nature portfolio

Peer Review File

Genetic architecture in Greenland is shaped by demography, structure & selection

Corresponding Author: Dr Anders Albrechtsen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
The manuscript by Staeger et al describes the genetic architecture of Greenlanders using a population specific WGS reference panel of 448 individuals and GWAS data from 5996 individuals. The manuscript is very interesting providing a real deep dive in the genetic landscape of Greenland. It nicely summarizes new analyses, the current knowledge and results from previous work from the modern Greenland population. It also expands our understanding of the population structures and history worldwide.

The main contribution is the generation of a Greenlandic reference genome data and an imputation panel. This reviewer wonders if this reference genome data is available to other resaerchers, in the “Data Availability” section only MEGA-chip data is listed.

The main limitation of the manuscript is the somewhat modest novelty of the results. As the main results are contributed by the GWAS data that has been reported by the same, as well as other groups in a number of previous papers (many of them referenced in a recent review by the group Dan Med J 2023, 70(12)), quite expectedly the new WGS, although valuable, provides limited opportunities for new discoveries. Another limitation is the limited phenotype information of the participants and thus no new disease or health associations are listed.

Minor comments:

Line 254: There is a typo, it seems that part of the sentence is missing “…resulted in 0.7 mio. more reliably…”

Supplementary Figure 1: The lines in some of the graphs are hard to distinguish from one another, please consider improving the contrast between the lines.

Supplementary Figure 4: similar problem as in Figure 1, lines are hard to separate from each other

Referee #2

(Remarks to the Author)
I appreciate the opportunity to review this fascinating story that was more than two decades in the making by this international group of researchers, largely based in Greenland and Denmark. It is of note that the research was approved by the Scientific Ethics Committee in Greenland, and through public engagement it was determined that research to investigate hereditary causes of health and disease is broadly supported. Through a health equity lens, the authors emphasize the importance of inclusion of Indigenous populations in genetic health care and research.

Combining several cohort studies (dating from 1999) a total of 5996 Greenlanders participated. Whole genome sequencing was carried out on 448 samples from a recent cohort and an additional 5548 samples with SNP-array genotyping allowed

for imputation, providing a robust sample size for representation of a population of ~56,000. There is further strength in the recruitment methods reflecting random sample collection throughout Greenland.

Understanding the genetic architecture in the context of specific population dynamics and known clinical variants is not only interesting it is useful information clinically. The work described in this manuscript is a model for understanding the clinical effects of population bottlenecks on complex disease and the impact of recently arising variants on clusters of autosomal recessive disease, both informing precision medicine.

It is well known that polygenic risk scores are less relevant in populations poorly represented in genomic population databases, however less well known to what degree addition of population specific data can improve the predictions. This work contributes to that discussion. Although it is also well known that lack of representation within genomic reference databases results in an excessive number of potential disease causing variants after filtering, this group has shown that adding Greenlandic specific genomic data reduced the number of non-causal candidate variants by six-fold, an important finding for those involved in rare disease diagnosis.

The manuscript is well written, and accessible to the non-expert reader. The concepts are well explained and the methods are clear. I appreciate the figures presented, the associated legends and the text referring to them. However, there are some figures that could use clarification/expansion. Specifically, Figure 3c might be presented with more clarity, and perhaps is better in the supplementary data. On the other hand, Figure 3 in the supplementary data, is broadly relevant and could be presented in the main manuscript.

Presumably long read sequencing was not carried out. Do the authors see this as a limitation?
Over-all, I believe this group has provided a unique and relevant contribution demonstrating the clinical and scientific benefit of including Indigenous populations, including those that are small, in genomic studies. Based on this work, the Greenland population may be especially suited to genomic based clinical care/precision medicine. Would the authors like to comment on this potential?

Referee #3

(Remarks to the Author)
Isolated populations that have undergone founder events (i.e. prolonged bottlenecks) are wellsprings for discovering genetic variants with large phenotypic effects. This is because the population bottleneck allows variants that would have been keep at low frequency are allowed to drift to higher frequencies because the strength of genetic drift is much higher during a bottleneck. Here, the authors leverage the population history of Greenlanders to discover large effect common variants that affect a variety of traits and diseases, including Type 2 Diabetes as well as plasma proteins. They find that demographic history has dramatically altered the genetic architecture of complex traits in this population, skewing the allele frequency spectrum towards common variants. They apply these insights toward improving polygenic risk scores and improving Mendelian disease screening. The authors also study the impacts of fine-scale structure on patterns of genetic relatedness, which has implications for genetic mapping.

This study is novel and compelling and I believe it is an excellent candidate for publication in Nature. It provides a clear case for why population history is important in the context of medical genetics and provides a foundation for future genetic work not only in Greenland, but also across diverse populations around the world. The manuscript is well-written and concise and the article appropriately references previous work. Error bars/confidence intervals are provided where needed. Below I provide some comments on specific aspects of the work that I believe can be improved.

Major comments:
- Line 191: the authors compare rare pLoF variants when using Gnomad as the reference versus including 448 Greenlandic WGS samples. Could the authors confirm that there are no individuals with monogenic diseases in their cohort to ensure there are no false negatives?

- The authors argue that burden tests are not very useful in their cohort because the entire site frequency spectrum is shifted towards common variants due to the population bottleneck. This has the effect of making the genetic architecture at a single gene skewed towards a single common variant explaining most of the burden, rather than many rare variants explaining the genetic architecture. The latter case is when burden tests are expected to be most informative. I'm confused by this analysis because when there are common variants at a gene, I believe that a single marker test will always be more powerful than a burden test, so it's not clear to me that this analysis is adding much. I would be happy for the authors to provide an explanation about the value of this analysis. I am also unsure of the arbitrary 50% and 90% cutoffs for when the gene burden is informative or uninformative. Perhaps a continuous scale could be used here?

- Line 163: Given the rate of increase of SNP discovery based on sample sizes (Fig. 1f), I wonder if the authors could project out when SNP discovery is expected to no longer yield large gains? This would be useful for future study design in Greenland.

- The authors perform GWAS using standard corrections for population structure/relatedness. That is, using a genetic relatedness matrix as a random effect in a linear mixed model. Recent work suggests that fine-scale population structure may be unaccounted for by these methods (e.g. see PMID 35995948, 32691046). Given the authors' focus on fine-scale

population structure and disease burden, I would be interested to know if the GWAS results change substantially when the covariance matrix based on IBD from NGSremix is used rather than the standard common variant GRM.

- In Fig 2f, what are the error bars denoting? Is the difference between Arctic-specific and PGS derived in Europeans significant? If the authors have the space to do so, I would be interested to see a breakdown of how this varies by genetic architecture of the trait (that is, SNP-heritability and polygenicity). It seems that the main thing going on is adding in large effect common variants, which is quite different from the factors affecting polygenic risk score transferability across other populations (that is, LD differences causing causal allele tagging to differ between populations). Is it true that the traits that improve the most are those that have large effect common variants that are Arctic-specific?

-The authors study positive selection in their sample using Relate. It was not clear to me if the P-values reported in Figure 3f were corrected for multiple hypothesis test? If so, it should be stated. If not, please revise.

Minor comments:
- Line 122: Could the authors note the average coverage for the WGS samples and note the genotyping array for the 5548 individuals? This would help readers.
- Figure 1b caption: which population is the minimum allele frequency computed in?
- Line 188: Are the plus minus values standard error? standard deviation?
- Line 254: "mio" is not a standard abbreviation. Please revise this.
- Line 304: does the phrasing "mainly" additive and recessive mean something specific? If so, this should be clarified. If not, this should be removed.

Referee #4

(Remarks to the Author)
In this article the authors conduct a variety of population genetic analyses on 448 Inuit-ancestry Greenlanders via WGS data. With the increased call and necessity for greater ancestral diversity in the human genetics field, this paper is timely in that it focuses on both an underrepresented population and one that has a unique demographic history; as the authors explain, this Greenland population has experienced both a prolonged bottleneck (versus the shorter bottlenecks of Finland and Iceland) and admixture between Native American and European ancestries. As a result, and also as the authors show, the only way to truly represent such ancestries in human genetic analyses is by incorporating samples from the populations themselves. Specifically, the authors show that their Greenland cohort contains more high impact variants than a European-ancestry cohort for both metabolic and circulating-protein traits. They also show that European-trained metabolic polygenic scores and clinical screening with public references both perform poorly for their intended purposes, but that including Arctic-specific variants (for the former) or Greenland reference samples (for the latter) improves performance in both instances. Lastly, the authors also discuss the fine population structure they identify and the impact this has on current and future disease risk within Greenland.
I think this is a strong manuscript that shows a variety of interesting results that are properly supported with appropriate evidence. I do not have any major points; however, I do have a few minor points to help improve the accessibility and readability of the paper for a broader audience like Nature.

Minor Points:
1) For Figure 1b, I would either use ‘dbSNP’ instead of ‘rsID’ in the figure legend or say in the caption “…only found in dbSNP (ie has an rsID)…” to help ensure readers understand that by being in dbSNP there is or should be an rsID.

2) At the end of the ‘Clinical screening to diagnose monogenic diseases’ section, I would include a statement that explicitly links the drop in non-causal pLoF variants to more Arctic-ancestry variants being filtered out due to passing the allele frequency threshold now. I’m not sure if this will be as clear to a non-population genetics audience without an additional note explicitly stating that having Greenlander individuals as part of the reference data, vs. just gnomAD, leads to more SNPs being revealed as common.

3) On lines 197-167, for “The pattern is consistent independent of the maximum MAF threshold used for filtering (Supplementary Fig. 3).”, I think you need to show the full range of MAFs in the supplementary figure to fully support the statement that the pattern is independent of maximum MAF. I’d either add a second, 3b supplementary figure that isn’t zoomed in on the rare MAF thresholds (which in application are indeed the most relevant) or adjust the statement to something like “The pattern is consistent independent of a range of maximum MAF thresholds…”.

4) On lines 221-222, there should be a reference for this statement: “This approach is more powerful than GWAS in the presence of multiple causal rare variants.”

5) On line 227, I do not think ‘constrained gene’ has been defined yet in the main text. How ‘constrained’ is being defined here should be mentioned.

6) On line 240 there should be commas before and after “…consistent with previous findings…”

7) At the end of the “Linkage disequilibrium related consequences” section looking at the impact on imputation, is there any way to give a frame of reference for how meaningful .7M extra variants are? Does adding .7M more SNPs to a set of already

>6M really imply "…high quality imputation is only possible with representation of Greenlanders in the reference panel"? Looking at Figure 2d, it seems like there is a greater improvement in identifying high-quality rare variants when adding in the Greenland data than 1kG alone – is that something worth highlighting to help support the above point?

8) Two points on the "Comparative GWAS and polygenic score performance on metabolic traits" section. First, I would define incremental R2 here if it is going to be used. Second, I wonder if using relative vs. absolute R2 values here will be confusing for readers. I'm guessing relative is being used so that you can average across all 13 traits? I see in the supplementary figure you're using the absolute values, which is more straight forward in terms of interpretation. I would maybe include a note explicitly stating that relative is being used so that all traits can be compared and/or that readers should look at the supplement to see the absolute values. One of my concerns is that for readers not familiar with the PRS/PGS literature, they may misinterpret these "mean relative incr. R2" values as variance explained, whereas we know that PRS/PGS do not reach values anywhere close to 100%.

9) For lines 342-344 in the "Recent change in mobility" section, could the authors provide a frame of reference for these PO/FS numbers to further help show how they are higher in the Greenland cohort? For instance, maybe you could pull these same numbers from the UKB or FinnGen? I just think the raw numbers may stand out more with something to compare against.

10) There's something about the "Predicted consequences for disease susceptibility" section that I am having troubles following. Is the main reason we do not expect to see a similar reduction in future disease risk for CSID and monogenic diabetes is because they have more similar regional allele frequencies than the other three? Is there a metric you could use to show this across the five variants, like average and SD of the allele frequency across each region for each variant?

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors have responded to the critique adequately.

Referee #2

(Remarks to the Author)
I appreciate the comprehensive responses and changes to the manuscript based on the reviewer comments.

I believe this will be an important contribution, highlighting the genetic architecture and journey of this unique population while integrating the impact of key variants affecting health currently, with projections for clinical and population relevance in the future.

As stated in their rebuttal on novelty, 'The quantification of these differences between populations is novel, and important for designing and understanding the value of medical genetics studies in isolated populations'. I agree with this statement, and look forward to the final publication.

Referee #3

(Remarks to the Author)
The authors have addressed the points I raised in my review and I am happy with the manuscript in the current form.

Referee #4

(Remarks to the Author)
Thanks to the authors for their timely revisions and resubmission. All of my previous points have been addressed. In my view the manuscript is now ready for publication.

First, we would like to thank both the editor and reviewers for a thorough review of the manuscript. Your comments and suggestions have improved the work.

Point-by-point response to editorial suggestions

One thing we would like to highlight is the need for a more detailed and transparent data availability statement, in which any restrictions on data sharing are clearly explained. Please highlight all changes in the manuscript text file.

We have now started the process of uploading the remaining genetic data to the European Genome-phenome Archive and included additional information on how to apply for ethical approval:

Original sentence: "Genetic data from the MEGA-chip for 4,607 of the Greenlandic individuals is available at the European Genome-phenom Archive, accession number EGAD00010002057."

New sentence: "Genetic data is archived at the European Genome-phenome Archive; MEGA-chip data in IHIT/B99 (EGAD00010002057), MEGA-chip in B2018 (Accession number will be inserted when received from EGA), and whole genome sequencing data (Accession number will be inserted when received from EGA). Use of the data is contingent on approval by the Research Ethics Committee of Greenland (nun@nanoq.gl) and subsequent acceptance by the dataowner (Department of Health, Greenland Government; pn@nanoq.gl)."

Point-by-point response to referees' comments

Referee #1. Human genetics

The manuscript by Staeger et al describes the genetic architecture of Greenlanders using a population specific WGS reference panel of 448 individuals and GWAS data from 5996 individuals. The manuscript is very interesting providing a real deep dive in the genetic landscape of Greenland. It nicely summarizes new analyses, the current knowledge and results from previous work from the modern Greenland population. It also expands our understanding of the population structures and history worldwide.

We thank the reviewer for reviewing our work and the positive comments on the manuscript.

The main contribution is the generation of a Greenlandic reference genome data and an imputation panel. This reviewer wonders if this reference genome data is available to other resaerchers, in the "Data Availability" section only MEGA-chip data is listed.

We have now started the process of uploading the remaining genetic data to the European Genome-phenome Archive and included additional information on how to apply for ethical approval:

Original sentence: "Genetic data from the MEGA-chip for 4,607 of the Greenlandic individuals is available at the European Genome-phenome Archive, accession number EGAD00010002057."

New sentence: "Genetic data is archived at the European Genome-phenome Archive; MEGA-chip data in IHIT/B99 (EGAD00010002057), MEGA-chip in B2018 (Accession number will be inserted

when received from EGA), and whole genome sequencing data (Accession number will be inserted when received from EGA). Use of the data is contingent on approval by the Research Ethics Committee of Greenland (nun@nanoq.gl) and subsequent acceptance by the dataowner (Department of Health, Greenland Government; pn@nanoq.gl).”

The main limitation of the manuscript is the somewhat modest novelty of the results. As the main results are contributed by the GWAS data that has been reported by the same, as well as other groups in a number of previous papers (many of them referenced in a recent review by the group Dan Med J 2023, 70(12)), quite expectedly the new WGS, although valuable, provides limited opportunities for new discoveries. Another limitation is the limited phenotype information of the participants and thus no new disease or health associations are listed.

It is true that we have not analysed any new phenotypes in this cohort and thus have not reported novel associations. In this manuscript, we have focused on analysing the overall genetic architecture and its consequences for medical genetics in Greenland. To our knowledge, this is the first report and analysis of whole genome sequencing data from any Inuit population. Through a range of analysis we describe how the historic bottleneck of the Greenlandic population has shaped the genetic architecture of disease, and also how it differs from the genetic architecture of all other investigated populations of the world. The quantification of these differences between populations is novel, and important for designing and understanding the value of medical genetics studies in isolated populations.

Minor comments:

Line 254: There is a typo, it seems that part of the sentence is missing “…resulted in 0.7 mio. more reliably…”

We agree with the reviewer that this sentence is difficult to read. We have tried to make it more clear in the revised manuscript:

Original sentence: “We imputed the genomes of 5,548 SNP-chipped Greenlandic individuals which resulted in 0.7 mio. more reliably called common variants when including the Greenland WGS individuals in the reference panel compared to using only the 1KG populations (INFO>0.8 & MAF>5%; Only 1KG=6.2 mio., Greenland+1KG=6.9 mio.)”

New sentence: “We imputed the genomes of 5,548 SNP-chipped Greenlandic individuals, with the Greenland WGS individuals added to the reference panel. This resulted in an additional 1.7 million accurately called (INFO score>0.8) variants including 0.7 million common (MAF>5%).”

Supplementary Figure 1: The lines in some of the graphs are hard to distinguish from one another, please consider improving the contrast between the lines.

We agree with the reviewer that some populations are difficult to distinguish. To improve this, we have annotated each panel with the population labels. We have chosen to keep the 1000 genomes colour scheme such that the super populations are easy to distinguish in panel a. and such that the colours are directly comparable to the 1000 genomes paper (PMID: 26432245).

Supplementary Figure 4: similar problem as in Figure 1, lines are hard to separate from each other

We have changed the Supplementary Figure 4 (now Supplementary Fig. 6) in the same way as supplementary figure 1 mentioned directly above.

Referee #2. Medical genetics

I appreciate the opportunity to review this fascinating story that was more than two decades in the making by this international group of researchers, largely based in Greenland and Denmark. It is of note that the research was approved by the Scientific Ethics Committee in Greenland, and through public engagement it was determined that research to investigate hereditary causes of health and disease is broadly supported. Through a health equity lens, the authors emphasize the importance of inclusion of Indigenous populations in genetic health care and research.

We thank the reviewer for highlighting the ethics part of the study. This is something we continuously work on improving.

Combining several cohort studies (dating from 1999) a total of 5996 Greenlanders participated. Whole genome sequencing was carried out on 448 samples from a recent cohort and an additional 5548 samples with SNP-array genotyping allowed for imputation, providing a robust sample size for representation of a population of ~56,000. There is further strength in the recruitment methods reflecting random sample collection throughout Greenland.

Understanding the genetic architecture in the context of specific population dynamics and known clinical variants is not only interesting it is useful information clinically. The work described in this manuscript is a model for understanding the clinical effects of population bottlenecks on complex disease and the impact of recently arising variants on clusters of autosomal recessive disease, both informing precision medicine.

It is well known that polygenic risk scores are less relevant in populations poorly represented in genomic population databases, however less well known to what degree addition of population specific data can improve the predictions. This work contributes to that discussion. Although it is also well known that lack of representation within genomic reference databases results in an excessive number of potential disease causing variants after filtering, this group has shown that adding Greenlandic specific genomic data reduced the number of non-causal candidate variants by six-fold, an important finding for those involved in rare disease diagnosis.

We thank the reviewer for accurately summarising the study design and highlighting the importance of population specific data.

The manuscript is well written, and accessible to the non-expert reader. The concepts are well explained and the methods are clear. I appreciate the figures presented, the associated legends and the text referring to them. However, there are some figures that could use clarification/expansion. Specifically, Figure 3c might be presented with more clarity, and perhaps is better in the supplementary data. On the other hand, Figure 3 in the supplementary data, is broadly relevant and could be presented in the main manuscript.

We thank the reviewer for the encouraging words and agree with the comments on the two figures. As suggested by the reviewer we have moved the original supplementary figure 3 into figure 2b and expanded the analysis with a figure showing different thresholds from the Greenlandic reference in

what is now supplementary figure 5 (see answer to reviewer 3, major comment 1). We have also tried to both simplify and clarify figure 3c by removing the analysis about 'within regions', changing the title from originally "Parent-offspring and Sibling relations" to "Number of Parent-offsprings and Siblings per 100k pairs", and adjusted the text accordingly:

Original sentences in main text: "The connectivity between regions differed quite a bit and geographically adjacent regions were better connected than regions further apart (Fig. 3c) with the exception of Nuuk that was well connected with longer distances. In three of the regions, we had sample locations within both towns and villages and from the PO pairs, we observed an asymmetry between locations indicating that people tend to move from smaller villages into the larger towns (Fig. 3c) whereas movement between towns was symmetric. Taken together, the strong genetic clustering indicates that the regions have been very isolated historically, but within the last few generations the regions have become very connected, suggesting that the population is becoming more panmictic."

New sentences in main text: "The connectivity between regions, measured as the number of close relative pairs per 100k pairs of individuals, differed quite a bit and geographically adjacent regions were better connected than regions further apart (Fig. 3c) with the exception of Nuuk that was well connected across longer distances. Taken together, the strong genetic clustering indicates that the regions have been very isolated historically, but within the last few generations the regions have become very connected, especially Nuuk, suggesting that the population is becoming more panmictic."

Original sentences in figure caption: "Inferred parent-offspring and full sibling relations per 100k possible relationships between regions. Grey lines indicate siblings, and coloured lines represent parent offspring, where the colour indicates from which region the parent was sampled. Within regions is a similar plot but for regions where we had sample locations from both town and village and one example with two towns."

New sentences in figure caption: "Number of parent-offspring and full sibling relations inferred from genetic data per 100k possible relationships between each pair of regions. Grey lines represent siblings with the line width indicating the inferred number of sibling relationships per 100k possible relationship pairs. Coloured lines represent parent-offspring, where the colour indicates from which region the parent was sampled and the line width indicates the inferred number of sibling relationships per 100k possible relationship pairs."

Presumably long read sequencing was not carried out. Do the authors see this as a limitation?

In the manuscript we only analysed short read data. However, we agree with the reviewer that investigating the architecture of larger structural variants using data from long read sequencing would be of interest and that the structural variants also contribute to the genetics architecture. We have added the following sentence to the conclusion, to acknowledge this:

"Our analyses of genetic architecture were limited to SNPs and small indels identified through short read sequencing, but we expect the genetic architecture to affect other inherited variants similarly."

Over-all, I believe this group has provided a unique and relevant contribution demonstrating the clinical and scientific benefit of including Indigenous populations, including those that are small, in

genomic studies. Based on this work, the Greenland population may be especially suited to genomic based clinical care/precision medicine. Would the authors like to comment on this potential?

Precision medicine is of great importance across populations, in order to identify individuals in greatest need of treatment, and to ensure the most efficient treatment for each individual. We agree with the reviewer that the current and previously reported results suggest that genomic-based precision medicine could be particularly effective in the Greenlandic population. Clinical screenings would be markedly improved by including the Greenlandic reference panel, and this could be implemented with limited costs. Also, relatively common single variants with large effects in the Greenlandic population provide the potential for early prevention and treatment for diabetes (*TBC1D4* p.R684*; *HNF1A* c.1108G>T), hypercholesterolemia (*LDLR* p.G137S), and Congenital sucrase-isomaltase deficiency (*SI* c.273_274delAG), respectively.

We have added the following sentence to the conclusion to highlight the potential for precision medicine in Greenland:

"Importantly, the variants, like the previously reported *TBC1D4, LDLR*, and *SI* variants, not only have large effects, but are also common. This means that the variants are easier to study further than rare or low effect variants and that they provide great potential for genetic-based early prevention and treatment at population level."

Referee #3. Population and evolutionary genomics

Isolated populations that have undergone founder events (i.e. prolonged bottlenecks) are wellsprings for discovering genetic variants with large phenotypic effects. This is because the population bottleneck allows variants that would have been keep at low frequency are allowed to drift to higher frequencies because the strength of genetic drift is much higher during a bottleneck. Here, the authors leverage the population history of Greenlanders to discover large effect common variants that affect a variety of traits and diseases, including Type 2 Diabetes as well as plasma proteins. They find that demographic history has dramatically altered the genetic architecture of complex traits in this population, skewing the allele frequency spectrum towards common variants. They apply these insights toward improving polygenic risk scores and improving Mendelian disease screening. The authors also study the impacts of fine-scale structure on patterns of genetic relatedness, which has implications for genetic mapping.

This study is novel and compelling and I believe it is an excellent candidate for publication in Nature. It provides a clear case for why population history is important in the context of medical genetics and provides a foundation for future genetic work not only in Greenland, but also across diverse populations around the world. The manuscript is well-written and concise and the article appropriately references previous work. Error bars/confidence intervals are provided where needed. Below I provide some comments on specific aspects of the work that I believe can be improved.

We thank the reviewer for the encouraging words on the manuscript and the important comments which we believe have improved the manuscript as described below.

Major comments:

- Line 191: the authors compare rare pLoF variants when using Gnomad as the reference versus including 448 Greenlandic WGS samples. Could the authors confirm that there are no individuals with monogenic diseases in their cohort to ensure there are no false negatives?

The reviewer raises an important question which requires a bit of elaboration. The individuals used for estimating allele frequencies in all the cohorts (gnomAD, 1000 genomes, and the Greenlandic WGS) were not screened for all monogenic diseases. Some of the cohorts used in gnomAD may have been screened and their findings reported in ClinVar. However, this will only be for a small number of diseases and thus individuals with monogenic disease will indeed be present in all of the cohorts. Likewise, the Greenlandic cohort has only been screened for monogenic forms of diabetes (PMID: 36649380), which is also reported in ClinVar, but there will be causal variants affecting other monogenic genetic diseases.
This is why the allele frequency threshold used in screening is not exactly 0% but slightly above (0.1% in main analysis). The appropriate threshold to use in each population depends entirely on the prevalence of the disease in the respective population, but 0.1% is a common choice for rare dominant and additive diseases. In general, to ensure no false negatives, the choice of MAF threshold will be close to half of the disease prevalence since high penetrant alleles at higher frequency would result in a higher disease prevalence.
In the revised manuscript figure 2b and supplementary figure 5a, we now show that the difference in effectiveness of screening is consistent across MAF thresholds. Note that we use the same threshold on allele frequencies from all populations. However, since disease prevalence can be different between populations, the thresholds should ideally be adjusted per population accordingly when screening for a specific disease. This is particularly important for small populations since genetic drift has a stronger effect and disease prevalence can differ to a larger extent from the larger populations represented in the databases. Therefore, we have now also included a figure (supplementary figure 5b) where the MAF threshold used in the Greenlandic reference panel is different to the threshold used in the other reference populations.

We have made the following changes to address both this comment, reviewer 2 first comment, and reviewer 4 minor point 3:

1. We have moved the original supplementary figure 3 into main figure 2b. (reviewer 2 first comment)
2. We have included the full range of MAF thresholds in supplementary figure 5a. (reviewer 4 minor point 3)
3. We have added results from using different Greenlandic reference MAF thresholds than in the other reference populations in supplementary figure 5b.
4. We have tried to clarify this in the text by adding the following sentences at the end of the 'Clinical screening to diagnose monogenic diseases' results section:
   "Note that these analyses were carried out with the same MAF threshold across all populations, but for a specific disease, the thresholds could be adjusted according to the different disease prevalences of each population. Results from varying the thresholds in the Greenlandic reference while keeping the gnomAD threshold fixed is presented in Supplementary Fig. 5b."

- The authors argue that burden tests are not very useful in their cohort because the entire site frequency spectrum is shifted towards common variants due to the population bottleneck. This has the

effect of making the genetic architecture at a single gene skewed towards a single common variant explaining most of the burden, rather than many rare variants explaining the genetic architecture. The latter case is when burden tests are expected to be most informative. I'm confused by this analysis because when there are common variants at a gene, I believe that a single marker test will always be more powerful than a burden test, so it's not clear to me that this analysis is adding much. I would be happy for the authors to provide an explanation about the value of this analysis. I am also unsure of the arbitrary 50% and 90% cutoffs for when the gene burden is informative or uninformative. Perhaps a continuous scale could be used here?

The reviewer's summary and reasoning is completely correct. Our goal with the analysis was simply to illustrate and quantify to what extent the genetic architecture at a single gene is skewed towards a single common variant. The somewhat arbitrary thresholds were picked to give an example of how to quantify it and to make this quantification easier to digest for the reader.
We realise that we did not get these points across as clearly as we wanted and have now moved the text about gene burden to the end of the section called "The genetic architecture of Greenlandic Inuit" to make the natural connection to the site frequency spectrum clearer. We have also rephrased the text so it is now mainly about how in Inuit most genes will have the majority of the gene burden coming from a single variant (without discussing gene burden testing) and so it is explicitly stated that the arbitrary thresholds are simply there to provide quantification examples. Finally, we have moved fig 2b to figure 1g and added supplementary figure 4 with additional categories for completeness as well as results then using a 80% threshold to illustrate that we get similar results for other (arbitrary) thresholds than 90%.

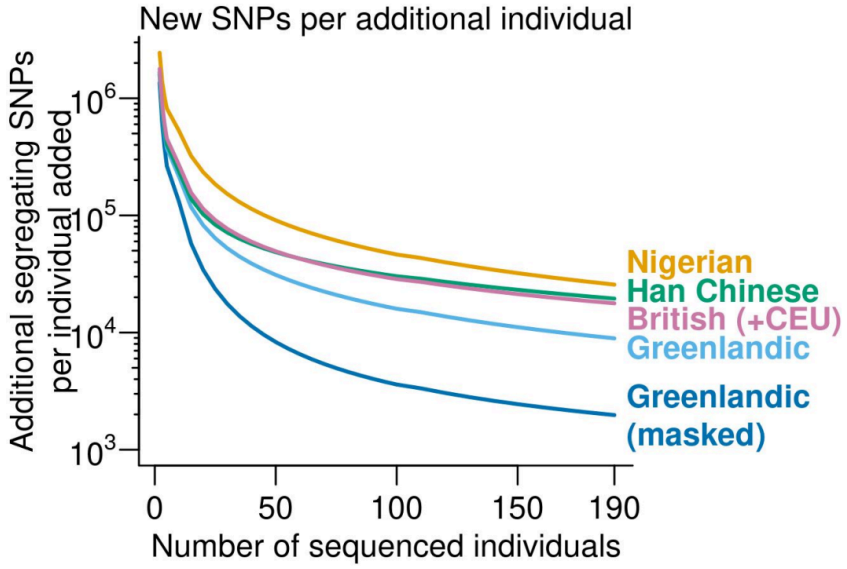
Original text:
"Gene burden testing
One approach to disease mapping is gene burden testing where the combined effect of multiple predicted deleterious variants within a gene is tested. This approach is more powerful than GWAS in the presence of multiple causal rare variants. To fairly compare the ability of performing burden tests in Greenland, East Asia, and Europe we constructed a gene burden based only on predicted missense or pLoF SNPs that are rare in African populations. We classified a gene burden as 'informative' if the most common SNP contributed less than 50% of the total gene burden and 'uninformative' when the most common SNP contributed more than 90% of the burden. Out of 9,533 constrained genes, we found an informative gene burden in 1.4% (95%CI: 1.2%-1.7%) of constrained genes for the unadmixed Greenlandic samples compared to 11.7% (95%CI: 11.1%-12.4%) and 12.7% (95%CI: 12.0%-13.3%) in the European and East Asian populations, respectively (Fig. 2b). In contrast, 20.0% (95%CI: 19.2%-20.8%), 13.8% (95%CI: 13.1%-14.5%), and 12.5% (95%CI: 11.9%-13.2%) of the constrained genes were dominated by a single common SNP in the unadmixed Greenlandic, European, and East Asian population, respectively (Fig. 2b). In the remaining genes, the most common SNP contributed between the two categories (50-90%), or no missense or pLoF variants were found. Combined with the overall distribution of allele frequencies, we saw that rare variants contributed less to diseases in Greenland compared to in Europe or East Asia and thus gene-burden tests have a limited potential in Greenland."

New text:
"In line with the relatively high proportion of common SNPs in Greenlandic (masked) Inuit, we found that for unadmixed Greenlandic individuals, a large proportion of constrained genes had a gene burden that is dominated by a single common SNP. E.g. if we estimate gene burden as the number of individuals that carry at least one predicted deleterious allele (predicted missense or pLoF SNPs), then

in 20.0% (95%CI: 19.2%-20.8%) of the constrained genes the most common predicted deleterious SNP contributed more than 90% to the gene burden (Fig. 1g and Supplementary Fig. 4). And in only 1.4% (95%CI: 1.2%-1.7%) of the constrained genes, the most common predicted deleterious SNP contributed less than 50% of the gene burden. For comparison, based on WGS from British(+CEU) and Han Chinese 1KG samples these fractions were 13.8% (95%CI: 13.1%-14.5%) and 11.7% (95%CI: 11.1%-12.4%) in Europeans and 12.5% (95%CI: 11.9%-13.2%) and 12.7% (95%CI: 12.0%-13.3%) in East Asians. Hence, for most genes rare variants contribute less to diseases in Greenland compared to in Europe and East Asia and as a consequence gene-burden tests will have limited power in Greenland. ”

- Line 163: Given the rate of increase of SNP discovery based on sample sizes (Fig. 1f), I wonder if the authors could project out when SNP discovery is expected to no longer yield large gains? This would be useful for future study design in Greenland.

We do not believe that it is possible to confidently project the number of segregating SNPs for a larger number of samples based on the trajectory in Fig. 1. As an example, we know that there will be fewer singletons in African populations compared to European populations with large sample sizes (>10,000) due to the super-exponential growth of the European population and in fact when sequencing many individuals (>>10000) more than half all of SNPs in a large European sample will be singletons(PMID: 22582263). Thus adding new European individuals will continue to add many new variants. This is not apparent from our figures. However, we can clearly show that each new sequenced Inuit individual will only add a limited number of new SNPs per additionally sequenced individual:



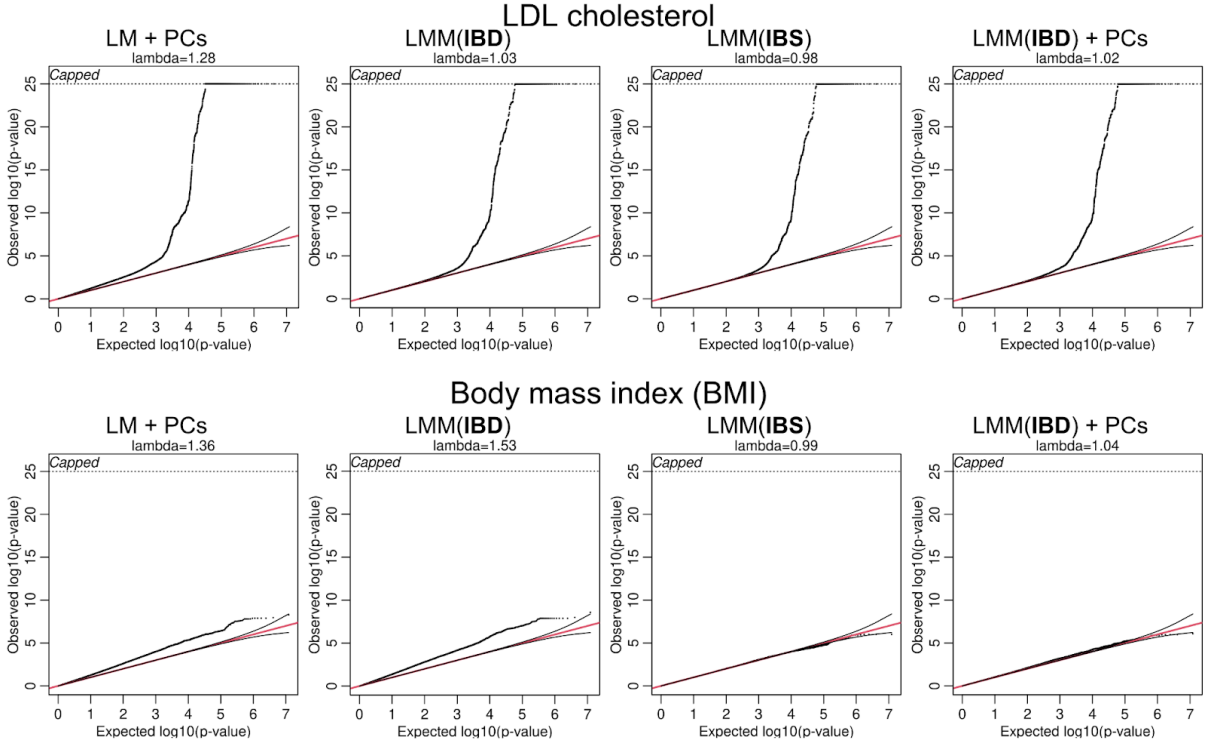
We have added this figure to the supplementary and made the following changes to the text:

Original sentence: “This is also reflected in the number of segregating SNPs observed in a given number of sequenced individuals (Fig. 1d) where Inuit (Greenlandic masked) had markedly fewer SNPs than other populations.”

New sentence: “This is also reflected in the number of segregating SNPs observed in a given number of sequenced individuals (Fig. 1d) where additional sequenced Inuit (Greenlandic masked) individuals add markedly fewer new SNPs than other populations (Supplementary Fig. 4).”

- The authors perform GWAS using standard corrections for population structure/relatedness. That is, using a genetic relatedness matrix as a random effect in a linear mixed model. Recent work suggests that fine-scale population structure may be unaccounted for by these methods (e.g. see PMID 35995948, 32691046). Given the authors' focus on fine-scale population structure and disease burden, I would be interested to know if the GWAS results change substantially when the covariance matrix based on IBD from NGSremix is used rather than the standard common variant GRM.

We thank the reviewer for this interesting question. We have done a lot of testing on which method is best for analysis of the Greenlandic data which differs from most other cohorts that have been used for GWAS, because of the strong population structure and large amount of relatedness.

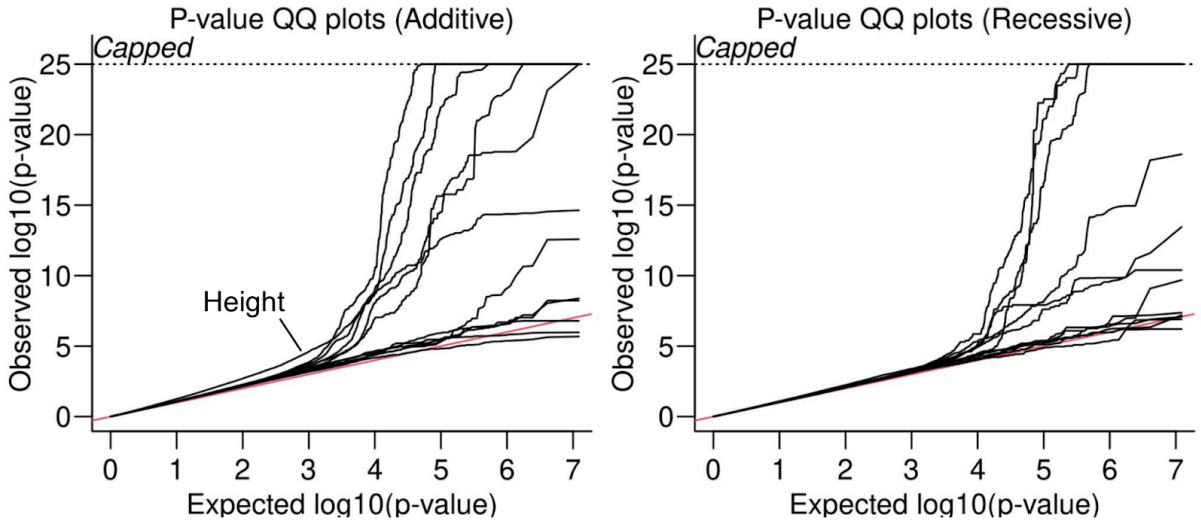


In the figure above, we illustrate the performance of several methods (linear model (LM) with 10 PCs, linear mixed model (LMM) based on an IBD matrix, LMM based on an IBS matrix and LMM based on an IBD matrix with 10 PCs) using two traits: LDL cholesterol (LDLC) and BMI with p-values capped at 10^{-30}. For LDLC there is a large amount of highly significant associations, regardless of method, but there is little difference in LDL values between Europeans and Inuit, so population structure will not inflate the test statistic a lot. As can be seen in the plots, we only observe a small amount of inflation when using LM with 10 PCs and using a LMM alleviates this inflation regardless of whether an IBD matrix or an IBS matrix is used, suggesting that both matrices take relatedness into account. Both have similar association strength for the top associated SNP (the known LDLR missense variant):

LDLC (GEMMA IBS): beta=6.244537e-01, P=1.620147e-102
LDLC (GEMMA IBD): beta=6.216129e-01, P=2.873901e-102

Of interest we also tried the LMM in GCTA including both the IBD and IBS matrix in the same analysis. However, we did not observe any difference in effect size or significance compared to running GCTA using only an IBS matrix.

For BMI, where there is a large difference in phenotypic values between Inuit and Europeans, using a LM with PCs leads to a high degree of inflation, which is also present when using a linear mixed model with the IBD matrix. This suggests that both relatedness and population structure confounds the analysis. If PCs are added to the LMM based on IBD, most of the inflation is gone. However, using just a LMM with the IBS matrix takes both population and relatedness into account and removes all the inflation. This is why we chose the standard use of GEMMA with an IBS matrix. In the manuscript we have now included Supplementary Fig. 9 with QQplots of the metabolic traits to show that we control the false positive rate of the GWAS:



The only noticeable deviation for the expected is observed for height which is known to be highly polygenic.

- In Fig 2f, what are the error bars denoting? Is the difference between Arctic-specific and PGS derived in Europeans significant? If the authors have the space to do so, I would be interested to see a breakdown of how this varies by genetic architecture of the trait (that is, SNP-heritability and polygenicity). It seems that the main thing going on is adding in large effect common variants, which is quite different from the factors affecting polygenic risk score transferability across other populations (that is, LD differences causing causal allele tagging to differ between populations). Is it true that the traits that improve the most are those that have large effect common variants that are Arctic-specific?

The error bars denote standard error of the mean (SEM), which we have now added to the y-axis label of figure 2f for clarity. And yes, as the reviewer correctly says; the traits that improve the most are the traits with large effect common Arctic specific variants, which can be seen in supplementary figure 10 and supplementary table 4. We have tried to clarify this in the text:

Original sentence: “For LDL-cholesterol, total cholesterol, 2-hour glucose, and type 2 diabetes, the PGS in unadmixed Greenlanders outperforms that of the UK biobank baseline after adding the Arctic-specific variants (Supplementary Fig. 7 and Supplementary Table 4)”

New sentence: “For the traits that are affected the most by the few added Arctic-specific variants, i.e. LDL-cholesterol, total cholesterol, 2-hour glucose, and type 2 diabetes, the PGS in unadmixed Greenlanders outperforms that of the UK biobank baseline after adding these variants (Supplementary Fig. 10 and Supplementary Table 4)”

The uncertainty shown in Fig. 2f only reflects the uncertainty of the mean relative incremental R^2 and cannot be used to assess significance differences between analyses, because it does not reflect the paired nature of the analysis (same trait with and without the Arctic variants).

We have now tested whether the improvement in the PGS is significantly better when adding the Arctic specific variants. The test results have been added to the supplementary figure 10 legend:

“To test whether the incremental R^2 is improved with the added Arctic specific variants, we performed a two-sided paired t-test on the incremental R^2 value in Greenland with and without the Arctic specific variants in the PGS. The improvement was significant both when the test was both done on all traits (p-value=0.04067) and only on the traits with Arctic specific variants (p-value=0.03647).”

-The authors study positive selection in their sample using Relate. It was not clear to me if the P-values reported in Figure 3f were corrected for multiple hypothesis test? If so, it should be stated. If not, please revise.

The 5 p-values obtained when testing the 5 variants in Fig. 3f are the unadjusted p-values. We have now added the FDR-adjusted (BH) p-values in accompanying supplementary table 9 and in the results. The variants that show significant signal of positive selection after FDR adjustment are highlighted in bold in Fig 3f. To clarify this, we have added the following sentence to the figure legend:

“P-values are unadjusted and bold p-values indicate significance after FDR (BH), also see Supplementary table 9.”

Minor comments:
- Line 122: Could the authors note the average coverage for the WGS samples and note the genotyping array for the 5548 individuals? This would help readers.

We agree with the reviewers that this information should be presented immediately and have added the following two parenthesis in the paragraph:

“[...]we whole-genome sequenced (WGS) 448 individuals (average sequencing depth of 35X)”

and

“[...]were genotyped using genome-wide SNP arrays (MEGA chip, Illumina)”

- Figure 1b caption: which population is the minimum allele frequency computed in?

The minimum allele frequency is computed in the Greenland population, which we have now added to the figure 1b x-axis label.

- Line 188: Are the plus minus values standard error? standard deviation?

The plus minus values are standard error of the mean. We have added “SEM” to each of the numbers reported to clarify this.

- Line 254: "mio" is not a standard abbreviation. Please revise this.

Correct. We have removed the abbreviation and just written out “million”.

- Line 304: does the phrasing "mainly" additive and recessive mean something specific? If so, this should be clarified. If not, this should be removed.

Yes, we chose the model yielding the lowest p-value. Since some signals were genome-wide significant in both models and formally testing the inheritance pattern is tricky, we chose the more careful wording “mainly additive/recessive”. We have tried to clarify this by adding the sentence:

“Signals were classified as mainly additive if the p-value was lowest under the additive model and mainly recessive if the p-value was lowest under the recessive model”

Referee #4. Population and statistical genetics

In this article the authors conduct a variety of population genetic analyses on 448 Inuit-ancestry Greenlanders via WGS data. With the increased call and necessity for greater ancestral diversity in the human genetics field, this paper is timely in that it focuses on both an underrepresented population and one that has a unique demographic history; as the authors explain, this Greenland population has experienced both a prolonged bottleneck (versus the shorter bottlenecks of Finland and Iceland) and admixture between Native American and European ancestries. As a result, and also as the authors show, the only way to truly represent such ancestries in human genetic analyses is by incorporating samples from the populations themselves. Specifically, the authors show that their Greenland cohort contains more high impact variants than a European-ancestry cohort for both metabolic and circulating-protein traits. They also show that European-trained metabolic polygenic scores and clinical screening with public references both perform poorly for their intended purposes, but that including Arctic-specific variants (for the former) or Greenland reference samples (for the latter) improves performance in both instances. Lastly, the authors also discuss the fine population structure they identify and the impact this has on current and future disease risk within Greenland.
I think this is a strong manuscript that shows a variety of interesting results that are properly supported with appropriate evidence. I do not have any major points; however, I do have a few minor points to help improve the accessibility and readability of the paper for a broader audience like Nature.

We thank the reviewer for spending their time reviewing our manuscript.

Minor Points:

1) For Figure 1b, I would either use 'dbSNP' instead of 'rsID' in the figure legend or say in the caption "…only found in dbSNP (ie has an rsID)…" to help ensure readers understand that by being in dbSNP there is or should be an rsID.

We agree that it is incorrect to use rsID as a name for the database and have updated the figure legend to "dbSNP" as suggested.

2) At the end of the 'Clinical screening to diagnose monogenic diseases' section, I would include a statement that explicitly links the drop in non-causal pLoF variants to more Arctic-ancestry variants being filtered out due to passing the allele frequency threshold now. I'm not sure if this will be as clear to a non-population genetics audience without an additional note explicitly stating that having Greenlander individuals as part of the reference data, vs. just gnomAD, leads to more SNPs being revealed as common.

We agree with the reviewer, and have added the following sentence to the "Clinical screening to diagnose monogenic diseases" section:

"This marked reduction in non-causal pLoF variants is a result of filtering out variants common in Greenlanders, but rare in gnomAD, facilitated by the Greenlandic reference panel."

3) On lines 197-167, for "The pattern is consistent independent of the maximum MAF threshold used for filtering (Supplementary Fig. 3).", I think you need to show the full range of MAFs in the supplementary figure to fully support the statement that the pattern is independent of maximum MAF. I'd either add a second, 3b supplementary figure that isn't zoomed in on the rare MAF thresholds (which in application are indeed the most relevant) or adjust the statement to something like "The pattern is consistent independent of a range of maximum MAF thresholds…".

We agree with the reviewer and have added the full range of MAF in what is now supplementary figure 5a.

4) On lines 221-222, there should be a reference for this statement: "This approach is more powerful than GWAS in the presence of multiple causal rare variants."

Also as part of the response to reviewer 3 major comment 2, we have re-written the gene burden section and this sentence has now been removed.

5) On line 227, I do not think 'constrained gene' has been defined yet in the main text. How 'constrained' is being defined here should be mentioned.

We have now added the definition of constrained gene in the methods:

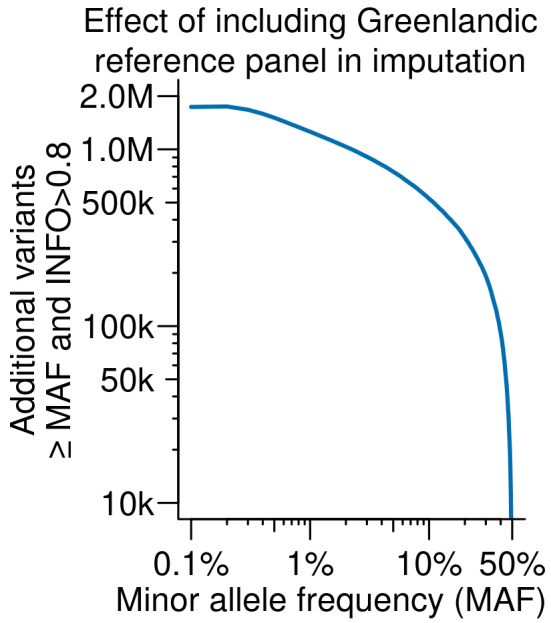
"Constrained genes were defined to be genes with expected number of pLoF and LOEUF score, estimated by Karczewski, K. J. et al., being less than 10 and 1, respectively. This resulted in 9,533 constrained genes from canonical transcripts."

6) On line 240 there should be commas before and after "…consistent with previous findings…"

Commas have been added, thank you.

7) At the end of the "Linkage disequilibrium related consequences" section looking at the impact on imputation, is there any way to give a frame of reference for how meaningful .7M extra variants are? Does adding .7M more SNPs to a set of already >6M really imply "…high quality imputation is only possible with representation of Greenlanders in the reference panel"? Looking at Figure 2d, it seems like there is a greater improvement in identifying high-quality rare variants when adding in the Greenland data than 1kG alone – is that something worth highlighting to help support the above point?

The 0.7 million extra common (MAF > 5%) variants are mainly variants that are not present in the 1000 genomes and are thus the most important when Arctic specific variants are of interest. It is true that more, ~1 million, high-quality rare/intermediate variants (MAF ≤ 5%) are also added. To show the number of added variants for different MAF-thresholds more clearly, we have added the following figure to supplementary figure 8.

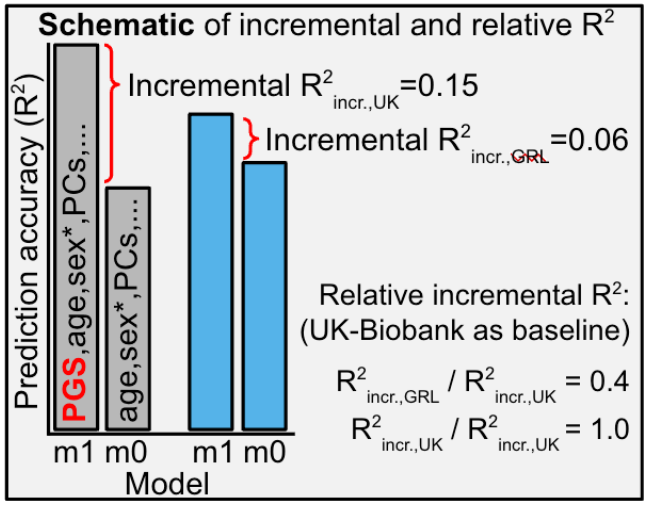


And added all the additional variants to the text so it now reads:
"We imputed the genomes of 5,548 SNP-chipped Greenlandic individuals, with the Greenland WGS individuals added to the reference panel. This resulted in an additional 1.7 million accurately called (INFO score>0.8) variants including 0.7 million common (MAF>5%)"

8) Two points on the "Comparative GWAS and polygenic score performance on metabolic traits" section. First, I would define incremental R2 here if it is going to be used. Second, I wonder if using relative vs. absolute R2 values here will be confusing for readers. I'm guessing relative is being used so that you can average across all 13 traits? I see in the supplementary figure you're using the absolute values, which is more straight forward in terms of interpretation. I would maybe include a note explicitly stating that relative is being used so that all traits can be compared and/or that readers should look at the supplement to see the absolute values. One of my concerns is that for readers not familiar with the PRS/PGS literature, they may misinterpret these "mean relative incr. R2" values as variance explained, whereas we know that PRS/PGS do not reach values anywhere close to 100%.

We thank the reviewer for the two important comments on the presentation of PGS results which we both agree with. Yes, we only use the relative R^2 so that we can summarise results from the 13 traits as a single value. We have tried to clarify both points:

1. We have added the sentence: "Incr. R2 is the increase in R2 when including the PGS in the model (see schematic in Supplementary Fig. 7)." And added the following schematic to the supplementary:



2. To make it more clear to the reader why we use relative R^2, we have added the sentence: "To compare across traits, we normalised the incr. R^2 to the incr. R^2 of the UK biobank (see Supplementary Fig. 7 for unnormalised values)"

9) For lines 342-344 in the "Recent change in mobility" section, could the authors provide a frame of reference for these PO/FS numbers to further help show how they are higher in the Greenland cohort? For instance, maybe you could pull these same numbers from the UKB or FinnGen? I just think the raw numbers may stand out more with something to compare against.

We agree and have now added the equivalent numbers for UK biobank for comparison:

Original sentence: We further utilised the fact that many participants have close relatives in the cohort with 7.2 parent offspring relationships (PO) and 13.4 full sibling relationships (FS) per hundred thousand (100k) pairs of individuals.

New sentence: "We further utilised the fact that many participants have close relatives in the cohort with 7.2 parent-offspring relationships (PO) and 13.4 full sibling relationships (FS) per hundred thousand (100k) pairs of individuals, which are high rates compared to UK biobank 0.0035 PO and 0.0178 FS per 100K pairs of individuals)." PMID: 36335127

10) There's something about the "Predicted consequences for disease susceptibility" section that I am having troubles following. Is the main reason we do not expect to see a similar reduction in future disease risk for CSID and monogenic diabetes is because they have more similar regional allele frequencies than the other three? Is there a metric you could use to show this across the five variants, like average and SD of the allele frequency across each region for each variant?

Yes, the regional allele frequencies are much less variable for the variants TBC1D4(diab.) and SI(CSID). To quantify this, we have now included the coefficient of variation for the homozygous frequency (AF^2) in the result section “Predicted consequences for disease susceptibility”

“Some of the variants had much bigger regional differences than others (Fig. 3d, Coefficient of Variation (AF2): ATP8B1=149%, PCCB=164%, and ADCY3=251%, SI=74%, and TBC1D4=42%)”