Identification of copy number variations and candidate genes for reproduction traits in Finnish pig populations

Iso-Touru, Terhi

2022

Iso-Touru , T , Uimari , P , Elo , K , Sevon-Aimonen , M-L & Sironen , A 2022 , ' Identification of copy number variations and candidate genes for reproduction traits in Finnish pig populations ' , Agricultural and Food Science , vol. 31 , no. 3 , pp. 149-159 . https://doi.org/10.23986/afsci.116081

http://hdl.handle.net/10138/350681
https://doi.org/10.23986/afsci.116081

AGRICULTURAL AND FOOD SCIENCE

Agricultural and Food Science (2022) 31: 149–159
https://doi.org/10.23986/afsci.116081

Identification of copy number variations and candidate genes for reproduction traits in Finnish pig populations

Terhi Iso-Touru[1], Pekka Uimari[2], Kari Elo[2], Marja-Liisa Sevon-Aimonen[1] and Anu Sironen[1]

[1]Production systems, Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland

[2]Department of Agricultural Sciences, University of Helsinki, 00014 Helsinki, Finland

e-mail: anu.sironen@luke.fi

Animal breeding programs can be improved by genetic markers associated with production and reproduction traits. Reproduction traits are important for economic success of pig production and therefore development of genetic tools for selection is of high interest to pig breeding. In this study our objective was to identify genomic regions associated with fertility traits in two Finnish pig breeds using large scale SNP genotyping and genome wide association analysis and characterization of copy number variations (CNV). Since CNVs are structural variations of the genome they potentially have a large effect on gene expression and protein function. We analyzed 1265 genotyped boars for nine different reproduction traits and identified 46 CNV regions encompassing 13 genes. 11 of the CNV regions were shared between the breeds, 20 were unique to the Finnish Yorkshire and 15 to the Finnish Landrace. The genome wide association (GWAS) analysis identified zero to five reproduction associated genomic regions per trait. Furthermore, we identified 23 genomic regions with 20 candidate genes associated with fertility traits using GWAS analysis. The identified CNV regions were compared against GWAS regions to detect candidate regions with an effect on reproduction traits. This study reports candidate genes and genomic regions within two Finnish pig breeds for reproduction traits, which can be utilized in breeding programs.

Key words: swine, fertility, genetics, GWAS, CNV

Introduction

Reproductive traits are important factors in pig breeding programs to increase the productivity of pig populations. Although the heritability of reproductive traits is often low, genomic tools can help to improve the selected fertility traits by identification of associated genomic regions and genes with a role in reproduction pathways. Previous studies have reported a large number of QTLs and candidate genes for reproduction traits, but the complex genetic mechanisms affecting reproductive performance are still largely unknown. Genome wide association analysis (GWAS) is a powerful tool for identification of economically important genomic regions for polygenic traits (Guo et al. 2016, Wang et al. 2018, Wu et al. 2018). Our earlier studies have identified reproduction related genomic region in Finnish landrace using GWAS and high throughput SNP chip. This study underlined significant association of genomic regions for total number of piglets born in first and later parities and piglet mortality between birth and weaning in later parity (Uimari et al. 2011). GWAS combined with structural genomic variants can provide wider understanding of economically important complex traits and novel tools for animal breeding. Analysis of copy number variations (CNV) provides knowledge of structural genomic changes within a population and their possible effects on important traits in animal breeding programs. Due to their large size (usually >1 Kb) CNVs can be expected to have a bigger effect on gene function than single nucleotide polymorphisms (SNPs). SNP arrays are a cost effective and useful tool for identification of CNVs in different pig breeds (Peiffer et al. 2006, Winchester et al. 2009, Wang et al. 2013), because the GWAS analysis and CNV calling can be done from the same dataset. The aim of this study was to identify genomic regions associated with reproduction traits and their overlap with CNVs within Finnish pig breeds.

Materials and methods

Animal material

The data consisted of 1843 AI-boars born between 1992 and 2013 and genotyped with 60K porcine chip (Illumina). 1021 samples met the quality criteria (described later) for CNV mapping (605 Finnish Yorkshire and 414 Finnish Landrace boars). The genotypes were produced in two different laboratories (FIMM and GeneSeek). In total 1265 boars (443 Finnish Landrace and 822 Finnish Yorkshire) had deregressed breeding values for reproduction traits: total number of piglets born in firs parity (TNB1), total number of piglets born in later parities (TNB2), number of stillborn piglets in first parity (NSB1), number of stillborn piglets in later parities (NSB2), piglet mortality between

Received 12 April 2022 / Accepted 9 September 2022
The Scientific Agricultural Society of Finland

birth and weaning in first parity (PM1), piglet mortality between birth and weaning in second/later parity (PM2), age at first farrowing (AFF), first farrowing interval (FFI), second farrowing interval (SFI). Estimated breeding values were obtained from the national breeding value estimation that considered the effects of herd, year, month, type of insemination, breed of the litter, age at farrowing, sire of the litter, and the permanent environmental of the animal while estimating the additive genetic effects of the animal. Deregression was done using Jamrozik method (Jamrozik et al. 2000).

Genome wide association analyses (GWAS)

Quality control

Genotypes from boars with deregressed breeding values for reproduction traits (n = 1265) were used for GWAS analysis. First, markers located in X and Y chromosomes were removed as well as markers with unknown position (Sscrofa 10.2). The quality parameters used for selection of single nucleotide polymorphisms (SNPs) were minimum call rate of 90% for individuals and for loci. SNPs with minor allele frequencies below 5% and deviation from Hardy-Weinberg proportion ($p < 0.000001$) were excluded. The number of SNPs remaining after quality control was 34,255 SNPs for Yorkshire and 34,878 SNPs for Landrace. Altogether 916 Yorkshire and 405 Landrace boars remained for further analyses. To reduce the deviation and obtain more reliable phenotypes, we applied the cut-off value of 30 for the number of descendants. This reduced the number of boars available for GWAS to 534 and 317 for Yorkshire and Landrace, respectively.

GWAS analysis

Association analysis between phenotypes and genotypes was done using GWAS analysis and separately for Yorkshire and Landrace populations using single-locus mixed model GWAS (EMMAX) method (Kang et al. 2010) implemented in the software package SVS 8.6.0 (Golden Helix). GWAS mixed linear model analysis uses a kinship matrix to correct for cryptic relatedness as a random effect. An identity-by-state (IBS) kinship matrix was computed and used as a random effect in the model. The genome-wide significance was tested using a Bonferroni correction. However, Bonferroni correction is very conservative leading to many true associations being discarded, simply because the correction is performed for all SNPs in the panel, even if many are in linkage disequilibrium. For this study, we considered all variations with $-\log_{10}(p\text{-value}) \geq 4.0$ (p-value < 0.0001) as potential candidates. For identification of functional effect of the SNPs variations having $-\log_{10}(p\text{-value}) \geq 4.0$ were annotated with the variant effect predictor tool using the Ensembl database, Release 87 (McLaren et al. 2010). The prediction whether an amino acid substitution caused by missense variation affects protein function was estimated by SIFT analysis (Ng and Henikoff 2003) implemented in VEP tool (McLaren et al. 2010). The SIFT prediction is based on sequence homology and the physical properties of amino acids. Manhattan plots were created with software package SVS 8.6.0 (Golden Helix).

Copy number variation (CNV) analyses

Quality control

All genotyped animals that met the quality criteria (n = 1021) were used for CNV analysis. CNVs were called from the log R ratio values produced during genotyping, which are calculated by comparing the observed normalized probe signals in each sample with an expected signal intensity calculated from the Illumina defined reference sample cluster. Several different quality control steps were done prior to CNV analyses. Firstly, markers located on X and Y chromosomes were removed as well as markers with unknown genomic position (according to Sscrofa 10.2). The differences between sample batches and breeds were analysed by principal component analysis (PCA) with the Log R ratio values (Fig. 1). Analysis revealed clear batch effect most likely caused by the laboratory (Fig. 1A), since data used for CNV calling is intensity data and varies between protocols and chips used. Breed clustering was used as an internal control, which showed clear separation between Finnish Yorkshire and Landrace (Fig. 1B).

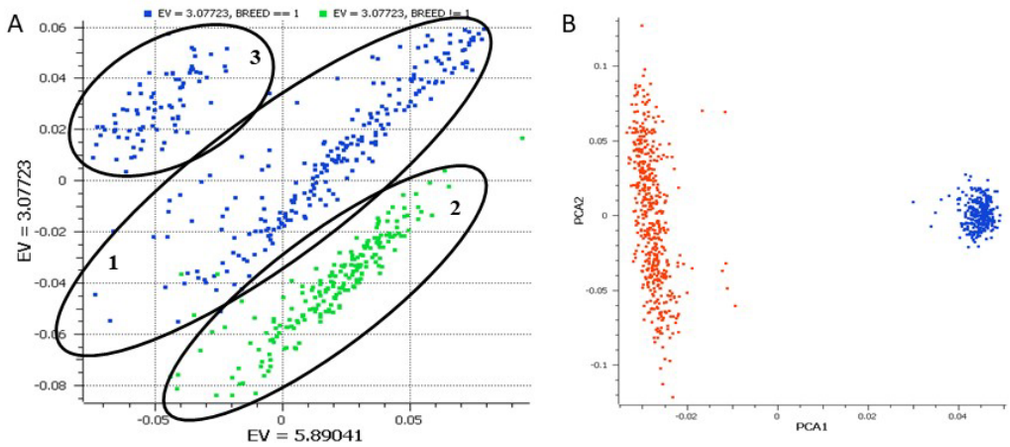

Fig. 1. PCA analysis from the combined Log R ratio data. A. Batch effect on the genotypes. Blue dots = Finnish Yorkshire, green dots = Finnish Landrace. B. Clustering of different breeds. Red dots = Finnish Yorkshire and blue dots = Finnish Landrace. Each dot represents one individual.

Due to the strong batch effect the data sets were analyzed separately for CNV calling. According to the PCA analysis (Fig. 1) three different data sets were established: 1) Finnish Yorkshire samples genotyped in FIMM (n=317), 2) Finnish Landrace samples genotyped in FIMM (n=386) and 3) Finnish Yorkshire samples genotyped in GeneSeek (n=288). Finnish Landrace samples (n=28) genotyped in GeneSeek were discarded due to small sample size.

Secondly, we calculated median upper outlier threshold for derivative log ratio spread (DLRS) and excluded samples having higher DLRS than the upper limit. The limit varied between the datasets. The outliers were removed leaving 292 samples on dataset 1, 384 samples on dataset 2 and 272 samples to dataset 3. The next quality control step was the wave detection, which was done for each sample and samples exceeding the abs wave factor 0.05 were filtered out. Data set sizes after this step were 215, 193 and 147 for datasets 1, 2 and 3, respectively. The last quality control step was to correct remaining batch effect among the datasets. It was done by PCA analysis and datasets were corrected with the varying number (2 to 3) of PCAs.

CNV analyses

The PCA corrected datasets were used for CNV calling (datasets 1, 2 and 3, Fig. 1A). The calling was done with the CNAM univariate segmentation algorithm implemented in the software package SVS 8.6.0 (Golden Helix). The univariate segmentation method considers only one sample at a time and is ideal for detecting rare and/or large CNVs. After CNV detection, the copy numbers were distinguished for gains, neutrals and losses. Segment means were filtered with the dataset specific values (< –1 for losses and > 0.5 for gains in datasets 1 and 2, < –2.2 for losses and > 0.5 for gains in dataset 3). This creates three state covariates where –1 is for potential copy number losses, 0 for potential copy number neutrals and 1 for potential copy number gains. These were used as covariates for association testing between the phenotypes and CNVs. Association analysis was done using single-locus mixed model GWAS (EMMAX) method (Kang et al. 2010) implemented in the software package SVS 8.6.0 (Golden Helix). Association tests were done for each datasets separately and results were compared between datasets 1 and 3 as well with the results obtained from the genome-wide association analyses done with the SNP markers. For GWAS analysis we used the cut-off value of 30 for the number of descendants as explained in the GWAS methods section. Copy number variations regions (CNVRs) were defined by merging overlapping CNVs to CNVRs using bedtools (Quinlan and Hall 2010). CNVRs were annotated with the Biomart tool embedded to Ensembl database (Release 87) and genes found within the regions were listed.

Genomic locations of all pig QTLs were downloaded from the Animal QTLdb (Release 31, 30 December 2016). Associated regions identified in GWAS analysis and downloaded QTLs were compared with the CNV results using BedTools package (Quinlan and Hall 2010).

Results and discussion

GWAS identified potential reproduction related markers for selection in the Finnish Yorkshire

For the Finnish Yorkshire population, the most promising candidate QTL for reproduction traits is located on chromosome 7 for TNB1 (total number of piglets born in first parity, Table 1, Fig. 2B). The QTL region is located approximately between positions 63,074,243 – 66,304,912bp. One associated SNP in the region (rs80,969,873) within a gene C15orf39 is a missense variation potentially deleterious to the gene's protein product. This region contained four associated SNPs and the lead SNP in this region was rs80,906,873 at position 7:65,928,035 (log10(p) 5.18, Table 1). The function of the gene is not known, but based on the GO ontology it has a role in the cytoplasm and is relatively highly expressed in the ovary and uterus. The variant has also been associated with sperm quality in Duroc breed (Wang et al. 2022). With further analysis this SNP could be a potential marker for selection. Another interesting candidate region for TNB-trait is located on chromosome 5 (4,410,900 – 4,517,270bp). Both TNB1 and TNB2 (total number of piglets born in first parity, total number of piglets born in later parities) are associated to that region. Region has three genes (*L3MBTL2*, *POLR3H*, *ACO2*) with associated SNPs. *ACO2* is a regulatory enzyme of the tricarboxylic acid (TCA) cycle, which is crucial for ATP production in sperm mitochondria and therefore mutations in *ACO2* may have an effect on sperm motility (Tang et al. 2014). *L3MBTL2* has been suggested to play a role in chromatin remodeling during meiosis and spermiogenesis and its depletion led to decreased sperm counts and increased number of abnormal sperm in mice (Meng et al. 2019). However, the best candidate gene for TNB1 and TNB2 traits is *POLR3H.* Loss-of-function mutation in mouse *Polr3h* gene causes embryonic lethality, but missense mutations result in decreased fertility, low number of follicles and small litter sizes (Franca et al. 2019).

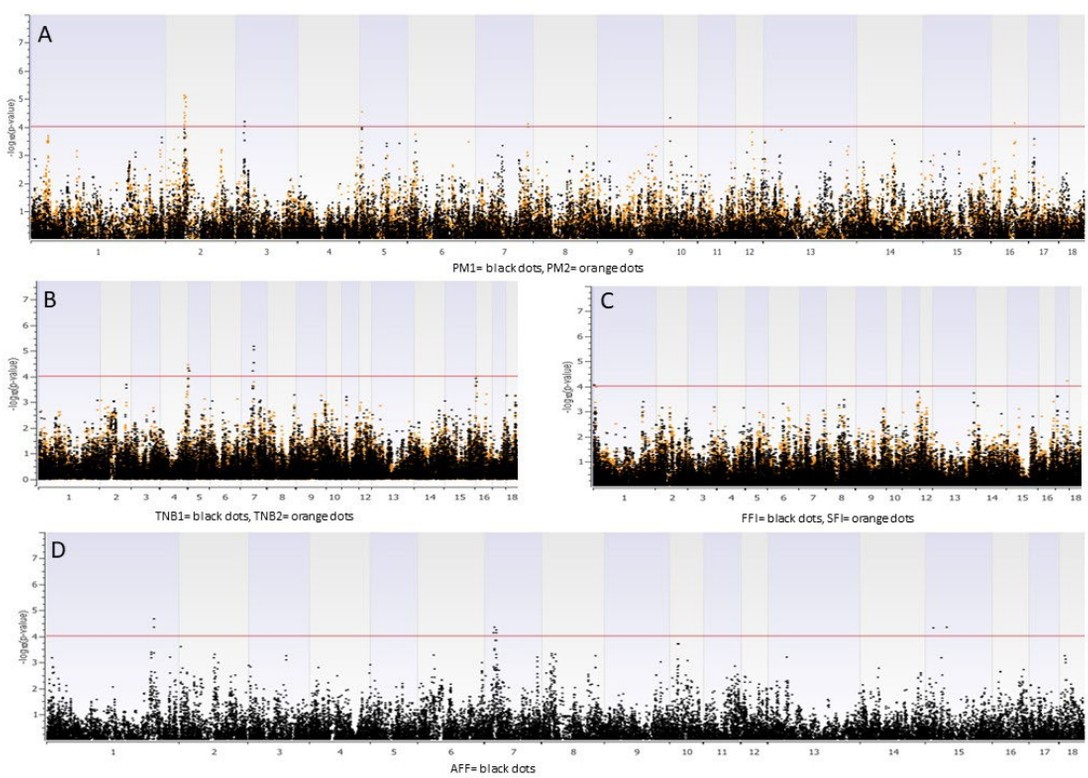

Fig. 2. Genome-wide Manhattan plots for Finnish Yorkshire. A. PM1 (back dots) and PM2 (orange dots), B. TNB1 (black dots) and TNB2 (orange dots), C. FFI (black dots) and SFI (orange dots) and D. AFF. Red line indicates the genome-wide significance level –log10(p) = 4.0.

Table 1. Potential QTL regions for each tested fertility trait for the Finnish Yorkshire. Top SNP for each QTL region are shown including position (Sscrofa 10.2), -log10(p)-value, allele substitution effect (R.beta), standard error for R.beta value, proportion of variance explained by SNP in question, minor allele frequency (MAF), annotation of the top SNP and information about which gene top SNP is located. Lead SNP positions in genome build Sscrofa 11.1 are listed in Supplementary Table 1.

Chr	Start	End	Trait	Lenght	No. of associated SNPs in region	No. of genes with associated SNPS within the QTL region	Top SNP	Position of the top SNP	-log10(p)	R.Beta	Beta SE	Proportion of variance explained	MAF	Annotation of the top SNP	Top SNP in gene	Gene ID
1	255487386	255682541	AFF	195155	2	-	rs80891802	255487386	4.65	3.298	0.771	0.033	0.38	intergenic variant	-	-
7	22301965	22987329	AFF	685364			rs80989881	22987329	4.34	2.828	0.688	0.031	0.45	intergenic variant	-	-
7	27225485	27778363	AFF	552878	4	1	rs80975214	27778363	4.23	-3.139	0.775	0.03	0.25	non coding transcript variant	*C2*	ENSSSCG00000001422
15	17898207	17898207	AFF	-	1		rs81478853	17898207	4.31	-2.633	0.643	0.031	0.38	intergenic variant	-	-
15	50746568	50746568	AFF	-	1	-	rs80966936	50746568	4.32	-4.731	1.154	0.031	0.07	intron variant	-	
5	24769365	24769365	FFI	-	1	1	rs81316127	24769365	4.52	0.907	0.215	0.032	0.36	intron variant	*KIF5A*	ENSSSCG00000000439
12	42868692	42868692	FFI	-	1	-	rs81268186	42868692	4.05	-0.91	0.23	0.028	0.22	intergenic variant	-	-
5	24769365	24769365	SFI	-	1	-	rs81316127	24769365	4.45	0.775	0.186	0.032	0.36	intron variant	*KIF5A*	ENSSSCG00000000439
6	137335517	137335517	SFI		1		rs81392580	137335517	4.18	0.728	0.181	0.029	0.38	intergenic variant	-	-
9	23437153	23826749	SFI	389596	2	-	rs80981360	23437153	4.2	-0.808	0.2	0.03	0.35	intergenic variant	-	-
1	4936025	4979417	NSB1	43392	2	-	rs80804050	4936025	4.03	-0.088	0.022	0.028	0.38	intergenic variant	-	-
17	57886538	57886538	NSB2	-	1	1	rs81467086	57886538	4.21	-0.078	0.019	0.03	0.5	intron variant	*UBE2V1*	ENSSSCG00000030632
3	20318891	20778770	PM1	459879	3	-	rs81341840	20318891	4.16	0.08	0.02	0.029	0.4	intergenic variant	-	-
10	14802676	14802676	PM1	-	1	-	rs81281284	14802676	4.31	-0.106	0.026	0.031	0.14	intergenic variant	-	-
2	42912421	44805085	PM2	1892664	9	4	rs81358018	44246660	5.11	0.123	0.027	0.037	0.41	intron variant	*SERGEF* *USH1C* *ABCC8* *INSC*	ENSSSCG00000042767 ENSSSCG00000013377 ENSSSCG00000013378 ENSSSCG00000013385
2	47055959	47166445	PM2	110486	4	-	rs81225422	47155005	5.04	-0.114	0.025	0.036	0.49	intergenic variant	-	-
5	4384938	4384938	PM2	-	1	2	rs80842388	4384938	4.52	-0.088	0.021	0.032	0.38	3 prime UTR variant	*POLR3H* *ACO2*	ENSSSCG00000000063 ENSSSCG00000000064
7	123169296	123169296	PM2	-	1		rs80961242	123169296	4.1	-0.136	0.034	0.029	0.11	intergenic variant	-	-
16	56029991	56029991	PM2	-	1		rs81459856	56029991	4.13	-0.104	0.026	0.029	0.26	intergenic variant	-	-
5	4410900	4517270	TNB1	106370	2	1	rs81235592	4410900	4.33	-0.166	0.04	0.031	0.47	intron variant	*ACO2*	ENSSSCG00000000064
7	63074243	66304912	TNB1	3230669	4	1	rs80906873	65928035	5.18	-0.238	0.052	0.037	0.46	intergenic variant	*C15orf39*	ENSSSCG00000001885
5	4384938	4517270	TNB2	132332	2	3	rs80842388	4384938	4.46	-0.189	0.045	0.032	0.38	3 prime UTR variant	*POLR3H* *ACO2* *L3MBTL2*	ENSSSCG00000000063 ENSSSCG00000000064 ENSSSCG00000000066
7	65904152	65904152	TNB2	-	1		rs80956245	65904152	4.52	-0.232	0.055	0.032	0.31	intergenic variant	-	-

For piglet mortality between birth and weaning in second/later parity (PM2) a promising association was found from chromosome 2 spanning the region 42,912,421 – 44,805,085 bp (Table 1, Fig. 2A). Several associated SNPs are found covering four genes (*SERGEF, USH1C, ABCC8, INSC*). The list does not contain any obvious candidate genes, but many of them are related to gene regulation. Thus, the identification of the possible causal mutation requires further studies, but based on the haplotype within this region a selection tool could be developed.

Age at first farrowing (AFF) showed association with three different chromosomes, chromosome 1 (255,487,386 – 255,682,541 bp), chromosome 7 (27,225,485 – 27,778,363 bp) and chromosome 15 (positions 17,898,207 and 50,746,568 bp Table 1, Fig. 2D). First and second farrowing intervals (FFI and SFI) were associated with a region on chromosome 5 (24,769,365 – 24,769,365 bp, Table 1, Fig. 2C). The peak SNP is located within *KIF5A*, which is related to axon guidance and microtubule-based movements that are needed in intracellular transportation.

EIF2B2 and *ESRRG* identified as potential candidate genes for reproduction traits in GWAS for Finnish Landrace

For Finnish Landrace the strongest association for TNB1 and TNB2 was located in an intergenic region on chromosome 5 (30,252,056 bp) (Table 2, Fig. 3A). NSB1 and NSB2 had a QTL peak on chromosome 12 (40,395,494 – 41,093,290 bp, Table 2, Fig. 3B). The top SNP was again located in an intergenic region. The SNPs in intergenic regions can be linked to the causal variation or can be located on a regulatory region. Another interesting region was found on chromosome 7, where the SNP associated with NSB1 was located within *EIF2B2* (eukaryotic translation initiation factor 2B subunit beta), which has been linked to ovarian follicle development (Fogli et al. 2003) and is therefore a good candidate gene for sow reproduction.

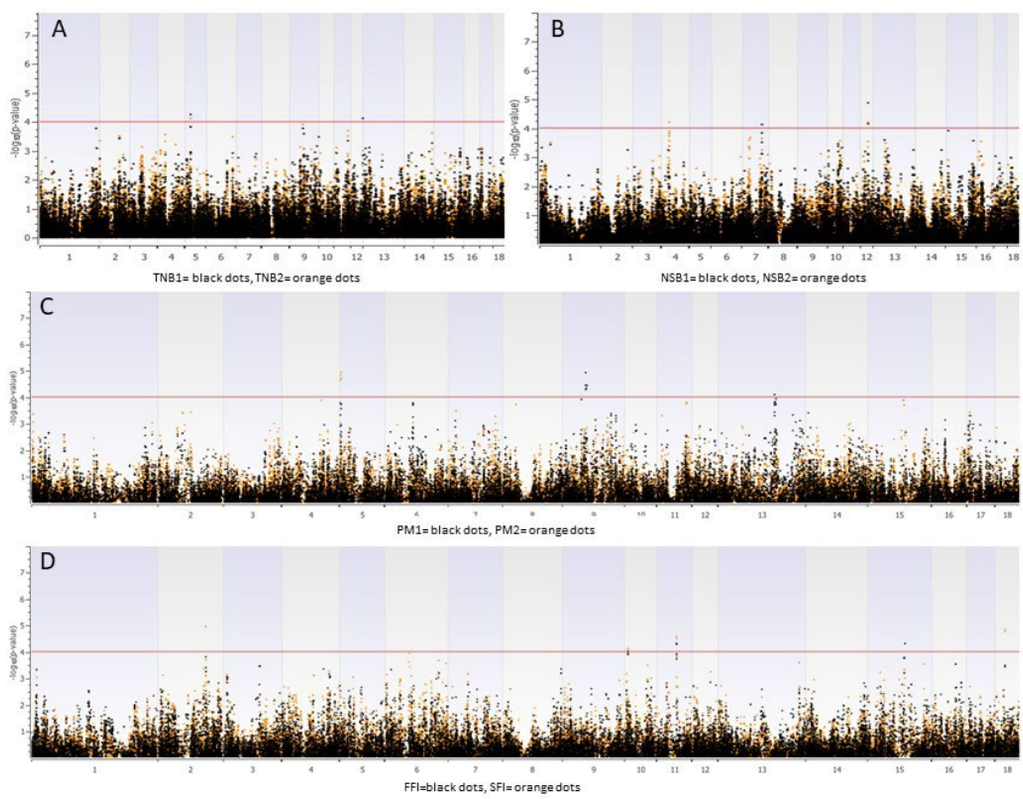

Fig. 3. Genome-wide Manhattan plots for the Finnish Landrace. A. TNB1 (black dots) and TNB2 (orange dots), B. NSB1 (black dots) and NSB2 (orange dots), C. PM1 (black dots) and PM2 (orange dots), D. FFI (black dots) and SFI (orange dots). Red line indicates the genome-wide significance level –log10(p) = 4.0.

A region on chromosome 13 (around position 141,229,923 bp) was associated with PM1 (Table 2, Fig. 3C) with strongest association within gene *LSG1*, which functions in nuclear export and ribosome biogenesis (Malyutin et al. 2017). The beginning of the chromosome 5 seemed to harbor a QTL for PM2, but no obvious candidate genes were found (Fig. 3C, Table 2).

Table 2. Potential QTL regions for each tested fertility trait for the Finnish Landrace. Top SNPs for each QTL region are shown including position (Sscrofa 10.2), -log10(p)-value, allele substitution effect (R.beta), standard error for R.beta value, proportion of variance explained by SNP in question, minor allele frequency (MAF), annotation of the top SNP and information about in which gene the top SNP is located. Lead SNP positions in genome build Sscrofa 11.1 are listed in Supplementary Table 2.

Chr	Start	End	Trait	Lenght	No. of associated SNPs in region	No. of genes with associated SNPS within the QTL region	Top SNP	Position of the top SNP	log10(p)	R.Beta	Beta SE	Proportion of variance explained	MAF	Annotation of the TOP SNP	Top SNP in gene	Gene ID
6	15265862	15265862	AFF	0	1	-	rs81247212	15265862	5.07	5.313	1.173	0.0611	0.12	intergenic variant	-	-
16	74299091	74388983	AFF	89892	2	1	rs81461694	74299091	4.95	3.799	0.85	0.0596	0.24	intron variant	*FAXDC2*	ENSSSCG00000017068
11	50602385	50621724	FFI	19339	2	-	rs81323090	50621724	4.31	1.236	0.3	0.0511	0.33	intergenic variant	-	-
15	91795690	91795690	FFI	0	1	-	rs81478992	91795690	4.3	1.558	0.379	0.0509	0.22	intergenic variant	-	-
2	119897088	119897088	SFI	0	1	1	rs81362981	119897088	4.94	1.54	0.345	0.0594	0.23	intron variant	*TMEM232*	ENSSSCG00000014196
10	8362152	8664907	SFI	302755	4	1	rs81225607	8664907	4.11	-1.198	0.299	0.0485	0.22	intron variant	*ESRRG*	ENSSSCG00000010814
11	50602385	50621724	SFI	19339	2	-	rs81323090	50621724	4.59	1.241	0.291	0.0547	0.33	intergenic variant	-	-
18	25179268	25239804	SFI	60536	2	-	rs81467776	25239804	4.83	2.318	0.527	0.0579	0.06	intergenic variant	-	-
7	104051815	104051815	NSB1	0	1	1	rs81269236	104051815	4.11	-0.114	0.028	0.0485	0.27	intron variant	*EIF2B2*	ENSSSCG00000002377
12	40395494	41093290	NSB1	697796	2	-	rs81347276	41093290	4.86	-0.105	0.024	0.0584	0.44	intergenic variant	-	-
4	41648344	41648344	NSB2	0	1	-	rs81381778	41648344	4.2	-0.139	0.034	0.0496	0.14	intergenic variant	-	-
12	40395494	40395494	NSB2	0	1	-	rs81434700	40395494	4.2	-0.097	0.024	0.0497	0.41	intergenic variant	-	-
9	57362823	60935929	PM1	3573106	6	-	rs81411156	57362823	4.92	-0.132	0.03	0.0591	0.29	intergenic variant	-	-
13	141229923	141229923	PM1	0	1	1	rs80968940	141229923	4.1	0.114	0.029	0.0482	0.5	synonymous variant	*LSG1*	ENSSSCG00000011827
5	1385843	3844970	PM2	2459127	4	-	rs80858480	3844970	4.93	0.121	0.027	0.0592	0.38	intergenic variant	-	-
5	30252056	30252056	TNB1	0	1	-	rs80933580	30252056	4.26	-0.266	0.065	0.0504	0.19	intergenic variant	-	-
13	426118	426118	TNB1	0	1	-	rs80822529	426118	4.12	-0.215	0.054	0.0486	0.24	intron variant	-	
5	30252056	30252056	TNB2	0	1	-	rs80933580	30252056	4.11	-0.275	0.069	0.0484	0.19	intergenic variant	-	-

No associations were identified for AFF, but FFI and SFI had some interesting associations to chromosomes 2, 10, 11 and 15 (Table 2). On chromosome 2 a SNP within the gene *TMEM232* showed association with FFI, but the function of the gene is unknown. A region of 83Mb on chromosome 10 was associated with SFI (Table 2). One known gene, *ESRRG* (estrogen related receptor gamma), is located within that region. *ESRRG* is a good candidate gene for reproduction traits, since it is involved in steroid hormone mediated signaling pathway. Other associated regions were intergenic on chromosome 11 (FFI and SFI) and 15 (FFI).

CNVs in Finnish pig breeds

Altogether, 46 copy number variation –regions (CNVRs) were found among the two breeds (Fig. 4). 11 CNVRs were found to be shared between the breeds (Fig. 4, Table 3), average length of the shared CNVRs being 72,029 bp. All shared segments were copy number losses. The Finnish Yorkshire had 20 breed specific CNV regions including two gains and 18 losses (Fig. 4, Table 4) and the average length was 77,741 bp. Altogether 15 CNVR losses were unique to the Finnish Landrace and average length was 82,279 bp (Table 5). In total 17 genes were localized within the CNVRs.

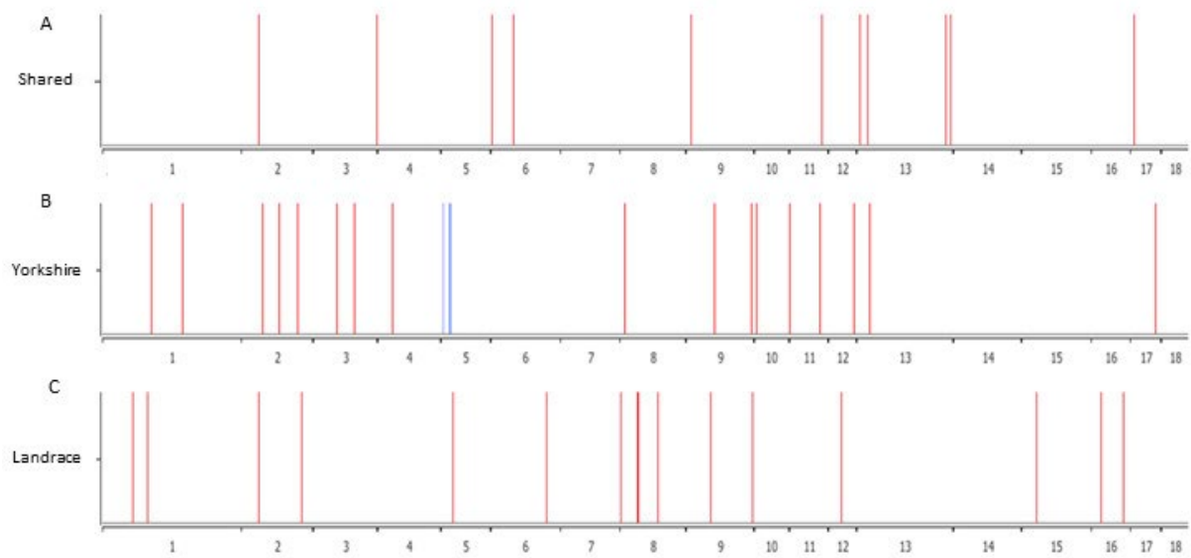

Fig. 4. The locations of identified CNVRs. Red bars marks copy number losses and blue bars copy number gains. A. Locations of the CNVRs shared between the breeds. B. CNVRs locations unique to the Finnish Yorkshire. C. CNVRs locations unique to the Finnish Landrace.

Table 3. CNVRs shared between the Finnish Yorkshire and Landrace

CHR	start (bp)	end (bp)	lenght (bp)	state	no Yorkshire	no Landrace	no total	no genes	genes	gene acc
2	40232829	40246783	13954	loss	11	5	16	-	-	-
3	144231979	144740773	508794	loss	1	4	5	2	*ZNF316*	ENSSSCG00000023378 ENSSSCG00000008671
6	52574745	52595445	20700	loss	21	6	27	-	-	-
6	5812388	5816473	4085	loss	50	7	57	-	-	-
9	12069732	12094268	24536	loss	13	1	14	-	-	-
11	74825658	74827003	1345	loss	1	3	4	-	-	-
13	7810100	7901494	91394	loss	3	1	4	-	-	-
13	25898589	25919479	20890	loss	6	6	12	-	-	-
13	202736300	202798825	62525	loss	8	1	9	1	*N6AMT1*	ENSSSCG00000028724
13	213450408	213489484	39076	loss	1	1	2	1	*IGSF5*	ENSSSCG00000012071
17	9916840	9921867	5027	loss	1	4	5	-	-	-

Table 4. CNVRs unique to the Finnish Yorkshire

CHR	start (bp)	end (bp)	lenght (bp)	state	no total	no genes	genes	gene acc
1	113553257	113593991	40734	loss	58			
1	182623403	182642549	19146	loss	1			
2	49156831	49278605	121774	loss	18	1	*BTBD10*	ENSSSCG00000013395
2	87102362	87106233	3871	loss	1	-		
2	129950230	129999370	49140	loss	30			
3	54849268	54954971	105703	loss	1	1	*IL1R2*	ENSSSCG00000028331
3	94706101	94868661	162560	loss	1			
4	34382638	34469345	86707	loss	1	1	*ZFPM2*	ENSSSCG00000006038
5	4645402	4739534	94132	loss	2	1	*EP300*	ENSSSCG00000000068
5	5559455	5562688	3233	gain	2			
5	21238084	21251973	13889	gain	1			
8	10249007	10371511	122504	loss	2			
9	65218233	65242035	23802	loss	1	1	*NTM*	ENSSSCG00000015253
9	148691522	148795560	104038	loss	1			
10	6048687	6116151	67464	loss	2			
11	1309196	1462483	153287	loss	1			
11	70685626	70775677	90051	loss	2	1		ENSSSCG00000009498
12	58613497	58821700	208203	loss	2	1		ENSSSCG00000024467
13	31394191	31398134	3943	loss	1			
17	58450650	58531296	80646	loss	2	2	*PTPN1* *RIPOR3*	ENSSSCG00000007469 ENSSSCG00000007470

Table 5. CNVRs unique to the Finnish Landrace

CHR	start (bp)	end (bp)	lenght (bp)	state	no total	no genes	genes	gene acc
1	71212047	71264177	52130	loss	9			
1	102858476	102917151	58675	loss	2			
2	40774796	40920860	146064	loss	2			
2	138281538	138308699	27161	loss	6			
5	26724414	26895376	170962	loss	7	1	*SLC16A7*	ENSSSCG00000000456
6	129045981	129046349	368	loss	2			
8	3351928	3354915	2987	loss	2	1	*TBC1D14*	ENSSSCG00000027349
8	41935601	42031291	95690	loss	16			
8	87292781	87520544	227763	loss	2	2	*SLC10A7* *POU4F2*	ENSSSCG00000009035 ENSSSCG00000026015
9	57234372	57282676	48304	loss	4			
9	152609200	152685081	75881	loss	5			
12	30846585	30905303	58718	loss	1			
15	35674169	35675096	927	loss	3			
16	23516947	23754408	237461	loss	1			
16	72234796	72273598	38802	loss	2			

No overlaps between CNVRs and GWAS regions were identified in this study, thus we expanded our analysis to QTLs in the Animal Genome Database. Genomic regions harboring QTLs for reproduction were underrepresented in CNVR regions (Fisher's Exact Test p-value < 3.387e-08 and 0.5 fold under enrichment). This may partly explain the lack of overlaps between CNVRs and GWAS regions and poor associations between CNVRs and reproduction traits.

However, some interesting genomic regions were identified between CNVRs and reproduction phenotypes (Supplementary Figs. S1–S3). For example the CNVR in chromosome 13 found from the Finnish Yorkshire population shows an association with NSB1. However, due to the limited dataset further studies are needed. Several CNVRs were only identified in few (1–4) pigs and may represent an interesting genomic region for association, but require studies with larger data set. Both analysed breeds showed separate sets of associated genomic regions, which is expected due to separate breeding of these pig populations and small population size. It can be assumed that the identified CNVs are relatively recent and do not descent from common ancestors in these pig breeds.

Conclusions

This study was designed to characterize reproduction related genomic modifications in Finnish pig breeds. We have identified copy number variations and reproduction related candidate genes within these regions, which provide tools for pig breeding programs. Furthermore, the reported data inform about the genomic structural changes in Yorkshire and Finnish Landrace pig breeds enabling also further studies to reveal the functional importance of specific genomic regions.

Acknowledgements

The technical assistance in DNA extraction and sample processing by Tiina Jaakkola and Tarja Hovivuori, Luke, Finland is greatly appreciated. The study was funded by the Ministry of Agriculture and Forestry (Helsinki, Finland) and Finnish Pig Breeding Foundation (Suomen Sianjalostuksen Säätiö, Hollola, Finland). The pedigree data was provided by Faba (The Finnish Animal Breeding Association, Hollola, Finland) and phenotypes by Figen Oy (https://www.figen.fi/). The funding bodies have not participated in the design of the study, collection, analysis, or interpretation of data or in writing the manuscript.

Supporting information

Supplementary Figure 1. Genome-wide Manhattan plots for the association between Animal Genome database retrieved reproduction QTLs and CNVRs in Finnish Yorkshire dataset 1 (FIMM).

Supplementary Figure 2. Genome-wide Manhattan plots for the association between Animal Genome database retrieved reproduction QTLs and CNVRs in Finnish Landrace dataset 2 (FIMM).

Supplementary Figure 3. Genome-wide Manhattan plots for the association between Animal Genome database retrieved reproduction QTLs and CNVRs in Finnish Yorkshire dataset 3 (Geneseek).

Supplementary Tables 1 and 2. Genomic locations of identified lead SNPs in Sscrofa11.1.

References

Fogli, A., Rodriguez, D., Eymard-Pierre, E., Bouhour, F., Labauge, P., Meaney, B.F., Zeesman, S., Kaneski, C.R., Schiffmann, R. & Boespflug-Tanguy, O. 2003. Ovarian failure related to eukaryotic initiation factor 2B mutations. American Society of Human Genetics 72: 1544–1550.https://doi.org/10.1086/375404

Franca, M.M., Han, X., Funari, M.F.A., Lerario, A.M., Nishi, M.Y., Fontenele, E.G.P., Domenice, S., Jorge, A.A.L., Garcia-Galiano, D., Elias, C.F. & Mendonca, B.B. 2019. Exome Sequencing Reveals the *POLR3H* Gene as a Novel Cause of Primary Ovarian Insufficiency. The Journal of Clinical Endocrinology & Metabolism 104: 2827–2841. https://doi.org/10.1210/jc.2018-02485

Guo, X., Su, G., Christensen, O.F., Janss, L. & Lund, M.S. 2016. Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish Landrace and Yorkshire pigs. BMC Genomics 17: 468. https://doi.org/10.1186/s12864-016-2806-z

Jamrozik, J., Schaeffer, L. & Jansen, G. 2000. Approximate accuracies of prediction from random regression models. Livestock Production Science 66: 85–92. https://doi.org/10.1016/S0301-6226(00)00158-5

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C. & Eskin, E. 2010. Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42: 348–354. https://doi.org/10.1038/ng.548

Malyutin, A.G., Musalgaonkar, S., Patchett, S., Frank, J. & Johnson, A.W. 2017. Nmd3 is a structural mimic of eIF5A, and activates the cpGTPase Lsg1 during 60S ribosome biogenesis. EMBO J 36: 854–868. https://doi.org/10.15252/embj.201696012

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. & Cunningham, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26: 2069–2070. https://doi.org/10.1093/bioinformatics/btq330

Meng, C., Liao, J., Zhao, D., Huang, H., Qin, J., Lee, T.L., Chen, D., Chan, W.Y. & Xia, Y. 2019. L3MBTL2 regulates chromatin remodeling during spermatogenesis. Cell Death and Differentation 26: 2194–2207. https://doi.org/10.1038/s41418-019-0283-z

Ng, P.C. & Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research 31: 3812–3814. https://doi.org/10.1093/nar/gkg509

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C.A., Belmont, J., Cheung, S.W., Shen, R.M., Barker, D.L. & Gunderson, K.L. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Research 16: 1136–1148. https://doi.org/10.1101/gr.5402306

Quinlan, A.R. & Hall, I.R. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

Tang, M., Liu, B.J., Wang, S.Q., Xu, Y., Han, P., Li, P.C., Wang, Z.J., Song, N.H., Zhang, W. & Yin, C.J. 2014. The role of mitochondrial aconitate (ACO2) in human sperm motility. Systems Biology in Reproductive Medicine 60: 251–256. https://doi.org/10.3109/19396368.2014.915360

Uimari, P., Sironen, A. & Sevon-Aimonen, M.L. 2011. Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed. Genetics Selection Evolution 43: 42. https://doi.org/10.1186/1297-9686-43-42

Wang, J., Wang, H., Jiang, J., Kang, H., Feng, X., Zhang, Q. & Liu, J.F. 2013. Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. PLoS One 8: e68683. https://doi.org/10.1371/journal.pone.0068683

Wang, Y., Ding, X., Tan, Z., Xing, K., Yang, T., Wang, Y., Sun, D. & Wang, C. 2018. Genome-wide association study for reproductive traits in a Large White pig population. Animal Genetics 49: 127–131. https://doi.org/10.1111/age.12638

Wang, T., Feng, Y., Chen, D., Bai, R., Tang, J., Zhao, Y., Zhu, L., Ye, L., Li, F. & Li, J. 2022. Nonsynonymous SNPs within C7H15orf39 and NOS2 are associated with boar semen quality. Animal Biotechnology 27: 1–5. https://doi.org/10.1080/10495398.2022.2077213

Winchester, L., Yau, C. & Ragoussis, J. 2009. Comparing CNV detection methods for SNP arrays. Briefings in Functional Genomics 8: 353–356. https://doi.org/10.1093/bfgp/elp017

Wu, P., Yang, Q., Wang, K., Zhou, J., Ma, J., Tang, Q., Jin, L., Xiao, W., Jiang, A., Jiang, Y., Zhu, L., Li, X. & Tang, G. 2018. Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. Genomics 110: 171–179. https://doi.org/10.1016/j.ygeno.2017.09.009