**Supplementary Table 1**: Iterative steps for model reduction to predict RFI class using different machine learning algorithms

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | Nb VIP | Average (%) | Overall Correct (%) | ROC OOB | K-S Lean |
| Random Forest | 328 | 95.98 | 95.95 | 0.989 | 0.919 |
| 100 | 95.98 | 95.95 | 0.993 | 0.947 |
| 50 | 97.46 | 97.47 | 0.993 | 0.945 |
| 25 | 97.30 | 97.30 | 0.996 | 0.945 |
| 10 | 97.30 | 97.30 | 0.996 | 0.945 |
| Gradient Tree Boosting | 391 | 100 | 100 | 1 | 1 |
| 100 | 100 | 100 | 1 | 1 |
| 50 | 100 | 100 | 1 | 1 |
| 25 | 100 | 100 | 1 | 1 |
| 10 | 100 | 100 | 1 | 1 |

Random forest (RF) and gradient tree boosting (GTB) algorithms were applied on a transcriptomic dataset containing 26,687 molecular probes measured in whole blood sampled from 148 pigs. Dataset was split into training (n=74) and validation test (n=74) subsets to evaluate models performance in classifying pigs into low or high residual feed intake (RFI) groups. Success rate (%) was evaluated for different iterative steps used to reduce the initial dataset into the most relevant probes (so called very important variables in prediction, VIP) able to attribute the right class for each pig. Whatever the number of retained VIP, the rate of success was better with the GTB procedure than with the RF algorithm.

**Supplementary Table 2**. List of probes identified as important to classify pigs in low or high RFI using random forest algorithm on transcripts levels of genes in the whole blood

|  |  |  |  |
| --- | --- | --- | --- |
| Probe name | Gene symbol | Full name | Score |
| A\_72\_P304024 | PSEN1 | presenilin 1 | 24.11 |
| A\_72\_P359418 | WDHD1 | WD repeat and HMG-box DNA binding protein 1 | 4.56 |
| A\_72\_P418319 | HTRA1 | HtrA serine peptidase 1 | 3.61 |
| A\_72\_P763826 | CYP24A1 | cytochrome P450, family 24, subfamily A, polypeptide 1 | 3.55 |
| A\_72\_P742047 | SERPINF1 | serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 | 3.52 |
| O12841 | PARVG | parvin, gamma | 3.29 |
| A\_72\_P387418 | C6orf221 | chromosome 6 open reading frame 221 | 2.67 |
| A\_72\_P008221 | SERPINF1 | serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 | 2.66 |
| A\_72\_P723043 | SERPINF1 | serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 | 2.61 |
| A\_72\_P146401 | SERPINF1 | serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 | 2.60 |
| O12773 | PCIF1 | PDX1 C-terminal inhibiting factor 1 | 1.97 |
| A\_72\_P548816 | HMG20A | high mobility group 20A | 1.93 |
| A\_72\_P337333 | CD1A | CD1a molecule | 1.79 |
| A\_72\_P250342 | RPS18 | ribosomal protein S18 | 1.79 |
| A\_72\_P035801 | EPAS1 | endothelial PAS domain protein 1 | 1.63 |
| A\_72\_P585246 | PCIF1 | PDX1 C-terminal inhibiting factor 1 | 1.45 |
| A\_72\_P633086 | CD1A | CD1a molecule | 1.40 |
| A\_72\_P185296 | CLU | Clusterin | 1.39 |
| O12605 | HMG20A | high mobility group 20A | 1.17 |
| A\_72\_P131741 | SLC46A3 | solute carrier family 46, member 3 | 1.16 |
| A\_72\_P006091 | PLA2G4A | phospholipase A2, group IVA (cytosolic, calcium-dependent) | 1.14 |
| A\_72\_P609509 | CD1A | CD1a molecule | 0.88 |
| O8180 | ARID3B | AT rich interactive domain 3B (BRIGHT-like) | 0.76 |
| A\_72\_P121746 | SLCO2B1 | solute carrier organic anion transporter family, member 2B1 | 0.75 |
| A\_72\_P177616 | DCT | dopachrome tautomerase (dopachrome delta-isomerase, tyrosine-related protein 2) | 0.74 |
| A\_72\_P671275 | GPX3 | glutathione peroxidase 3 (plasma) | 0.72 |
| A\_72\_P440086 | GPX3 | glutathione peroxidase 3 (plasma) | 0.59 |
| O10865 | BMPR2 | bone morphogenetic protein receptor, type II (serine/threonine kinase) | 0.55 |
| A\_72\_P337268 | HEATR4 | HEAT repeat containing 4 | 0.54 |
| A\_72\_P473804 | WWP1 | WW domain containing E3 ubiquitin protein ligase 1 | 0.49 |
| O12495 | TM7SF2 | transmembrane 7 superfamily member 2 | 0.45 |
| A\_72\_P000006 | ZNF644 | zinc finger protein 644 | 0.42 |
| A\_72\_P570814 | ATP5O | ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit | 0.40 |

A random forest (RF) algorithm was applied on transcriptomic dataset (26,687 molecular probes) from the whole blood of 148 pigs. Data were split into training (n=74) and validation test (n=74) subsets to evaluate model performance in predicting feed conversion ratio (FCR). A subset of 50 molecular probes were retained by the algorithm as important for FCR prediction with a good accuracy (R²=0.80; RMSE=0.23; RMSEP=0.15). Corresponding identified genes were listed by the order of importance (score).

**Supplementary Figure 1**. Partition of molecular probes expressed in the whole blood between trained and validation datasets to analyze traits related to feed efficiency in pigs



**Training data sets (3 merged data sets)**

**26,322 probes**

**N =74 pigs**



**Validation data sets (RFI)**

**778 probes**

**N =74 pigs**



**Validation data sets (FCR)**

**1,393 probes**

**N =74 pigs**

Three microarrays dataset generated from the whole blood of 148 growing pigs were merged into a single dataset (26,322 common expressed annotated probes). Randomly selected bootstrap pig samples (n = 74) were used for learning, whereas the remaining pig samples (n = 74) were used for validation. Subsets of molecular probes (<5% of the all probes) were selected as important to predict class of residual feed intake (RFI) and value of feed conversion ratio (FCR) by using machine learning algorithms.