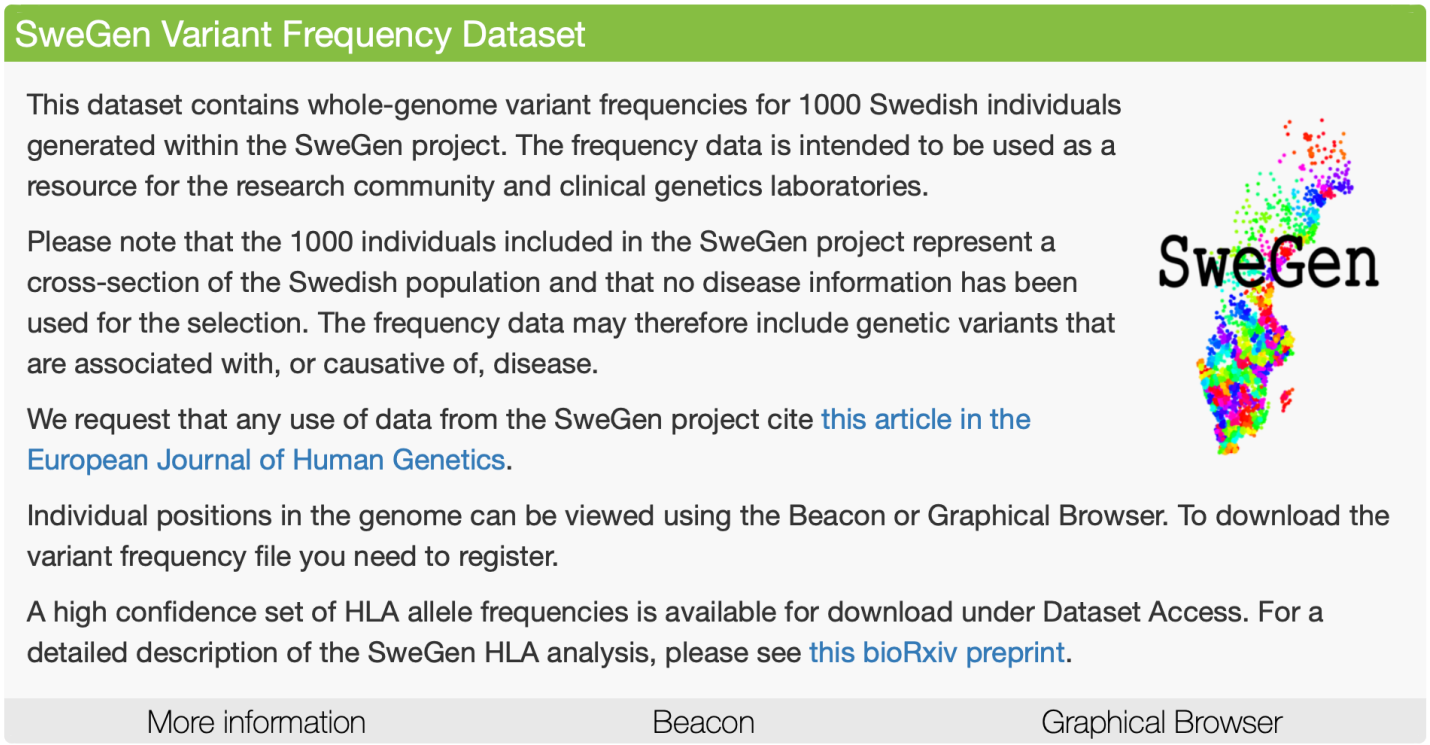
Open Access | Published: 23 August 2017

SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population

Adam Ameur, Johan Dahlberg, Pall Olason, Francesco Vezzi, Robert Karlsson, Marcel Martin, Johan Viklund, Andreas Kusalananda Kähäri, Pär Lundin, Huiwen Che, Jessada Thutkawkorapin, Jesper Eisfeldt, Samuel Lampa, Mats Dahlberg, Jonas Hagberg, Niclas Jareborg, Ulrika Liljedahl, Inger Jonasson, Åsa Johansson, Lars Feuk, Joakim Lundeberg, Ann-Christine Syvänen, Sverker Lundin, Daniel Nilsson, ... Ulf Gyllensten + Show authors

European Journal of Human Genetics **25**, 1253–1260 (2017) | Cite this article

Making genome data available



- Aggregated frequencies available from: ***swefreq.nbis.se***
- Possible to access individual genotype data through secure server

SweGen resulted in many follow-up studies

Completed projects

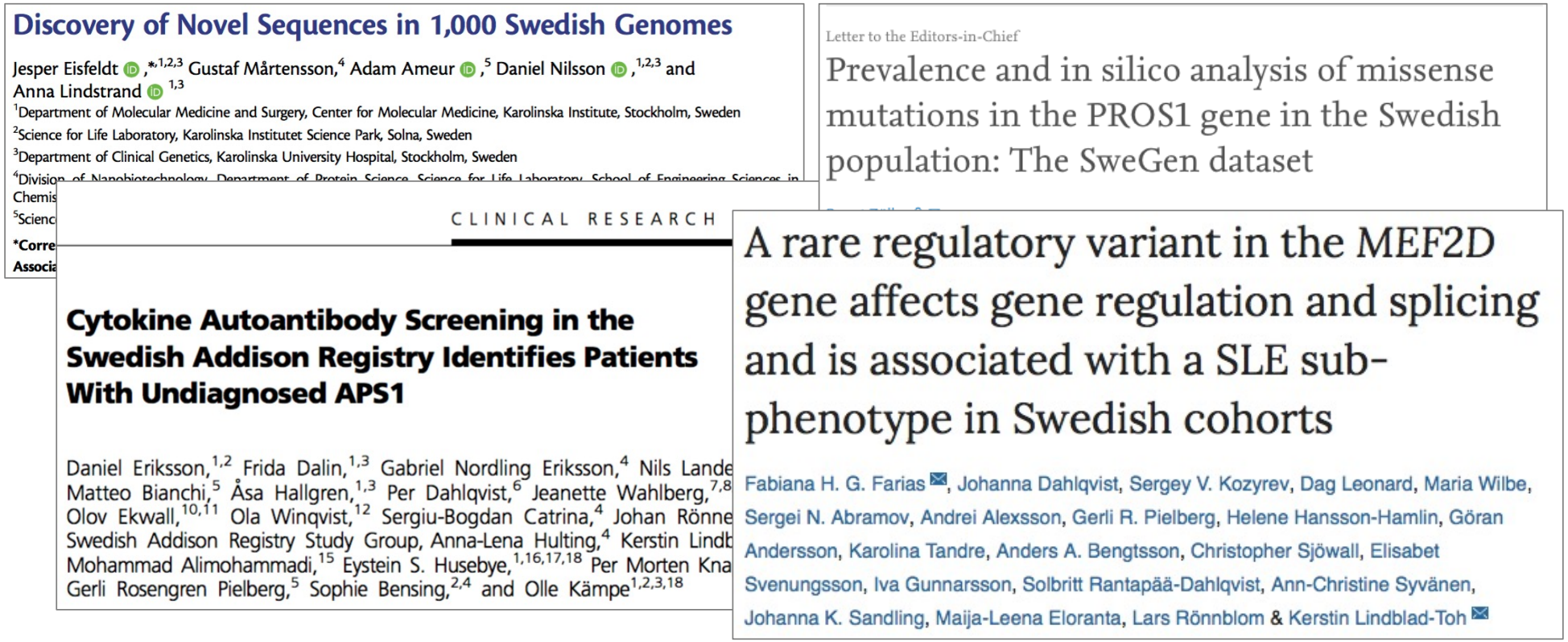
- HLA genotyping (Jennifer Meadows, UU)
- Detection of “novel sequences” (Anna Lindstrand, KI)
- Study mitochondrial DNA (Marie Allen, UU)
- Study chromosome Y (Monika Karmin, Estonia)
- Pharmacogenomics (Mia Wadelius, UU)
- Mosaic loss of chromosome Y (Lars Forsberg, UU)
- Transposable elements (Anna Lindstrand, KI)
- And much more…

Ongoing projects

- Reanalyze against T2T Genome Reference (Åsa Johansson, UU)

Impact of the SweGen dataset

About 200 publications have made use of the data…



Discovery of Novel Sequences in 1,000 Swedish Genomes

Jesper Eisfeldt, Gustaf Mårtensson, Adam Ameur, Daniel Nilsson and Anna Lindstrand

[1]Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institute, Stockholm, Sweden
[2]Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden
[3]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

Letter to the Editors-in-Chief

Prevalence and in silico analysis of missense mutations in the PROS1 gene in the Swedish population: The SweGen dataset

CLINICAL RESEARCH

Cytokine Autoantibody Screening in the Swedish Addison Registry Identifies Patients With Undiagnosed APS1

Daniel Eriksson, Frida Dalin, Gabriel Nordling Eriksson, Nils Lande
Matteo Bianchi, Åsa Hallgren, Per Dahlqvist, Jeanette Wahlberg,
Olov Ekwall, Ola Winqvist, Sergiu-Bogdan Catrina, Johan Rönne
Swedish Addison Registry Study Group, Anna-Lena Hulting, Kerstin Lindb
Mohammad Alimohammadi, Eystein S. Husebye, Per Morten Kna
Gerli Rosengren Pielberg, Sophie Bensing, and Olle Kämpe

A rare regulatory variant in the MEF2D gene affects gene regulation and splicing and is associated with a SLE sub-phenotype in Swedish cohorts

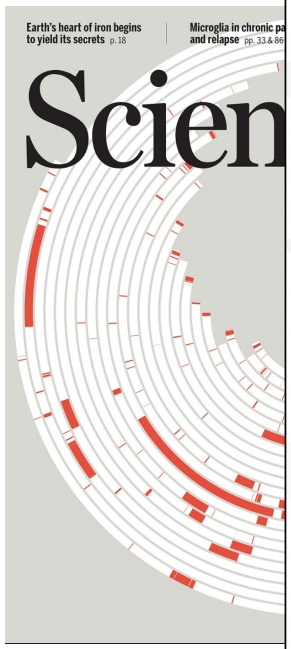
Fabiana H. G. Farias, Johanna Dahlqvist, Sergey V. Kozyrev, Dag Leonard, Maria Wilbe, Sergei N. Abramov, Andrei Alexsson, Gerli R. Pielberg, Helene Hansson-Hamlin, Göran Andersson, Karolina Tandre, Anders A. Bengtsson, Christopher Sjöwall, Elisabet Svenungsson, Iva Gunnarsson, Solbritt Rantapää-Dahlqvist, Ann-Christine Syvänen, Johanna K. Sandling, Maija-Leena Eloranta, Lars Rönnblom & Kerstin Lindblad-Toh

...but more importantly, SweGen is used in clinical routine diagnostics!

Limitations with the SweGen dataset

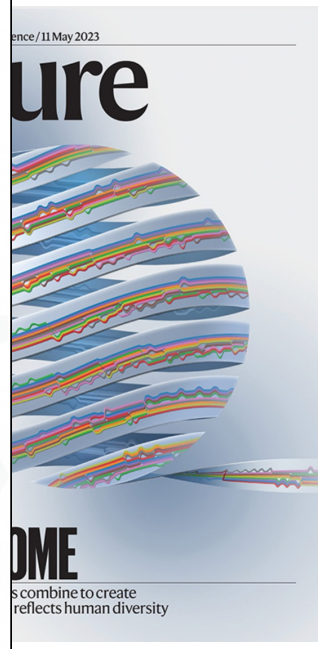
- There now is a better technology – Long Read Sequencing (LRS)!

2022: Gives <u>complete</u> infor



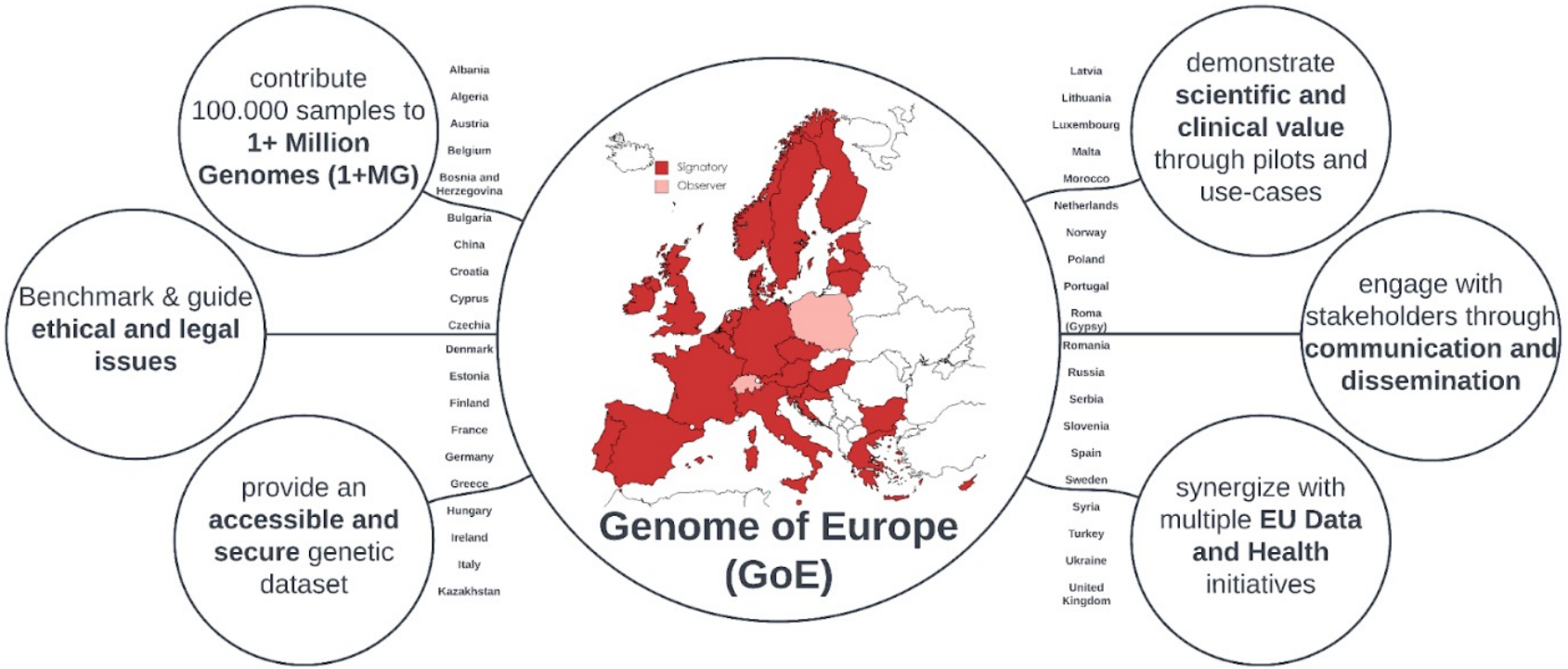
And epigenetic information (methylation) that can be associated to environmental factors!

tion of new genetic variation!



What will we do next?

- LRS of 500+ individuals, as Sweden's contribution to the Genome of Europe



- A multi-year project, planned start for sequencing in 2025!

Thank you again!

We couldn’t have done this work without Bianca…

NATIONAL GENOMICS INFRASTRUCTURE

NBIS

SciLifeLab

Uppsala Multidisciplinary Center for Advanced Computational Science

UPPMAX