ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

A genome-wide association study identified a major QTL affecting the red colour in nitrate free hams

Jacopo Vegni [a], Martina Zappaterra [a], Roberta Davoli [a], Roberta Virgili [b], Nicoletta Simoncini [b], Cristina Schivazappa [b], Alessio Cilloni [c] & Paolo Zambonelli [a].

[a] Department of Agri-food science and technologies (DISTAL), University of Bologna, Italy
[b] Experimental Station for the Food Processing Industry (SSICA), Parma, Italy
[c] Fratelli Galloni S.p.A, Langhirano (Parma), Italy

STAZIONE SPERIMENTALE PER L'INDUSTRIA DELLE CONSERVE ALIMENTARI - FONDAZIONE DI RICERCA

Programma di Sviluppo Rurale dell'Emilia-Romagna
2014-2020

UNIONE EUROPEA
Fondo Europeo Agricolo per lo Sviluppo Rurale

Regione Emilia-Romagna

L'Europa investe nelle zone rurali

PQ INN

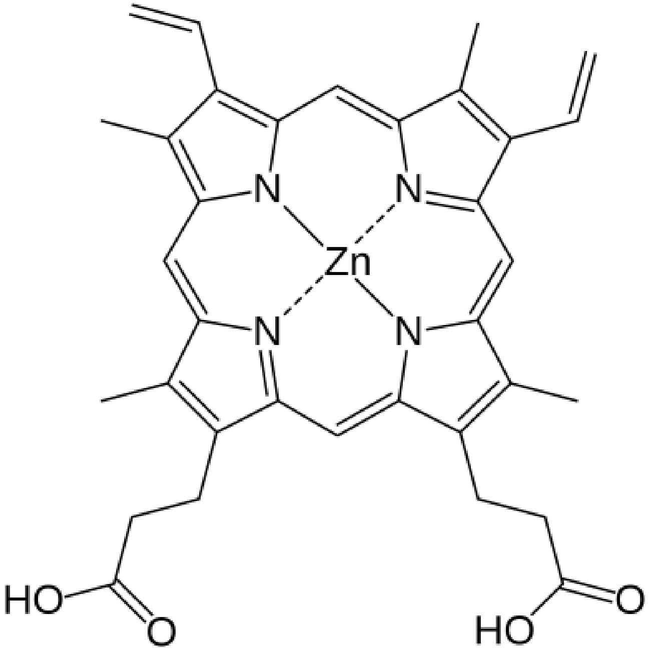
Pilot project Pork Quality Innovation (PQ-Inn) for the evaluation of meat quality parameters of heavy pigs for protected designation of origin (PDO) ham production using innovative genomic and process technologies.

(https://agricoltura.regione.emilia-romagna.it/progetti-innovazione/raccolta-progetti-innovazione/competitivita-delle-filiere-agroalimentari/16.2/bando-2017/progetto-pilota-qualita-tecnologica-carne-suino-pesante-prosciutti-tecniche-innovative-genomiche-pork-quality-innovation)

Introduction

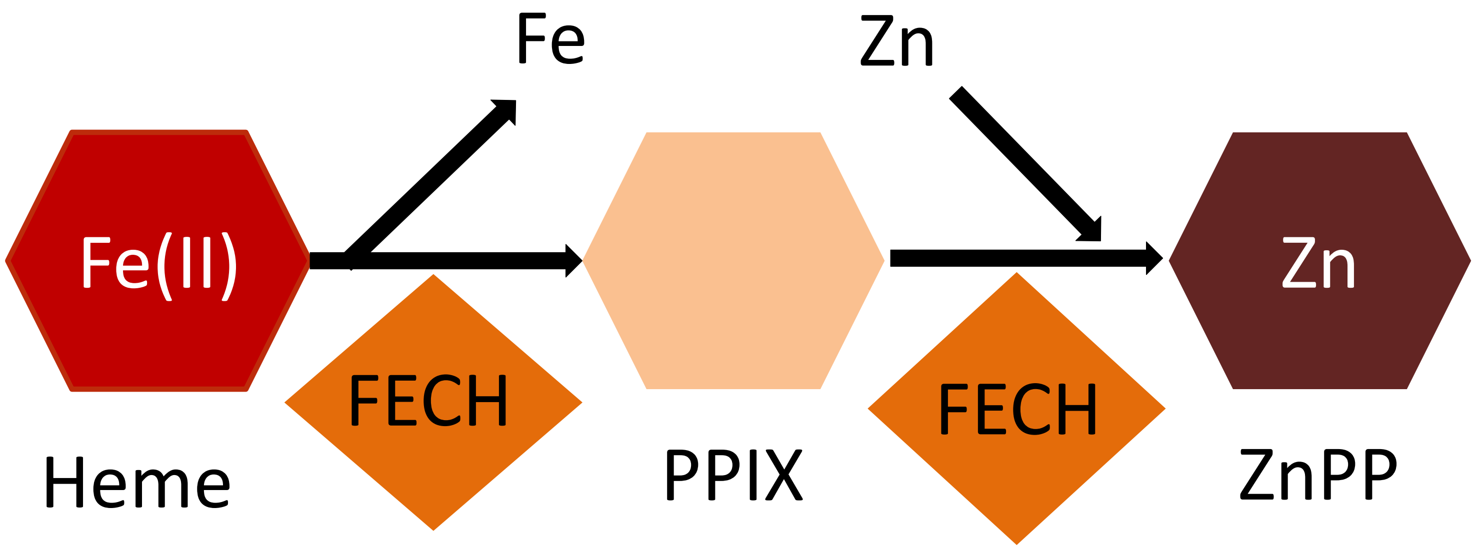
In **Parma ham** that is free of nitrites/nitrates, the colour is **bright red** (lipophilic stable red pigment).

(Images derived from Wakamatsu, 2022)



- Antioxidant activity.
- Stable against light, heat and low oxygen concentration.

Wakamatsu, J. (2022). Evidence of the mechanism underlying zinc protoporphyrin IX formation in nitrite/nitrate-free dry-cured Parma ham. Meat Science, 192, 108905. https://doi.org/10.1016/j.meatsci.2022.108905.



Aim of the study

To identify **genomic regions related to variation of the phenotype, activity of the enzyme ferrochelatase (FECHA)**. A GWAS was employed as a genetic method for the study, on samples derived from the *Semimembranosus* muscle of hybrid pigs

Materials and methods

Animals

- Commercial hybrid pigs.
- No pedigree data and nor litter information.
- Piglets from different litters are mixed.
- 238 fresh hams (from 114 barrows and 124 gilts).

(Images source: https://www.qualigeo.eu/prodotto-qualigeo/prosciutto-di-parma-dop/)

Phenotypic analysis

- **FECHA** was expressed as nmol of ZnPP formed min^{-1} 100 g^{-1} dry matter (Parolari *et al*., 2016 with slightly modified).

Genotyping

- DNA extraction from *Semimembranosus* muscle.
- Extracted DNA samples were analyzed with **GeneSeek® Genomic Profiler porcine HD genotyping array (68,516 markers).**

Parolari, G., Aguzzoni, A. & Toscani, T. (2016). Effects of Processing Temperature on Color Properties of Dry-Cured Hams Made without Nitrite Foods, 5, 33. https://doi.org/10.3390/foods5020033

Statistical analysis

- Quality control (MAF < 0.05, HWE < 0.001, call rate > 10%).
- Genetic distance population & vector extrapolation.
- PCA.
- Cleaning of samples (IBS > 0.9).
- GWAS

SNPs significant for *P*-values below the genome or chromosome-wide threshold **of 1 and 5%** (corrected for the genomic inflation factor λ, close to 1).

Test significant for **FDR < 0.01.**

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

- Detection of the nearest genes, **± 500 kb region flanking** the associated SNPs, and isoform characterization (**FECH-201** and **FECH-202**).

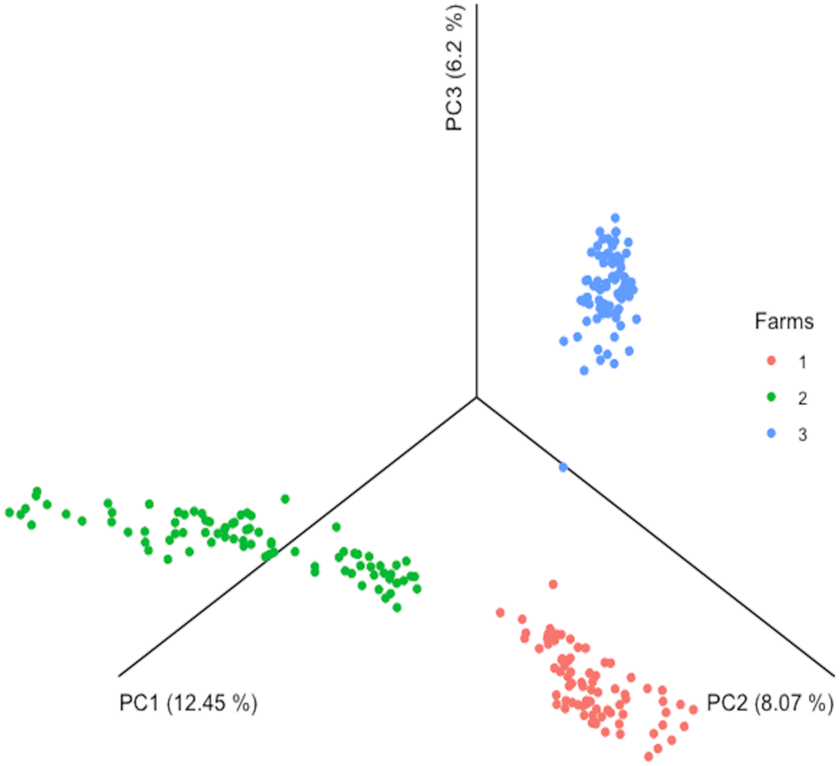
- Post GWAS:

$$y_{ijkl} = \mu + G_i + \text{sex}_j + \text{slau}_k + \text{farm}_l + e_{ijkl}$$

y = observation vector for the ith trait;
μ = mean of the population;
G = fixed effect of each SNP (i = 1, 2, 3);
sex = two levels for barrows and gilts;
slau = five levels for slaughter day;
farm = three levels;
e = error represent random effects of residues

- **Linkage disequilibrium analysis** performed using **Haploview 4.2 software**

Results



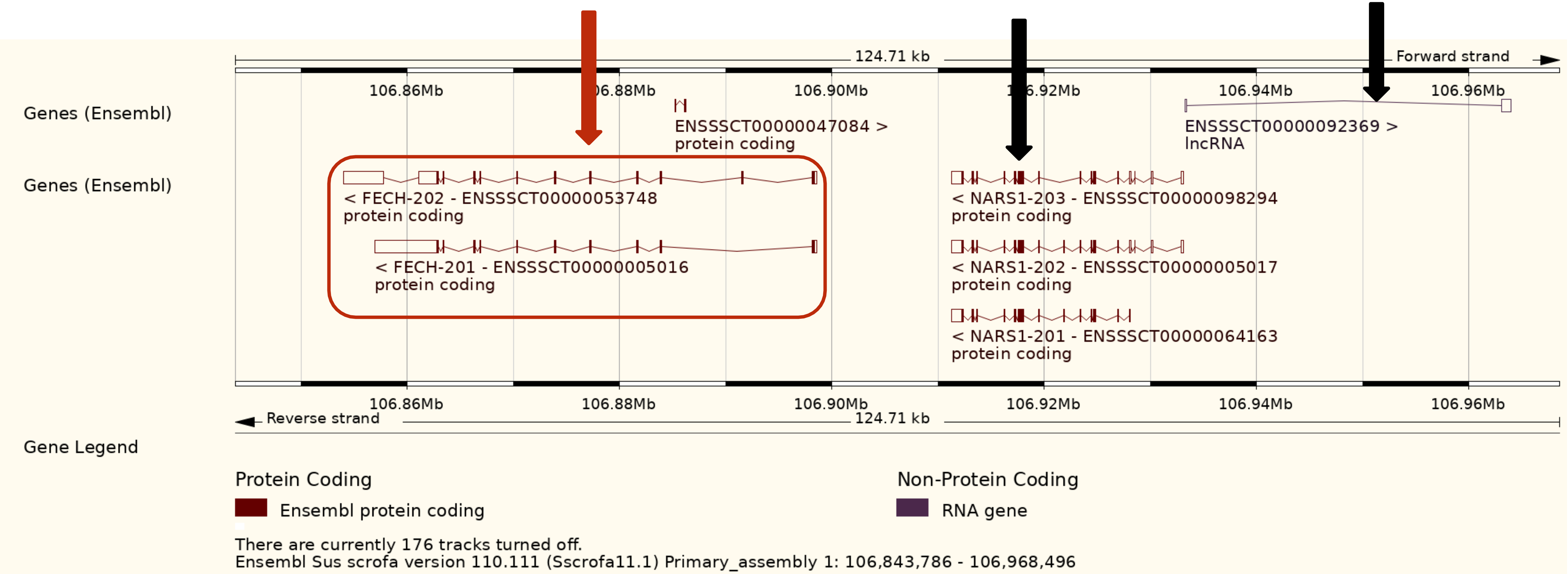
SNPs		ALGA0005395[1]	ASGA0004152[1]	DIAS0002366[1]
Reference n° of SNPs		rs81355515	rs81216562	rs81216057
nt on S.scrofa chr 1		106,952,113	106,877,209	106,917,692
effB (SEB)		-4.85 (0.75)	-4.92 (0.76)	-4.92 (0.76)
Eff (AB)		-5.61	-5.46	-5.46
Eff (BB)		-9.44	-9.73	-9.73
FDR		4.02E-11	4.02E-11	4.02E-11
Allele frequencies	**1**	0.51	0.55	0.41
	2	0.49	0.45	0.59
Genotype frequencies	**11**	0.27 (65)	0.32 (77)	0.2 (48)
	12	0.48 (113)	0.46 (108)	0.41 (98)
	22	0.25 (59)	0.22 (52)	0.38 (91)

[1] 1% genome-wide significant markers.

ASGA0004152
(Missense variant,
Val > Ala on *FECH*)

DIAS0002366
(Synonymous variant
on *NARS1*)

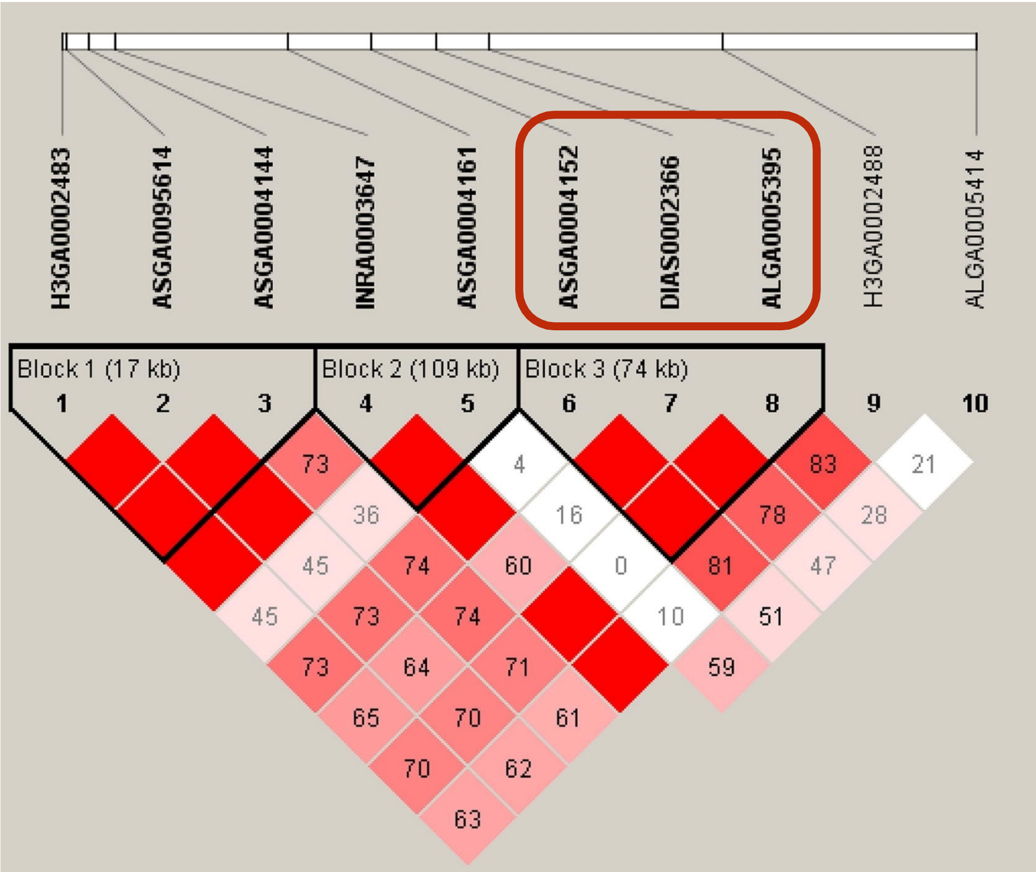
ALGA0005395
(Intronic variant,
on *ENSSSCG00000057874*)



There are currently 176 tracks turned off.
Ensembl Sus scrofa version 110.111 (Sscrofa11.1) Primary_assembly 1: 106,843,786 - 106,968,496

(https://www.ensembl.org/Sus_scrofa/Location/View?r=1:106843786-106968496;db=core;time=1692716964902.902)

- **Three linkage blocks** in the region 106.6-107.3 Mb of porcine chr 1



SNPs		ALGA0005395	ASGA0004152	DIAS0002366
***P*-value**		< 2.2E-16	< 2.2E-16	< 2.2E-16
EMM	**11**	49.7	50.2	48.9
	12	34	31.6	29.8
	22	16.1	14.5	14.1
Contrasts[1]	**11-12**	15.70^{***}	18.60^{***}	15.80^{***}
	11-22	33.50^{***}	35.80^{***}	34.80^{***}
	12-22	17.80^{***}	17.10^{***}	19.00^{***}
Additive effect[2]		***	***	***
Dominance effect		0.65	0.74	0.49

[1] *** significant for $P \le 0.001$.

[2] *** significant for $P < 0.0001$.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

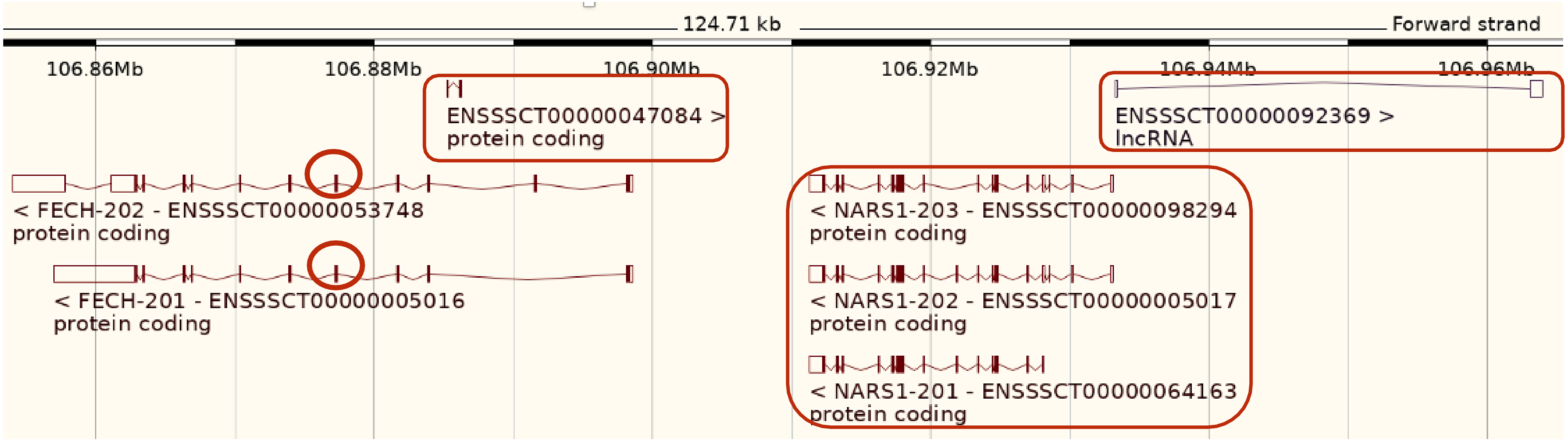
Discussion

- The animals showed highest and lowest FECHA were homozygous for **ALGA0005395, ASGA0004152** and **DIAS0002366** markers most significantly associated with this parameter.

Important effect exerted by the alleles

- Now there is no indication whether these **two transcripts (FECH-201 and FECH-202) may be responsible for a different FECHA (or just in linkage).**

A large difference was noticed between the EMMs for the opposite homozygotes of **ASGA0004152** was observed.



- **SIFT** (https://sift.bii.a-star.edu.sg/) reveals that variant ASGA0004152 (rs81216562) is **not** deleterious for both transcripts (FECH-201 and FECH-202).

Compatible with the function of the coded protein

- ***NARS1***, ***ENSSSCG00000033063*** and ***ENSSSCG00000057874*** genes can be also considered positional candidates for FECHA.

Conclusion

- **FECHA** is a trait **has not yet been considered for GWAS** or other association analysis.

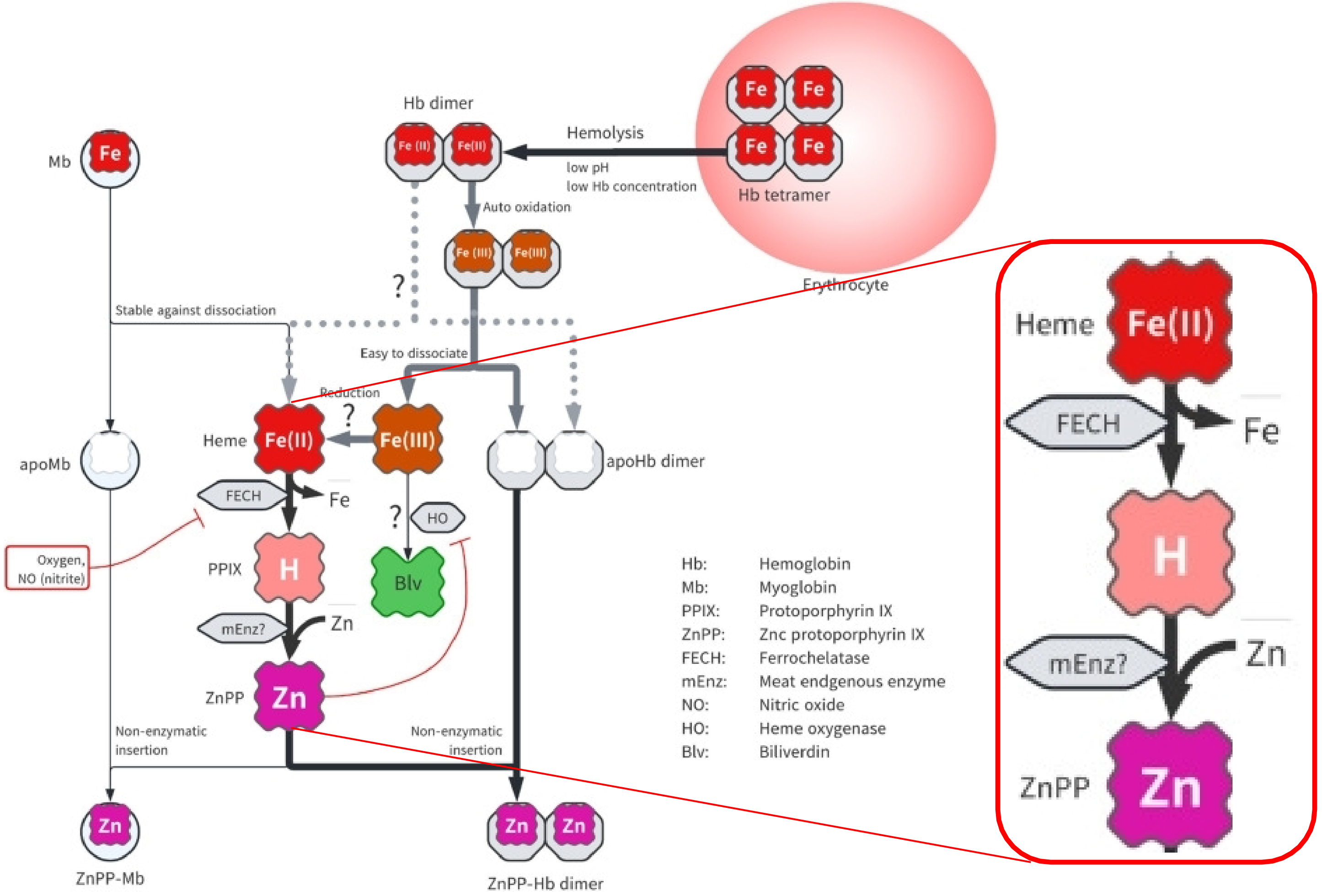
- **ASGA0004152** can be considered an interesting candidate for further studies**.**

- **ASGA0004152 influences the activity of the translated peptide and**, as a consequence, **the efficiency of the enzymatic reaction catalyzed.**

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Thank you for your attention!

Jacopo Vegni
PhD in Zootechnical Sciences
jacopo.vegni2@unibo.it
Department of Agricultural and Food Sciences- DISTAL
viale Fanin 46, 40127, Bologna



Hb:	Hemoglobin
Mb:	Myoglobin
PPIX:	Protoporphyrin IX
ZnPP:	Znc protoporphyrin IX
FECH:	Ferrochelatase
mEnz:	Meat endgenous enzyme
NO:	Nitric oxide
HO:	Heme oxygenase
Blv:	Biliverdin

(Image modified from the one in Wakamatsu, 2022)

Statistical analysis

- Descriptive statistics and Pearson correlations:
  - **All statistics on phenotypic traits** related to pig carcass, fresh and salted hams, and processing weight losses were obtained by **SPSS ver. 28 statistical package** (SPSS Inc., Chicago, USA). Person phenotypic correlations were evaluated between the ham-assayed quality traits and the technological traits measured over processing.

- Genome-wide association study:
  - Markers mapped in sex chromosomes, on the mitochondrial genome, or unmapped, were removed (5,497 markers were escluded).

Quality control

63,019 SNPs
gPLINK 1.9 (based PLINK 1.07; Purcell et al., 2007)
SNPs were excluded:

- **Call rate for SNP < 90%**
- **Minor allele frequency (MAF) < 0.05**
- **Hardy-Weinberg equilibrium with *P*-value < 0.001**

Individuals with call rate **> 10% of missing data were removed**

Genetic distance population & vector extrapolation

PLINK

Principal components analysis

Factor: **farm 1, 2, 3**

Cleaning of samples

check.marker function of **GenABEL** (R version 3.4.4)
and the samples with the:
Identity by state > 0.9 were removed
51,166 SNPs e 237 individuals obtained

GWAS

- ***polygenic*** (grammar-gamma) and ***qtscore*** functions GenABEL (Nicolazzi et al., 2015).
- **Statistical model** included **sex, the day of slaughter, and the farm** (each farm was characterised by a genetic type) as **fixed effects**.
- **Genomic kinship matrix** calculated using the ***ibs*** function in the GenABEL and considering the employed SNPs included in the statistical model to infer the pedigree relationships among the pigs
- Additive polygenic model fitted with a genomic relationship matrix:

$$\mathbf{Y_i = X_i\beta_i + Z_i a_i + e_i}$$

Y_i = observation vector for the ith trait;
β = vector of effects for three factors (sex = barrows and gilts; slaughtering date = d1, d2, d3, d4, d5 and farm = f1, f2 and f3);
a = animal (random factor)
e = residuals (random factor)

They assumed to be normally distributed as $a_i \sim N(0,G\sigma_a^2)$ and $e_i \sim N(0,I\sigma_e^2)$:
G = genomic relationship matrix;
σ_a^2 = additive genomic variance
σ_e^2 = residual variance

- **SNPs significant for *P*-values** below the genome or chromosome-wide threshold **of 1 and 5%** (corrected for the genomic inflation factor λ, close to 1 with minimum value of 0.9999805 for all traits, a good correction for the stratification of the samples).
- Test significant for **FDR < 0.01.**

- Detection of the nearest genes an isoform characterization:
  - Genes present within **± 500 kb region flanking** the associated markers used for identification of candidate genes (**ENSEMBL pig genome viewer**).
  - **Characterise** the two swine Ferrochelatase (FECH) mRNA isoforms (**FECH-201**: ENSSSCT00000005016; **FECH-202**: ENSSSCT00000053748).
  - To define **where the mutant SNP located** within the two mRNAs (**blastn online tool**)

- Post genome-wide association study:
  - For significant markers, an **association study for each SNP** performed to estimate the effect of the alleles with a linear model to obtain **the Estimated Marginal Means (EMM)**. The fixed effects were sex, day of slaughter, and farm:

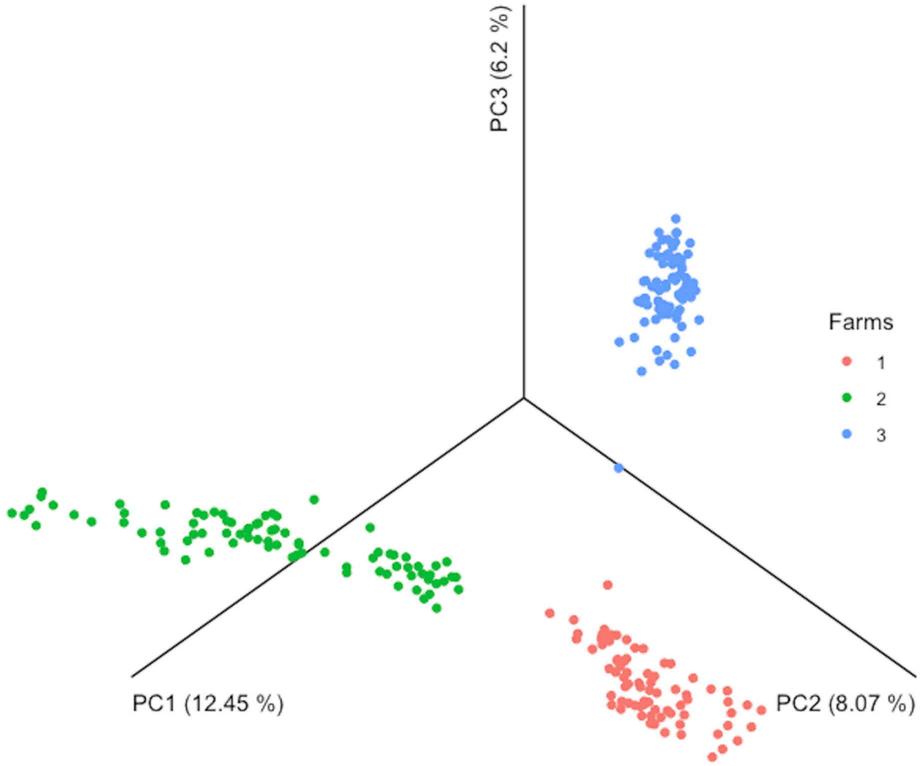
$$y_{ijkl} = \mu + G_i + sex_j + slau_k + farm_l + e_{ijkl}$$

y = observation vector for the ith trait;
μ = mean of the population;
G = fixed effect of each SNP (i = 1, 2, 3);
sex = two levels for barrows and gilts;
slau = five levels for slaughter day;
farm = three levels;
e = error represent random effects of residues

- **Linkage disequilibrium analysis** performed using **Haploview 4.2 software** with default settings and using the most significant markers (i.e the markers showing the three genotypes or having a *P*-value and/or the contrasts significant) located on *Sus scrofa* chromosome 1

Results

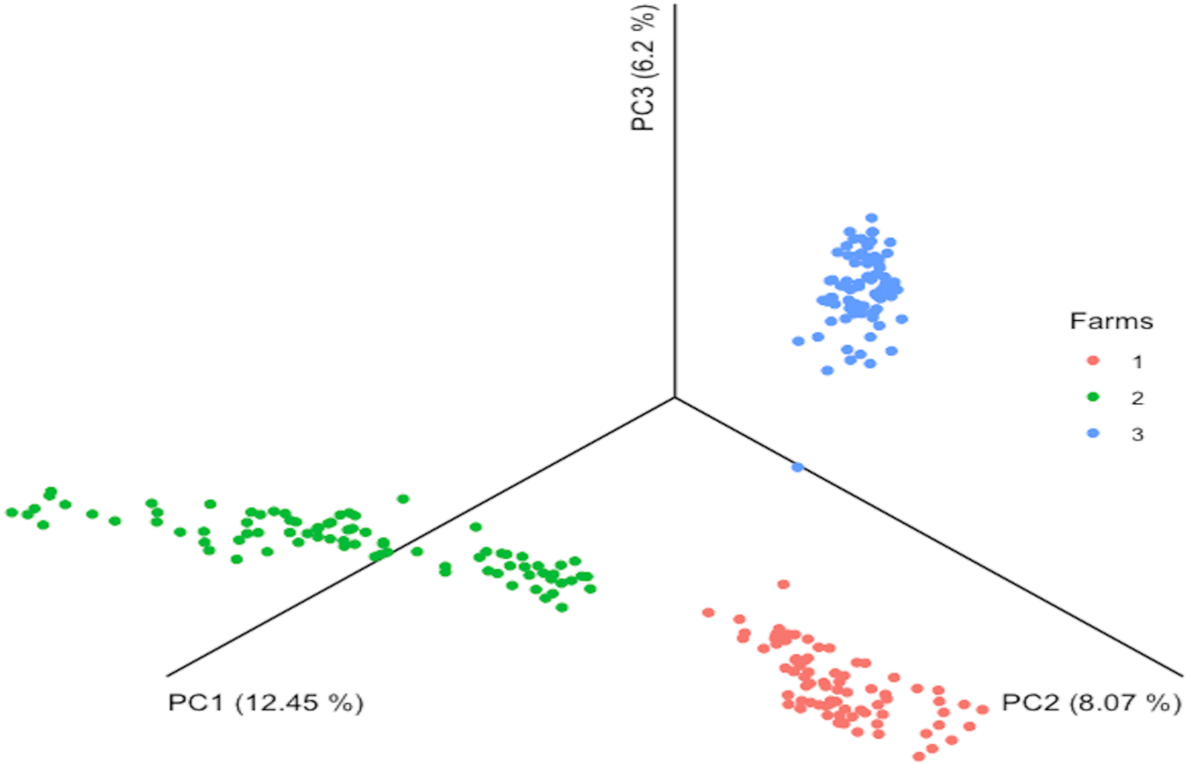
- **14 genome-wide association markers** were associated with the **FECHA** on the limited region of porcine chromosome 1 (105.75-116.36 Mb).



SNP markers	Reference n° of SNPs	nt on *S.scrofa* chr 1	A1	A2	effB(SEB)[1]	effAB[2]	effBB[3]	FDR	Allele frequencies		Genotype frequencies[4]		
									1	2	11	12	22
ALGA0005395[5]	rs81355515	106,952,113	G	C	-4.85(0.75)	-5.61	-9.44	4.02E-11	0.51	0.49	0.27 (65)	0.48 (113)	0.25 (59)
ASGA0004152[5]	rs81216562	106,877,209	A	G	-4.92(0.76)	-5.46	-9.73	4.02E-11	0.55	0.45	0.32 (77)	0.46 (108)	0.22 (52)
DIAS0002366[5]	rs81216057	106,917,692	A	G	-4.92(0.76)	-5.46	-9.73	4.02E-11	0.41	0.59	0.20 (48)	0.41 (98)	0.38 (91)
H3GA0002488[5]	rs81355534	107,100,243	G	A	-4.49(0.77)	-4.19	-8.99	1.39E-08	0.49	0.51	0.22 (53)	0.53 (126)	0.24 (58)
INRA0003647[5]	rs321031460	106,714,123	C	A	-4.74(0.84)	-5.63	-8.51	3.51E-08	0.07	0.93	0.00 (0)	0.15 (35)	0.85 (202)
ASGA0095614[5]	rs81474204	106,683,184	A	G	-4.50(0.81)	-5.05	-8.55	6.03E-08	0.66	0.34	0.45 (106)	0.43 (102)	0.12 (29)
ASGA0004144[5]	rs81355510	106,697,548	A	C	-4.50(0.81)	-5.05	-8.55	6.03E-08	0.32	0.68	0.10 (24)	0.43 (102)	0.47 (111)
ASGA0004161[5]	rs81355527	106,823,633	G	A	-4.85(0.94)	-5.79	-7.17	1.57E-06	0.25	0.75	0.06 (14)	0.38 (90)	0.56 (133)
ALGA0005414[5]	rs80808933	107,261,641	A	G	-4.18(0.82)	-3.21	-8.45	2.21E-06	0.42	0.58	0.13 (30)	0.58 (137)	0.30 (70)
ALGA0005524[5]	rs80820778	110,947,652	G	A	4.24(0.93)	4.68	7.51	9.42E-05	0.70	0.30	0.48 (113)	0.46 (108)	0.07 (16)
INRA0003923[5]	rs318548067	116,359,904	G	A	-3.39(0.80)	-3.01	-7.09	8.39E-04	0.70	0.30	0.51 (121)	0.39 (92)	0.10 (24)
INRA0003610[6]	rs319996306	105,751,534	A	C	-3.88(0.95)	-4.61	-5.83	2.23E-03	0.85	0.15	0.71 (168)	0.28 (66)	0.01 (3)
H3GA0002483[7]	rs80929711	106,679,629	G	A	2.96(0.74)	5.44	5.66	4.33E-03	0.34	0.66	0.12 (29)	0.43 (102)	0.45 (106)
ALGA0005538[8]	rs80800059	111,184,327	G	A	3.13(0.86)	4.50	4.64	2.85E-02	0.79	0.21	0.61 (144)	0.37 (87)	0.03 (6)

[1] Effect of the B allele in the allelic test; SEB: SE of effB reported in brackets.
[2] Effect of the AB genotype in the genotypic test.
[3] Effect of the BB genotype in the genotypic test.
[4] The number of samples is reported in brackets.
[5] 1% genome-wide significant markers.
[6] 5% genome-wide significant marker.
[7] 1% chromosome-wide significant marker.
[8] 5% chromosome-wide significant marker.

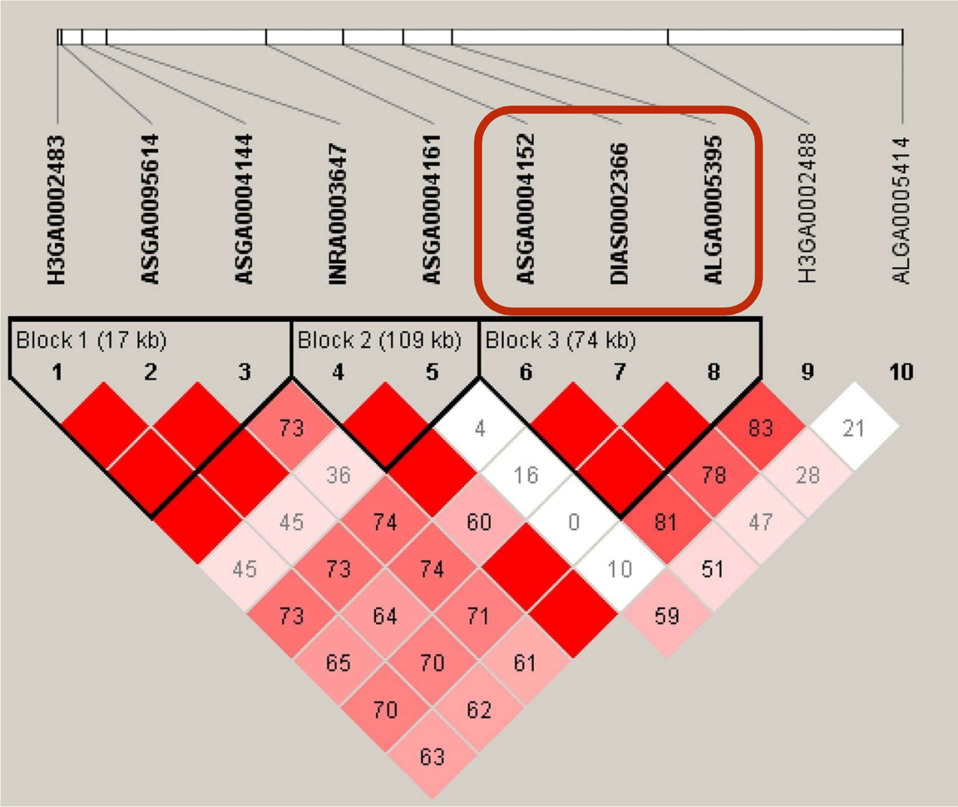
Results



SNP	Reference n° of SNPs	nt on *S.scrofa* chr 1	A1	A2	effB (SEB)	Eff AB	Eff BB	FDR	Allele frequencies		Genotype frequencies		
									1	2	11	12	22
ALGA0005395[1]	rs81355515	106,952,113	G	C	-4.85 (0.75)	-5.61	-9.44	4.02E-11	0.51	0.49	0.27 (65)	0.48 (113)	0.25 (59)
ASGA0004152[1]	rs81216562	106,877,209	A	G	-4.92 (0.76)	-5.46	-9.73	4.02E-11	0.55	0.45	0.32 (77)	0.46 (108)	0.22 (52)
DIAS0002366[1]	rs81216057	106,917,692	A	G	-4.92 (0.76)	-5.46	-9.73	4.02E-11	0.41	0.59	0.20 (48)	0.41 (98)	0.38 (91)

[1] 1% genome-wide significant markers.

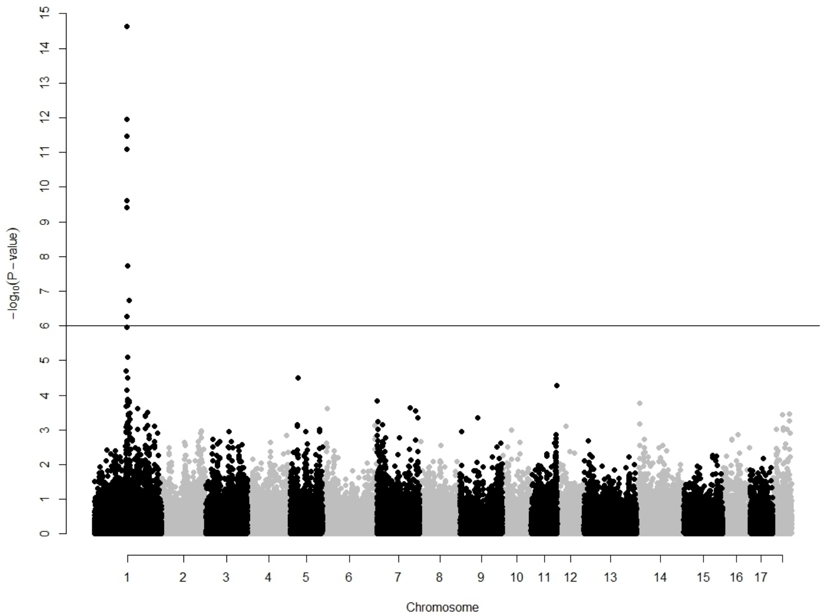
- **Three linkage blocks** in the region 106.6-107.3 Mb of porcine chr 1

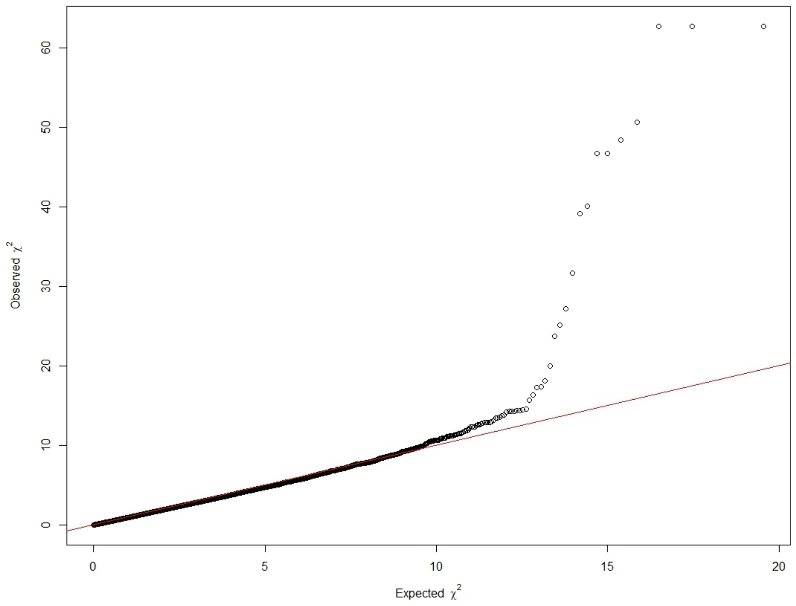


SNP	P-value	EMM			Contrasts[1]			Additive effect[2]	Dominance effect
		11	12	22	11-12	11-22	12-22		
ALGA0005395	< 2.2E-16	49.70	34.00	16.10	15.70^{***}	33.50^{***}	17.80^{***}	***	0.65
ASGA0004152	< 2.2E-16	50.20	31.60	14.50	18.60^{***}	35.80^{***}	17.10^{***}	***	0.74
DIAS0002366	< 2.2E-16	48.90	29.80	14.10	15.80^{***}	34.80^{***}	19.00^{***}	***	0.49

[1] *** significant for $P \le 0.001$.

[2] *** significant for $P < 0.0001$.





ASGA0004152

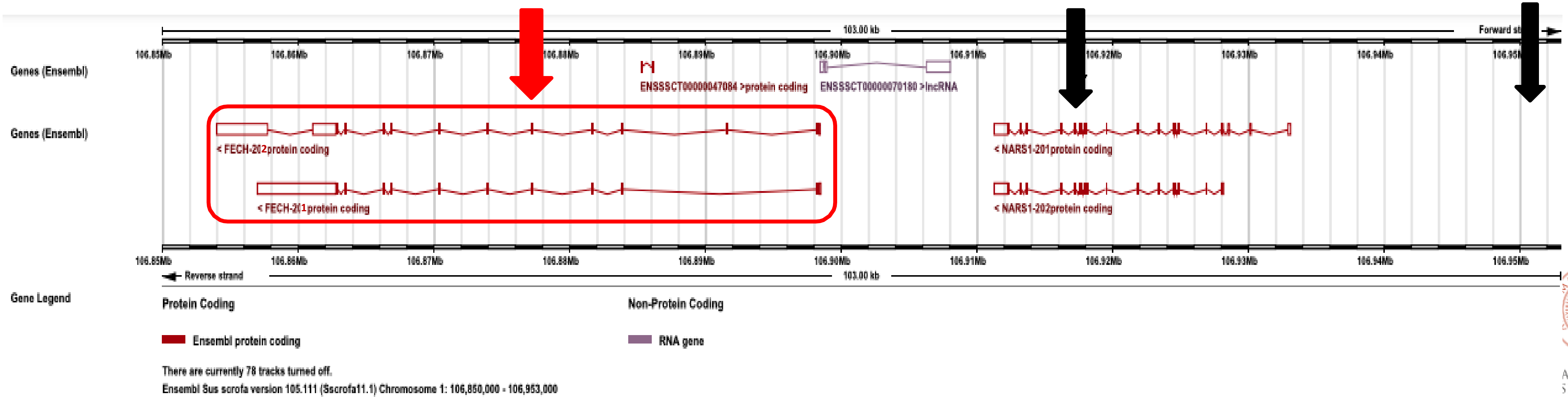
(Val > Ala)

DIAS0002366

(Synonymous variant)

ALGA0005395

(between *NARS1* and *ENSSSCG00000033063*)



There are currently 78 tracks turned off.

Ensembl Sus scrofa version 105.111 (Sscrofa11.1) Chromosome 1: 106,850,000 - 106,953,000

SNP markers	P-value	EMM[1]			Contrasts[2]			Additive effect[3]	Dominance effect[3]
		11	12	22	11-12	11-22	12-22		
ALGA0005395	< 2.2E-16	49.70	34.00	16.10	15.70***	33.50***	17.80***	***	0.65
ASGA0004152	< 2.2E-16	50.20	31.60	14.50	18.60***	35.80***	17.10***	***	0.74
DIAS0002366	< 2.2E-16	48.90	29.80	14.10	15.80***	34.80***	19.00***	***	0.49
H3GA0002488	2.36E-15	46.70	35.20	15.70	19.50***	31.00***	11.60***	***	0.11
INRA0003647[d]	3.67E-05	-	20.00	36.10	-	-	-16.10***	-	-
ASGA0095614	5.62E-13	43.0	29.90	12.40	13.10***	30.50***	17.50***	***	0.42
ASGA0004144	1.21E-12	42.8	28.60	12.60	16.10***	30.30***	14.20***	***	0.75
ASGA0004161	0.09	43.7	30.50	34.70	13.22	9.00	-4.22	0.17	0.03
ALGA0005414	1.71E-05	51.2	33.50	26.80	17.66***	24.36***	6.69***	***	0.07
ALGA0005524	0.10	31.9	34.20	44.60	-2.36	-12.77	-10.40	0.03	0.26
INRA0003923	0.57	33.4	35.10	30.00	-1.68	3.40	5.08	0.50	0.30
INRA0003610	0.14	39.5	38.30	31.80	6.55	7.78	1.23	0.61	0.74
H3GA0002483	3.61E-13	12.40	29.90	43.10	-17.50	-30.70	-13.20	***	0.44

[1] All significant Estimated Marginal Means; adjusted means of the considered trait according to the statistical model used effects are significant for $P < 0.001$.

[2] *** means that the contrasts are significant for $P \leq 0.001$.

[3] *** means that the effects are significant for $P < 0.0001$.