Environmental and genetic influences on whole-blood gene expression levels in Large White/Creole backcross pigs

Arthur Durante

Rapport de stage

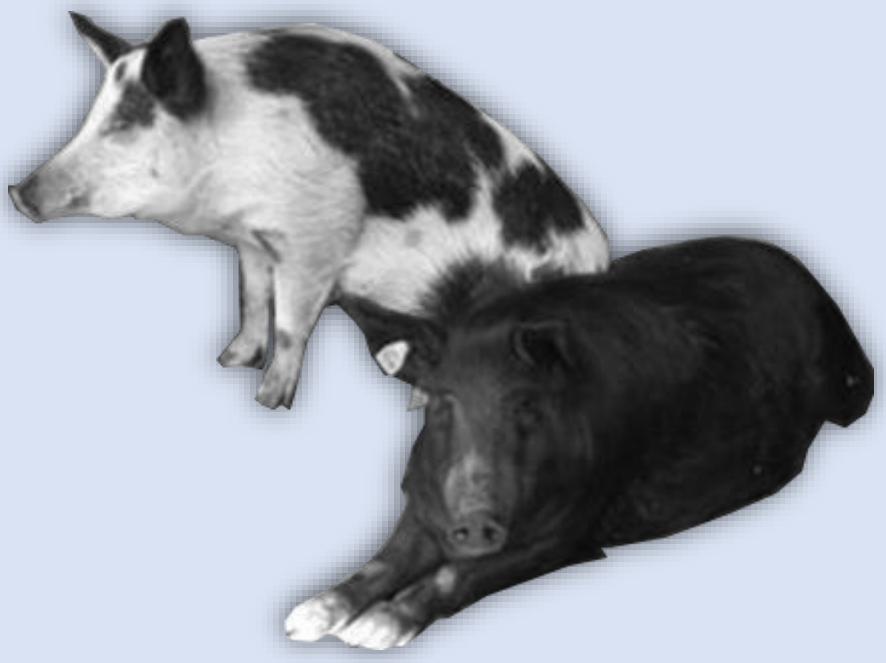
Janvier – Juin 2023

Environmental and genetic influences on whole-blood gene expression levels in Large White/Creole backcross pigs

Durante Arthur

Aix-Marseille Université
Master Sciences et Technologie
Mention Biochimie structurale, Génomique
Parcours Génomique et analyse des données
2022-2023

Tuteur de stage : Guillaume DEVAILLY
Professeur référent : Christophe BORDI
Organisme d'accueil : UMR 1388 GenPhySE INRAE
24 Chemin de Borde Rouge
31320 Auzeville-Tolosane



Acknowledgements

I am grateful to **Guillaume Devailly** for his guidance and help during this internship, as well as for welcoming me to the GENEPI team during the internship period. I am also grateful for his willingness to continue mentoring me during the thesis project set as a follow-up to the internship.

I am grateful to **Yann Labrune** and **Laurence Liaubet** for their help in acquiring different data sets needed during this internship, as well as for introducing me to analysis tools used within this internship. I would also like to thank **Denis Milan** for multiple interesting and intriguing insight given during the internship.

I am grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing help for computing and storing resources.

I am grateful to **Noémien Maillard**, **Lea le Berre** (GenPhySE M2 interns), as well as **Elora Bohuon** (CDD technician) and **Mathilde Perret** (PhD student) for their insight in certain thought processes during my internship.

I am also grateful for the warm welcome given by the entire GENEPI team and the GenPhySE unit researchers.

Table of contents

Acronyms Table

Acronym	Full form
HS	Heat stress
BC	Backcross
QTL	Quantitative trait locus
eQTL	Expression quantitative trait locus
DEG	Differentially expressed gene
SNP	Single-nucleotide polymorphism

Table of figures

Introduction

Meat consumption has increased with the enrichment of the global world population, and is expected to increase up to 76% by 2050 (Godfray *et al.*, 2018), including increases up to 42% in pork, 69% in beef and 100% increase in poultry (Alexandratos and Bruinsma, 2012). Ongoing efforts are being made to limit the environmental impact of livestock production, including through the reduction of per-capita meat consumption in rich countries. Nonetheless, nourishing the growing population will require livestock production systems adapted to withstand a global temperature increase from +2.5°C to +4.5°C expected in 2080(Coffel, Horton and De Sherbinin, 2018). The increasing amount of heatwaves expected to come with this increase of global temperature will drastically affect livestock production, as the animals might not be able to cope with their surrounding temperature. Heat stress has already caused the loss of around US$1200 million in the dairy industry, US$370 million in the beef industry and around US$300 million in the pig industry (St-Pierre, Cobanov and Schnitkey, 2003).

Pigs are the second most consumed livestock meat worldwide (FAO, 2021). Unlike most other livestock, pigs are particularly unable to cope with an increase of environmental temperatures, therefore, heat stress (HS) acts as limiting factor for pig growth and production (Renaudeau *et al.*, 2012). The low amount of sweat glands present in their dermis and epidermis decreases the capacity to cope with heat stress through cutaneous evaporation. Pigs need to compensate through other means such as evaporation through the respiratory tract (Renaudeau, Leclercq-Smekens and Herin, 2006). Pigs also decrease the amount of heat produced by their metabolism, through a significant decrease in feeding intake, inherently leading to a slowdown of the animal growth (Liu *et al.*, 2022).

Commercial pork meat comes from crossbred pigs, mainly for the genetic improvements brought by the heterosis effect, increasing their growth and resilience capacities (Cassady, Young and Leymaster, 2002). For instance, heterosis increased heat stress resilience throughout multiple crossbreeding systems (Cassady, Young and Leymaster, 2002) as well as in other livestock (Menéndez-Buxadera *et al.*, 2012; Nguyen *et al.*, 2017). Genetics might compensate the rise of temperature, but this will require the identification of alleles enabling a good adaptation and resilience to heat stress, notably including alleles from tropical breeds. Thus, an experimental cross has been set-up between two different pig races: Creole and Large White.

The Creole pigs are particularly useful in this experimental approach, as they originate from a tropical environment, making them particularly resistant to heat stress compared to those bred solely for their growth capacity (Waltz *et al.*, 2014). In contrast, the Large White race is cosmopolitan, and particularly efficient for its growth capacity, but pigs from this race are less resilient to HS than their Creole counterparts (Gourdine *et al.*, 2006; Poullet *et al.*, 2023).

The Creole x Large White crossbreeding strategy aims to produce genetic and phenotypic diversity between production capacity and heat resilience within their offspring, thus also increasing the statistical power available for association studies. The backcross pigs studied were sired from Large White mothers and a F1 hybrid father (½ Créole, ½ Large White) using artificial insemination. Each Large White mothers are related to the same Large White grand-sire, therefore sharing their genetic backgrounds. Large White mothers either live in tropical environment in French Guadeloupe or in tempered environment in Charente (France) (**Figure 1**). This process allows the determination of segregated alleles within the backcrossed (**BC**) pigs, between their ¼ Creole genetic background and their ¾ Large White genetic background. The ability of these BC pigs to withstand HS has been tested for multiple conditions, either by continuously living in a tropical environment, or in a tempered environment where they experience an artificially induced heat wave over the span of two weeks.

This research plan was widely studied among INRAE teams, as numerous datasets have been produced from the experiment. Such datasets include phenotyping, genotyping, metabolome, microbiota and blood transcriptome data. Studies conducted on phenotypic traits resulted in a low to moderate heritability rate of thermoregulatory traits and the discovery of the impact of the environment on the genetic mechanisms underlying the traits (Renaudeau, Huc and Noblet, 2007; Rosé *et al.*, 2017; Gourdine *et al.*, 2019). Metabolome studies showed the promising results capable of being obtained by using blood samples as a reference for biomarker detection, as correlations were observed between the pig sensitivity predicted from blood metabolome and a composite sensitivity index computed through *t* statistics applied to phenotypic responses to HS (Dou *et al.*, 2017). Microbiota analysis resulted in two distinct enterotypes, each correlated with a different physiological response to HS, with the enterotype specific to pigs living in tropical environment showing a better adaptability to HS (Le Sciellour *et al.*, 2019).

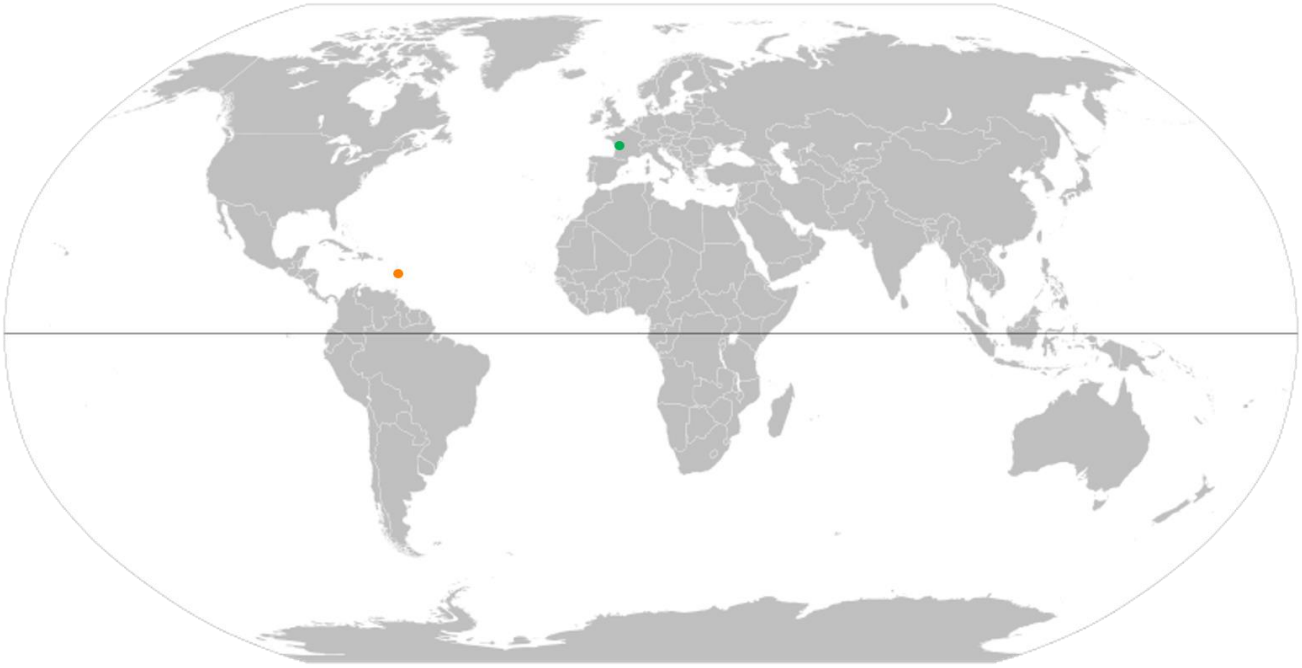


***Figure 1**: **Location of the two experimental farms**. The orange dot locates the farm in French Guadeloupe (pigs living in tropical environment). The green dot locates the farm in Poitou-Charantes, France (pigs living in tempered environment)*

The data used for this particular study includes the aforementioned transcriptomic and genotypic data, obtained from an Agilent SurePrint G3 Microarray and Porcine SNP60v2 BeadChip (60K SNPs chip), respectively 361 and 1 263 animals.

The aim of this study is to detect genes differentially expressed during a HS event, whether they be by continuous exposition (pigs living in tropical region) or immediate stress (pigs subject to an experimental heat wave). The second aim of this study is the detection of genetic variants capable of affecting the expression levels of genes within the BC pigs. Such variants could then be used to provide plausible mechanisms linked with the phenotypical thermoregulatory traits identified by others (unpublished) in this experiment.

Materials and Methods

Animal production and data generation

The experimental BC population (Rosé *et al.*, 2017) comprises of two different pig populations. First are 634 pigs reared from 60 Large White sows, raised in a temperate environment (INRA-Génétique, Expérimentation et systems Innovants, Poitou-Charantes, France, 46° N, 0.45° W). Second are 664 BC pigs reared from 70 Large White sows, raised in a tropical environment (INRA-Plateforme Tropicale d'Expérimentation sur l'Animal, Guadeloupe, French West Indies, 16° N, 0.45° W) (Gourdine *et al.*, 2019). A group of 10 F_1 Creole × Large White boars were sired with genetically related Large White sows (themselves sired from the same Large White fathers), resulting in related BC pigs with approximately similar genetic profiles in both living environments (**Figure 2**). The test period in this study comprises of 2 weeks of artificially induced HS for the pigs living in tempered area and no specific HS period for pigs living in the tropical area, with a daily mean Temperature Humidity Index difference of 2.4°C between the two living environments (Rosé *et al.*, 2017; Gourdine *et al.*, 2019).

Whole blood transcriptomic data was generated for a subset of 361 animals: 181 pigs living in tropical environment and 180 pigs living in tempered environments. Whole blood transcriptomic analysis was performed on the Agilent SurePrint G3 microarray designed for the pig genome. The TRiX sequencing platform performed multiple quality control steps before releasing the data. They proceeded to remove aberrant values caused by spots and scratches on the array, as well as transcripts with little to no values. Each probe was then annotated to match *Sus scrofa* 11.1 Ensembl annotation (Warr *et al.*, 2020). More precisely, we used the original annotation of the probe (on the *Sus scrofa* 10.2 Ensembl genome), corrected and augmented by a newer annotation made by the SIGENAE team on the *Sus scrofa* 11.1 Ensembl genome.

The 180 pigs living in temperate conditions were exposed to an experimentally induced 30°C HS for a period of two weeks. Samples were taken before HS (d0), two days during HS (d2) and two weeks during HS (d15). Quality controls and filters were applied to the resulting transcriptomic data before the different analysis presented in this study.

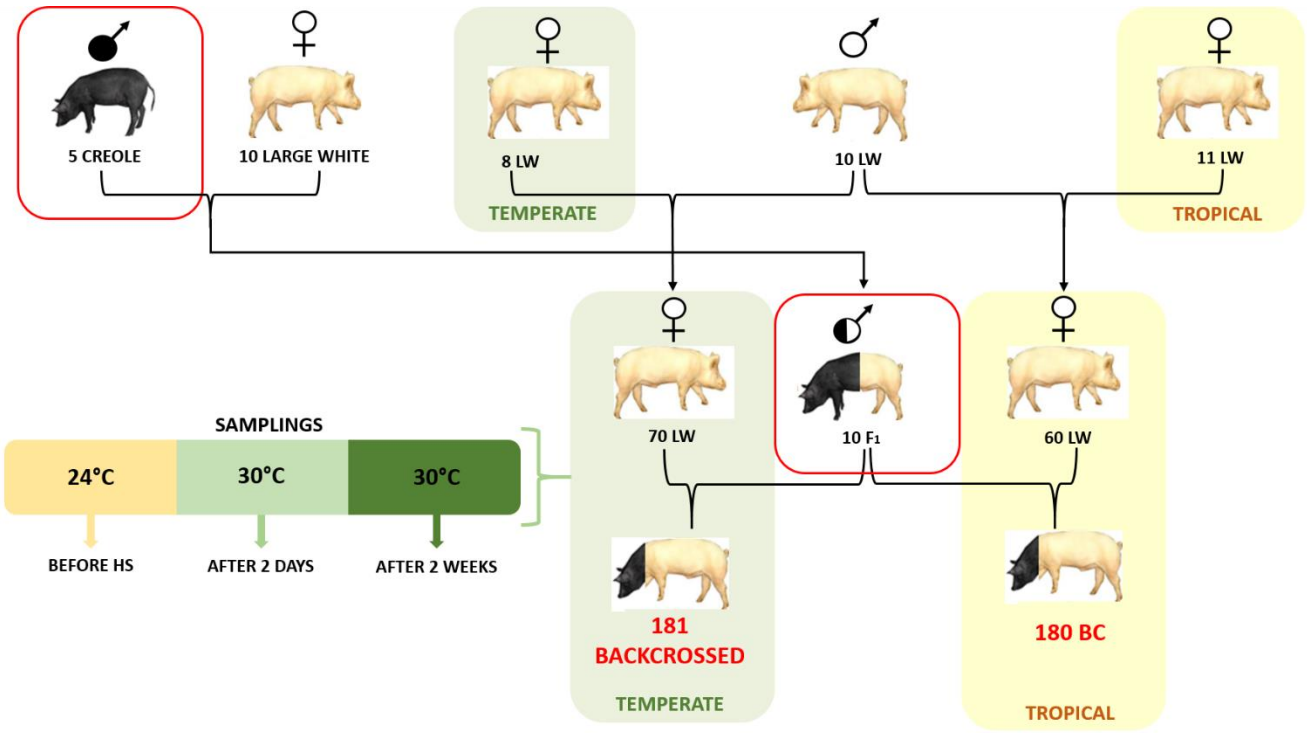


***Figure 2**: **Backcross design** with the number of tested castrated male and female backcross pigs sired from 10 F_1 in tempered or tropical environments. ♀ = female; ♂ = male; LW = Large White; CR = Creole; BC = backcross; HS = heat stress. Backcross were castrated.*

Genotypic data were obtained for 1 263 pigs using Porcine SNP60v2 BeadChip (60K SNPs chip), 358 of which were also analysed by the transcriptomic microarray mentioned above. Data quality control was carried out before my internship, by excluding single-nucleotide polymorphisms (SNPs) that failed to reach a call rate of 95%. During the genome-wide association step, we excluded SNPs with a minor allele frequency below 1% (set as default in Gemma), resulting in a total of 48 498 exploitable SNPs.

Data Processing

During most parts of the analysis process, I computed and analysed the data by using R studio (version 4.2.2). For heavier tasks requiring more computational power, I used the Genotoul cluster (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12).

Transcriptomic analysis

Data Normalization

Samples present on the microarray comprise of the 181 tropical samples, as well as 180 temperate samples (before HS, d0), 177 temperate samples (2 days along the HS, d2) and 180 temperate samples (2 weeks along the HS, d15) for a total of 718 whole blood samples. We processed the transcriptomic data through multiple quality control selections before the analysis step. We first removed probes with missing values for more than 1% of samples. Next, we removed samples with more than 1% missing probe values (removing 7 484 probes and one sample, respectively), resulting in 35 802 exploitable probes expressions.

A quantile normalization was applied per sample to homogenize value distributions for each samples (Qiu, Wu and Hu, 2013). We removed samples containing more than 10% aberrant probe values, defined as values of more than 4 times the standard-deviation above or bellow the mean probe expression for a sample. Beside the removal of samples, aberrant values were changed into NAs for the remaining samples. This filter resulted in the removal of two samples, leading to a total of 715 exploitable pig blood samples.

We conducted multiple PCA coloured by different metadata columns using the mixOmics R package (version 6.22.0) (Rohart *et al.*, 2017). The metadata file comprised of the sample

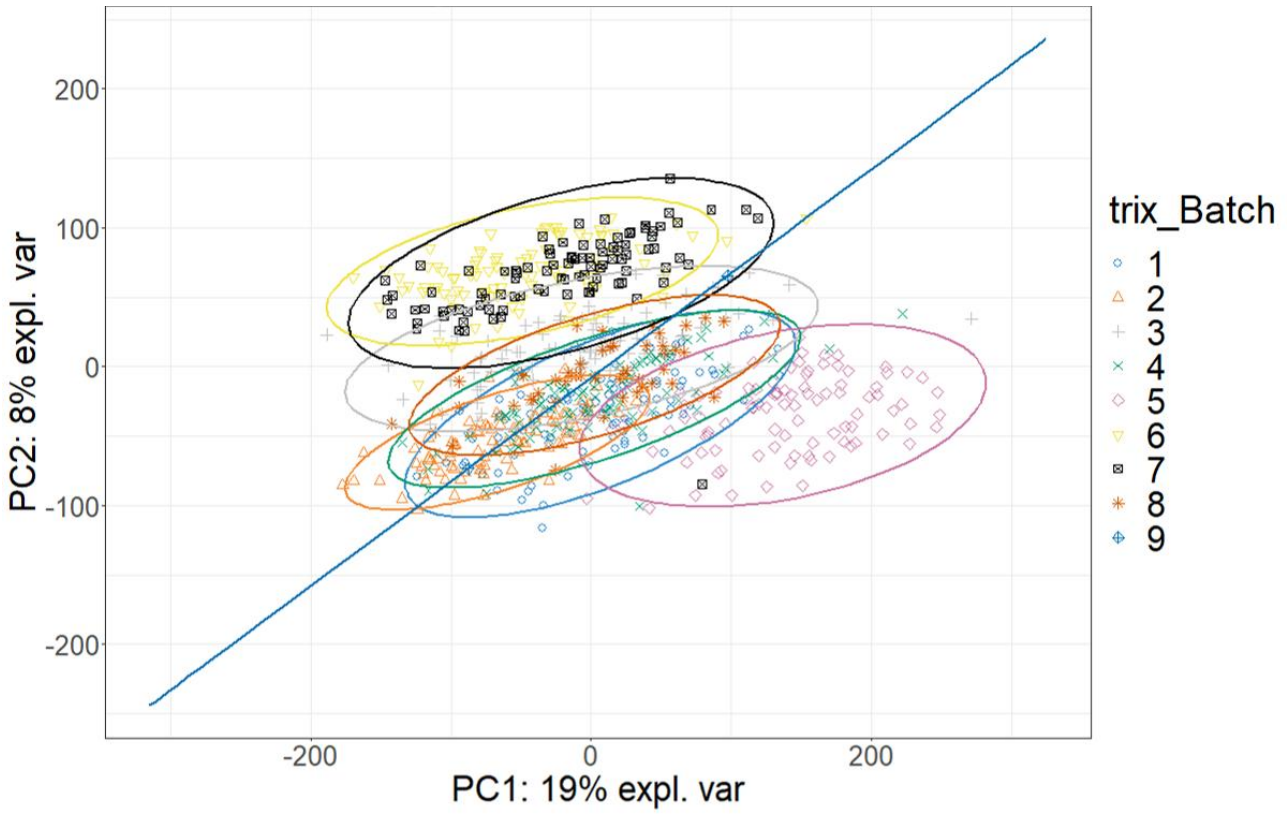


***Figure 3**: **Principal component analysis of the transcriptomic dataset**. Each dot is a sample coloured according to its batch on the transcriptomic facility. Ellipsis are group samples of the same batch. The batch number 9 comprises of only 2 values, the resulting ellipsis thus appears as a segment. Comparison between the first (PC1) and second (PC2) components.*

ID, animal ID, condition, sex, breeding strip (a group of animals raised together), and transcriptomic batch. **Figure 3** notably highlights the unfortunate effect of the transcriptomic batches (Lazar *et al.*, 2013).

A total of 88% of exploitable expression probes were annotated to a gene. Some genes were measured by several probes, some of them measuring closely similar values (**Figure 4A**). We deduplicated the transcriptomic dataset using the following approach: correlation were computed between each expression probes measuring the same gene. Probes with a spearman correlation above 0.9 were merged, leading to a single synthetic probe using the average expression value (**Figure 4**). The deduplication process resulted in the deduplication of 1 638 probes, leading to a final count of 34 164 exploitable probe expressions.

Differential expression analysis

Differentially expressed genes (**DEG**) were identified through the use of the limma R package version 3.54.0 (Ritchie *et al.*, 2015).

In order to compare living environments (tropical or temperate), we subjected each probe expressions to a regression model, with predictors set as the condition, sex, and transcriptomic batch (experimental processing of the sample). We used the breeding strip as a block within the regression model (since they are inherently living in different conditions), using the *duplicateCorrelation* function (limma package).

The comparison between pigs living in tempered environment subject to experimental heat stress was performed through a similar process. The linear regression model used for this comparison used predictors set as the condition (sampling day), the sex, and the transcriptomic batch. Within this model, we used the animal ID as a block for repeated observations (since the animals are the same during the three samplings).

P-values were adjusted using the Bonferroni method. DEGs were defined as having a Fold-Change above 1.25 (or bellow -1.25) and an adjusted p-value under 0.01. Volcano plots were obtained using the EnhancedVolcano package version 1.16.0 available on Bioconductor.

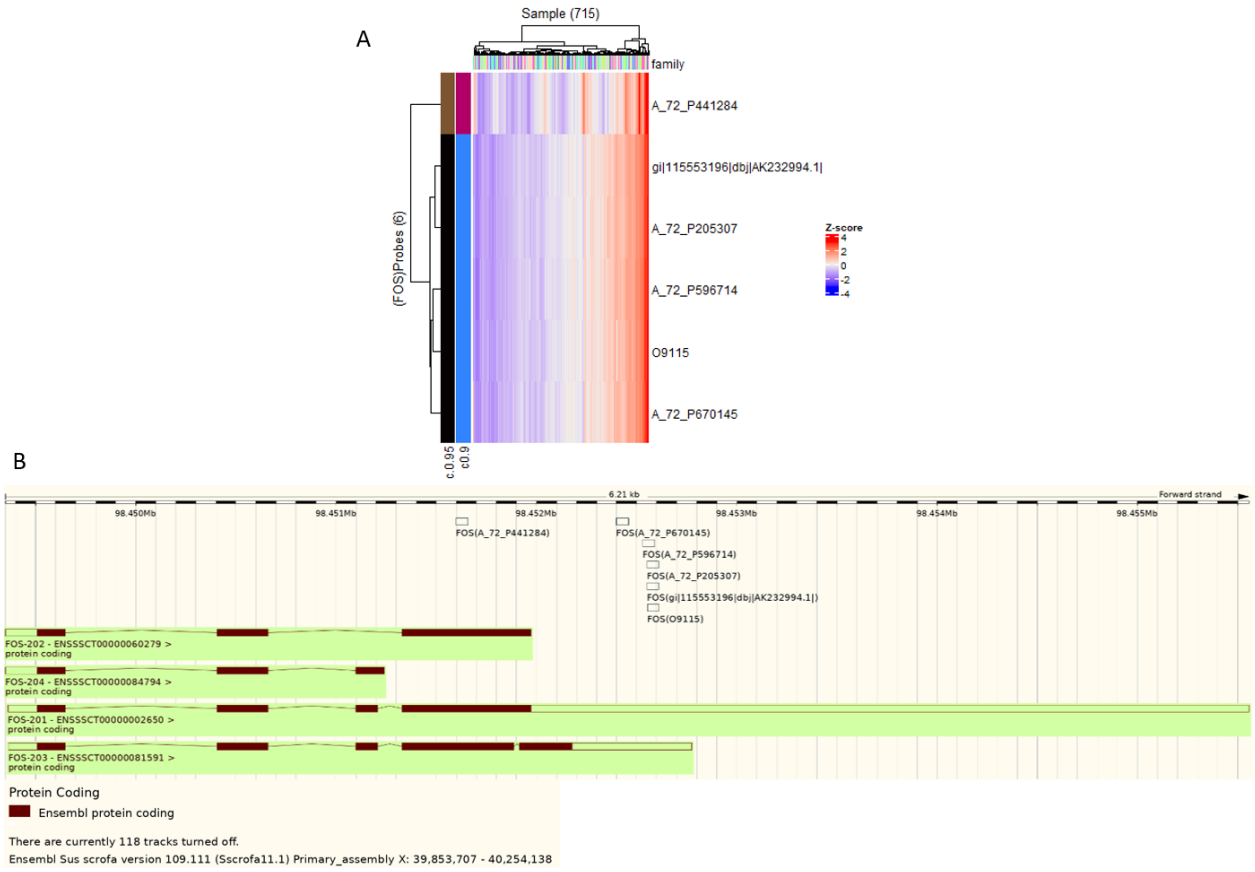


***Figure 4**: **Deduplication of probes with similar expression profiles throughout samples** (example through the FOS gene). (**A**) Expression matrix of each probe annotated to FOS. Each column represents a sample; each row represents a probe annotated to this gene. Left annotation bars represent different correlation thresholds tested (0.9 or 0.95), and the resulting deduplication is summarised as vertical color bars. Top annotation bar represents the family to which the sample belongs (based on F_1 ID names). (**B**) Visual representation of the position of each probe annotated to FOS (Ensembl genome browser). Each green bar represents an isoform of FOS, each line above the bars represent a probe. Not all probes are measuring the same set of FOS transcripts.*

Heatmaps were obtained using the ComplexHeatmap package version 2.14.0 available on Bioconductor (Gu, Eils and Schlesner, 2016). We ordered the heatmap dendrograms using the optimal leaf ordering method (Bar-Joseph, Gifford and Jaakkola, 2001), applied by the seriation 1.4.1 package (Hahsler, Hornik and Buchta, 2008).

Pseudo-autosomal regions of the genome were retrieved by plotting the X and Y chromosomes via the D-Genies web interface (Cabanettes and Klopp, 2018), and used to annotate probes originating from this region.

We conducted functional enrichments of each DEG detected in each analysis that passed the adjusted p-value threshold, with no fold-change filter, to determine the significant enrichment in each environment. The enrichment analysis was computed using the gprofiler2 package version 0.2.1 (Kolberg *et al.*, 2020). We set the reference gene list as all the genes measured in our dataset. The functional enrichment was carried out using the *gost* function, with a default enrichment significance p-value of 0.05 and the g_SCS multiple testing correction (specific to gprofiler2). Only the five most significant enrichments from each source (GO:MF, GO:CC, GO:BP, KEGG, REAC, MIRNA, HP) were shown to not overcrowd the graphic representation. We did not subject smaller enrichment lists to this visual filter (with less than 10 enrichment detected in a single source).

GWAS & eQTL

Relatedness Matrix

A genomic relatedness matrix was obtained from genotypes with GEMMA version 0.98.4 (Zhou and Stephens, 2012) (**Figure 5**). We used the full set on SNP including gonosomes, filtered through MAF and call-rate (48 498 total SNPs) in order to compute the relatedness matrix.

Linear mixed model for Genome-Wide association study

For genome wide associations, we decided to remove transcripts and SNPs annotated to gonosomes, as well as transcripts and SNPs not assigned to an assembled chromosome, resulting in a total of 21 927 autosome-annotated expression probes and 47 549 SNPs.

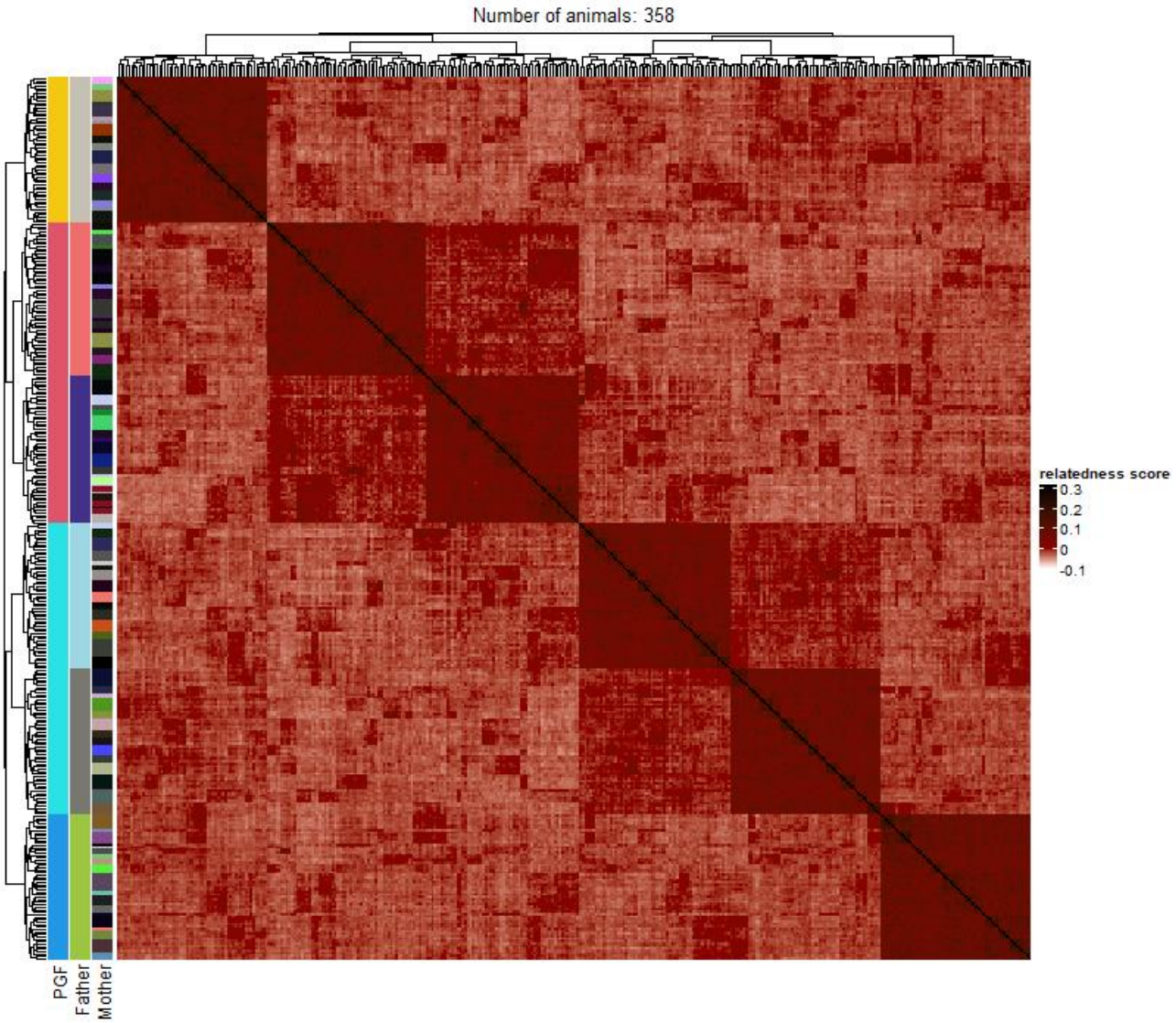


***Figure 5**: **Relatedness matrix** between each pig with whole blood transcriptome and genotype data. Each column and row represent an animal (in similar order). Left bar annotation represent paternal grandsire, father and mother of each animal. PGF = Paternal grandfather.*

Each probe expression was subjected to a linear regression model, with predictors set as the sex, condition, transcriptomic batch and breeding strip. The resulting residuals were passed as phenotypes for each 21 927 GWAS.

We used the GEMMA tool to perform association tests between genotypes and each gene expression values. The Wald test was used to compute the summary statistics.

The Bonferroni multiple-testing correction method supposes no correlation between values. As many of our SNPs are correlated due to linkage-disequilibrium, and as many transcripts are correlated due to co-expression, Bonferroni corrections would have been over-conservative. We estimated the number of independent observations within our data sets using the hscovar package version 0.4.2 (*simpleM* function) available at CRAN, over the correlation matrix of SNPs present on each chromosome and the correlation matrix of residual expression values of each probe for each animal. We decided to the set the variance threshold at 99.5% in order to search for the number of independent variables representing 99.5% of the expression/genetic variance. We detected 640 independent probe expression values, as well as around 173 to 213 independent SNPs for each chromosome. Hence, the adjusted threshold was determined as 2.10×10^{-8}.

The expression quantitative trait locus (eQTL) window was set as the 10Mb area around each significant SNPs. Overlapping eQTL windows were merged into one, using the GenomicRanges R package version 1.50.2 (available on Bioconductor). We used these eQTL windows to determine the number of eQTL per transcript. We excluded eQTL with only one significant SNP above the genome-wide significance threshold.

Code and data availability

R scripts used to process the data are available at the gitlab repository: https://forgemia.inra.fr/arthur.durante/pigheat_transcriptome. Expression data will be deposited to ArrayExpress.

Results

Transcriptome expression analysis

Pigs raised in tropical and temperate environment have distinct transcriptome

Whole blood transcriptomes were compared between genetically related pigs raised in temperate and tropical environments.

We detected a total of 516 differentially expressed transcripts (over 32 076 total), amongst which 306 were significantly overexpressed in pigs living in tempered environment and 210 which were overexpressed in pigs living in tropical environment (**Figure 6**, **Figure 7**). We decided to visualize the DEGs expressions through a heatmap, so expression groups may be detected (**Figure 8**). In fact, a clear dichotomy is identifiable between up-regulated and down-regulated transcripts between pigs living in tropical and tempered conditions.

We then performed a functional enrichment of DEGs. We used the DEGs only filtered through p-values, to enable the detection on further DEGs significantly detected in specific conditions but not especially over-expressed (**Annex 1**). Without fold-change filters, we detected 2 203 DEGs specific to pigs living in tropical environment and 1 328 DEGs specific to pigs living in tempered environment. 32 biological pathways are enriched in genes significantly expressed in the blood of pigs from the tropical farm. TGF-beta receptor signalling pathway, RHO GTPase cycle, leucocytosis, B and T cell signalling pathways are amongst the most enriched biological pathways. Only the RNA processing pathway was detected as enriched amongst genes significantly expressed in pigs living in tempered environment (**Figure 9**).

Individual variations masking the transcriptomic response to experimental heat stress

We then compared the whole blood transcriptome of pigs raised in a temperate environment at three time points during an experimental heat wave (d0, d2 and d15). Thereby, gene expression is capable of changing between the three different time points.

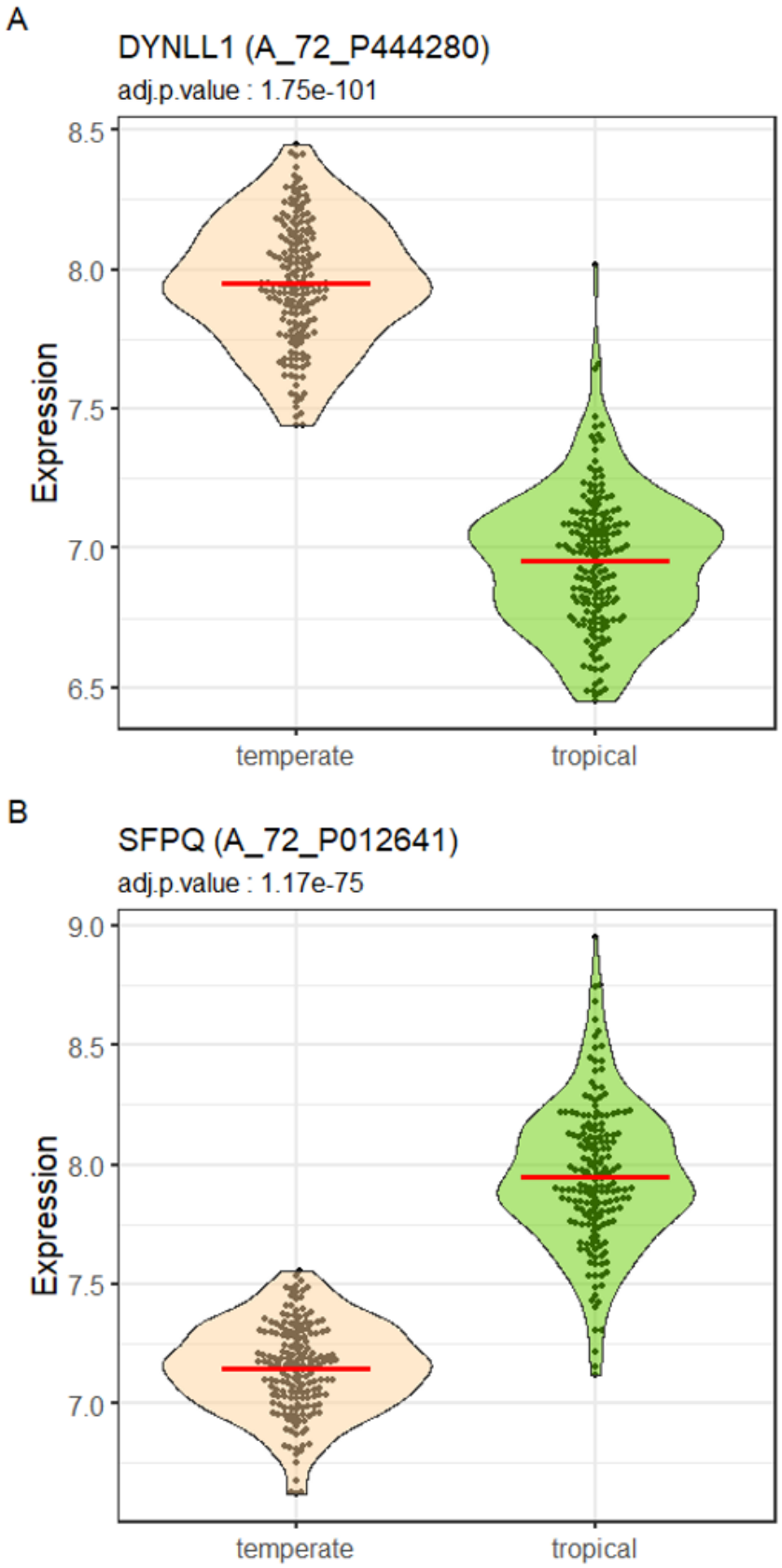


Figure 6: ***Distribution of gene expression***. (***A***) *Example of a gene overexpressed in pigs living in temperate environment.* (***B***) *Example of a gene overexpressed in pigs living in tropical environment. Dots represent the expression value of each animal. The red line represents the mean value of each group.*

We were able to detect multiple expression patterns, representing either a gradual decrease (or increase) of expression, or a time point specific expression (**Figure 10**).

Differential expression analysis resulted in 181 DEGs spread between the three different sampling conditions. This includes 121 DEGs detected between the d15 and d0 samples, 15 DEG detected between d2 and d15 samples, and 11 DEG detected between d2 and d0 samples (**Figure 11**). Few genes are shared amongst each comparison, with the highest amount being 25 genes, between the comparisons comprising of the d15 samples.

Furthermore, the detailed distribution of expression profiles of samples taken during immediate HS results in a distinct decomposition of specific sampling times (**Figure 12**), rather than ordered groups as seen in the living environment comparison (pigs living in either tropical or tempered conditions). Samples taken 2 weeks (d15) during the HS period tend to have the same expression profiles, as d2 and d0 samples are shuffled together. Moreover, most DEGs were detected within the comparison between samples from d0 and d15, presuming a major difference in expression between the two sampling times

Removing the fold-change filter, we detected a total of 588 DEGs, with a similar DEG distribution than the DEGs detected with the filter (**Annex 2**). The functional enrichment of these DEGs between each comparisons leads to the detection of fewer biological pathways, among which are the ssc-let-7e microRNA, the immune system, Osteoclast differentiation process and NOD-like receptor-signalling pathway (**Figure 13**).

When comparing the different list of DEGs detected between the two previous comparisons (living environment and immediate HS), a particular pattern unfolds (**Figure 14**). A lot more DEGs were detected between the living environment comparison (compared to those detected in immediate HS), and almost none of them are detected during the immediate HS.

Effect of sex on the pig response to heat stress

We performed DEG analysis based on sex on each of the previous comparisons (immediate HS or living environment comparison). Primarily, analysis of DEG based on sex between pigs living in tropical and temperate environments resulted in 49 DEG amongst which DEG located in either Y or X chromosome disjointed from each other, with the exception of a single DEG located on the X with an expression profile similar to DEG on the Y chromosome. A clear dichotomy is visible between DEGs specific to males and females (**Annex 3**).

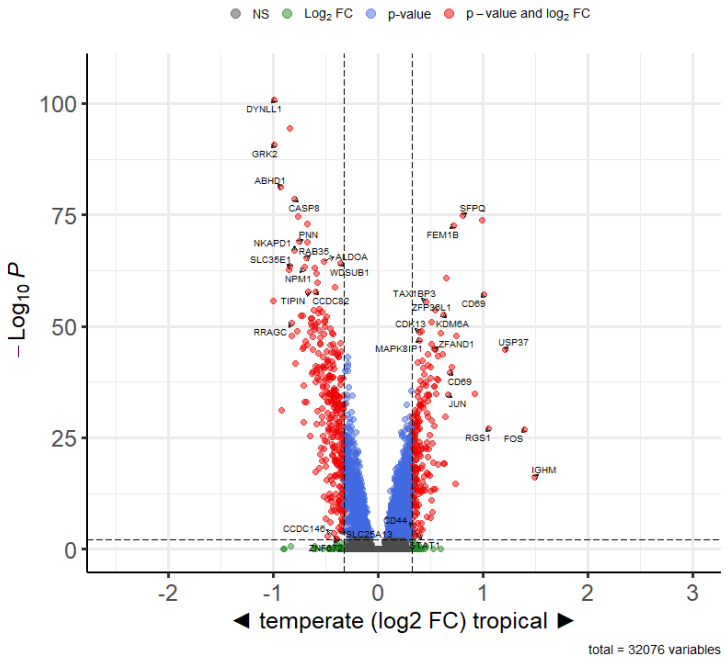


***Figure 7**: **Volcano plot representing DEGs specific to pigs living in tempered or tropical conditions** (red dots). Each dot is an expression probe. Blue dots = sufficient p-value but non-significant fold-change; green dots = sufficient fold-change but non-significant p-value; grey dots = non-significant fold change and p-value.*

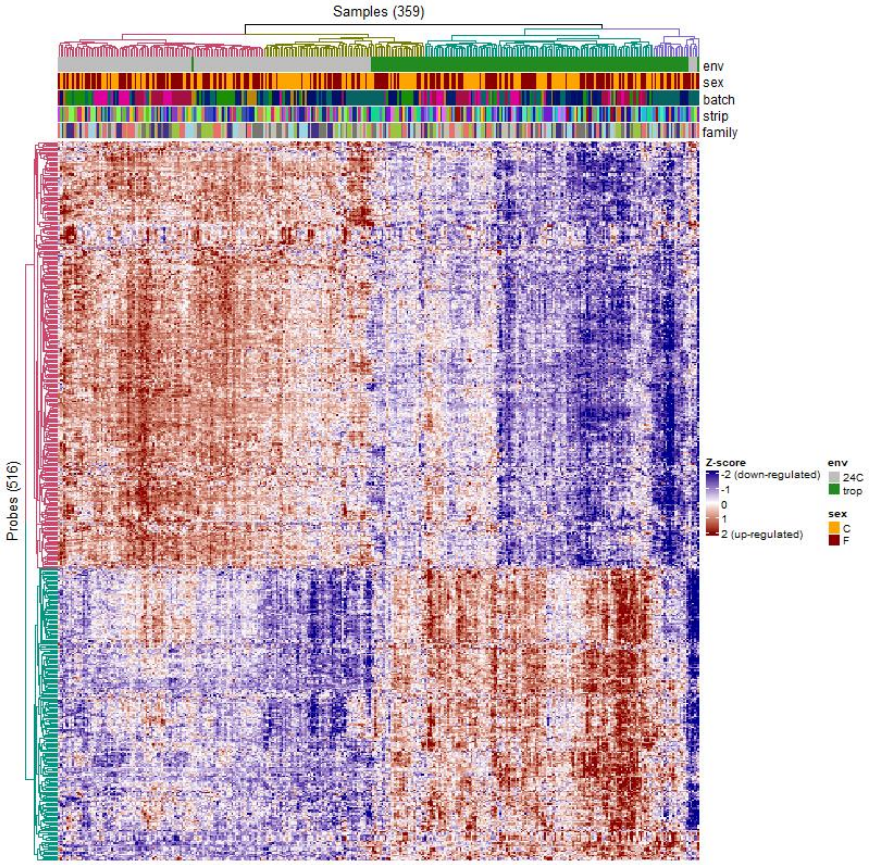


***Figure 8: Expression heatmap of DEGs detected between pigs living in tropical and temperate environment**. Each column represents a sample, each row represent the expression of a DEG across samples. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; 24C = pig living in temperate environment; trop = pig living in tropical environment; C = castrated male; F = female.*

The sex-based DEG analysis of pigs subject to a 2 week HS resulted in 52 different DEG (**Annex 4**), with a well maintained sample group of expression specific to each sex.

Impact of genetic variants on gene expression

The previous approach focused on genes with similar profiles across animals of each experimental conditions. We then looked into individual variations of gene expression levels. This analysis process aims to reveal the genetic variants associated with gene expression variations between individual pigs.

Effect of alleles on gene expression

We carried out association tests across 47 549 SNPs and 21 927 transcripts on 358 animals. Different examples of associations were detected amongst genes whose expression are associated with SNPs with p-values above the genome-wide significant threshold (**Figure 15**). Four examples are displayed. VPS39 is a gene located on the chromosome 1. A variant located on the chromosome 1 is associated with the expression level of this gene. Thereby, this variant is in an area of the genome associated with the expression level of a gene, referred more commonly as an expression quantitative trait locus (eQTL). More particularly, since this eQTL is located near the gene it regulates, it is referred as a cis-eQTL. CCDC136 is the gene with the highest association detected in this study. This gene is located on the chromosome 18, and its expression is associated with a cis-eQTL. MLLT11 is a gene located on the chromosome 4. This particular gene holds the example of having its expression associated with a cis-eQTL as well as a variant located on chromosome 14. Such variations associated with the expression of a distant gene are referred as trans-eQTLs. IL15 is a gene located on the chromosome 8, whose expression is associated with a cis-eQTL. IL15 highlights the example of a gene whose cis-eQTL barely surpasses the threshold. Each Manhattan plot displays its respective top-SNP expression change based on the alleles detected within the 358 animals.

Genome-wide association matrix

We decided to look for the distribution of eQTL throughout the entire genome by visualizing the p-values of each association based on their genomic-location (**Figure 16**). Each SNP with a high p-value present on the diagonal line represent the cis-eQTLs while other high p-value SNPs visible outside this diagonal represent the different trans-eQTLs.

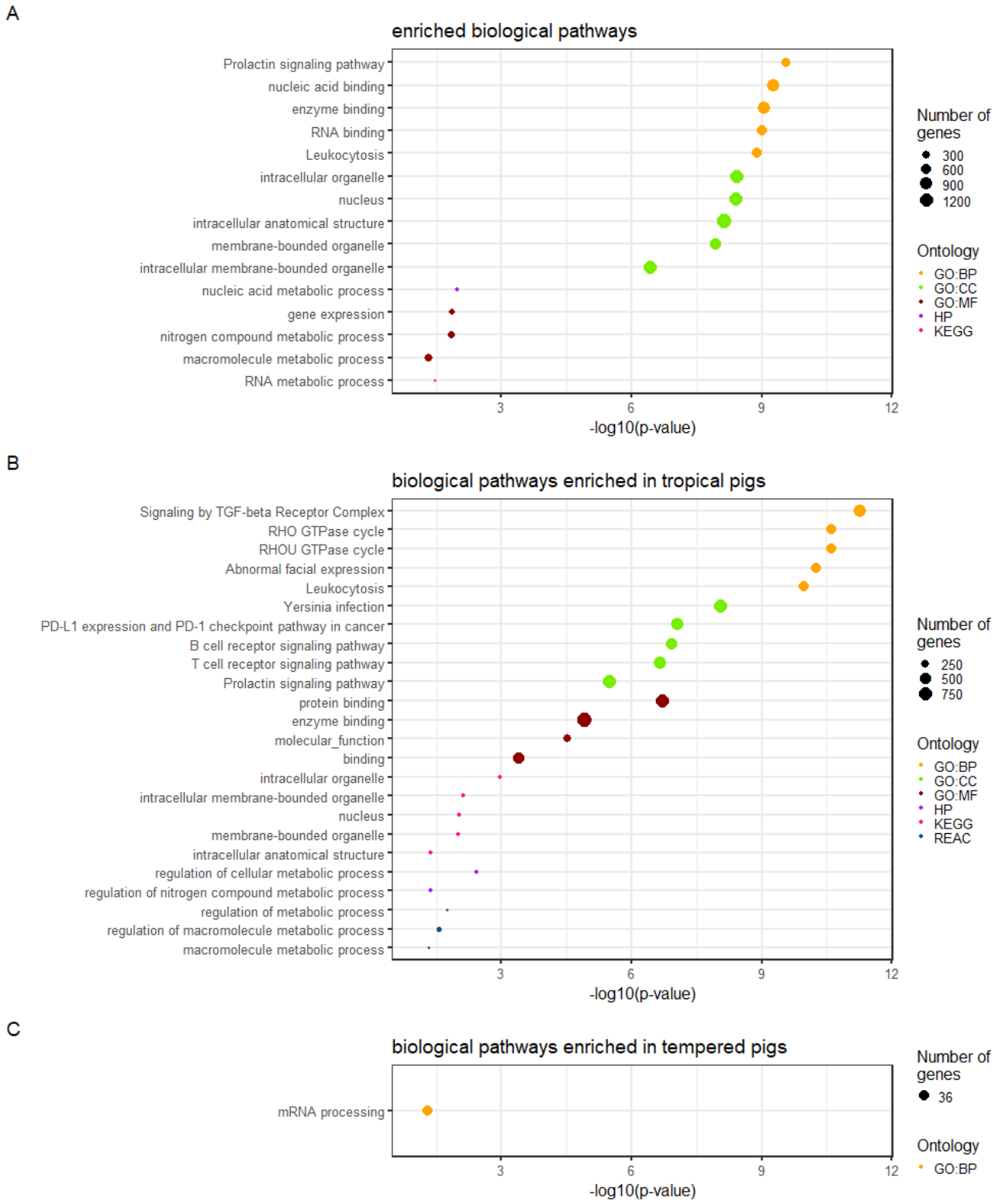


***Figure 9**: **functional enrichment of biological pathways detected from DEGs between pigs living in tropical and temperate environments**. Enrichment computed with the gprofiler2 package, using the list of DEGs with p-values above 0.01 and no fold-change filter. 3 531 DEGs are used as reference for the enrichment. (**A**) Biological pathways enriched through the entire DEG list. (**B**) Biological pathways enriched within DEGs detected in pigs living in tropical environment. (**C**) Biological pathway enriched within DEGs detected in pigs living in tropical environment.*

Overall, 23 704 transcripts did not have a detectable eQTL(eQTL with a least two significant SNP), 7 284 transcripts had a single eQTL and 1 080 transcripts had two to six eQTLs (**Figure 17**)

Effect of genetic variance on expression

For each genes, we also estimated the proportion of expression variance that was due to genetic variations. Results varies greatly between genes, from near complete genetic origin of the expression variance, to genes with no genetic components in their expression variances. Genetic variance was the main source of expression variance for 26% of the genes (**Figure 18**).

Discussion and perspectives

Multiple studies were conducted before my internship on the same Creole x Large White pigs. The objective of this project is to determine the different aspects at play that impact their capacity to withstand HS, as the usual crossbred growing pigs used for consumption cannot cope, thusly impacting their welfare and growing efficiency (due to a decrease in feeding intake).

Previous studies focused on the phenotypic aspect of the response to HS, in which distinct difference were observed between the pigs living in tropical and temperate environments, suggesting that pigs living in tropical environment experienced chronic HS (Rosé *et al.*, 2017). Multiple aspects of HS resilience deciphered as traits (such as rectal temperature) were detected as different between sire families during HS, suggesting a correlation between genetics and resilience to HS.

Although characteristics were detected between phenotypic resilience and growth capacity within the experimental design, direct difference in gene expression and the underlying eQTLs has yet to be uncovered. Our analysis process was aimed at two objectives:

1) Detecting genes whose expression differed significantly between comparisons, either between an everyday environment under heat stress or not, or between sampling times during an immediate heat stress event.
2) Detecting QTLs linked with gene expression levels of Creole x Large White crossbred pigs.

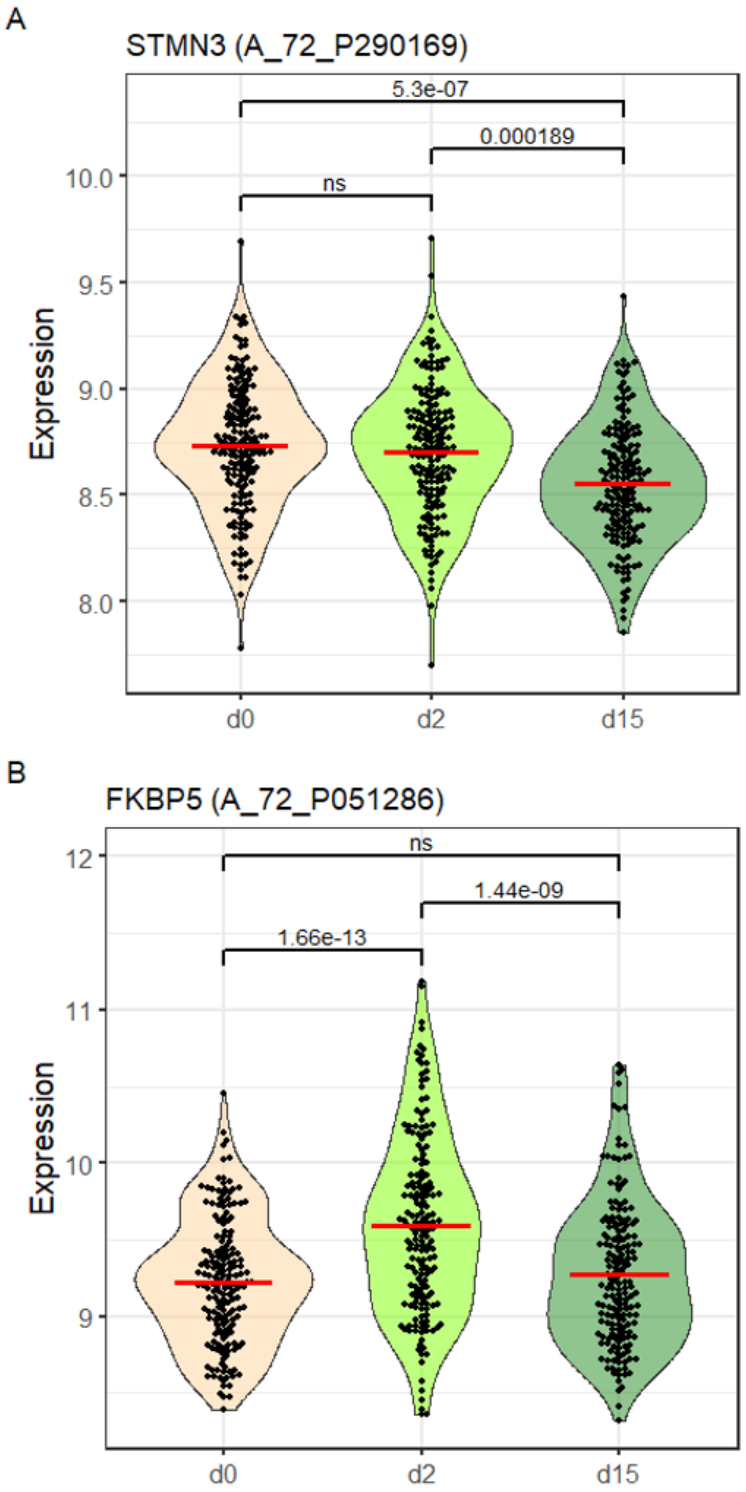


***Figure 10**: **Distribution of gene expression** within pig living in tempered environment before and during an experimental HS event. Values represent the p-value computed by Limma to set if the difference is significant. Dots represent the expression value of each animal. The red line represents the mean value of each group. ns = not-significant; d0 = samples taken before experimental HS; d2 = samples taken 2 days during experimental HS; d15 = samples taken 2 weeks during experimental HS.*

Main limitations

The annotation accessible for *Sus.scrofa* is still imperfect. This study is limited by the capacity and the accuracy of the annotation. During this intership, the transcript array annotation was augmented so that the most number of annotation was available, by combining the annotations made on reference genomes 10.2 and 11.1.

While using whole blood transcriptome is highly advantageous, it leads to multiple biases. Whole blood transcriptome does not reflect the full transcriptomic scope of an animal, in addition, some specific cell types, such as leukocytes are specifically abundant in blood, compared to organs. This abundance of specific cells needs to be taken into consideration when analysing the whole blood transcriptome and the resulting enrichments.

During the filtering of our data, we processed a deduplication of probe expression, on the basis on removing redundant expression profiles. This approach is limited as the deduplication does not take into consideration the spacing between the transcripts on the genome. Tools that would allow to considerate the genomic spacing between probes in order to adjust correlations would alleviate this problem.

Some transcriptomic probes were not associated to any known pig genes while still being differentially expressed or associated with an eQTL. Some transcriptomic probes display a strong trans-eQTL, but no cis-eQTL. It is possible, and even likely that such probes were annotated to the wrong gene.

Similarly, some SNP were associated to gene expression levels with high significance, but were not surrounded by other significant SNP as it should be the case, due to linkage disequilibrium. This likely reflects errors of annotations of the genotyping microarray. We decided to remove eQTLs with only a single significant SNP to alleviate this issue, although smarter filtering approach based on LD could be deployed.

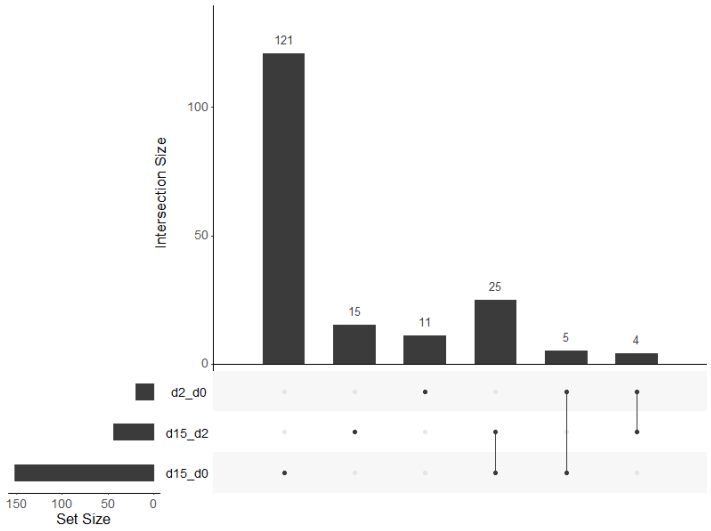


***Figure 11**: **Upset plot representing the intersection of DEGs** between sampling time points of pigs exposed to experimental heat stress. With d0 = before HS; d2 = 2 days during HS; d15 = 2 weeks during HS. d2_d0 = DEGs specific to the comparison between d0 and d2 samples; d15_d2 = DEGs specific to the comparison between d2 and d15 samples; d15_d0 = DEGs specific to the comparison between d0 and d15 samples.*

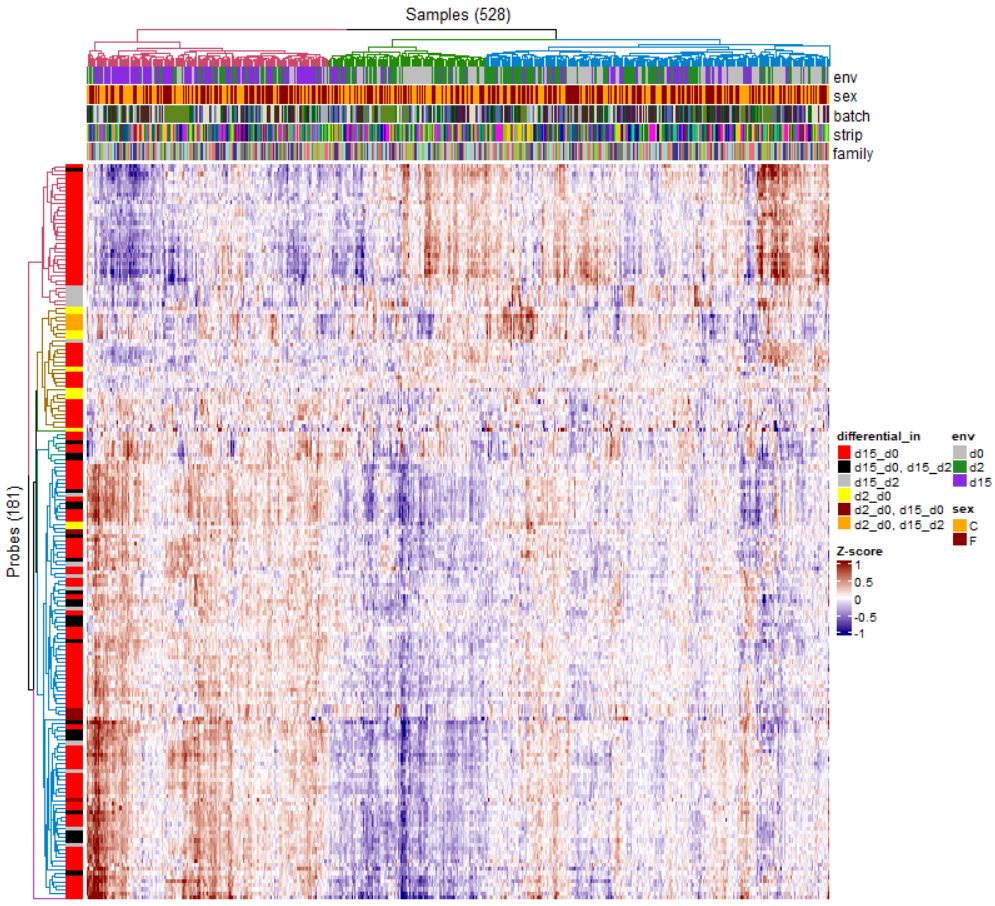


***Figure 12**: **Expression heatmap of DEGs detected between different time points in pigs living in tempered environment subject to an experimental heat stress**. Each column represents a sample, each row represent the expression of a DEG across samples. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; C = castrated male; F = female; differential_in = time point comparisons, with d0 = before HS, d2 = 2 days during HS, d15 = 2 weeks during HS.*

Differential expression of genes in response to Heat stress

Effect of acclimatization on the pig transcriptome

Differential expression analysis allowed us to detect different genes which were differentially expressed in one environment or another. In fact, we found a distinct difference of DEGs between pigs living in tempered and in tropical environments. This clear difference is visible through the dichotomy of expression profiles seen in **Figure 8**, as genes overexpressed in an environment are under-expressed in the other.

Using the larger list of detected DEGs (fold-change set at 0), the number of DEGs go from 306 to 1 328 in pigs living in tempered environment and 201 to 2 203 in pigs living in tropical environment (**Figure 8**, **Annex 1**).

This implies that more genes are expressed in the blood of pigs raised under the tropics than in the blood of pigs raised in metropolitan France. The ensuing enrichment based on the large DEGs list resulted in the detection of a handful of biological pathways. The enrichment was performed by using gprofiler2, which was the package selected after trying other enrichment packages, which did not provide results. The other packages previously tried were the ClusterProfiler R package version 4.6.0 (Yu *et al.*, 2012), as well as PANTHER.db R package version 1.0.11 (available on Bioconductor).One limiting factor of gprofiler2 was likely the difference in native support of the pig annotations.

The Gene Ontology enrichment done by the ClusterProfiler package as well as the PANTHER.db enrichments did not result in significant enrichments of DEGs, as most enrichments detected did not lead to concluding information. The reason why gprofiler was selected for the enrichment analysis was due to its rich catalogue of data sources, which includes Gene Ontology, previously used through ClusterProfiler. Moreover, we selected gprofiler2 as its annotation is reliable, since gprofiler follows Ensembl's quarterly update cycles. Enrichment significance threshold was computed from the g:scs method tailored by gprofiler, as they confirm the reliability of this method over Bonferroni and FDR.

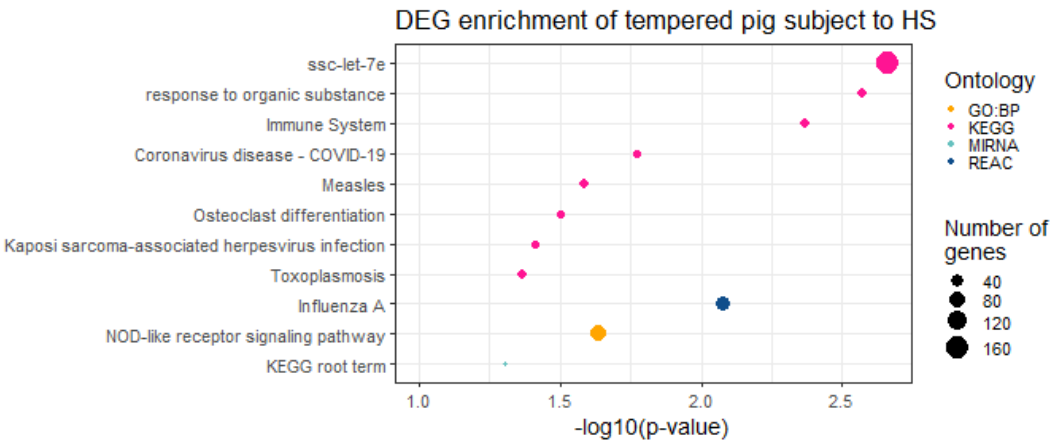


***Figure 13**: **functional enrichment** of differentially expressed genes (DEGs) detected amongst tempered pigs subject to 2 weeks experimentally induced HS.*

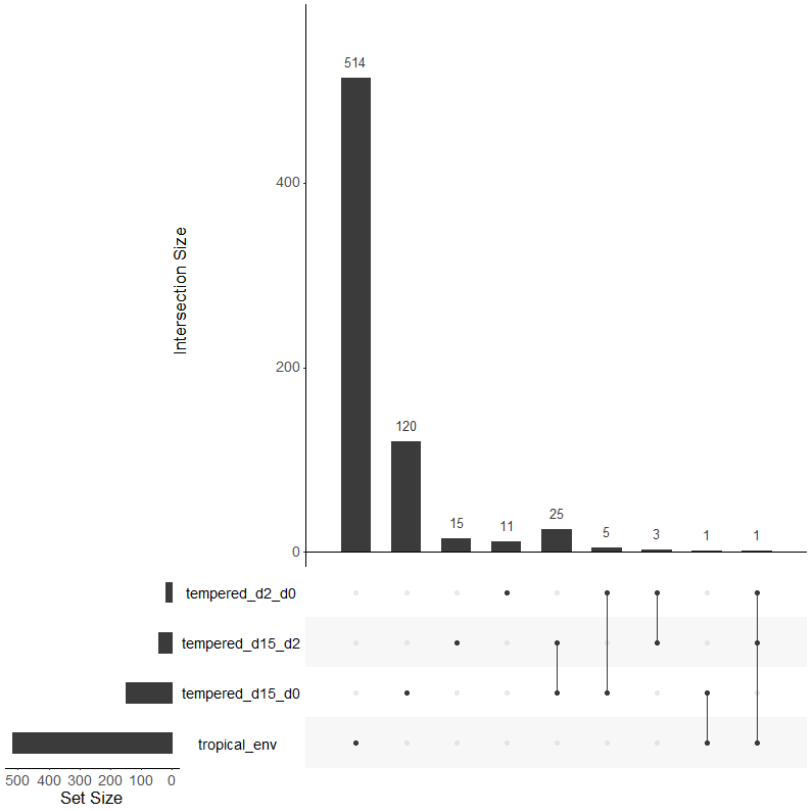


***Figure 14: Upset plot representing the intersection of DEGs between different comparisons**. Tempered_d2_d0 = DEGs specific to the comparison between d0 and d2 samples; tempered_d15_d2 = DEGs specific to the comparison between d2 and d15 samples; tempered_d15_d0 = DEGs specific to the comparison between d0 and d15 samples; tropical_env = DEGs specific to the comparison of pigs living in tempered (d0) and tropical environments.*

Determination of biological pathways specific to the living environment

We searched further information on the best enrichments detected in pigs living in tropical conditions (**Figure 9 B**). TGF-beta receptors as well as the RHO GTPase cycle contributes to the smooth process of cell differentiation (Ridley, 2015; Tzavlaki and Moustakas, 2020). Leucocytosis is linked with a high white blood cell count, which is coherent with the fact the samples studied are from whole blood, which also correlates to the observation of B and T cells pathways. This enrichment may be underlying of the difference in the farm design between the temperate and the tropical environment. In fact, pigs living in tempered environment lived inside a closed pigpen, whereas pigs living in tropical environment lived inside an open pigpen, thereby exposing them to more sanitary challenges, which might have caused their immune system to strengthen. Only a single enrichment was detected for DEGs specific to pigs living in temperate environment, even though the enrichment was based on a list of 1 328 DEGs (**Figure 9 C**) . This enrichment is composed of 36 genes associated with the RNA processing pathway. This suggests that pigs living in temperate environment are relatively more able to process the expression of their genes to grow relatively faster, as they do not experience any particular stress.

This also confirms the lack of enrichments detected, as pigs living in temperate environment are capable of expressing a wider range of genes from different biological pathways, resulting in numerous DEGs not particularly representing specific biological pathways.

Effect of individual variations on gene expression during heat stress

Through differential expression analysis of tempered pigs subject to a two weeks experimental HS, we obtained a fewer number of DEGs than the living environment comparison (**Figure 11**). From a total of 181 DEGs, most of them were specifically detected between the d0 to d15 time points (121 DEGs), with only a handful of DEGs shared with the other two comparisons. When looking into the direct response to HS (span of two days during HS), only a few DEGs were detected, confirming the low capacity of pigs to cope with HS events, as their organism express only a handful of genes to compensate the stress. The longer they are subject to HS, the better their organism adapts to the HS event. This particular difference can suggest the start of the acclimatization of pigs to the experimental HS.

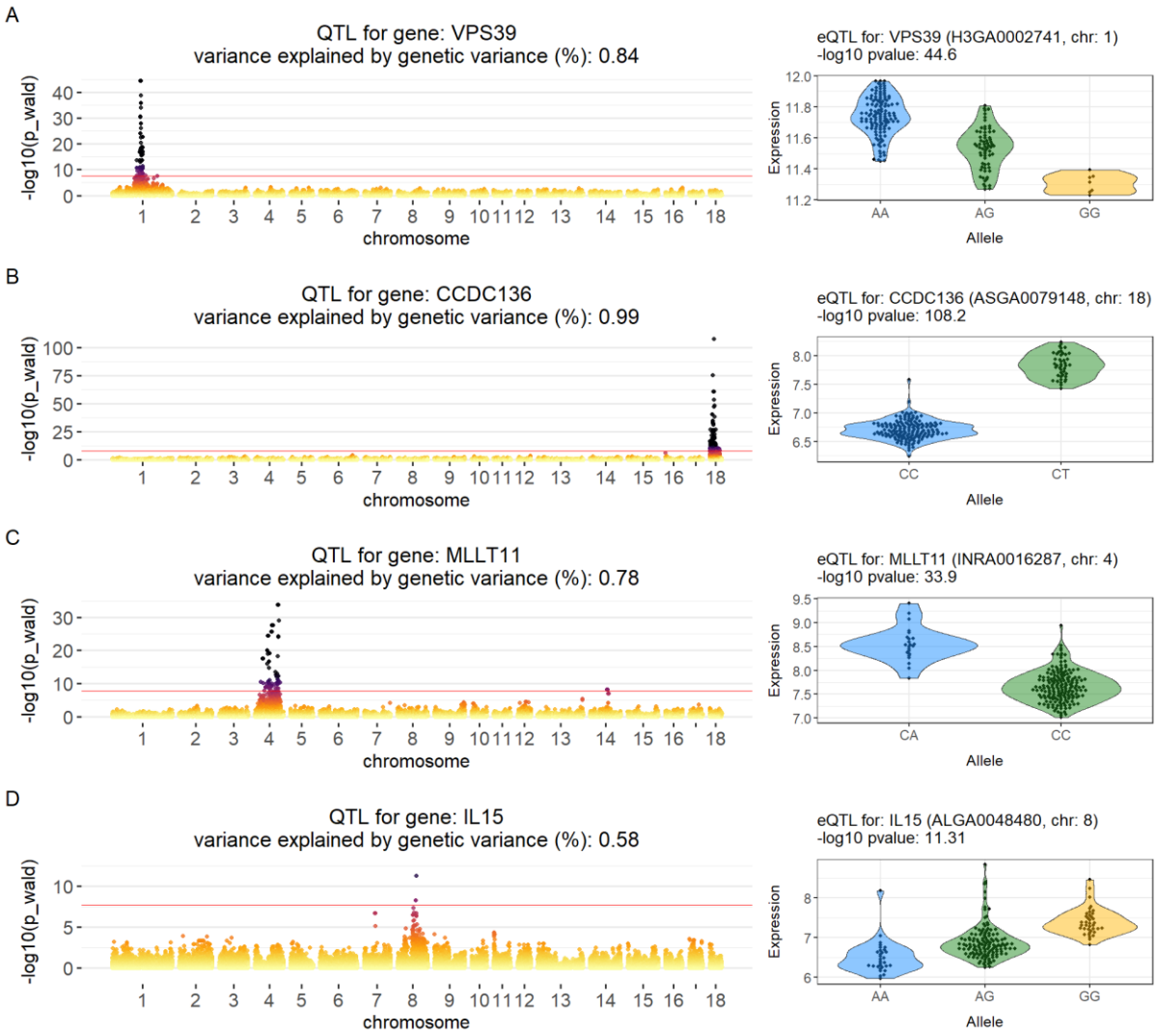


***Figure 15**: **Manhattan plots and boxplot of their corresponding top-SNP** from Genome-wide association study (GWAS) of whole blood transcriptome in response to HS. The colour gradient corresponds to the significance level of each SNP. Red line = adjusted threshold (p-value: 2.10 x 10^{-8}). Boxplots represent the expression level of the gene based on the allele detected in each animal (black dots). (**A**) eQTL of VPS39 through the o8931 transcript, and visualization of its top SNP (H3GA0002741). (**B**) eQTL of CCDC136 through the A_72_P340723 transcript, and visualization of its top SNP (ASGA0079148). (**C**) eQTL of MLLT11 through the A_72_P475107 transcript, and visualization of its top SNP (INRA0016287). (**D**) eQTL of IL15 through the A_72_P004386 transcript, and visualization of its top SNP (ALGA0048480).*

Moreover, we revealed that the global response to HS was masked by individual variations. In fact, DEGs detected throughout the different samples and time points (**Figure 12**, **Annex 2**) do not group themselves based on the time points. Therefore, the transcriptomic response to acute HS seems of a lesser strength than the transcriptomic acclimatization resulting from growing in tropical environment.

Biological pathways affected by immediate heat stress

Deeper looks into the DEGs using a less stringent filter (fold-change set at 0) led to the detection of 11 enriched biological pathways related to the immediate response to HS in pigs living in tempered environment (**Figure 13**). Among the 588 DEGs obtained, ¼ of them belong to the ssc-let-7e microRNA. This microRNA is from the let-7 microRNA family, which are well known precursors negatively linked the development of aggressive cancers in humans (Esquela-Kerscher and Slack, 2006). This would suggest that during the HS event, the pig organism will compensate the HS event by expressing genes associated with the mechanism underlying the cell differentiation. The role of the let-7e microRNA family is not fully known yet, but might be discovered in the future. Nevertheless, this microRNA plays a major role in the acclimatization process of pigs subject to immediate experimental HS. Most other detected enriched terms are non-informative, which may be caused by the high number of annotation associated with experiments not related to HS.

Difference between acclimatization and immediate response

When comparing the different DEGs detected between each comparison, we are able to observe a particular pattern (**Figure 14**). In fact, genes detected as differentially expressed between environments (temperate and tropical) are not detected amongst the comparisons made between the different time points during the experimental HS. Exceptions exist, as two genes are shared: FKBP5 and SCML4. These genes might be the corner stones to the possible discovery of more potential genes capable of being associated with the shared response to HS.

The lack of shared DEGs between comparisons implies that the transcriptomic response to HS is different in pigs born and raised in a HS environment, compared to pigs subjected to a two week HS event. This suggests that the DEGs specific to the comparison between environments may be associated with the acclimatization of the animals to HS, while DEGs specific to the immediate HS response are associated with the inability to cope efficiently, gradually taken over by DEGs associated with short-term acclimatization.

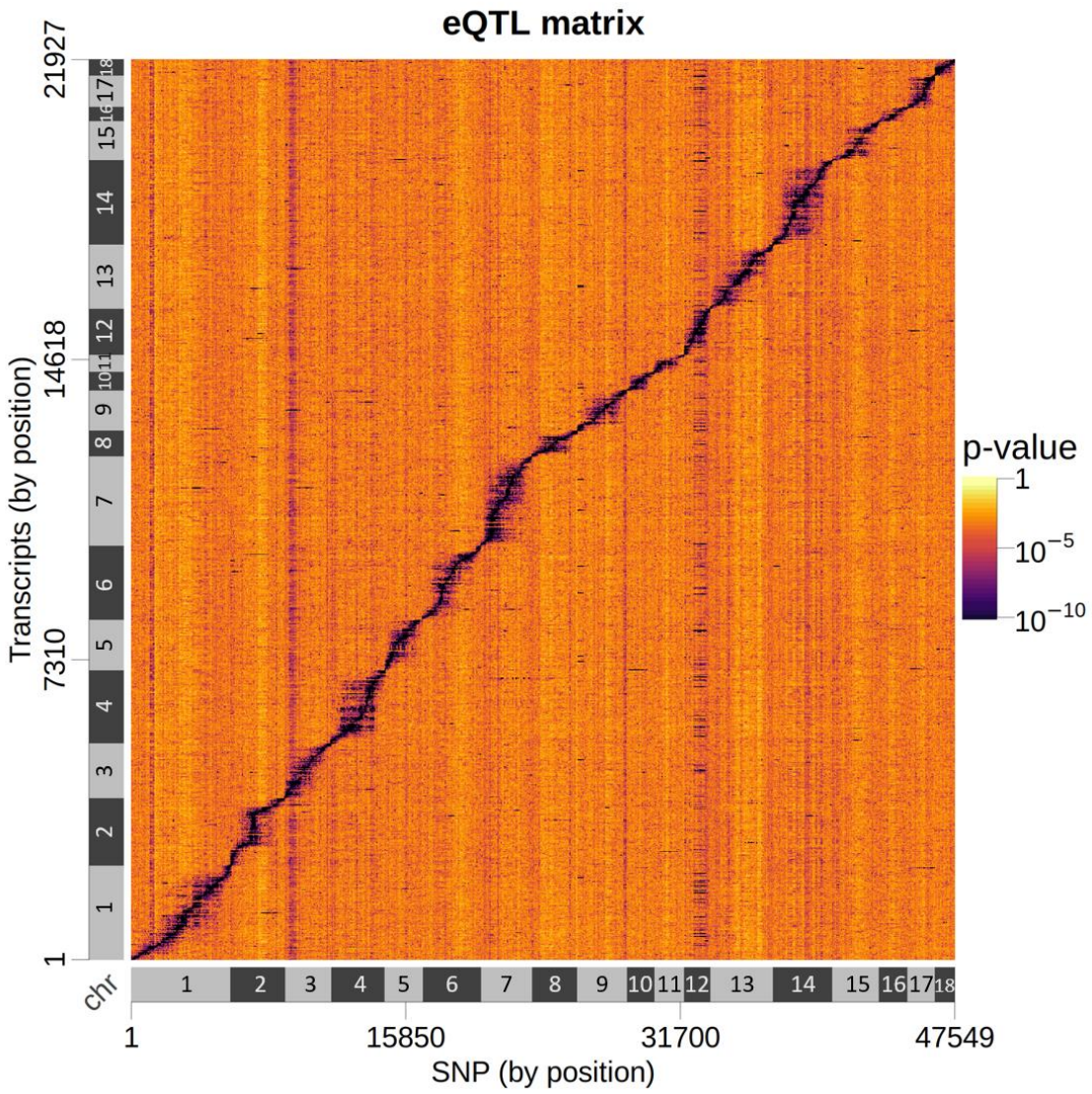


***Figure 16**: **Matrix representing each 21 927 Manhattan plot** ordered based on their respective probe location on the genome.*

eQTL detection in Creole x Large White crossbred pigs

The genome-wide association study led us to uncover the multiple SNPs associated with variations of expression levels of genes in the whole blood of pigs. Genetic polymorphisms can either come from a polymorphic site in Large White pigs, from polymorphic site in Creole pigs, and from polymorphic site in both populations. It can also come from monomorphic site in Large White and Creole pigs where each breed had fixated a different allele (**Figure 15**).

Most genes have their expression associated with a SNP on the same chromosome as the gene it regulates (cis-eQTL), while only one shows the example of a gene whose expression is regulated by a cis-eQTL as well as a trans-eQTL in **Figure 15**. As the number of SNP used is relatively small (40k), the top-SNPs can be considered as associated with the gene expression but cannot be defined as causal variants. They are, nonetheless, in high linkage disequilibrium with the causal SNP.

The representation of eQTLs assigned to each transcript enabled us to have the overview of the distribution of eQTL throughout the pig genome (**Figure 16**). With the diagonal representing the cis-eQTL, we can now determine that most eQTLs are located near the gene they regulate, implying that the majority of gene expression is regulated through cis-eQTLs. Nonetheless, we obtained multiple genes whose expressions are regulated by trans-eQTLs. These eQTLs have to be taken into consideration with caution, as they may be caused by miss-annotations. Multiple high p-value smears are visible on the chromosome 1, 3 and 12. This suggests that the SNPs on those chromosomes might have a high association with causal SNPs whose variants influence the entire genome.

For 26% of the genes, the genetic variance could explain more than half of the gene expression variance (**Figure 18**).

In this study, we decided to set the eQTL range at 10Mb, since the backcross pigs studied only experienced one meiosis the Creole x Large White crossbreeding. Therefore, eQTL that are coming from polymorphisms fixated in Large White and Creole are very broad due to the structure of this experimental cross.

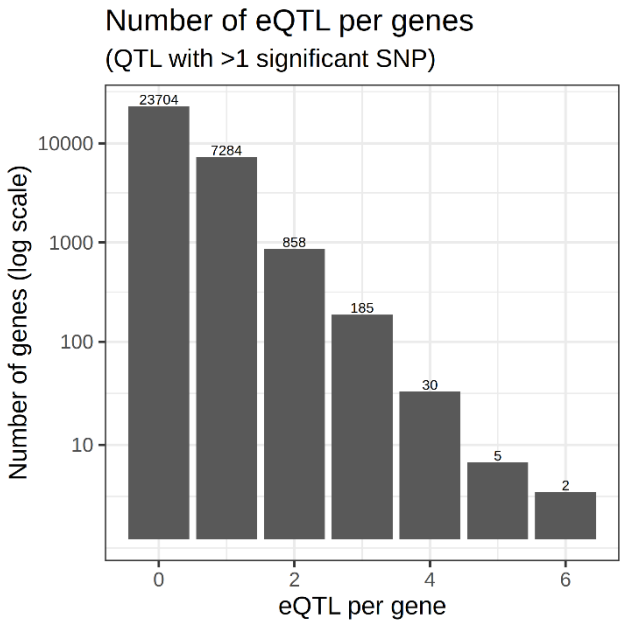


***Figure 17**: Number of eQTL detected per expression probe.*

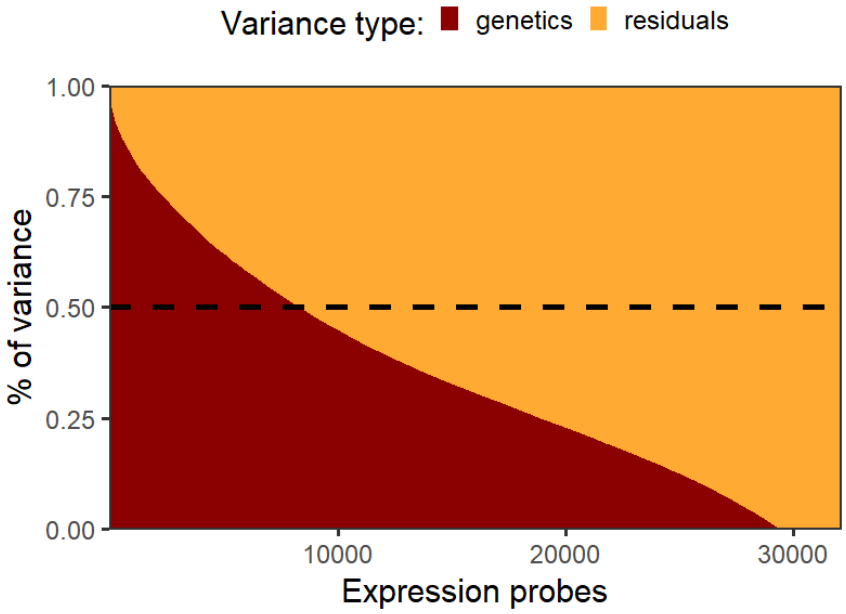


***Figure 18**: **main source of expression variance** throughout the 32 076 probes (number of probes before selection of autosomes). Dark red area represent the genetic variance; orange area represent to residual variance. Black dotted line indicates the threshold of 50% percentage.*

Among the 32 068 transcripts, 26% are associated with an eQTL, and only 3.4% total transcripts are associated with more than two eQTLs (**Figure 17**)**.** Within those 3.4%, multiple eQTLs can be detected on the same chromosome as the gene they regulate, since *Sus.scrofa* chromosomes have sizes ranging from 274Mb to 43Mb. Furthermore, within this percentage, some eQTLs are trans-eQTL, like the MLLT11 gene (**Figure 15C**).

Perspectives

During my internship, I analysed different aspects of the pig response to HS based on whole blood transcriptomic data as well as the underlying QTLs capable of regulating the expression of genes. Nevertheless, further analysis is required, and improvements can be added into the already processed analysis. We may improve the analysis regarding the transcriptomic response to HS by possibly using other tissues. Yet, this procedure is contrary to the design of the samplings, as these methods are particularly invasive, and would not allow the sampling of the same animal at multiple time points.

eQTL detection may be improved by using a wider range of SNPs, definitively enabling us to detect the possible causal variant within each eQTL. Furthermore, cis-eQTLs still need to be further detailed, as the window of a cis-eQTL is highly dependent on the experimental design. Commonly, the cis-eQTL window is defined as the 1Mb area next to the gene (Aguet *et al.*, 2023). We cannot use this window, as our experimental design did not experience enough genetic recombination, making the most significant peaks as wide as a chromosome. Moreover, it is possible for a gene to be regulated by multiple causal variant within a same cis-eQTL, hence, we may look into QTLs associated with the expression level of multiple genes (Rat Genome Sequencing and Mapping Consortium *et al.*, 2013).

The next step of my internship consists of detecting eQTLs associated with the immediate HS. Furthermore, we will analyse the possible genotype x environment interactions associated with the different HS environments studied. We may detect multiple genes whose expression levels are associated with the environment the animals live in.

During the upcoming 3 years I will spend as a PhD candidate at INRAE, under the supervision of Denis Milan and Guillaume Devailly, I will also have the opportunity to determine the impact of methylation on the pig genome, using the same experimental design. We may uncover possible mechanism capable of regulating the gene expression in pigs subject to heat stress.

References

Aguet, F. *et al.* (2023) 'Molecular quantitative trait loci', *Nature Reviews Methods Primers*, 3(1), pp. 1–22. Available at: https://doi.org/10.1038/s43586-022-00188-6.

Alexandratos, N. and Bruinsma, J. (eds) (2012) *World agriculture towards 2030/2050: the 2012 revision*. (ESA Working Papers 12-03). Available at: https://doi.org/10.22004/ag.econ.288998.

Bar-Joseph, Z., Gifford, D.K. and Jaakkola, T.S. (2001) 'Fast optimal leaf ordering for hierarchical clustering', *Bioinformatics (Oxford, England)*, 17 Suppl 1, pp. S22-29. Available at: https://doi.org/10.1093/bioinformatics/17.suppl_1.s22.

Cabanettes, F. and Klopp, C. (2018) 'D-GENIES: dot plot large genomes in an interactive, efficient and simple way', *PeerJ*, 6, p. e4958. Available at: https://doi.org/10.7717/peerj.4958.

Cassady, J.P., Young, L.D. and Leymaster, K.A. (2002) 'Heterosis and recombination effects on pig growth and carcass traits', *Journal of Animal Science*, 80(9), pp. 2286–2302. Available at: https://doi.org/10.1093/ansci/80.9.2286.

Coffel, E.D., Horton, R.M. and De Sherbinin, A. (2018) 'Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century', *Environmental Research Letters*, 13(1), p. 014001. Available at: https://doi.org/10.1088/1748-9326/aaa00e.

Dou, S. *et al.* (2017) '1HNMR-Based metabolomic profiling method to develop plasma biomarkers for sensitivity to chronic heat stress in growing pigs', *PLOS ONE*. Edited by M.F.W. Te Pas, 12(11), p. e0188469. Available at: https://doi.org/10.1371/journal.pone.0188469.

Esquela-Kerscher, A. and Slack, F.J. (2006) 'Oncomirs — microRNAs with a role in cancer', *Nature Reviews Cancer*, 6(4), pp. 259–269. Available at: https://doi.org/10.1038/nrc1840.

Godfray, H.C.J. *et al.* (2018) 'Meat consumption, health, and the environment', *Science (New York, N.Y.)*, 361(6399), p. eaam5324. Available at: https://doi.org/10.1126/science.aam5324.

Gourdine, J.L. *et al.* (2006) 'Effects of breed and season on performance of lactating sows in a tropical humid climate1', *Journal of Animal Science*, 84(2), pp. 360–369. Available at: https://doi.org/10.2527/2006.842360x.

Gourdine, J.-L. *et al.* (2019) 'Genotype by environment interactions for performance and thermoregulation responses in growing pigs1,2', *Journal of Animal Science*, 97(9), pp. 3699–3713. Available at: https://doi.org/10.1093/jas/skz245.

Gu, Z., Eils, R. and Schlesner, M. (2016) 'Complex heatmaps reveal patterns and correlations in multidimensional genomic data', *Bioinformatics*, 32(18), pp. 2847–2849. Available at: https://doi.org/10.1093/bioinformatics/btw313.

Hahsler, M., Hornik, K. and Buchta, C. (2008) 'Getting Things in Order: An Introduction to the R Package seriation', *Journal of Statistical Software*, 25, pp. 1–34. Available at: https://doi.org/10.18637/jss.v025.i03.

Kolberg, L. *et al.* (2020) 'gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset gprofiler', *F1000Research*, 9, p. ELIXIR-709. Available at: https://doi.org/10.12688/f1000research.24956.2.

Lazar, C. *et al.* (2013) 'Batch effect removal methods for microarray gene expression data integration: a survey', *Briefings in Bioinformatics*, 14(4), pp. 469–490. Available at: https://doi.org/10.1093/bib/bbs037.

Le Sciellour, M. *et al.* (2019) 'Effect of chronic and acute heat challenges on fecal microbiota composition, production, and thermoregulation traits in growing pigs1,2', *Journal of Animal Science*, 97(9), pp. 3845–3858. Available at: https://doi.org/10.1093/jas/skz222.

Liu, F. *et al.* (2022) 'Review: What have we learned about the effects of heat stress on the pig industry?', *animal*, 16, p. 100349. Available at: https://doi.org/10.1016/j.animal.2021.100349.

Menéndez-Buxadera, A. *et al.* (2012) 'Genetic variation of adaptation to heat stress in two Spanish dairy goat breeds', *Journal of Animal Breeding and Genetics*, 129(4), pp. 306–315. Available at: https://doi.org/10.1111/j.1439-0388.2011.00984.x.

Montaldo, H.H. *et al.* (2017) 'Genotype × environment interaction for fertility and milk yield traits in Canadian, Mexican and US Holstein cattle', *Spanish Journal of Agricultural Research*, 15(2), p. e0402. Available at: https://doi.org/10.5424/sjar/2017152-10317.

Nguyen, T.T.T. *et al.* (2017) 'Short communication: Implementation of a breeding value for heat tolerance in Australian dairy cattle', *Journal of Dairy Science*, 100(9), pp. 7362–7367. Available at: https://doi.org/10.3168/jds.2017-12898.

Poullet, N. *et al.* (2023) 'Comparative analysis of whole blood transcriptomics between European and local Caribbean pigs in response to feed restriction in a tropical climate', *BMC genomics*, 24(1), p. 292. Available at: https://doi.org/10.1186/s12864-023-09381-7.

Qiu, X., Wu, H. and Hu, R. (2013) 'The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis', *BMC bioinformatics*, 14, p. 124. Available at: https://doi.org/10.1186/1471-2105-14-124.

Rat Genome Sequencing and Mapping Consortium *et al.* (2013) 'Combined sequence-based and genetic mapping analysis of complex traits in outbred rats', *Nature Genetics*, 45(7), pp. 767–775. Available at: https://doi.org/10.1038/ng.2644.

Rauw, W.M. and Gomez-Raya, L. (2015) 'Genotype by environment interaction and breeding for robustness in livestock', *Frontiers in Genetics*, 6. Available at: https://doi.org/10.3389/fgene.2015.00310.

Renaudeau, D. *et al.* (2012) 'Adaptation to hot climate and strategies to alleviate heat stress in livestock production', *Animal: An International Journal of Animal Bioscience*, 6(5), pp. 707–728. Available at: https://doi.org/10.1017/S1751731111002448.

Renaudeau, D., Huc, E. and Noblet, J. (2007) 'Acclimation to high ambient temperature in Large White and Caribbean Creole growing pigs1', *Journal of Animal Science*, 85(3), pp. 779–790. Available at: https://doi.org/10.2527/jas.2006-430.

Renaudeau, D., Leclercq-Smekens, M. and Herin, M. (2006) 'Differences in skin characteristics in European (Large White) and Caribbean (Creole) growing pigs with reference to thermoregulation', *Animal Research*, 55(3), pp. 209–217. Available at: https://doi.org/10.1051/animres:2006012.

Ridley, A.J. (2015) 'Rho GTPase signalling in cell migration', *Current Opinion in Cell Biology*, 36, pp. 103–112. Available at: https://doi.org/10.1016/j.ceb.2015.08.005.

Ritchie, M.E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), p. e47. Available at: https://doi.org/10.1093/nar/gkv007.

Rohart, F. *et al.* (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS computational biology*, 13(11), p. e1005752. Available at: https://doi.org/10.1371/journal.pcbi.1005752.

Rosé, R. *et al.* (2017) 'Interactions between sire family and production environment (temperate vs. tropical) on performance and thermoregulation responses in growing pigs1,2', *Journal of Animal Science*, 95(11), pp. 4738–4751. Available at: https://doi.org/10.2527/jas2017.1611.

St-Pierre, N.R., Cobanov, B. and Schnitkey, G. (2003) 'Economic Losses from Heat Stress by US Livestock Industries1', *Journal of Dairy Science*, 86, pp. E52–E77. Available at: https://doi.org/10.3168/jds.S0022-0302(03)74040-5.

Tzavlaki, K. and Moustakas, A. (2020) 'TGF-β Signaling', *Biomolecules*, 10(3), p. 487. Available at: https://doi.org/10.3390/biom10030487.

Uffelmann, E. *et al.* (2021) 'Genome-wide association studies', *Nature Reviews Methods Primers*, 1(1), pp. 1–21. Available at: https://doi.org/10.1038/s43586-021-00056-9.

Waltz, X. *et al.* (2014) 'Effect of heat stress on blood rheology in different pigs breeds', *Clinical Hemorheology and Microcirculation*, 58(3), pp. 395–402. Available at: https://doi.org/10.3233/CH-131722.

Warr, A. *et al.* (2020) 'An improved pig reference genome sequence to enable pig genetics and genomics research', *GigaScience*, 9(6), p. giaa051. Available at: https://doi.org/10.1093/gigascience/giaa051.

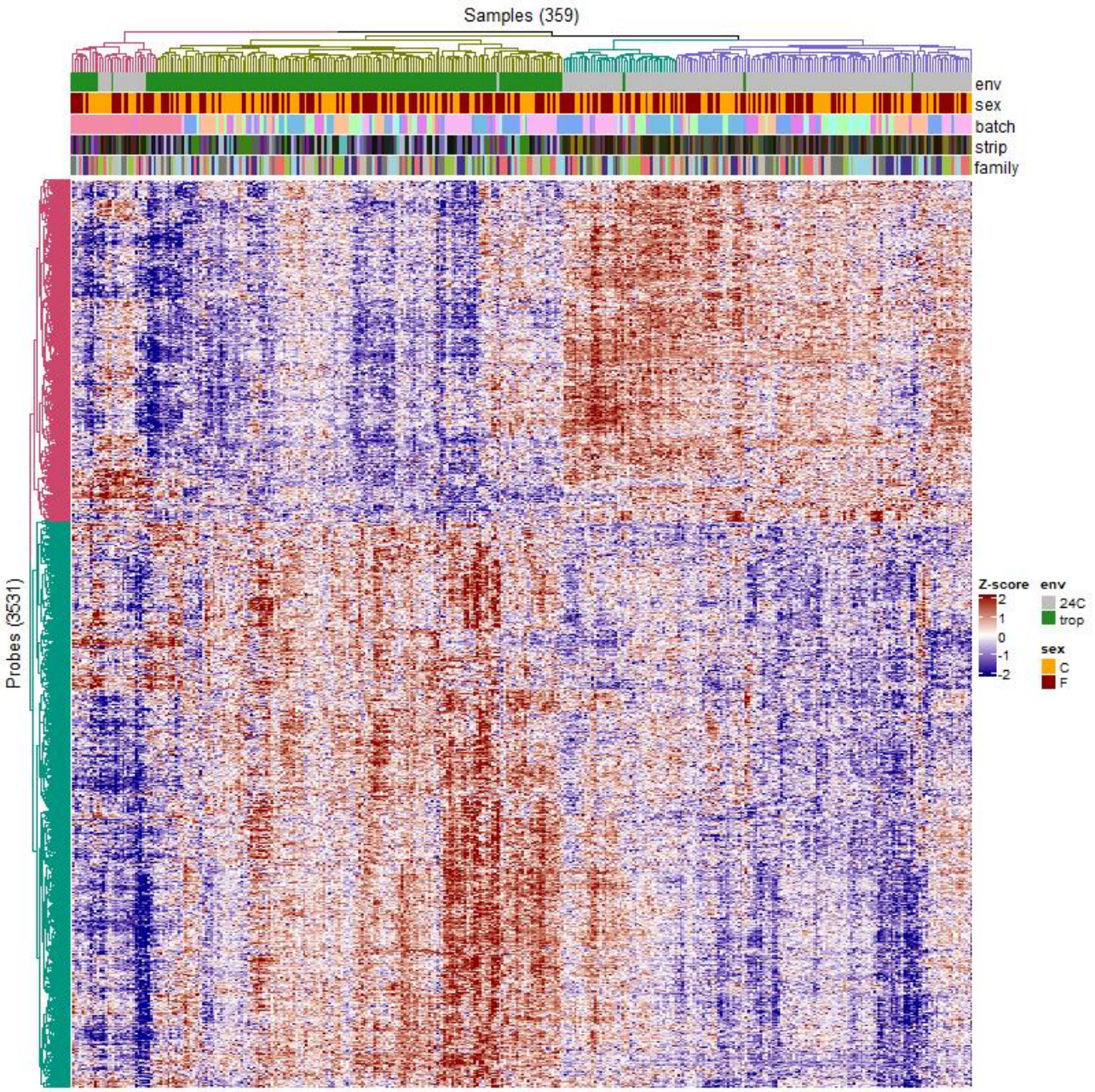
Yu, G. *et al.* (2012) 'clusterProfiler: an R package for comparing biological themes among gene clusters', *Omics: A Journal of Integrative Biology*, 16(5), pp. 284–287. Available at: https://doi.org/10.1089/omi.2011.0118.

Zhou, X. and Stephens, M. (2012) 'Genome-wide Efficient Mixed Model Analysis for Association Studies', *Nature genetics*, 44(7), pp. 821–824. Available at: https://doi.org/10.1038/ng.2310.

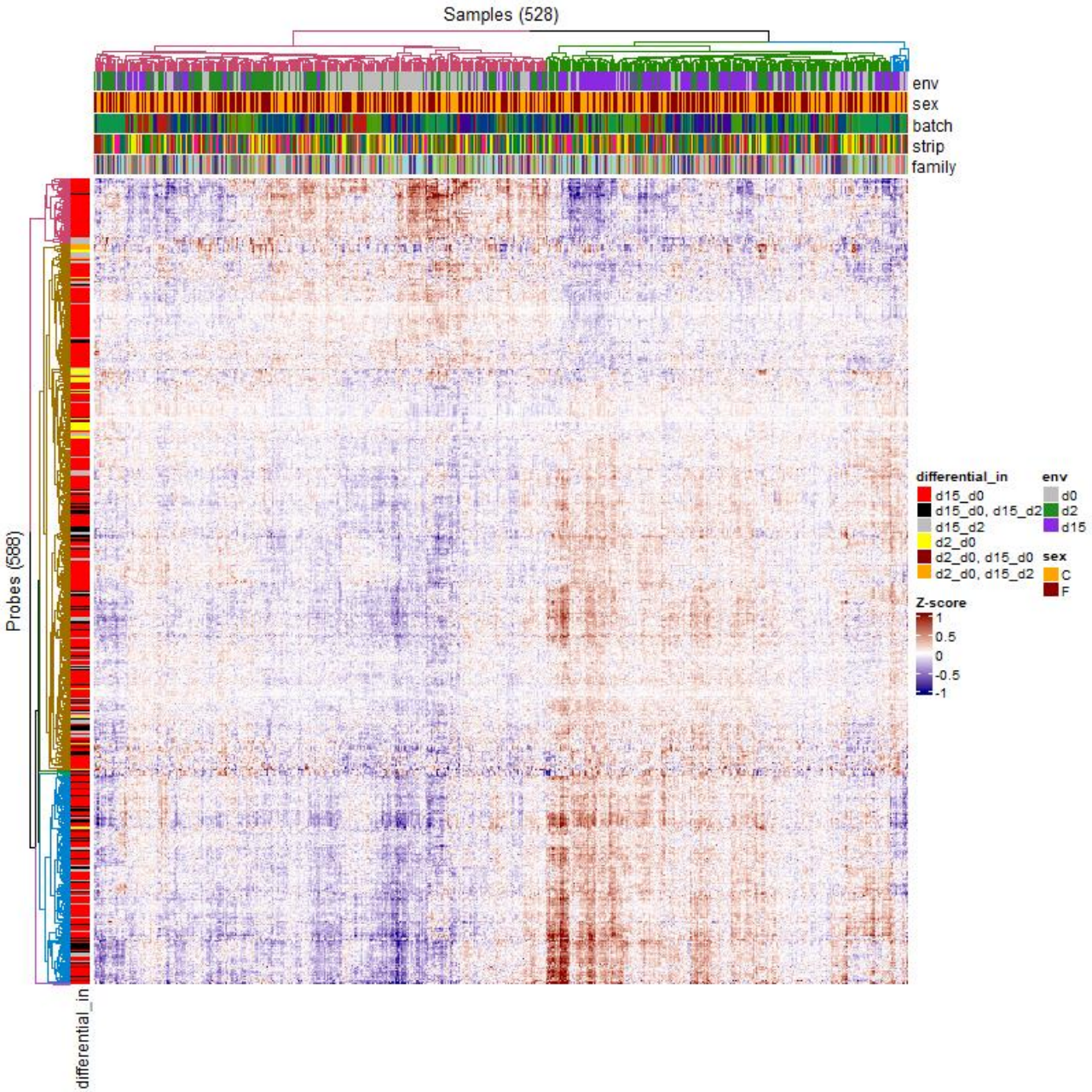
Annex table

Page

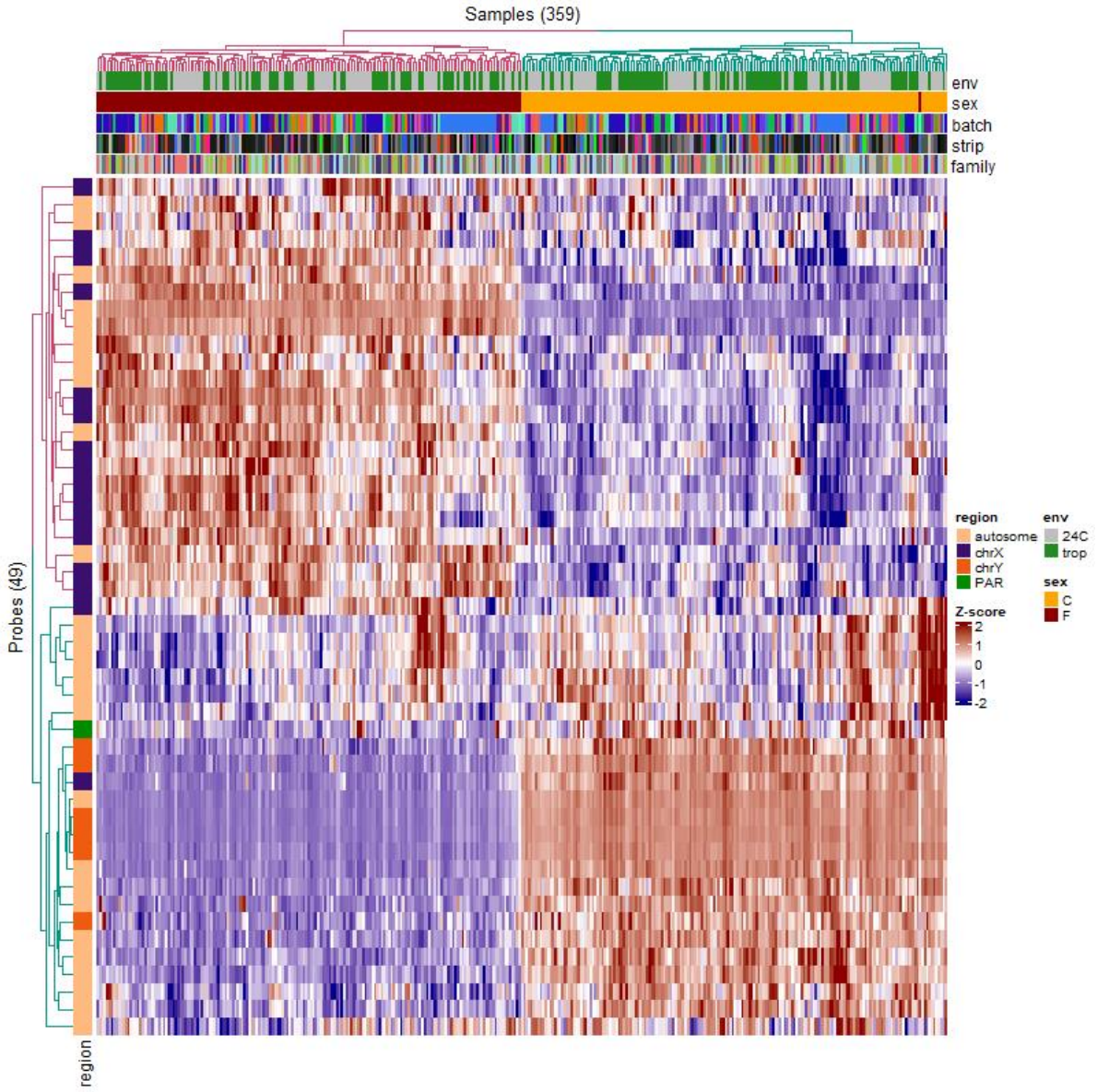
Annexes



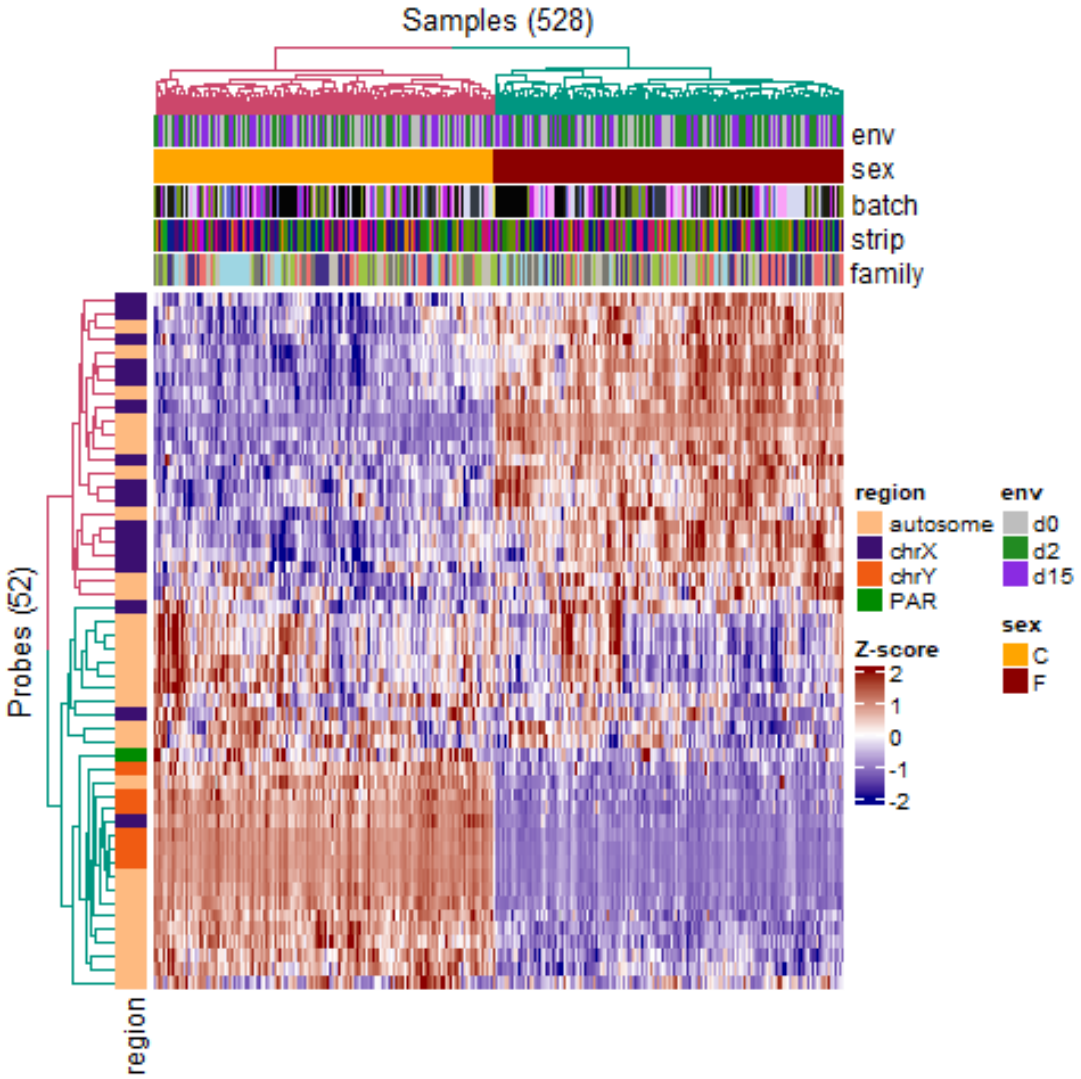
***Annex 1**: Expression heatmap of DEGs detected between pigs living in tropical and temperate environment. Fold-change set as zero. Each column represents a sample, each row represent the expression of a DEG across. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; C = castrated male; F = female.*



***Annex 2**: Expression heatmap of DEGs detected between different time points in pigs living in tempered environment subject to an experimental heat stress. Fold-change set as zero. Each column represents a sample, each row represent the expression of a DEG across samples. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; C = castrated male; F = female; differential_in = time point comparisons, with d0 = before HS, d2 = 2 days during HS, d15 = 2 weeks during HS.*



***Annex 3**: Expression heatmap of DEGs detected between pigs living in tropical and temperate environment based on their sex. Each column represents a sample, each row represent the expression of a DEG across. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; C = castrated male; F = female; chrX = chromosome X; chrY = chromosome Y; PAR = pseudo-autosomal region.*



***Annex 4**: Expression heatmap of DEGs detected between different time points in pigs living in tempered environment subject to an experimental heat stress, based on their sex. Each column represents a sample, each row represent the expression of a DEG across. The annotation bars above the heatmap represent different information regarding the sample. Red marks are up-regulated transcripts; blue marks are down-regulated transcripts. Env = living environment; batch = transcriptomic batch; strip = group of animals raised together; family = kinship depending on the ID of the F_1 sire; C = castrated male; F = female; chrX = chromosome X; chrY = chromosome Y; PAR = pseudo-autosomal region. D0 = before HS; d2 = 2 days during HS; d15 = 2 weeks during HS.*

Abstract

Meat consumption has been increasing with the enrichment of the global pupation. The rising frequency of heat stress puts the livestock industry at risk, as animals are affected by the rise of temperatures. Pigs are particularly affected, causing both discomfort to the animal and an economical loss to the farmers. Tropical pig breeds such as the Creole pigs are well adapted to high temperatures. A crossbreeding Creole pigs and Large White cosmopolitan pigs (known for their high growth rate) was setup at INRAE to investigate the genetic determinism of heat adaptation in pigs.

During my internship, I studied the whole blood transcriptome of the backcrossed offspring of a Creole x Large White cross. We processed a differential expression analysis between the whole blood of genetically related pigs living in tropical environment (French Guadeloupe) and in temperate environment (Charente, France), as well as between different time points along a 2 weeks experimental heat stress of the pigs living in the temperate environment. We identified multiple differences: pigs living in a tropical environment develop acclimatization capacities, with most differentially expressed genes focusing on the response to cellular stress, while pigs living in a temperate environment are capable of expressing a wider range of genes, which can be associated with their optimal living conditions. Moreover, when pigs living in temperate environment experience an acute heat stress event, they are hardly capable of coping with the event within two days, but are capable of expressing particular genes associated with what we may call an immediate acclimatization mechanism.

The next analysis step consisted of detecting the genetic variants associated with the expression levels of genes. Multiple quantitative trait locus affecting the expression (eQTLs) were detected, with most being located near the gene they regulate, and a few detected as distant to the gene they regulate. 26% of transcripts were associated with an eQTL, and only 3.4% of total transcripts were associated with two or more eQTLs. Further analysis are needed to detect eQTLs specific to the acute heat stress response, as well as potential genetic × environment interactions on gene expression.

Keywords: Heat stress, backcrossed pigs, transcriptome, DEG, eQTL