Roadmap for a Precision Medicine Initiative in the Nordic Region

Pål Rasmus Njølstad[1,2,3*], Ole Andreas Andreassen[4,5], Søren Brunak[6], Anders D. Børglum[7,8,9*], Joakim Dillner[10,11], Tõnu Esko[12,13], Paul W. Franks[14,15,16,17*], Nelson Freimer[18,19], Leif Groop[14,20], Hakon Heimer[21], David M. Hougaard[22,23], Eivind Hovig[24,25,26], Kristian Hveem[27,28,29], Anu Jalanko[30], Jaakko Kaprio[20], Gun Peggy Knudsen[31], Mads Melbye[32,33,34], Andres Metspalu[35], Preben Bo Mortensen[36,37], Juni Palmgren[38,39], Aarno Palotie[3,20,40,41,42,43*], Wenche Reed[44], Hreinn Stefánsson[45,46], Nathan O. Stitziel[47,48,49], Patrick F. Sullivan[50], Unnur Thorsteinsdóttir[47,48], Marc Vaudel[1], Eero Vuorio[51], Thomas Werge[52,53,22], Camilla Stoltenberg[30,54*] & Kári Stefánsson[45,46]

[1]Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway.

[2]Department of Pediatrics, Haukeland University Hospital, Bergen, Norway.

[3] Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Cambridge, MA ,USA.

[4]NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.

[5]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

[6]Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[7]Department of Biomedicine, Aarhus University, Aarhus, Denmark.

[8]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark.

[9]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark.

[10]Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

[11]Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden.

[12]The Estonian Genome Center, University of Tartu, Tartu, Estonia.

[14]Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

[14]Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre , Lund University , Malmö, Sweden.

[15]Department of Public Health and Clinical Medicine, Section for Medicine, Umeå University, Umeå, Sweden.

[16]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA.

[17]Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

[18]Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA.

[19]UCLA Center for Neurobehavioral Genetics, Los Angeles, CA, USA.

[20]FIMM Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

[21]School of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark .

[22]Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark.

[23]iPSYCH - The Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark.

[24]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.

[25]Institute of Cancer Genetics and informatics. Oslo University Hospital, Oslo, Norway.

[26]Department of Informatics, University of Oslo, Oslo, Norway.

[27]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway.

[28]HUNT Research Center, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway.

[29]Department of Medicine, Levanger Hospital, Levanger, Norway.

[30]Genomics and Biomarkers Unit, National Institute for Health and Welfare, Helsinki, Finland.

[31]Norwegian Institute of Public Health, Oslo, Norway.

[32]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

[33]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.

[34]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

[35]The Estonian Genome Center, University of Tartu, Tartu, Estonia.

[36]National Centre for Register-Based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark.

[37]iPSYCH - The Lundbeck Foundation Initiative for Psychiatric Research, Aarhus, Denmark.

[38]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[39]Swedish e-Science Research Center, SeRC, Stockholm, Sweden.

[40]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[41]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[42]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[43]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA.

[44]Department of Research, Innovation and Education, Oslo University Hospital, Oslo, Norway.

[45]deCODE genetics, Reykjavik, Iceland.

[46]University of Reykjavik, Iceland.

[47]Cardiovascular Division, Department of Medicine, Washington University, Saint Louis, MO, USA.

[48]Department of Genetics, Washington University, Saint Louis, MO, USA.

[49]McDonnell Genome Institute, Washington University, Saint Louis, MO, USA.

[50]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[51]Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland.

[52]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark.

[53]Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

[54]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway.

*Members of the writing group

Address correspondence to: Professor Pål Rasmus Njølstad, M.D., Ph.D., Center for Diabetes Research, Hospital for Children and Adolescents, Haukeland University Hospital, Haukelandsbakken 15, NO-5021 Bergen, Norway (pal.njolstad@uib.no).

Abstract, 142 words; Text, 3,044 words; Figures, 1; Boxes, 3; Supplementary Material, 1

Abstract

The Nordic region, comprised here primarily of Denmark, Estonia, Finland, Iceland, Norway, and Sweden, has many of the characteristics necessary to be at the forefront of genome-based precision medicine. These include egalitarian and universal health-care, expertly-curated patient and population registries, biobanks, large population-based prospective cohorts linked to registries and biobanks, and a widely embraced sense of social responsibility that motivates public engagement in biomedical research. However, this can only be achieved through coordinated action involving all actors of the healthcare sector. Now is an opportune time to organize scientists in the Nordic region, together with other stakeholders including patient representatives, governments, pharmaceutical companies, academic institutions, and funding agencies, to initiate a Nordic Precision Medicine Initiative. We present a roadmap for how this can be done. The Initiative will facilitate research, clinical trials, and knowledge transfer to meet regional and global health challenges.

Background

Complex disease is caused by environmental exposures perturbing multiple biochemical pathways. Determining the impact of these exposures on the individual patient, and the extent to which these can be mitigated by therapy, are fundamental objectives of contemporary medicine. One of the major goals is to tailor medicine to the patient, not only by understanding the disease, but also by understanding the specific characteristics of the person. Another important goal is to accurately determine a person's risk of developing disease, and to use this information to optimize the timing,

147 delivery and type of preventive action. Collectively, these goals represent
148 *precision* medicine[1]: individual preventive and therapeutic interventions
149 incorporating a detailed understanding of human diversity.

150 Human diversity is determined by the joint effects of environmental
151 exposures and DNA sequence variation[2]. Genotyping millions of common
152 DNA variants in very large human populations has been made possible by
153 recent major technological advances. These technologies have been used to
154 discover and characterize many thousands of independent association
155 signals between common variants and disease traits. In general, common
156 single nucleotide polymorphisms (SNPs) individually confer relatively little
157 risk, but jointly, account for a substantial proportion of the overall
158 heritability of many complex disorders; for instance, in mental disorders,
159 the proportion of variance attributable to genome-wide SNPs (liability-
160 scale SNP heritability) is between a third and a quarter of the overall
161 heritability[3,4]. Methods have been developed to use a large number of
162 sequence variants to construct what is often termed a *polygenic risk score*
163 for phenotypes[5]. Polygenic risk scores can be useful in assessing both
164 relative and absolute risks of diseases. The frequencies and effects of
165 common variants tend to be similar in most ethnic groups with some
166 exceptions. The most recent breakthroughs in understanding the
167 relationships between human DNA sequence variation and phenotypic
168 diversity have been made possible by the development of technologies for
169 sequencing whole genomes or whole exomes in very large populations; this
170 in turn has identified rare risk variants, and expanded the catalogue of
171 verified polymorphic genomic regions related to disease[6,7]. Many of the
172 recently discovered rare variants cluster in the coding part of genes that
173 confer high risk of various diseases[8]. It is, however, important to remember
174 that the rare variants have arisen recently and tend to be population-, or

even family-, specific. Nevertheless, even *de novo* mutations, affecting the probands but not their parents, have proven relevant for public health[9,10].

Although these major technical advances in human genetics have been a driving force in the evolution of precision medicine, there are many practical obstacles hindering the implementation of these technologies at scale. We argue that the Nordic countries (**Fig. 1**) offer a priviledged environment to negotiate these obstacles making it likely that, with a well-functioning organizational structure and adequate funding, the Nordic countries will likely play a very significant role in precision medicine over the coming years, at least as it relates to the diseases common within the region. Indeed, given the region's unique resources, we feel obligated to ensure that this possibility is realized.

Although major progress has been made in disease genomics, further characterization will require very large prospective cohorts with harmonized genetic and phenotypic data, yet such datasets are uncommon in most parts of the world. Over recent decades, the Nordic countries have individually collected very large, carefully phenotyped, prospective cohorts, often for the sake of disease monitoring and surveillance. These cohorts and registries contain data from the majority of Nordic citizens, following many of them throughout parts of the life-course; data are often collected through primary care clinics, hospitals, and post-mortems. This process has resulted in the availability of datasets comprising some 27 million Nordic citizens (**Fig. 1**), many of whom have undergone repeated sampling over many decades. Notably, a considerable amount of today's knowledge on the epidemiology of human disease is based on research from Nordic cohorts and registries (**Supplementary Table 1**).

The Nordic countries include Denmark, Finland, Iceland, Norway, and Sweden, with their associated territories (Greenland, the Faroe Islands, and the Åland Islands). The 27 million people of the Nordic countries are ancestrally mainly Scandinavian or Finnish, with Greenlandic Inuit (around 56,000) and the Sami (50-80,000) as indigenous peoples. Estonians share language and historical roots with the Finns and identify very closely with Nordic culture. Furthermore, with its well-established national biobank, Estonia has been an active partner in the Nordic Biobank Network. Despite the fact that the region for centuries has had immigration from neighboring, continental countries, and more recently from elsewhere, immigrants still comprise only a minor proportion of the Nordic population. In 2012 for instance, 10.9%, 7.9%, and 14.9% of the Norwegian, Danish, and Swedish populations were immigrants, respectively (Statistics Norway). The Nordic countries have much in common in their way of life, history, social structures, and languages. These countries do not form a truly united political entity, but co-operate closely on many levels, including within the Nordic Council.

Access to unique data needed to apply precision medicine

One of the assets needed for the implementation of precision medicine is collections of data on individual phenotypes and genotypes that allow detailed studies of the causal effects of genetic variants in disease. Such datasets can be used for the crucial replication of published genetic associations with disease and other relevant phenotypes, as well as for determining population-specific frequencies and the extent to which specific variants impact disease (effect size or predictive accuracy). These datasets are, however, of greatest value for i) discovering genomic variants

that pinpoint druggable pathways, ii) aiding in the reclassification of disease diagnoses (to generate new taxonomies that can be treated more effectively), and iii) facilitating the stratification of populations based on risk factor susceptibility or therapuetic response (to optimize prevention or treatment), each of which are core features of precision medicine.

Although today there is no truly pan-Nordic database on phenotypes and genotypes, the required components exist. These include: (a) a long history of integrated healthcare, patient registries and biobanks, with existing assets of biological samples, patient records, and longitudinal follow-up; (b) population characteristics such as founder effects and stable, traditional societies with homogeneous environmental exposures; (c) strong public trust based on a history of social welfare and commitment to research for the public good; and (d) access to technology and expertise for generating, managing, storing and interpreting genomic and clinical data (**Supplementary Table 2**). What remains is to bring together the wealth of data and biomaterials under a common framework, and to make this accessible to the research community through federated data-access models.

No other countries currently have access to population-based registries of comparable size and with similar quality and detail of clinical information as those of the Nordic nations (**Box 1**). The unique features of this Nordic resource include complete nationwide social and health registers from about 1950 onwards, the world´s largest health studies with detailed phenotypes, biological samples, follow-up of 30-50 years, hospital diagnoses as well as prescription and treatment registries, including all inpatients and outpatients for all hospitals during the past decades, and newborn screening programs of live births with samples stored since the early 1980s. There are population-based biorepositories and data from at

least 8-10 million individuals available for research in the Nordic countries today.

Regarding the secure storage and use of data and the challenge of utilizing data while protecting the data donor's privacy, the Nordic region has advantages due to its traditions of equality and a strong public sector. The citizens generally commit to studies with broad consent. In all Nordic countries, there is an overall positive attitude toward health research, including genetic studies[11]. There is a healthy balance between privacy regulations and willingness to share data for research. New European Union regulations may, however, create new challenges (**Box 2**). As the European General Data Protection Regulation (GDPR) came into effect in May 2018, its full consequences remain unknown. There are, however, clear ways in which GDPR expectations can be fulfilled. Implied consent is sufficient for data that are not sensitive, while analysis of sensitive data requires opt-in consents specific for each research question addressed with the data.

deCODE genetics in Iceland and through its collaboration elsewhere in the Nordic region has been at the forefront of human genomic research for two decades (**Supplementary Table 1)**. This engagement corresponds with the generation of extensive genotype data from SNP arrays for some 650,000 Nordic participants. Whole exome or genome sequence data from over 95,000 Nordic participants have also been generated. deCODE genetics has genotyped half of the Icelandic nation and performed whole genome sequencing on ten percent of it[12]. In addition to work done in the Nordic countries[13], recent significant contributions to human genetics coming from the USA have been made through the use of clinical material, registers, cohorts, and biobanks in the Nordic countries (**Supplementary Table 1**)[14,15]. A pilot project on colorectal cancer supported by the Nordic

Council of Ministers has connected biobanks and registries in all Nordic countries, including transfer of personal data between Nordic countries as well as joint genotyping and whole genome sequencing of biospecimens from several Nordic countries, demonstrating that the infrastructure and regulations allow considering the Nordic countries as a single scientific region (NordForsk).

The Nordic countries have also been at the forefront of epidemiological studies linking environmental exposures to disease outcomes. These studies have identified the diversity in exposure to important external determinants of disease. Until now, these have been largely independent of genetic information, and genetic association studies have mostly been conducted independently of known environmental risk factors, despite many large-scale epidemiological studies in the Nordic countries during the past 50 years. Whilst the origins of precision medicine have come from population genetics research, the successful implementation of precision medicine to tackle common complex disease will almost certainly require consideration of the joint effects of genes and environment, as lifestyle has a major influence on most common diseases. Recently, DNA has been extracted, genotyped or sequenced in many of these outstanding Nordic cohorts, enabling adequately powered pioneer studies uncovering the interplay of the environment and genetics.

A special feature of the Nordic countries is the existence of multiple genetic isolates, some of which are contained within the large registers and biobanks. This unique resource has facilitated the discovery of variants that are private to specific population isolates or families. An example is detection of a rare loss-of-function (LoF) variant in the *SLC30A8* gene that is protective for type 2 diabetes and enriched in the Botnia region[14]. To explore the underlying mechanisms, we were able to go back to the families

carrying the LoF mutation and by sequencing additional family members, increased the number of mutation carriers 3-fold. We then selectively recruited participants by genotype for additional metabolic studies to pinpoint the mechanisms for protection. Although Finland is not technically an isolate, its history – small founder population, evolutionary bottleneck and then rapid expansion – makes it ideal for identifying rare mutations.

Several initiatives that pave the way for a Nordic precision initiative are now underway in individual countries: The Danish government and municipal authorities have recently launched a national project in precision medicine for the use of genetic analysis technologies in the prevention and treatment of many diseases. Novo Nordisk Foundation has recently announced its support this initiative with over 900 mill DKK (approx. 137 mill USD). In Finland, major changes in the legal framework are currently undergoing parliament hearings: National registers, genome, and biobank legislation will be reorganized to improve the secondary use of health data in research and development. The Research Council of Norway, as well as major universities and university hospitals, and the Norwegian Institute of Public Health, are increasingly supporting projects on precision medicine, and the legal framework is becoming more positive towards genetic studies utilizing population registers and biobanks. The Swedish government has commissioned the Swedish Research Council to support national infrastructure for register-based research, including clinical registers. Work is ongoing to create a single national entry point for research using registers, cohorts and biobanks. The NordForsk funding agency is further actively supporting development of mechanisms for secure sharing of person-sensitive data across the Nordic countries through Tryggve, and the Nordic ELIXIR nodes are similarly enabling secure data exchange through participation in national European Phenome-Genome repositories.

SNP array studies have shown a close correspondence between genetic and geographic distances in Europe and that the geographical map of Europe naturally arises as an efficient two-dimensional summary of genetic variation in Europeans[16,17]. Their descent can genetically, and hence geographically, be distinguished by drawing a line from the north to the south-east (northern Europe to the Balkans), with another east-west axis across Europe. Y-chromosome studies show three large haplogroups that account for most of Europe's patrilineal descent. Nordic populations overlap considerably, particularly in major cities and neighboring regions, but differ from other European populations in their genetic substructure, with Finns being especially distinct, to the extent that they are essentially a genetic isolate[18]. This is also reflected in the language groups (**Supplementary Fig. 1**). Hence, in addition to the wisdom of working with the biobanks in the individual Nordic populations and the ability to link them to population registries on healthcare information and other relevant demographic data, the shared ancestry of the Nordic peoples that is reflected in overlap of genomic sequences makes them ideal partners in genetic research and in the implementation of precision medicine (**Supplementary Fig. 2**).

Healthcare with universal access and societal acceptance

Another asset necessary for early implementation of precision medicine is government-funded healthcare systems with excellent records and universal access that are focused on longterm benefits to society rather than shortterm profit. An additional asset is having societies that are committed to protecting the rights of individuals to privacy while

recognizing the importance of using healthcare information for discoveries that improve health and care.

A number of factors predict that precision medicine may be implemented across the region without worsening health disparities. Indeed, we argue that this strategy is a necessary extension of a number of economic and cultural specialities of the region. The Nordic countries rank at the top in a range of metrics of national economic performance, including education, digitalization, economic competitiveness, civil liberties, quality of life, and human development[19]. Together, the economies of the Nordic countries have one of the best macroeconomic performances in the world, and are leaders in sustainable development. The Nordic countries also share many aspects of their economic systems and social structures: market economy is combined with relatively strong labor unions and a well-developed public welfare sector[20]. There is a high degree of income distribution and relatively little social unrest. Individual rights are secured legally and have an increasingly strong influence on the health care systems. In general, inhabitants of the Nordic countries are positive towards research and frequently consent to genetic research with a wide scope both as participants in disease-specific cohorts and population-based general health surveys. The Nordic countries all have single payer healthcare systems with good access and quality of services. Our precision medicine strategy hence appears as a natural extension of economic and cultural specificities of the region.

In spite of these qualities of the Nordic societies, extensive coordination and ambitious funding strategies are required to achieve the necessary societal support to enable the implementation of precision medicine. Efforts to introduce population genomics in the Nordic countries will rely

on a combination of public and foundation funds, as well as investments from biotechnology and pharmaceutical industries. This can induce concerns about the protection of the rights and privacy of citizens, which will need to be adequately addressed if public support is to be maintained. The European GDPR offers the opportunity for the Nordic countries to align processes for personal data used in research. A second issue is the need to find solutions to how the value that is generated in the international private sector using samples and data from the Nordic region shall be returned to the nations and citizens who funded and generated these resources. A third good example of the opportunities and challenges ahead is provided by the story of the *BRCA2* mutation in Iceland. Work done at deCODE genetics has brought insight into the whole genomes of most of the nation. Hence, all *BRCA2* mutation carriers in Iceland could in theory be identified *in silico* and offered interventions that mitigate the cancer risk conferred by the mutation[21,22]. This would be an excellent example of how precision medicine can contribute substantially to public health. This has, however, not been done yet, because the Icelandic society is still debating how to approach the mutation carriers with this clinically critical information. There are people who are deeply concerned about the right of the carriers not to know about their genetic risk. Shall the participants in research studies always have the right to learn the results, even if the medical consequences of the genetic discovery are not yet fully understood? As this situation illustrates, the obstacles to the implementation of precision medicine are not only financial, technical and scientific, but also societal and ethical. We believe that the people of the Nordic region are ready to tackle these challenges and offer a positive example for the rest of the world[11].

A roadmap for the way forward

Recently, the Nordic Society for Human Genetics and Precision Medicine was formed and launched a roadmap for the way forward (**Box 3**). The Society was created in order to: (a) accelerate discovery of disease susceptibility genes and genes protecting from disease through integrated analyses using multiple large-scale datasets and a range of experimental designs; (b) translate these findings so that they can be used for precision medicine to improve public health; (c) and uphold and promote the highest legal, regulatory, social, and ethical standards.

The Society will also be a vehicle to engage the many constituencies of precision medicine, ranging from research and clinical geneticists to data scientists and legal experts. It will also allow for the communication of accurate, up-to-date information to policymakers, research funders, and, most importantly, the public.

We believe our initiative will accelerate research, clinical trials, and transmission of knowledge to meet numerous local, regional, and global health challenges, taking advantage of the unique Nordic health-care system, patient and population registries and biobanks, as well as the social responsibility that has motivated public engagement in biomedical research.

URLs

Danish strategy for personalized medicine, http://www.sum.dk/English.aspx;

deCODE genetics, https://en.wikipedia.org/wiki/DeCODE_genetics;

European Commission, Joint Research Center, http://data.europa.eu/89h/jrc-

ghsl-ghs_pop_gpw4_globe_r2015a; genotyping in Nordic countries, https://www.nordforsk.org; Nordic Council of Ministers pilot project on colorectal cancer, https://www.nordforsk.org/en/policy/policy-briefs-1; Nordic countries, https://en.wikipedia.org/wiki/Nordic_countries; Nordic Society for Precision Medicine, http://nordicprecisionmedicine.org/; Nordic platform for collaboration on sensitive data, https://wiki.neic.no/wiki/Tryggve; Nordic region and sustainable development, http://blogs.worldbank.org/governance/among-wealthy-nations-nordic-countries-are-leading-pack-sustainable-development; publications from deCODE genetics, https://www.decode.com/publications/; Statistics Norway, https://www.ssb.no/en/befolkning

Acknowledgements

We thank the Banbury Center at Cold Spring Harbor Laboratory, New York for organizing and hosting the meeting "Studying the Genomic Variation that Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems" at the Banbury Center February 16-19, 2016. We are grateful to Drs. C.E. Jaquish, G. Tybring as well as D. Høybråten for helpful discussions. We acknowledge funding from the Norwegian Research Council (#223273 to O.A.A.) and NordForsk (to P.W.F.). The work conducted by P.W.F. that is germane to this paper was supported by the Swedish Research Council and ERC-2015-CoG_NASCENT_681742. P.R.N. was supported by grants from the European Research Council (AdG #293574), the Research Council of Norway (#240413), Helse Vest (PERSON-MED_DIA), and Stiftelsen Kristian Gerhard Jebsen (Center for Diabetes Research).

Author contributions

P.R.N., P.W.F., A.D.B., A.P., C.S, and K.S. wrote the manuscript with edits from O.A.A., S.B., J.D., T.E., N.F., L.G., H.H., D.M.H., E.H., K.H., J.K., G.P.K., M.M., A.M., P.B.M., J.P., A.P., W.R., H.S., N.O.S., P.F.S., U.T., M.V., E.V., and T.W. P.R.N., O.A.A., S.B., A.D.B., J.D., T.E., P.W.F., N.F., L.G., H.H., D.M.H., E.H., K.H., A.J., J.K., G.P.K., M.M., A.M., P.B.M., J.P., A.P., W.R., H.S., N.O.S., P.F.S., U.T., M.V., E.V., T.W., C.S., and K.S. conceived the project.

Competing financial interests

H.S., U.T., and K.S. are employed by deCODE Genetics/Amgen, Inc. The other authors declare no competing financial interests.

References

1. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. National Academies Press, Washington, DC, USA, 2011.

2. Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S., US National Human Genome Research Institute. A vision for the future of genomics research. *Nature* **422**, 835-847 (2003).

3. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984-994 (2013).

4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427(2014).

5. Khera, A.V. *et al*. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med*. **375**, 2349-2358 (2016).

6. Gudbjartsson, D.F. *et al*. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet*. 2015 May;47(5):435-44.

7. Marouli, E. *et al*. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).

8. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).

9. Acuna-Hidalgo, R., Veltman, J.A., Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome*

Biol. **17**, 241 (2016).

10. Kong, A. *et al*. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-475 (2012).

11. Ebbesen, M., Sundby A, Pedersen FS, Andersen S. A philosophical analysis of informed consent for whole genome sequencing in biobank research by use of Beauchamp and Childress' four principles of biomedical ethics. *J. Clin. Res. Bioeth.* **6**, 244 (2015).

12. Editorial. Letters from Iceland. *Nat. Genet*. **47**, 425 (2015).

13. Maretty, L. *et al*. Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* **in press** (2017).

14. Flannick, J. *et al*. Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. *Nat. Genet.* **46**, 357-63 (2014).

15. Fuchsberger, C. *et al*. The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47 (2016).

16. Novembre, J., *et al*. Genes mirror geography within Europe. *Nature* **456**, 98-101 (2008).

17. Palo, J.U., Ulmanen, I., Lukka, M., Ellonen, P., Sajantila, A. Genetic markers and population history: Finland revisited. *Eur. J. Hum. Genet.* **17**, 1336-1346 (2009).

18. Lao, O. *et al*. Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).

19. *Nordic Statistical Yearbook 2014* (ed. Haagensen, K.M.). Nordic Council of Ministers, **52**, 7-135. Nord, Copenhagen, 2014 (ISBN 978-92-893-3854-7).

20. Wooldridge, A. Northern lights. Special report: the Nordic countries.

The Economist Henry Thoreau and Enterprise **2**, 1-14 (2013).

21. Thorlacius, S. *et al*. A single *BRCA2* mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nat. Genet*. **13**, 117-119 (1996).

22. Thorlacius, S. *et al*. Study of a single *BRCA2* mutation with high carrier frequency in a small population. *Am. J. Hum. Genet.* **60**, 1079-1084 (1997).

23. Deloukas, P. *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease**.** *Nat. Genet*. **45**, 25-22 (2013).

24. Ripke, S. *et al*. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150-1159 (2013).

25. Fuchsberger, C. *et al*. The genetic architecture of type 2 diabetes. *Nature.* **536**, 41-47 (2016).

26. Willer, C.J. *et al*. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274-1283 (2013).

27. Melin, B.S. *et al*. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet*. **49**, 789-794 (2017).

28. Lambert, C.J. *et al*. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452-1458 (2013) 29. Law, M.H. *et al*. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat. Genet.* **47**, 987-995 (2015).

Figure legend

Figure 1: The Nordic countries, a geographical and cultural region in Northern Europe and the north Atlantic sea.

The Nordic region has around 27 million inhabitants. The population number is shown for Iceland, the kingdoms of Denmark, Norway, and Sweden, and the republic of Finland, and associated territories. These are Greenland and Faroe Islands (ruled by Denmark), Åland Islands (ruled by Finland), and Svalbard (ruled by Norway). Estonia is often associated with the Nordic countries as well. Population density as of 2015 was obtained from the Global Human Settlement project and displayed in shades of grey (**Supplementary Note**; European Commission, Joint Research Centre, Columbia University, NYC, NY). Abbreviations: k, thousand; mil, million; inhab, inhabitant.

574 Boxes

575

Box 1: Selected examples of population-based cohorts, biobanks and registers in the Nordic region

Denmark

The Danish National Biobank at Statens Serum Institut (www.ssi.dk) contains eight million specimens increasing by 0.5 million annually. Most are serum, plasma and DNA, although there also are spinal fluids, faeces and urine. Eleven robots are effectively performing pipetting, DNA purification, biomarker analysis, and storage. Detailed phenotypic data can be obtained via a linkage to national registers.

The National Biobank Register (www.nationalbiobank.dk) includes 22 million specimens from 5.4 million persons in Danish biobanks. Combining data with register information allows searching for diagnosis, age, sex and other variables as well as specific searches that e.g. will identify specimens taken at specific time points before, at, or after a certain diagnosis. Participation and use is free of charge.

Bio- and Genome Bank Denmark (www.regioner.dk/rbgben) is the entry to biological material at all hospitals. The material is professionally collected, nationally registered and regionally stored in five hospitals. Pathological material is stored locally for clinical use. Phenotypic data can be obtained via linkage to registers and clinical databases.

The Danish Neonatal Screening Biobank includes filter blood spots from two million subjects, almost all that are born since 1982. iPSYCH (www.ipsych.au.dk) is the largest study utilizing these samples, including GWAS of 80,000 individuals, of which 50,000 suffer from mental disorders, integrating genomic and national register data.

The Danish National Birth Cohort (www.ssi.dk) contains holds questionnaire data from pregnancy (100,000) and the offspring at six and 18 months as well as seven and 11-year follow-ups. Biological specimens have been collected twice during pregnancy together with cord blood (Danish National Biobank).

Estonia

The Estonian Biobank is population-based (www.biobank.ee) including medical history and current health status as well as extra data from psychiatric patients. All samples (155,000) have been genotyped (SNP arrays); while exome sequencing has been done on 2,500 and whole

genome sequence on 2,500 samples. The biobank can be linked with national health registries and hospital databases.

Finland

The National Institute of Health and Welfare (THL) hosts the majority of large epidemiological and disease specific cohorts (e.g. Finrisk, Health2000, Twins, Botnia, Migraine, GeneRisk) that contain blood samples from 200,000 individuals (www.thl.fi/biobank) of all ages. These include questionnaire, genomic, and biochemical data and link to National Health Register data (EHR) providing decades of disease follow-up data. All university hospitals have established biobanks linking sample data with EHR and EMR data. Hospital biobanks host tissue samples from almost 3 million participants, collected earlier as part of routine diagnostics and have recently been transferred to biobanks. The hospital biobanks as well as the Blood Service Biobank have recently collected prospective samples from 120,000 participants and continue rapidly expanding their collections. The rapid expansion is due Finnish Biobanks partnering with the public-private FinnGen project aiming to collect GWAs and national health register data from 500,000 participants by 2022 (https://www.finngen.fi/). The FinnGen research project has been a major facilitator for development of national biobanking and the accumulated genome data of FinnGen – currently 145 000 – is foreseen to serve as a major basis for GWAS and PheWAS analyses towards development of personalized medicine. For biobanks, the newly produced genome data will further enrich the EHR data with symptom-level information, pathology and biochemical data thus building more possibilities for excellent science and development. All biobanks are networked by BBMRI.fi and have harmonized broad consents and practices.

Iceland

deCODE genetics (www.decode.com) has gathered genotypic and medical data from over 160,000 participants, well over half of the adult population. Using Iceland's uniquely comprehensive genealogical records, deCODE has also a genealogy database covering the entire present day population stretching back to the founding of the country more than 1,000 years ago. The combination of size of the population, the participation of so many people in the discovery work, the genealogies, and high quality universal healthcare have made possible very large-scale studies of virtually any common disease. At the same time, deCODE´s work has minimized the selection bias that confronts research in larger, more stratified populations, enabling to impute or predict genotypes using the genealogies, multiplying many-fold the

amount of data that can derive from genotyping and sequencing.

Norway

The Norwegian Mother and Child Cohort Study (MoBa) is a pregnancy cohort (114,500 children, 95,000 mothers, 75,000 fathers) recruited 1999-2009 (www.fhi.no/MoBa). Information on health and exposures are collected from questionnaires during pregnancy and regularly after birth, and by linkage to registries. Biomaterials were collected from fathers and mothers at pregnancy week 17 and after birth, and from umbilical veins. DNA has been extracted. 25,000 triads have been genotyped (SNP arrays). Many outcome registries have been linked to the cohort.

The Health Survey of Northern Trøndelag (HUNT) contains medical histories and specimens (120,000) from a homogeneous population collected over 30 years (www.hunt.no). Three surveys include information on health-related lifestyle, prevalence and incidence of diseases, health determinants, and associations between disease phenotypes and genotypes. 70,000 samples have been genotyped (SNP arrays). Data can be linked to national health registries.

The Tromsø Study is prospective and population-based with six repeated health surveys after 1974 (www.tromsoundersokelsen.no) including questionnaire data, DNA, serum, and clinical measurements. 40,000 subjects attended at least once, and 15,000 in three surveys or more. In addition to national quality controlled disease registries, the study holds a validated endpoint registry of many well-defined diseases.

The Hordaland Health Studies (husk-en.b.uib.no) were conducted in 1992/93 (The Homocysteine study) and in 1997/99 (HUSK). The main focus is cardiovascular disease, cancer, osteoporosis, anxiety and depression. Some 36,000 residents of Hordaland county participated; 18,000 in 1992/93 and 26,000 in 1997/99. About 7,000 of those who participated in the 1992/93 survey also participated in 1997/99.

Sweden

The Genomic Aggregation Project in Sweden (GAPS) has around 30 Swedish cohorts within which existing detailed genetic and phenotypic data is available (170,000). Within these cohorts exists tens of thousands of additional data-points against which blood samples are stored, from which DNA will be extracted for future genotyping and sequencing. The cohorts already genotyped include the Malmö Diet and Cancer cohort (28,000) in southern Sweden, the Breast Cancer Studies

cohort (30,000) in the central Sweden, the GLACIER Study (20,000) in northern Sweden, and multiple case-control and cohort studies of severe psychiatric disorders (63,000). Aside from disease record linkage, cohorts in Sweden are frequently linked to the drug registry and demographic databases (allowing genealogies dating back to the 1700s to be linked with genetic and phenotypic data).

ANDIS is a large and well-phenotyped study comprising all new subjects with diabetes in Skåne, a complete-capture case series of >15,000 patients with diabetes, from which samples have been genotyped and linked to a range of cross-sectional and prospective registry databases.

Text box 2: The European Union's General Data Protection Regulation and the Nordic Countries

The General Data Protection Regulation (GDPR, Regulation (EU) 2016/679) of May 25, 2018 replaced the Data Protection Directive (officially Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data). The European Commission's (EC) objectives with this new legislation included the harmonization of 27 national data protection regulations into one unified regulation, the improvement of corporate data transfer rules outside the European Union (EU), and the improvement of user control over personal identifying data. The proposed new EU data protection regime thus extends the scope of the EU data protection law to all foreign companies processing data of EU residents. It provides for a harmonization of the data protection regulations throughout the EU, thereby making it easier for non-European companies to comply with these regulations; however, this comes at the cost of a strict data protection compliance regime with severe economic penalties.

The GDPR preserves the equilibrium between the necessity of effectively protecting the subject's rights in a digitalised and globalized world while allowing the processing of personal data, including sensitive data, for scientific research. It reinforces cooperation duties and transparency between the actors of the processing, internally and with regard to the supervisory authorities, which should create a more integrated EU data protection system and diminish some useless administrative costs by decentralising elements of the data protection governance towards data controllers and processors. Whilst the GDPR adopts new specific provisions to ensure adapted data protection in research, the field remains widely regulated at national level, in particular, regarding the application of research participants' rights,

which some could regret. However, the GDPR has the merit to set up clearer rules that will positively serve the research practices notably regarding consent, regarding the rules for reusing personal data for another purpose, assessing the risks of data processing in the context of data protection impact assessment, adopting accountable management system of processing operations and building or reinforcing internal data protection competencies with the data protection officer. In addition, for the first time, the GDPR refers to the respect of ethical standards as being part of the lawfulness of the processing in research, what must be welcomed as an effort for sector-specific consistency. Finally, the GDPR opens new possibilities for structuring data sharing in scientific research with measures encouraging self-regulation development.

576

577

Box 3: Roadmap for the precision medicine initiative in the Nordic region

Develop the Nordic Society for Human Genetics and Precision Medicine, among whose tasks will be to:

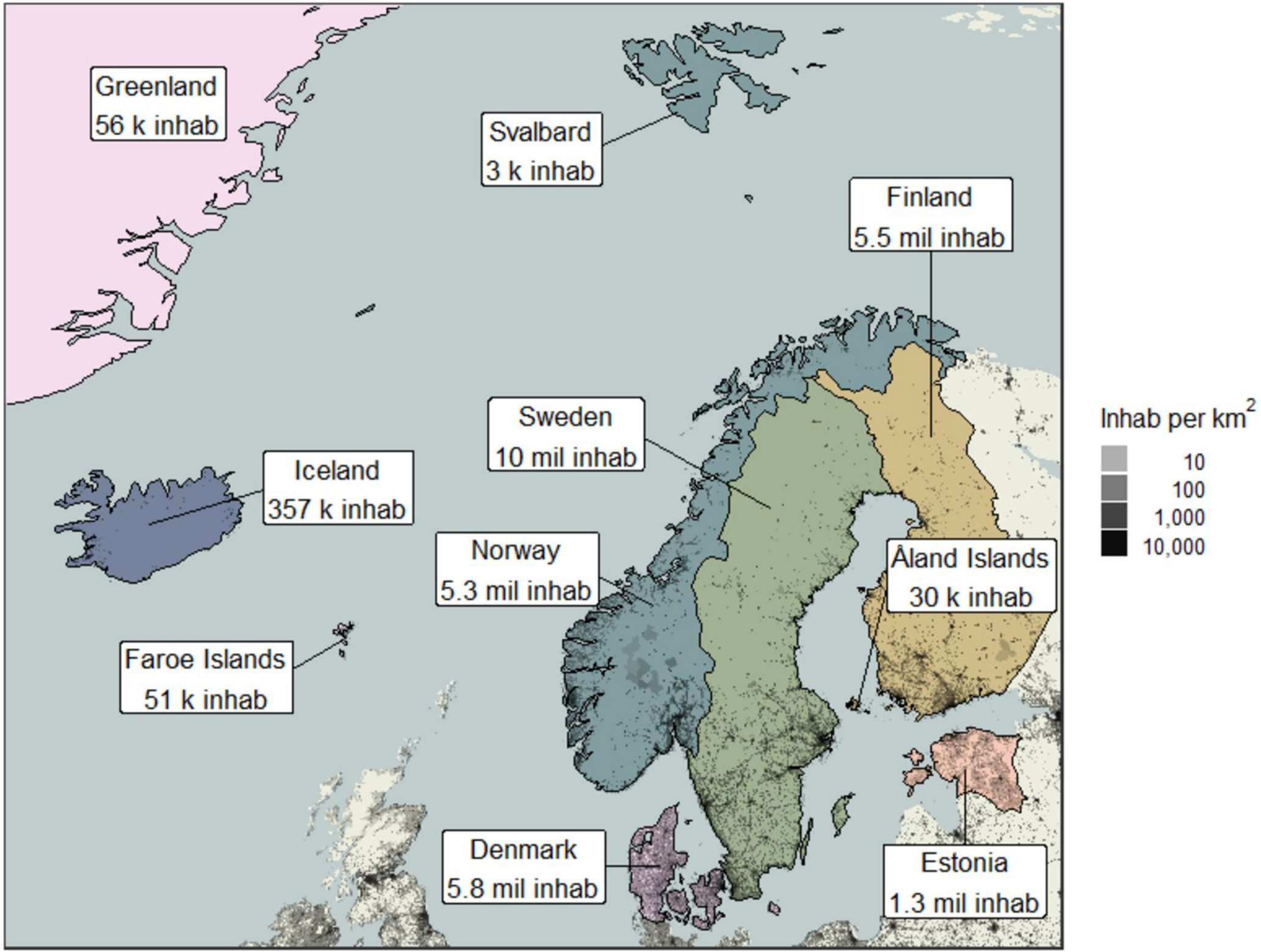
- Organize biannual large open scientific meetings
- Organize a series of workshops targeted to the constituencies of precision medicine, e.g., research geneticists, clinical geneticists, data scientists, legal scholars
- Write a white paper that summarizes the major needs
- Develop web-based resources, including a news feed, continuous updated overview of available cohorts, registers, and biobanks, as well as linked genomics and metabolomics information

Engage with important constituencies

- Policymakers
- The public
- Other organizations in this sphere

- Interact with funding partners
- NordForsk
- National research councils
- Private non-profit organizations and foundations
- Industry

578


Greenland
56 k inhab
Svalbard
3 k inhab
Finland
5.5 mil inhab
Sweden
10 mil inhab
Iceland
357 k inhab
Norway
5.3 mil inhab
Åland Islands
30 k inhab
Faroe Islands
51 k inhab
Denmark
5.8 mil inhab
Estonia
1.3 mil inhab
Inhab per km2
10
100
1,000
10,000

Roadmap for a Precision Medicine Initiative in the Nordic Region

Pål Rasmus Njølstad[1,2,3*], Ole Andreas Andreassen[4,5], Søren Brunak[6], Anders D. Børglum[7,8,9*], Joakim Dillner[10,11], Tõnu Esko[12,13], Paul W. Franks[14,15,16,17*], Nelson Freimer[18,19], Leif Groop[14,20], Hakon Heimer[21], David M. Hougaard[22,23], Eivind Hovig[24,25,26], Kristian Hveem[27,28,29], Anu Jalanko[30], Jaakko Kaprio[20], Gun Peggy Knudsen[31], Mads Melbye[32,33,34], Andres Metspalu[35], Preben Bo Mortensen[36,37], Juni Palmgren[38,39], Aarno Palotie[3,20,40,41,42,43*], Wenche Reed[44], Hreinn Stefánsson[45,46], Nathan O. Stitziel[47,48,49], Patrick F. Sullivan[50], Unnur Thorsteinsdóttir[47,48], Marc Vaudel[1], Eero Vuorio[51], Thomas Werge[52,53,22], Camilla Stoltenberg[30,54*] & Kári Stefánsson[45,46]

[1]Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway.

[2]Department of Pediatrics, Haukeland University Hospital, Bergen, Norway.

[3] Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Cambridge, MA ,USA.

[4]NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.

[5]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

[6]Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[7]Department of Biomedicine, Aarhus University, Aarhus, Denmark.

[8]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark.

[9]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark.

[10]Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

[11]Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden.

[12]The Estonian Genome Center, University of Tartu, Tartu, Estonia.

[14]Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

[14]Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre , Lund University , Malmö, Sweden.

[15]Department of Public Health and Clinical Medicine, Section for Medicine, Umeå University, Umeå, Sweden.

[16]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA.

[17]Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

[18]Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA.

[19]UCLA Center for Neurobehavioral Genetics, Los Angeles, CA, USA.

[20]FIMM Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

[21]School of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark .

[22]Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark.

[23]iPSYCH - The Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark.

[24]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.

[25]Institute of Cancer Genetics and informatics. Oslo University Hospital, Oslo, Norway.

[26]Department of Informatics, University of Oslo, Oslo, Norway.

[27]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway.

[28]HUNT Research Center, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway.

[29]Department of Medicine, Levanger Hospital, Levanger, Norway.

[30]Genomics and Biomarkers Unit, National Institute for Health and Welfare, Helsinki, Finland.

[31]Norwegian Institute of Public Health, Oslo, Norway.

[32]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

[33]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.

[34]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

[35]The Estonian Genome Center, University of Tartu, Tartu, Estonia.

[36]National Centre for Register-Based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark.

[37]iPSYCH - The Lundbeck Foundation Initiative for Psychiatric Research, Aarhus, Denmark.

[38]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[39]Swedish e-Science Research Center, SeRC, Stockholm, Sweden.

[40]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[41]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[42]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[43]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA.

[44]Department of Research, Innovation and Education, Oslo University Hospital, Oslo, Norway.

[45]deCODE genetics, Reykjavik, Iceland.

[46]University of Reykjavik, Iceland.

[47]Cardiovascular Division, Department of Medicine, Washington University, Saint Louis, MO, USA.

[48]Department of Genetics, Washington University, Saint Louis, MO, USA.

[49]McDonnell Genome Institute, Washington University, Saint Louis, MO, USA.

[50]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[51]Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland.

[52]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark.

[53]Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

[54]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway.

*Members of the writing group

Address correspondence to:

Professor Pål Rasmus Njølstad, M.D., Ph.D., Center for Diabetes Research, Hospital for Children and Adolescents, Haukeland University Hospital, Haukelandsbakken 15, NO-5021 Bergen, Norway (pal.njolstad@uib.no).

Contents

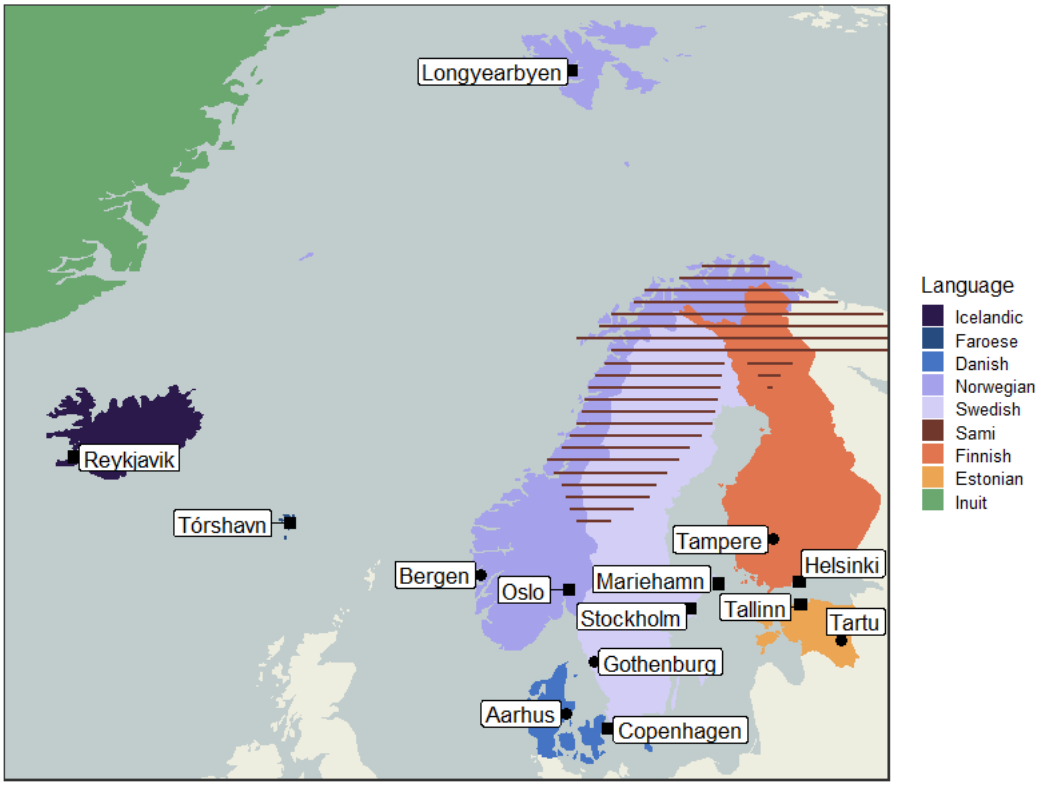
	Cardiovascular disease		Schizophrenia		Type 2 diabetes		Lipid Disorders		Glioblastoma		Alzheimer's disease		Melanoma		All	
	n (%)		*n (%)*		*n (%)*		*n (%)*		*n (%)*		*n (%)*		*n (%)*		*n (%)*	
Nordic	93,510	*(48%)*	7,051	*(40%)*	9,817	*(35%)*	42,094	*(22%)*	4,067	*(13%)*	6,094	*(8%)*	994	*(2%)*	163,627	*(28%)*
USA	7,018	*(4%)*	2,281	*(13%)*	3,767	*(13%)*	1,586	*(1%)*	18,914	*(59%)*	28,550	*(39%)*	8,951	*(21%)*	71,067	*(12%)*
UK	30,122	*(15%)*	4,407	*(25%)*	7,120	*(25%)*	22,746	*(12%)*	4,204	*(13%)*	12,010	*(16%)*	12,872	*(30%)*	93,481	*(16%)*
Other	63,777	*(33%)*	3,970	*(22%)*	7,646	*(27%)*	122,151	*(65%)*	4,769	*(15%)*	27,089	*(37%)*	20,390	*(47%)*	249,792	*(43%)*
Total	194,427		17,709		28,350		188,577		31,954		73,743		43,207		577,967	
References	23		24		25		26		27		28		29			

Supplementary Table 1: Contribution of Nordic populations to published GWAS meta-analyses

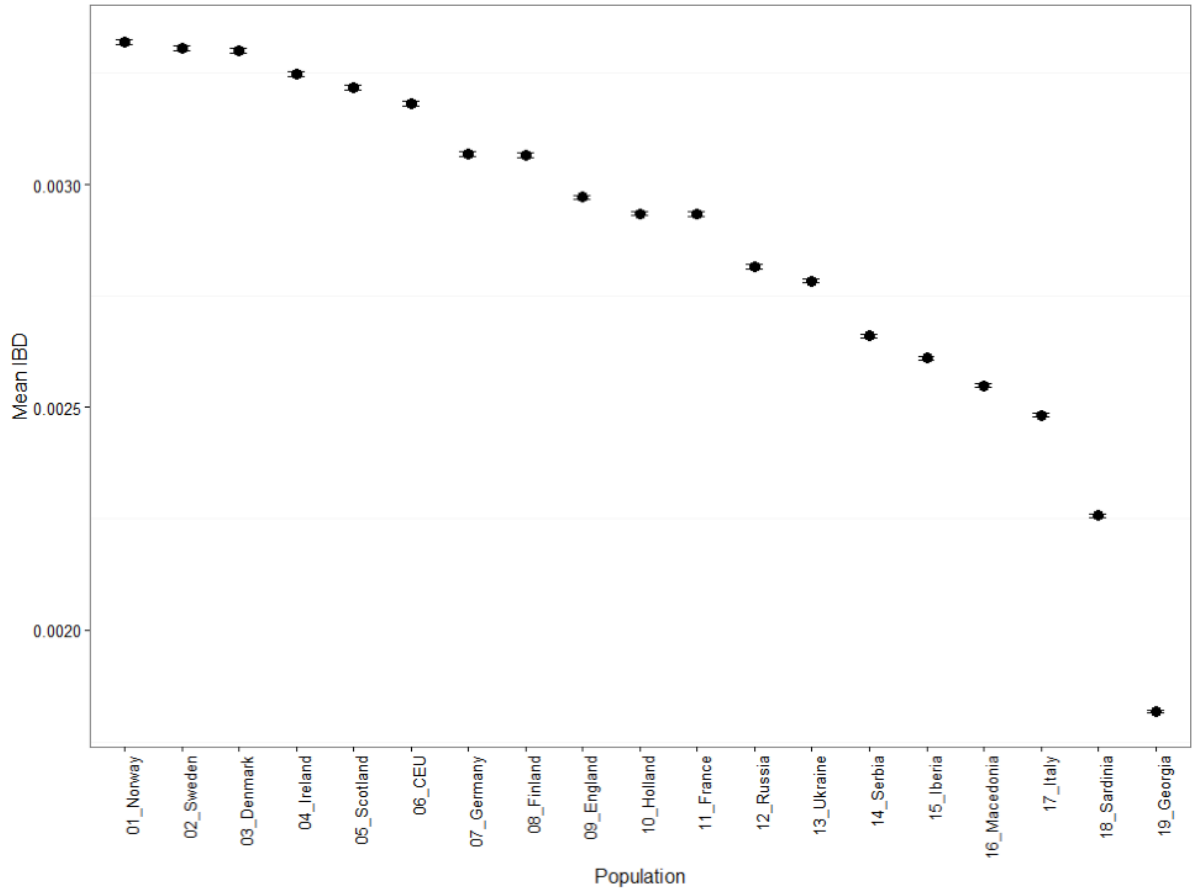
The extent to which patient samples are derived from the UK, USA, and Nordic countries for the major GWAS meta-analyses for seven diseases. Proportions were estimated from the European ancestry cohorts in each meta-analysis. In some rare instances, precise sample contributions could not be determined, as the contributing cohorts included participants from multiple countries; a best estimate was used in those cases based on the information provided in the supplementary materials and other cohort description papers.

Governance and legislation
Unique personal identification number
Public health care systems
National registries that can be linked
Large population-based and patient cohorts with deep phenotypes and biological samples
Biobanks with large databases based on analyses of the biobank samples

Supplementary Table 2: Nordic resources to address global health challenges



Supplementary Figure 1: Languages of the Nordic region.
The Nordic has three major linguistic families: 1) Danish, Faroese, Icelandic, Norwegian, and Swedish - rooted in the Old Norse language - belong to the North Germanic branch of the Indo-European languages and are displayed in shades of blue; 2) Finnish, Karelian, and Sami are part of the Finno-Ugric languages and displayed in a red to yellow gradient; 3) Inuit is a branch of the Eskimo-Aleut languages and is displayed in green. The capital of each country is shown by a filled square, while some the largest cities are shown by filled circles (**Supplementary Note**; European Commission, Joint Research Centre, Columbia University, NYC, NY).



Supplementary Figure 2: Common variant sharing shows similarity of Nordic populations

A comparison of 22,500 Icelanders typed for the Illumina Omni Express SNP chip against 25 to 100 individuals with SNP chip data from each of 19 different populations of European ancestry – based on a common set of 144,248 SNPs. For each pair of individuals, we calculated the proportion of the genome shared in fragments longer than 2cM that are identical by descent (IBD). The figure shows the mean proportion of the genome shared IBD between the Icelanders and each of the 19 populations, with 95% confidence intervals. As expected, the greatest degree of sharing is with populations from Scandinavia and the British Isles. The fragments shared between Icelanders and these populations are longer because of more recent common ancestry, which implies a greater propensity to share rare mutations.

Supplementary Note: Method for generation of Figure 1 in the printed article

Figure 1 of the published article: This map was generated using a script provided in the link below. In short, the world map using the Mollweide projection was plotted in light yellow over a light blue background. The countries of the Nordic region, composed here of Estonia, Finland, Denmark, Iceland, Norway, and Sweden, and associated territories, the Faroe Islands, Greenland, Svalbard, and the Åland Islands, were plotted in different colors and outlined in black.

Population density as of 2015 was obtained from the Global Human Settlement project at a resolution of 1 km (European Commission, Joint Research Centre [JRC]; Columbia University, Center for International Earth Science Information Network - CIESIN [2015]: GHS population grid, derived from GPW4, multitemporal [1975, 1990, 2000, 2015]. European Commission, Joint Research Centre [JRC]).

The density matrix was rasterized in 2000 bins in latitude and longitude over the Nordic region, aligned onto the world map, and displayed in shades of grey. Finally, the name along with the number of inhabitants was annotated for each country and associated territory. The number of inhabitants was the latest available in Wikipedia at time of writing.

Supplementary Figure 1: This map was generated using the script in the link below. In short, the world map using the Mollweide projection is plotted in light yellow over a light blue background as in the population map.

The Nordic countries and associated territories were colored according to the official language. The Sami language was annotated with horizontal segments. Capitals and major cities were annotated with points and names using their latitude and longitude as obtained from latlong.net. North Germanic languages were annotated using a gradient of blue, Finno-Ugric languages using a red to yellow gradient, and Inuit as part of Eskimo-Aleut languages in green.

URLs:
Script: https://github.com/mvaudel/Nordic-maps
Global Human Settlement project: https://ghsl.jrc.ec.europa.eu/ghs_pop.php
European Commission, Joint Research Centre (JRC): http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a)