Genome-Wide Association Testing for Haemorrhagic Bowel Syndrome in a Swiss Large White Pig Population

Arnav Mehrotra[1]*, Alexander S. Leonard[1], Cord Drögemüller[2], Alexander Grahofer[3], Negar Khayatzadeh[4], Andreas Hofer[4], Stefan Neuenschwander[5], Hubert Pausch[1]

[1] Animal Genomics, ETH Zürich, Universitätstrasse 2, Zürich, 8092, Switzerland

[2] Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern 3012, Switzerland

[3] Clinic for Swine, Department for Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern, Bremgartenstrasse 109a, 3012, Bern, Switzerland

[4] SUISAG, Allmend 10, 6204, Sempach, Switzerland

[5] Animal Genetics, ETH Zürich, Tannenstrasse 1, 8092, Zürich, Switzerland

*Corresponding author

Email addresses:

AM: arnav.mehrotra@usys.ethz.ch

ASL: alexander.leonard@usys.ethz.ch

CD: cord.droegemueller@unibe.ch

AG: alexander.grahofer@unibe.ch

NK: nkh@suisag.ch

AH: aho@suisag.ch

SN: stefan.neuenschwander@usys.ethz.ch

HP: hubert.pausch@usys.ethz.ch

ABSTRACT

Background

The porcine haemorrhagic bowel syndrome (HBS) is a multifactorial disease causing fatal gastrointestinal disturbances and sudden death in fattening pigs. HBS is the leading cause of deaths during fattening in Swiss pigs, with unclear etiology. Environmental and management factors are associated with HBS incidence, but recent findings also suggest a potential genetic predisposition. Pigs sired by a Swiss Large White (SLW) line appear more prone to HBS. Here we conduct genome-wide association studies (GWAS) for HBS between cases and controls to investigate potential genetic factors for the disease in Swiss fattening pigs.

Results

Our study included 1,036 HBS cases and 4,080 controls with available microarray genotypes or whole-genome sequencing data. Variant positions were determined according to the current porcine reference assembly (Sscrofa11.1) or a HiFi-based SLW haplotype assembly which we constructed using trio-binning. GWAS for HBS were conducted using 12.49 to 15.46 million biallelic variants in three mapping cohorts consisting of purebred animals from SLW sire and dam lines, or crosses between these two parental lines. The statistical model applied for the GWAS accounted for animal relatedness, population structure, and an imbalanced case/control ratio. No sequence variants significantly associated with HBS were identified, regardless of the cohort analysed and the reference sequence considered.

Conclusions

The lack of genetic associations despite a relatively large sample size suggests that susceptibility to HBS in the studied SLW population is not due to large effect variants but may be influenced by numerous small effect genetic variants, in addition to environmental and management factors.

MAIN TEXT

Background

Haemorrhagic Bowel Syndrome (HBS) in pigs is a multifactorial disease that has become a substantial threat to swine production. The disease manifests in fatal gastrointestinal disturbances and sudden death in finisher pigs [1]. Key characteristics of HBS include pallor and abdominal distention in carcasses, with intestinal torsion frequently observed during necropsy [2,3]. Despite its substantial impact on animal welfare and the economics of pig production, the etiology of HBS is not fully understood. Several factors contributing to the incidence of the disease have been identified, including feeding systems [4], the origins of the fattening pigs, cleaning frequency of feed distribution systems, and the width of feeding places [2,5]. Seasonal variations [1,3,6], antimicrobial usage in feed [2,7], gut microflora composition [3], and infectious agents such as *Clostridium perfringens* and *Enterobacteriaceae* [2,4] have also been implicated.

A genetic component to HBS susceptibility has been proposed by previous studies [1,4]. A recent study by Holenweger et al. [5] also revealed that fattening pigs descending from a Swiss Large White (SLW) sire line were overrepresented among affected animals. These recent findings suggest that genetic variants may contribute to the susceptibility to HBS and may explain at least part of the observed across-breed variability to develop the disease.

This study aimed at identifying genetic variants associated with HBS in Swiss pigs through genome-wide association testing, thereby contributing to a better understanding of the etiology of the syndrome and facilitating the development of breeding strategies to prevent disease incidence.

Methods

Tissue samples of 1036 pigs that died from HBS were collected on farm and at animal carcass collection points in Switzerland over a period of six months, through SUISAG, the service partner for Swiss pig producers. HBS was confirmed post-mortem based on the inspection of the intestinal tract by stock veterinarians and veterinarians from SUISAG-SGD. We prepared DNA from ear tissue samples of 1036 HBS-affected pigs using the Promega Maxwell RSC DNA System (Promega, Dübendorf, Switzerland) and sent it to Gencove for low-pass sequencing (https://gencove.com/). An average number of 945K (between 150K and 5.15M) read pairs (2 x 150bp) corresponding to an average genome coverage of approximately 1-fold were collected for the HBS cases. Genotypes for 45,100,556 biallelic sequence variants (SNP) corresponding to the Sscrofa11.1 (GCA_000003025.6) reference sequence [8] were provided by Gencove. To infer the population structure, we extracted a subset of 48,919 SNPs from the genotypes that are also present on the SNP arrays routinely used for genomic prediction in Switzerland. These SNP genotypes were combined with array-derived genotype data of 17,006 pigs that were genotyped for routine genomic prediction, resulting in a combined dataset of 18,042 animals. After removing 4,554 markers with minor allele frequency (MAF) less than 5% using PLINK (v1.90) [9], a genomic relationship matrix (GRM) was constructed based on 44k SNPs using GCTA (v1.94.1) [10] and the top principal components were extracted and visualized to assess the structure of the

genotyped populations and select three ancestry-matched control cohorts. The cohorts consisted of 1) purebred animals from a SLW sire line, (70 cases, 280 controls), 2) purebred animals from a SLW dam line (61 cases, 244 controls), and 3) a mixed population of animals from the sire and dam lines, their crosses, as well as crosses between either the sire or the dam line and Landrace animals (1,020 cases, 4,080 controls), maintaining a 1:4 case-to-control ratio in all cohorts. Boars and sows genotyped for routine genomic prediction whose progeny were registered in the national herdbook and therefore not affected by HBS before reaching reproductive age were considered as control animals.

Genotypes for the control animals were imputed to the whole-genome sequence level with Beagle v5.4 [11] using a sequenced reference panel of 421 pigs that included mostly SLW animals. The reference animals had genotypes at 22,018,148 autosomal and 350,478 X-chromosomal sequence variants [12]. We retained 15.43 million autosomal 192,930 X-chromosomal biallelic SNPs that had a model-based accuracy of imputation greater than 0.8. Sex of HBS cases was inferred based on allele frequency estimates of X-chromosomal markers (--impute-sex flag in PLINK).

We considered SNPs with a minor allele frequency greater than 5% and less than 10% missingness for the subsequent association tests. For the purebred cohorts we excluded SNPs from genome-wide association testing that deviated significantly ($p<1 \times 10^{-6}$) from Hardy-Weinberg proportions. After the quality control, we retained between 12.49 and 13.46 million SNP for association testing in the three cohorts.

The genome-wide association studies (GWAS) in the three cohorts were performed using a generalized linear mixed model implemented in SAIGE (v1.0.9) [13] which accounts for sample relatedness and an imbalanced ratio of cases and controls. The top 10 principal components derived from a GRM, and sex were included as fixed factors, and a GRM was fitted as random effect. A Bonferroni-corrected significance threshold ($p=3.1 \times 10^{-9}$) was applied to account for multiple testing.

A haplotype of the SLW breed was assembled through trio-binning [14], utilizing PacBio high-fidelity (HiFi) reads from a male offspring originating from a crossing between a purebred Swiss Large White boar and an Alpenschwein sow. High Molecular Weight (HMW) DNA was extracted from a liver tissue sample of the F1 using the Monarch® HMW DNA Extraction kit and sequenced on the PacBio Sequel IIe platform, employing three SMRT cells to generate 4.93 million HiFi reads with an average length of 17.73 kb. Additionally, DNA from maternal blood and paternal liver tissue samples was sequenced on an Illumina Novaseq6000 to generate short reads with an average coverage of 28.7x. Haplotype-resolved assemblies were then generated with hifiasm v0.16.1 [15] as outlined in [16]. The paternal haplotype (SLW) assembly size was 2.406 Gb across 998 contigs, with an average Quality Value (QV) of 49.96 and a contig N50 of 30.62 Mb. The assembly achieved a BUSCO single-copy score of 87.9%.

Paired-end short reads from our imputation reference cohort consisting of 421 samples were aligned to the herein generated SLW assembly using bwa-mem2 [17]. Variant calling was then performed with DeepVariant v1.4 and GLnexus v1.4.1 [18] resulting in 29.73

137 million autosomal variants. From these, an imputation reference panel was constructed
138 using 24.75 million biallelic SNPs using the bcftools v1.6 [19] view command with the '–m2 –
139 M2 –v snps' flags. We imputed the low-pass sequencing data of HBS cases to the sequence
140 level with Glimpse v1.1 [20]. Coordinates of SNP array genotypes of controls were lifted
141 over to the SLW assembly using nf-LO [21] and subsequently imputed up to the sequence
142 level with Beagle v5.4 [11]. The imputed case and control datasets were merged, retaining
143 19.78 million SNPs with a missingness of less than 10% and a MAF greater than 0.05 for
144 GRM preparation and principal component analysis. Following quality control performed
145 separately for each cohort, as previously detailed (--maf 0.05, --geno 0.1; --hwe 1 x 10^-6 for
146 purebred cohorts), we retained between 13.71 and 15.46 million SNPs for GWAS with
147 SAIGE, as described earlier.

148

149 **Results and discussion**

150 Genome-wide association studies with partially imputed sequence variant genotypes were
151 performed to identify genomic loci associated with susceptibility to HBS in Swiss fattening
152 pigs. A principal components analysis identified three distinct cohorts for association
153 testing: purebred animals from the SLW sire line, purebred animals from the SLW dam line,
154 and crosses (Figure 1A, 1D, 1G). The majority of HBS cases fell into the latter cohort, as it
155 included the fattening pigs which were produced by the crossing of purebred parental lines.

156

157 Our initial case/control association analyses relied on between 12.49 and 13.46 million
158 variants for which genotypes were called from alignments against the current Sscrofa 11.1
159 genome assembly which was assembled from a Duroc boar [8]. Neither the within-breed

nor the multi-breed GWAS revealed markers significantly associated with HBS (Figure 1B, 1E, 1H). The choice of the reference assembly can impact sequence read alignment, variant calling, and downstream genetic analyses [22], particularly if the target breed diverged from the reference sequence. We repeated the GWAS with between 13.71 and 15.46 million SNPs for which genotypes were called from a SLW assembly (Fig 1C, 1F, 1I), but again, found no significant genetic variants associated with HBS in either of the cohorts tested. The lack of significant associations is somewhat surprising as previous studies identified the SLW sire line as the main risk factor for developing HBS [5]. Given a relatively low effective population size of 72 and 44 for the sire and dam lines [23], respectively, and the fact that both populations diverged less than 20 years ago, we are confident that our cohort was large enough to identify large effect genetic variants for developing HBS. Therefore, our findings suggest that the previously reported breed-specific predisposition to HBS is mainly driven by small effect genetic variants, which are possibly modulated by environmental and management factors.

Incorporating environmental variables into the case/control GWAS was not possible as the case and control groups were reared under disparate conditions. Future research efforts to better understand the genetic architecture for the development of HBS in fattening pigs and to identify associated genetic variants should ideally select control pigs from the same fattening farms as the cases, ensuring that these controls are closely matched to the cases in terms of age and weight profile and have successfully reached slaughter. Unfortunately, implementing such an effort is not straightforward as fattening pigs are not routinely genotyped. Farms with high and low incidences of HBS identified before [5] could serve as candidates for undertaking such genotyping efforts. However, considering that an average

Swiss fattening farm produces approximately 500 fattening pigs per year, and assuming an HBS incidence of 1% [5], the sampling of a case/control cohort with sufficient statistical power to identify trait-associated small effect variants appears to be a major undertaking. Nevertheless, such a cohort may also serve as a reference for the implementation of genomic prediction.

The lack of pedigree information for the mostly cross-bred HBS cases, together with the structured and related case/control cohorts, makes it difficult to estimate heritability of HBS. Although the statistical methods implemented in SAIGE and fastGWA-GLMM [24] minimize type I errors attributable to relatedness and imbalances between cases and controls in GWAS, the variance components they estimate through penalized quasi-likelihood are not precise enough for calculating heritability [13, 24]. Documenting pedigree information for fattening pigs could contribute to a better understanding of the genetic architecture underlying the disease.

Finally, the potential impact of structural variations (SVs), which may not be adequately tagged by the SNPs used in our study, should also be considered. Although the newly built, breed-specific SLW assembly makes variants overlapping insertions with respect to the Duroc-based reference sequence amenable to association mapping, the short-read sequencing-based approach doesn't allow to comprehensively study SVs [25]. The establishment of a porcine SV imputation reference panel for pigs could enable future HBS research by integrating SV data for a more comprehensive association analysis.

Conclusions

We report the first GWAS for HBS in Swiss pigs reported to be at risk for HBS, a sporadically occurring disorder characterized by sudden death in fattening pigs. Our comprehensive genetic analysis, spanning several breeds and two different reference genomes, did not reveal any significant genetic markers associated with HBS. This finding suggests that the genetic susceptibility to HBS is likely to involve small effect genetic variants and/or more complex SVs that may interact with environmental and management factors, rather than large effect genetic variants. Future research should prioritize the selection of cases and controls from the same fattening farms, allowing a clearer distinction between genetic predisposition and environmental influences.

List of abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs

GRM: Genomic Relationship Matrix

GWAS: Genome-Wide Association Studies

HBS: Haemorrhagic Bowel Syndrome

HiFi: High Fidelity

HMW: High Molecular Weight

MAF: Minor Allele Frequency

QV: Quality Value

SLW: Swiss Large White

SNP: Single Nucleotide Polymorphism

SV: Structural Variation

Declarations

Ethics approval and consent to participate

Tissue samples for the HIS cohort were collected from dead pigs. The sampling of blood from the F1 and its parents used for constructing the de novo assembly was approved by the veterinary office of the Canton of Zurich (animal experimentation permit ZH077/2022).

Consent for publication

Not applicable

Availability of data and materials

Low-pass sequencing data of pigs affected by HBS have been deposited at the European Nucleotide Archive (ENA) of the EMBL at BioProject PRJEB62539. Raw sequence read data of pigs used to prepare the imputation reference panel are available at the European Nucleotide Archive (ENA) of the EMBL at BioProjects PRJEB38156 and PRJEB39374. Long and short sequencing reads from the trio used to generate the SLW assembly are at PRJEB74562.

Competing interests

The authors declare that they do not have any competing interests.

Funding

We received financial support from the Federal Office for Agriculture (FOAG), Bern, Fonds zur Anschubfinanzierung der Schweine Plus Gesundheitsprogramme, IP-SUISSE, and UFA AG. The funding bodies were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

252 **Authors' contributions**

253 AM performed the analyses and wrote the first draft of the manuscript with input from ASL
254 and HP. ASL assembled the SLW haplotype. SN collected samples from the trio used for SLW
255 assembly. HP and AH conceptualized the study. CD, AG, AH, NK and HP conceptualized the
256 project. All authors read and approved the final manuscript.

257

258 ## Acknowledgements

259 We acknowledge support from Hans-Peter Grünenfelder, Tanja Kutzer, Thomas Echtermann
260 and Xaver Sidler to produce the F1 cross. We thank the Functional Genomics Center Zurich
261 (Anna Bratus-Neuenschwander) for generating HiFi data. The assistance of Nathalie
262 Besuchet-Schmutz with DNA extraction is acknowledged.

263

264 ## References

265 1. Straw B, Dewey C, Kober J, Henry SC. Factors associated with death due to hemorrhagic
266 bowel syndrome in two large commercial swine farms. Journal of Swine Health and
267 Production. 2002 Mar 1;10(2):75-9.
268 2. Grahofer A, Gurtner C, Nathues H. Haemorrhagic bowel syndrome in fattening pigs.
269 Porcine health management. 2017 Dec;3(1):1-6.
270 3. Häni H, Huber A, Schmidt J, Zimmermann W. Das haemorrhagische Intestinalsyndrom
271 (HIS) des Schweines: klinische, pathologisch-anatomische und aetiopathogenetische
272 Aspekte. Schweizer Archiv fuer Tierheilkunde. 1993;135.

4. Labuscagne A, Spencer BT, Picard JA, Williams MC. An investigation to determine the cause of haemorrhagic enteritis in commercial pig grower units in the northern parts of South Africa. Journal of the South African Veterinary Association. 2012 Jan 1;83(1):1-6.

5. Holenweger F, Schüpbach G, Hofer A, Sidler X, Grahofer A. Housing and management factors and breed predisposition for haemorrhagic bowel syndrome in swine. Porcine health management. 2023 Oct 11;9(1):44.

6. Novotný J, Reichel P, Kovačocyová K, Cigánková V, Almášiová V, Šipoš D. Haemorrhagic bowel syndrome in Fattenig pigs. Acta Veterinaria. 2016 Mar 1;66(1):138-46.

7. Le MG. Syndrome de Distension Intestinale Porcin (SDIP) ("l'entérotoxémie"). Proceedings Journées Recherche Porcine. 2008;40:33–42.

8. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, Giron CG. An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience. 2020 Jun;9(6):giaa051.

9. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Dec 1;4(1):s13742-015.

10. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 2011 Jan 7;88(1):76-82.

11. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics. 2018 Sep 6;103(3):338-48.

12. Nosková A, Mehrotra A, Kadri NK, Lloret-Villas A, Neuenschwander S, Hofer A, Pausch H. Comparison of two multi-trait association testing methods and sequence-based fine mapping of six additive QTL in Swiss Large White pigs. BMC genomics. 2023 Apr 10;24(1):192.

13. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, Bastarache LA. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nature genetics. 2018 Sep;50(9):1335-41.

14. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TP, Phillippy AM. De novo assembly of haplotype-resolved genomes with trio binning. Nature biotechnology. 2018 Dec;36(12):1174-82.

15. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature methods. 2021 Feb;18(2):170-5.

16. Leonard AS, Crysnanto D, Fang ZH, Heaton MP, Vander Ley BL, Herrera C, Bollwein H, Bickhart DM, Kuhn KL, Smith TP, Rosen BD. Structural variant-based pangenome construction has low sensitivity to variability of haplotype-resolved bovine assemblies. Nature communications. 2022 May 31;13(1):3012.

17. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In2019 IEEE international parallel and distributed processing symposium (IPDPS) 2019 May 20 (pp. 314-324). IEEE.

18. Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics. 2020 Dec 15;36(24):5582-9.

19. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Feb;10(2):giab008.

20. Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics. 2021 Jan;53(1):120-6.

21. Talenti A, Prendergast J. nf-LO: a scalable, containerized workflow for genome-to-genome lift over. Genome Biology and Evolution. 2021 Sep 1;13(9):evab183.

22. Lloret-Villas A, Bhati M, Kadri NK, Fries R, Pausch H. Investigating the impact of reference assembly choice on genomic analyses in a cattle breed. BMC Genomics. 2021 May 19;22(1):363.

23. Nosková A, Bhati M, Kadri NK, Crysnanto D, Neuenschwander S, Hofer A, Pausch H. Characterization of a haplotype-reference panel for genotyping by low-pass sequencing in Swiss Large White pigs. BMC Genomics. 2021 Dec;22(1):1-4.

24. Jiang L, Zheng Z, Fang H, Yang J. A generalized linear mixed model association tool for biobank-scale data. Nature genetics. 2021 Nov;53(11):1616-21.

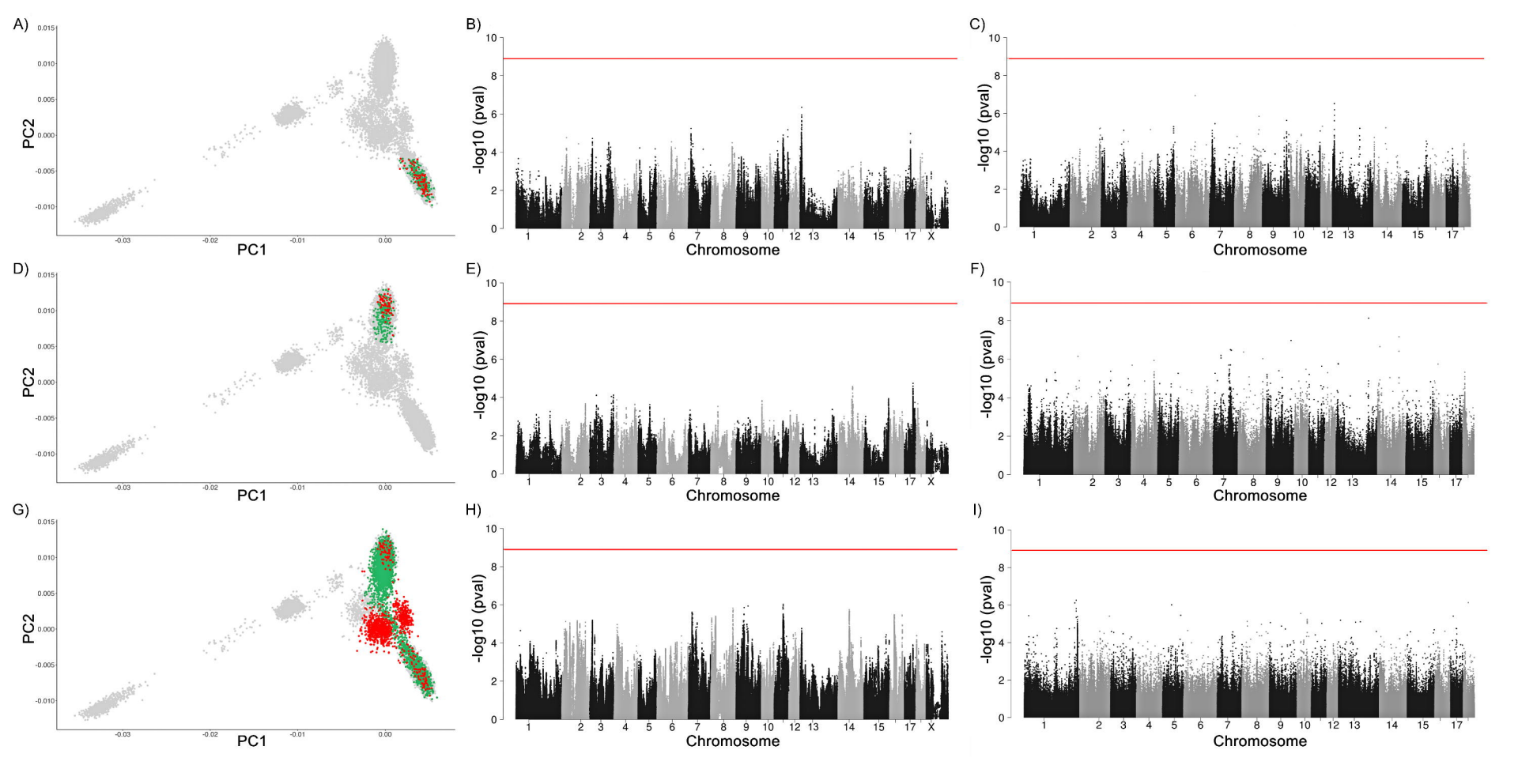
25. Chaisson MJ, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, Fan X. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nature Communications. 2019 Apr 16;10(1):1784.

Figures:

Figure 1 title: Principal component analysis (PCA) and genome-wide association study (GWAS) results across three cohorts. PCA plots highlighting cases (red) and controls (green) selected for the three cohorts: purebred Swiss Large White sire line (A), purebred Swiss Large White dam line (D), and a mixed population including these two lines their crosses, and crosses between either of those lines and Landrace (G). The corresponding Manhattan plots (B, E, H) illustrate the GWAS results using the Sscrofa11.1 reference genome. Plots (C, F, I) depict GWAS results utilizing the herein established Swiss Large White reference

343 assembly. The red line across all Manhattan plots represents the Bonferroni-corrected

344 significance threshold.


A)
B)
C)
D)
E)
F)
G)
H)
I)
PC1
PC2
-log10 (pval)
Chromosome