Which rabbit genes are different around Australia?

Final report for the Foundation for Rabbit Free Australia

By Dr Amy Iannella

Summary

This study aimed to determine whether the six known genetic groups of wild rabbits in Australia are associated with differences in genes that have been identified as potential contributors to RHDV resistance. Only one candidate RHDV resistance gene differentiated based on rabbit ancestry in this study, any variation in other resistance genes around the country is likely to have different causes, such as the presence of RCV-A1, and thus follow a different geographic pattern. The differentiating gene, called MHC class 1, is a major immune gene which could also impact resistance to other infectious diseases such as myxomatosis. The most common MHC class 1 variant in Western Australia and South Australia appears to differ from that of the eastern states. The methods developed in this study will allow rapid testing for geographic differentiation of any newly identified resistance genes as other studies across the globe continue to uncover the genetic causes of RHDV resistance.

Background

Rabbits around Australia have varied in susceptibility to rabbit haemorrhagic disease virus (RHDV) ever since the first outbreak (Mutze et al. 2002). Subsequent studies have found that the geographic patterns in RHDV susceptibility can be attributed to the presence of a benign calicivirus (RCV-A1) which causes partial cross-protection from RHDV (Cooke et al. 2017). Geographic variation in genetic resistance to RHDV has also been demonstrated experimentally (Elsworth et al. 2012). This could be caused by differences in selective pressure due to the presence/absence of RCV-A1, as the evolution of resistance would be expected to proceed more rapidly where mortality rates are higher.

Alternatively, or additionally, variation in RHDV resistance could result from the differing genetic profiles of rabbits in different areas. Iannella et al (2019) showed that rabbits in Australia form six genetic clusters, which are likely to result from multiple introduction events. The geographic spread of these clusters is indicated in Figure 1 below. If some of these clusters originally included gene variants that coincidentally provided a measure of protection against RHDV, those protective variants would have become increasingly common in those areas while remaining unavailable elsewhere.

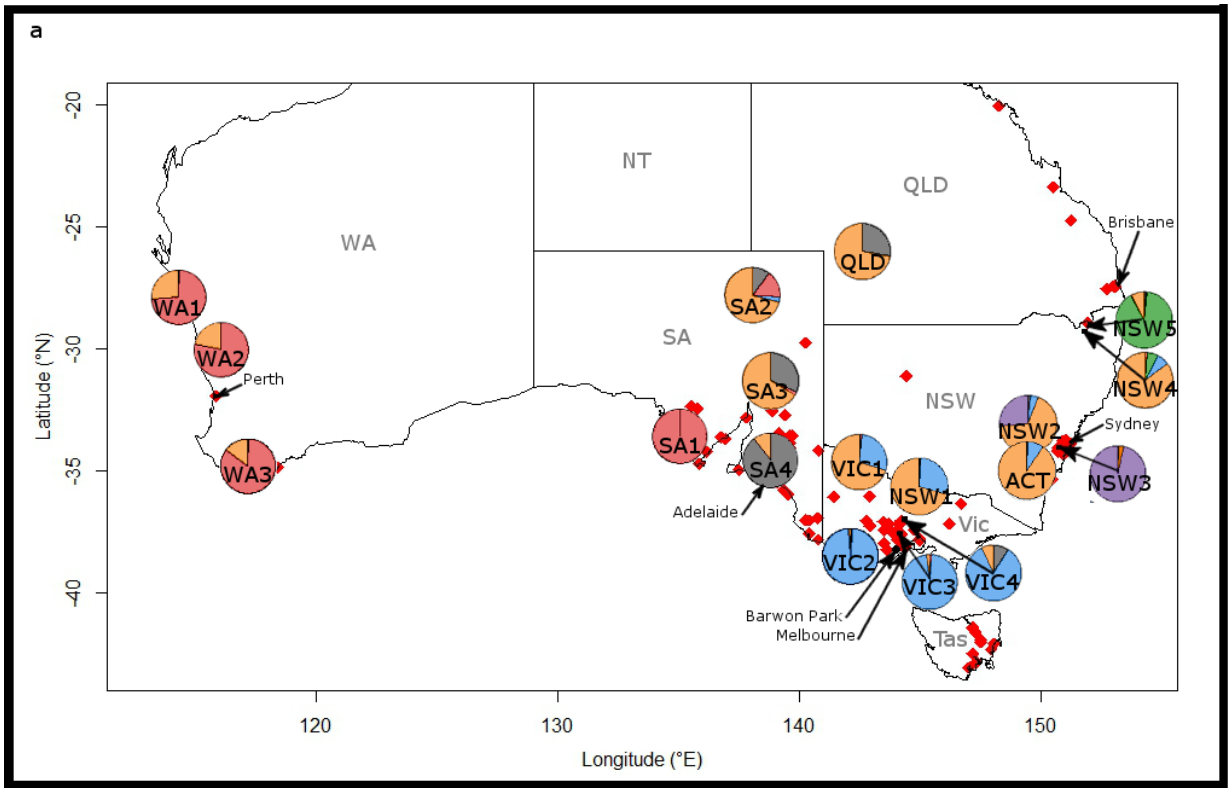

Figure 1: Map of rabbit sample sites taken from Iannella et al. (2019). Pie chart colours represent proportions of ancestry for rabbit sample sites as estimated by fastSTRUCTURE with K=6. Historical rabbit introduction records of successful or unknown outcome reported by Peacock and Abbott (2013) are represented as red diamonds, noting some may be hidden by pie charts. The Barwon Park release site in Victoria and state capital cities are specifically indicated.

In this study I test the hypothesis that some of the gene variants which are particularly different across the six genetic clusters of Australian rabbits are variants associated with RHDV resistance. Identifying RHDV resistance genes which vary notably around the country would aid predictions of rabbit response to RHDV biocontrol strategies, and open up the possibility of regionally customised RHDV strains for baiting in the distant future.

Methods in Brief

This project used genetic data obtained through a previous study, partially funded through the Foundation for Rabbit Free Australia, which was published by Iannella et al. (2019). This data included 7821 variable locations in the rabbit genome (known as SNPs) and covered 18 sites around Australia, including rabbits from all six of the genetic clusters identified by Iannella et al. (2019).

SNPs that were unexpectedly different (known as Fst outliers) among the six genetic clusters of rabbits were identified using three different software packages that each use different statistical approaches. SNPs that were supported as outliers by at least two of the three approaches were considered to be Fst outliers for the purposes of this study.

Each Fst outlier was tested for proximity to known genes. While SNPs located directly within functional genes are particularly interesting, they are not always available in this kind of genetic data which targets small sequences scattered around the genome. However, neighbouring regions of the genome are often involved in the regulation of those genes, and SNPs as far as 40kb from a gene have been shown to follow equivalent evolutionary trajectories (Lehne, Lewis and Schlitt 2011). Therefore, any gene lying within 40 kb of an Fst outlier in this study was considered likely to differ around Australia. The documented functions of these genes (if known to science) were noted, with emphasis on any genes previously identified as candidates for involvement in RHDV resistance by Schwensow et al (2017a) and Nyström et al. (2015).

Full technical details of methods are included in Appendix 1.

Findings

The three statistical approaches to finding Fst outliers each identified between seven and 106 outliers, for a total of 145 SNPs that differentiate the six groups of rabbits. Among these 16 were detected by two of the three programs and three by all programs, yielding 19 Fst outliers for investigation in this study. None of these 19 outliers was found to be within 40 kb proximity to any of the potential RHDV resistance genes found by Schwensow et al (2017a), nor to the SecI gene that Nyström et al. (2015) found to be correlated with RHDV1 survival. However, one outlier on chromosome 12 was just 16 kb from the major histocompatibility complex class 1 gene complex, some variants of which were associated with reduced survival from RHD by Schwensow et al. (2017b). This SNP had an Fst value of 0.404 across the six clusters of Australian rabbits, which indicates a substantial level of differentiation across the country. The SNP had two variants, of which one was most common in Western Australia and South Australia, while the other was common in New South Wales and central areas of Australia and extremely dominant in Queensland and Victoria (See Table 1 below).

Table 1: Australia-wide frequency comparison of variants in a rabbit SNP associated with the MHC class 1 gene complex. Row N indicates the number of rabbits used to generate frequency data for each ancestral genetic cluster of rabbits. Green shading indicates the more common variant present around each location.

	Adelaide	Perth	Sydney	Central	Melbourne	Brisbane
N	33	87	16	144	81	19
Variant 1	73%	78%	28%	21%	7%	0%
Variant 2	27%	22%	72%	79%	93%	100%

53 SNPs in this study were within 40kb of previously identified potential resistance genes but were not detected as Fst outliers. Seven of these had Fst values above 0.2 which indicates some level of differentiation around the country. The two with the highest Fst values (Fst 0.413 and 0.287, near the genes TBC1D22B and RALGPS1 respectively) were identified as statistical outliers by the software Arlequin, which used the least conservative statistical approach, but not by the other two programs in this study. This indicates that, despite observable differences, the statistical support for meaningful differentiation of these genes around the country is weak.

The 19 outlier SNPs found in this study were either in or within 40kb of a total of 34 other genes not previously associated with RHDV resistance, indicating that these other genes are likely to differ between Australia's six genetic groups of rabbits. The functions of these genes, where known, are summarised in Appendix 2 Table 1. Three of the genes belong to the major histocompatibility complexes, which are a set of key immune genes responsible for recognising pathogens, while several other genes have been associated with cancer or infectious diseases in humans.

Discussion

The most noteworthy finding in this study was the detection of four genes in the major histocompatibility complexes (MHC) which differ around Australia based on rabbit ancestral groups. MHC genes are key components in the immune response to viruses and other pathogens. They bind protein fragments from the pathogen and display them on the surface of an infected cell, allowing infected cells to be identified and destroyed by the immune system. Because MHC molecules form an interface between pathogens and the immune system the selective pressure on pathogens to evade binding by MHC molecules is extremely strong. This results in a tendency for MHC genes to be extremely variable within populations, because common variants are disadvantaged by pathogens developing resistance, and because this increases the chance of individuals bearing multiple variants, and thus having resistance to a greater variety of pathogens (Schwensow et al 2017b). Schwensow et al. (2017b) found that MCH class I genes (the most relevant class for viruses including RHDV) remain diverse in Australian rabbits despite repeated population bottlenecks caused by introduction, myxomatosis and RHDV. Although MHC classes II and III were not examined they did find that one group of common MHC class 1 genotypes was associated with reduced survival against the RHDV strain present in that year. The discovery of a SNP just 16 kb from the MHC class 1 gene complex which differs around Australia based on rabbit ancestry in this study is therefore particularly significant because it may directly impact resistance to RHDV.

This study indicates that the most common varieties of MHC class 1 genes may differ between WA/SA and the rest of the country (See Table 1 in Findings), which could in turn impact the strains of RHDV that rabbits are more susceptible to in each area. The potential for an East/West split in strain susceptibility may be worth keeping in mind as the impacts of new RHDV strains are monitored, it will remain important that monitoring activities not be restricted to research teams in the eastern states.

One intriguing possibility is that an East/West split in RHDV strain susceptibility may impact on natural RHDV evolution and circulation in Australia. For example, mutations to RHDV that help it to evade immune detection in Western Australia are likely to become dominant in that state, and cause outbreaks in South Australia as well, but may not fare so well in the eastern states where other MHC gene variants are more common. If this is occurring, then we could expect to see evidence for two separate systems of RHDV circulation. Studies of naturally occurring RHDV strain diversity around Australia have been limited so far, but Kovaliski et al. (2014) did note a clustering of Western Australian RHDV strains with South Australian strains in their phylogenetic work, which supports this idea. A more detailed exploration of this hypothesis may become more feasible if fly-based monitoring becomes widely established, allowing a full survey of RHDV diversity.

A geographic divide in MHC gene diversity has further ramifications for rabbit disease dynamics because the MCH gene complex is involved in detection of all pathogens, it is not RHDV exclusive. The above discussion of MCH variability impacting on RHDV susceptibility and circulation could thus equally apply to myxomatosis or any other infection. As pointed out by Schwensow et al (2017b), variants of MCH class 1 that are more susceptible to RHDV may nonetheless be common in some areas if they instead convey resistance to myxomatosis or other pathogens. Therefore, the East/West divide may impact which diseases are more effective in different areas, just as much as which strains of disease are effective.

There are two important caveats to this discussion. One is that this study did not sequence MHC class 1 genes directly. While SNPs do tend to evolve in tandem with neighbouring genes, and can have regulatory effects on those genes, to fully understand the impact of MHC class 1 diversity around Australia a future project would need to sequence that gene explicitly at several sites around the country, expanding on the work of Schwensow et al. (2017b). The second caveat is that MHC genes are likely not the only genes involved in RHDV resistance, as shown by Schwensow et al. (2017a). This study did not find evidence of other resistance genes being impacted by rabbit ancestry, so the impact of variations in the MHC genes may be overshadowed by other genes which may vary across a different geographic scheme.

The absence of statistically significant geographic differentiation around other putative RHDV resistance genes in this study makes it possible that any ancestral differences in capacity to adapt to RHDV (i.e. the presence of specific advantageous mutations in some rabbit populations but not others) have been eclipsed by the impact of RCV-A1. The presence or absence of RCV-A1 has a dramatic influence on the mortality rate of RHDV, impacting the strength of selective pressure that drives resistance to RHDV. This is hypothesised as a reason for increased RHDV genetic resistance in many areas without benign RCV-A1 presence (Elsworth et al. 2012), and the force of this selective pressure driving the evolution of other RHDV resistance genes could have more impact than the influence of rabbit ancestry on MHC gene diversity. To determine the dominant driver of geographic differences in resistance we would need to fully understand the distribution of MHC gene variants, RCV-A1 abundance and strength of RHDV resistance for comparison. Because such a study would require a very substantial research investment, a more practical approach would be to simply assume that both processes are likely to have some impact on disease dynamics across Australia.

The high Fst values found in two of the SNPs in this study which are near potential RHDV resistance genes previously identified by Schwensow et al (2017a) suggests that some impact of ancestry on the RHDV resistance capacity of these genes is possible, although differentiation was not strong enough to achieve statistical support in more than one of the three Fst outlier analyses used. Given the weak support for differentiation of these SNPs follow-up investigation would not be recommended at this stage. However, if compatible genetic data is obtained for more sites around Australia as part of any future studies the statistical power of this analysis would increase and may yield more definitive insights on these two genes.

Other than the MHC genes discussed earlier, the genes associated with SNPs that varied according to rabbit ancestry in this study were diverse in function and were not obvious targets for regional adaptation. The one exception to this was the gene POU5F1. POU5F1 is involved in embryo development and has been shown to increase expression when pregnant rabbit does were exposed to heat, which lead to an increase in gestational mortality (Marco-Jiménez et al 2013). It is therefore possible that variants of this gene would be advantageous for rabbits living in arid conditions. The presence of certain variants in some ancestral rabbit populations may have aided their spread into hotter regions of Australia. Subsequent natural selection would have further increased or decreased the frequency of these variants according to local climate, increasing the differentiation observed across the country.

It remains possible that more RHDV resistance genes do differentiate based on ancestral rabbit populations. This could be the case if the differentiating genes are not among the current candidates identified by Schwensow et al. (2017a) and Nyström et al. (2015). The workflow produced in this study will allow for simple and rapid testing of any new RHDV resistance candidates that are found in the future, to determine whether they differentiate on the basis of rabbit ancestry. Communications with Dr Nina Schwensow in Germany indicate that a new set of candidate genes is under production and will be ready to run through this new workflow early in 2020.

References

Cooke, B., Springer, K., Capucci, L., and Mutze, G. (2017) “Rabbit haemorrhagic disease: Macquarie Island rabbit eradication adds to knowledge on both pest control and epidemiology”. Wildlife Research **44**(2): 93-96.

Elsworth, P. G., Kovaliski, J., and Cooke, B. D. (2012) “Rabbit haemorrhagic disease: are Australian rabbits (Oryctolagus cuniculus) evolving resistance to infection with Czech CAPM 351 RHDV?”. Epidemiology & Infection **140**(11): 1972-1981.

Iannella, A., Peacock, D., Cassey, P., and Schwensow, N. (2019). “Genetic perspectives on the historical introduction of the European rabbit (Oryctolagus cuniculus) to Australia”. Biological invasions **21**(2): 603-614.

Kovaliski, J., R. Sinclair, G. Mutze, D. Peacock, T. Strive, J. Abrantes, P. J. Esteves and E. C. Holmes (2014). "Molecular epidemiology of Rabbit Haemorrhagic Disease Virus in Australia: when one became many." Molecular Ecology **23**(2): 408-420.

Lehne, B., Lewis, C., and Schlitt, T. (2011). “Exome localization of complex disease association signals”. BMC Genomics **12:** 92.

Marco-Jiménez, F., Naturil-Alfonso, C., Peñaranda, D. S., Jiménez-Trigos, E., García-Diego, F. J., & Vicente, J. S. (2013). “Maternal Exposure to High Temperatures Disrupts OCT 4 mRNA Expression of Rabbit Pre-Implantation Embryos and Endometrial Tissue”. Reproduction in Domestic Animals **48**(3): 429-434.

Mutze, G., Bird, P., Kovaliski, J., Peacock, D., Jennings, S., and Cooke, B. (2002) “Emerging epidemiological patterns in rabbit haemorrhagic disease, its interaction with myxomatosis, and their effects on rabbit populations in South Australia”. Wildlife Research **29**(6): 577-590.

Nyström, K., J. Abrantes, A. M. Lopes, B. Le Moullac-Vaidye, S. Marchandeau, J. Rocher, N. Ruvoen-Clouet, P. J. Esteves and J. Le Pendu (2015). "Neofunctionalization of the Sec1 alpha1,2fucosyltransferase paralogue in leporids contributes to glycan polymorphism and resistance to rabbit hemorrhagic disease virus." PLoS Pathogens **11**(4): e1004759.

Peacock, D. and I. Abbott (2013). "The role of quoll (Dasyurus) predation in the outcome of pre-1900 introductions of rabbits (Oryctolagus cuniculus) to the mainland and islands of Australia." Australian Journal of Zoology **61**(3): 206.

Schwensow, N. I., H. Detering, S. Pederson, C. Mazzoni, R. Sinclair, D. Peacock, J. Kovaliski, B. Cooke, J. Fickel and S. Sommer (2017a). "Resistance to RHD virus in wild Australian rabbits: Comparison of susceptible and resistant individuals using a genomewide approach." Molecular Ecology **26**(17): 4551-4561.

Schwensow, N., C. J. Mazzoni, E. Marmesat, J. Fickel, D. Peacock, J. Kovaliski, R. Sinclair, P. Cassey, B. Cooke and S. Sommer (2017b). "High adaptive variability and virus-driven selection on major histocompatibility complex (MHC) genes in invasive wild rabbits in Australia." Biological Invasions **19**(4): 1255-1271.

Appendix 1: Detailed methods

This project used genome-wide SNP data obtained through a previous ddGBS study published by Iannella et al. (2019). This data included 7821 SNP loci and covered 18 sites around Australia, including rabbits from all six of the genetic clusters identified by Iannella et al. (2019). Individual rabbits were assigned to populations based on Iannella et al. (2019) using the cluster of highest ancestral proportion for the site where the rabbit was caught. The software PGDspider was used for all necessary downstream file type conversions.

Fst outliers were detected using three different approaches, as implemented by the software packages Bayescan v2.1, Arlequin v3.5.2.2 and Fsthet v1.01. Bayescan models neutral gene frequencies with a multinomial Dirichlet distribution using a Bayesian approach. Bayescan was run with prior odds of 100 for the neutral model, sampling size of 5000, thinning interval of 10, 20 pilot runs of 5000 length and a burn-in of 50,000 runs. Outliers were accepted at a false discovery rate of 0.1. Arlequin's "Detection of loci under selection" module compares the joint distribution of FST and heterozygosity to a null distribution generated by a simulated genetic island model. Arlequin was run with 50,000 simulations and 100 simulated demes, using the "proportion of differences" distance estimator without hierarchical island model. Heterozygosity was set at 0-1, gamma at 0, and minimum DAF at 0.01 to match the minor allele frequency threshold used by Iannella et al. (2019). Outliers were accepted at $P<0.001$. Fsthet uses smoothed quantiles to plot locus Fst relative to expected heterozygosity in order to minimise assumptions about population demographics. Fsthet was run with alpha of 0.01, accepting outliers beyond the 99% confidence interval. SNPs that were supported as outliers by at least two of the three approaches were considered to be Fst outliers for the purposes of this study.

Each Fst outlier was tested for proximity to genes previously identified as candidates for involvement in RHDV resistance by Schwensow et al (2017a) and Nyström et al. (2015) using a custom python script, and for proximity to other known genes using the NCBI Genome Data Viewer. While SNPs located directly within functional genes are particularly interesting, they are not always available in this kind of genetic data which targets small sequences scattered around the genome. However, neighbouring regions of the genome are often involved in the regulation of those genes, and SNPs as far as 40kb from a gene have been shown to follow equivalent evolutionary trajectories (Lehne, Lewis and Schlitt 2011). Therefore, any gene lying within 40 kb of an Fst outlier in this study was considered likely to differ around Australia. Any documented functions of these genes in the NCBI database were noted.

Fst values across the six genetic clusters were calculated individually for each SNP locus within 40kb of the candidate resistance genes identified by Schwensow et al (2017a) and Nyström et al. (2015) using the GenAlEx v6.503 add-on software for Microsoft Excel.

Appendix 2: Gene functions

Table 1: Genes likely to differ between Australia's six genetic groups of rabbits. Gene functions are typically inferred from equivalent genes in humans or other highly studied species. Distance from SNP is measured in base pairs to the nearest end of a gene, "internal" indicates that the SNP was within the functional gene itself. Ensembl gene IDs are provided where available as a more reliable alternative to gene names which are not always static.

Gene name	Gene function summary	Distance from SNP	Ensembl Gene ID
LOC100338822	Major histocompatibility complex RLA class I antigen. Involved in various pathways related to infections and immunity (eg autoimmune thyroid disease, graft-vs-host disease, antigen processing and presentation, herpes simplex infection…)	16179	ENSOCUG00000010972
LOC100343144	Major histocompatibility complex SLA class II histocompatibility antigen, DQ haplotype D alpha chain. Involved in various pathways related to infections and immunity (eg asthma, influenza A, graft-vs-host, antigen processing and presentation, herpes simplex infection…)	15871	ENSOCUG00000002480

HLA-DQB1	Major histocompatibility complex, class II. Produces proteins that display foreign peptides to the immune system to trigger a response. Hundreds of alleles in humans involved in various diseases including coeliac.	39296	ENSOCUG00000002485
C2	Major histocompatibility complex class III gene. Involved in recruiting inflammatory and immunocompetent cells.	34258	ENSOCUG00000006999
POU5F1	Involved in embryo development by regulating pluripotency of stem cells. Expression is elevated when rabbit does are heat treated (25-35C) with increased gestational mortality.	18067	ENSOCUG00000017160
STEAP1	Mineral absorption - reduces ion complexes of Fe3+ and Cu2+. Strongly expressed in prostate cancer.	internal	ENSOCUG00000009088
LOC100350913	Endoplasmic reticulum lumen protein-retaining receptor 2 pseudogene	32846	NA
CCHCR1	Regulates mRNA metabolism and is involved in the psoriasis skin condition in humans	28470	ENSOCUG00000033705
TCF19	Transcription of genes. Associated with apoptosis, chronic hepatitis B and diabetes in humans.	24195	ENSOCUG00000017153
LOC100338571	Regulates telomere length	808	NA
LOC100339073	Required for 60S pre-ribosomal subunit export to the cytoplasm	21230	ENSOCUG00000010972
NEU1	Makes a lysosomal enzyme responsible for cleaving sialic acid	18590	ENSOCUG00000006970
SLC44A4	Transmembrane transport protein for choline and thiamine pyrophosphate.	4619	ENSOCUG00000006970
EHMT2	Gene regulation. Encodes a methyltransferase that recruits additional epigenetic regulators and represses transcription. Associations with cancer and chronic hepatitis B in humans.	internal	ENSOCUG00000006980
ABHD17C	Hydrolyses fatty acids. Involved in depalmitoylation of proteins for localization and signalling.	internal	ENSOCUG00000011140
SYT7	Ca^{2+} sensor, involved in Ca2+ mediated pathways like plasma membrane repair, synapse regulation and insulin/glucagon secretion. Triggers exocytosis of secretory and synaptic vesicles.	15174	ENSOCUG00000014577
LRRC10B	Possibly impacts lung cancer	11386	ENSOCUG00000014581
PPP1R32	Directs the substrate specificity of protein phosphatase 1, which is important in glycogen metabolism and several other regulatory pathways	4079	ENSOCUG00000000055
SDHAF2	Assists succinate dehydrogenase which links the citric acid cycle and oxidative phosphorylation in the mitochondria. SDHAF2 is also a tumour suppressor gene.	34852	ENSOCUG00000001683
KCNK10	Makes a potassium channel protein to modulate K+ concentrations across the cell membrane.	internal	ENSOCUG00000015561
FBN1	Fibrillin-1 is transported to the extracellular matrix to form microfibrils that support tissues and allow stretching of skin and ligaments.	internal	ENSOCUG00000003866
SMOC2	Promotes matrix assembly, stimulates endothelial cell proliferation and migration and formation of new blood vessels. Involved in wound healing.	1378	ENSOCUG00000012906
DEPDC5	Involved in regulating (by inhibition) the mTOR signalling pathway which is involved in cell growth and proliferation, and in nerve cell plasticity. Mutations can cause epilepsy in humans.	20358	ENSOCUG00000000594

YWHAH	A highly conserved protein that mediates signal transduction. Involved in many cellular processes, including apoptosis.	12460	ENSOCUG00000007090
FHIT	Involved in purine metabolism and a tumour suppressor. Damage is often implicated in human cancer.	internal	ENSOCUG00000011936
TMEM270	Transmembrane protein with unknown function.	8657	ENSOCUG00000005132
METTL27	A methyltransferase protein.	511	ENSOCUG00000039017
ABHD11	A protein containing an alpha/beta hydrolase fold domain. ABHDs are possibly involved in lipid metabolism.	3952	ENSOCUG00000011326
CLDN4	Channel-forming tight junction protein that mediates paracellular chloride transport in the kidney. Plays a critical role in the paracellular reabsorption of filtered chloride in the kidney collecting ducts. This protein is a high-affinity receptor for Clostridium perfringens enterotoxin (CPE) in humans and may play a role in internal organ development and function during pre- and postnatal life.	252	ENSOCUG00000010358
CLDN3	Plays a major role in tight junctions, creating a physical barrier to prevent solutes and water from passing freely through the paracellular space.	29523	NA
CRIM1	A transmembrane protein which may play a role in central nervous system development by interacting with growth factors implicated in motor neuron differentiation and survival. May play a role in capillary formation and maintenance.	12898	ENSOCUG00000001595
PCDHB1	These neural cadherin-like cell adhesion proteins are integral plasma membrane proteins. Their specific functions are unknown but they most likely play a critical role in the establishment and function of specific cell-cell neural connections.	13898	ENSOCUG00000007406
PCDHAC2	These neural cadherin-like cell adhesion proteins are integral plasma membrane proteins. Their specific functions are unknown but they most likely play a critical role in the establishment and function of specific cell-cell neural connections.	23070	ENSOCUG00000008769
LOC103349705	Uncharacterised gene	internal	ENSOCUG00000029761, ENSOCUT00000042382
ENSOCUG00000038861	Uncharacterised gene.	14524	ENSOCUG00000038861
ENSOCUG00000039509	Uncharacterised RNA gene.	31909	ENSOCUG00000039509
U2	Predicted small nuclear RNA	17990	ENSOCUG00000030650