Contributed paper

CURATION OF PIG TRAITS IN THE ONLINE MENDELIAN INHERITANCE IN ANIMALS (OMIA) DATABASE

S.M. Shields[1], S. Hermesch[2], F.W. Nicholas[1] and I. Tammen[1]

[1] Sydney School of Veterinary Science, University of Sydney, Sydney, NSW, 2006 Australia
[2] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2350 Australia

SUMMARY

Online Mendelian Inheritance in Animals (OMIA) is a freely available curated database that contains information on inherited traits and disorders (called phenes in OMIA) across more than 250 species. OMIA entries relating to pigs were reviewed, as a relatively low number of Mendelian phenes, as well as low number of phenes for which likely causal variants were listed, were noted when compared to other companion and livestock species. Of the 277 pig phenes recorded within the database at the beginning of this study in March 2020, 228 were classified as defects, 87 were Mendelian traits and for 37 of these, 45 likely causal variants were published. This study aimed to identify gaps in the information for pig phenes within OMIA. Changes to 30 pig phenes were made with a focus on updating information in OMIA's downloadable tables of known likely causal variants, One phene had previously been missed and was added, and 8 phenes were added as part of ongoing curation.

INTRODUCTION

Online Mendelian Inheritance in Animals (OMIA) (Nicholas 2020, Online Mendelian Inheritance in Animals 2021) is a freely available, curated, online database which provides researchers, veterinarians and breeders with up-to-date summary information on all the known harmful and beneficial variants in animals, together with background information on all known inherited disorders and beneficial traits. OMIA focuses on phenes with confirmed and suspected Mendelian modes of inheritance. However, phenes with unknown or complex modes of inheritance and phenes caused by somatic mutations, chromosomal abnormalities or genetic modifications or genome editing are also included. Furthermore, OMIA provides references for landmark and review articles and for papers describing genetic maps and reference genomes in animals (including 34 pig mapping and pig genome references). OMIA covers more than 3,500 phenes across more than 250 species. The vast majority of OMIA entries are for the major domesticated animals (Table 1).

One of the first causal variants identified in livestock was the variant causing malignant hyperthermia in pigs (OMIA 000621-9823, Fujii *et al.* 1991), and due to their anatomical similarities to humans, pigs are frequently used as models of human disease (Bassols *et al.* 2014). However, in comparison to other companion and livestock species, OMIA entries for pigs are relatively sparse. This is particularly evident for the number of Mendelian traits and numbers of likely causal variants. This study aimed to improve OMIA curation by adding new pig phenes and critically analysing and curating pig phenes currently recorded within OMIA, with a focus on updating variant information, as these data can be downloaded and used to develop DNA diagnostic tools.

* A joint venture of NSW Department of Primary Industries and the University of New England

Table 1. Numbers of phenes, Mendelian traits, traits with at least one likely known causal variant and total number of likely causal variants known in major companion and livestock species in OMIA (Online Mendelian Inheritance in Animals, 2021) at the time of writing (February 2021)

	Dog	Cattle	Cat	Pig	Sheep	Horse	Chicken	Goat	All species
Total phenes (traits/disorders)	784	555	362	286	257	242	223	90	3683
Total number of Mendelian traits	362	261	116	92	112	59	132	20	1553
Total number of Mendelian traits with at least one likely causal variant known	297	167	83	40	59	46	51	15	930
Total number of likely causal variants known	435	226	131	50	76	98	66	26	1268

MATERIALS AND METHODS

Identification of missing porcine information in OMIA. A literature search for phenes or publications that describe genetic conditions in pigs that are not currently listed in OMIA was performed via PubMed (PubMed 2021), using key search words, such as 'pigs', 'disease', 'inherited' and 'variant', in various combinations. Identified references and phenes were added to the database.

For Mendelian phenes where no gene or causal variant was recorded, the associated references were searched for variant information, including analysis of figures for clues, such as images of analysed sequence. Mendelian phenes without information on likely causal gene or likely causal variant were not investigated further.

Updating porcine variant information in OMIA. For phenes with at least one likely causal variant, the data in the downloadable tables of known likely causal variants was reviewed and updated to represent location information in the Sscrofa11.1 reference genome to facilitate development of diagnostic tools. Variant locations and predicted effects on proteins were determined through various methods depending on availability of published data, and included remapping and confirmation of the variant effect using *in silico* variant effect prediction.

Variant remapping. Any variants mapped in reference genomes other than the most recent, Sscrofa11.1, were remapped using the NCBI Genome Remap tool (National Center for Biotechnology Information 2021b). The input for the tool required selection of the source organism (Sus scrofa), source assembly, target assembly, variant location and chromosome number. No settings under the 'remapping options' or 'data' headings were altered. The input format for the variant location and chromosome followed the input guide provided by NCBI. New variant locations were confirmed using NCBI Genome Data Viewer (National Center for Biotechnology Information 2021a) which allowed for visualisation of the variant, determination of which strand (PLUS or MINUS) the variant was located on, and identification if the variant had an European Variant Achieve (EVA) ID (EMBL-EBI 2021). Variant locations for variants that lacked reference genome information or contained information from reference genomes not recognised by the NCBI Genome Remap tool were mapped manually, based on information provided in the original reference.

Variant effect prediction. Ensembl's variant effect predictor (VEP) (Ensembl 2021) tool was used to obtain further information on some of the recorded variants, including the variant consequence, allele, exon, intron and cDNA position. All analysed variants were located within Sscrofa 11.1 and were input in the format outlined by the VEP program (Ensembl 2021). To obtain all available results, the option of 'Ensembl and RefSeq transcripts' was selected within the

'Transcript database to use' component of the input settings. Results were confirmed by comparing the variant location generated to that previously published, recorded in OMIA or found via genome remapping.

RESULTS AND DISCUSSION

At the time of data collection (March 2020), the OMIA database contained 277 phenes in pigs. Of these, 87 were classified as Mendelian traits/disorders and 37 were Mendelian traits with known causal variants. Of the 37 Mendelian phenes with known causal variants, 17 were identified to be disease traits. The PubMed literature search resulted in the addition of one missed phene: Vitamin C deficiency (OMIA 002268-9823) as well as the addition of several references and additions to text curation fields to existing phenes.

Ongoing curation based on daily automated PubMed searches for all animal species resulted in the addition of further 8 porcine phenes. At the time of writing (February 2021) OMIA includes 286 pig phenes of which 92 are Mendelian traits. Of the 40 Mendelian traits with 50 known likely causal variants, 31 phenes are defects or disease- related (Table 2), 6 are coat-colour phenes, 2 are related to ear phenes and 1 is a blood-group phene (Table 3).

Updates made to OMIA's downloadable likely causal variant tables are summarised in Table 2 and Table 3. Chromosome (Chr), genomic DNA (g.), coding DNA (c.) and protein (p.) locations were added to 26, 26, 11 and 9 variants, respectively. Text was added to the verbal description field for 9 variants. Variant effect prediction was conducted for 16 variants, and EVA IDs were added for 13 variants.

The literature review has not been able to identify many additional porcine phenes or references and this suggests that the automated daily PubMed searches followed by manual curation have been an effective way to identify most genetic conditions in pigs. However, a broader literature search that searches journals that are not listed in PubMed may provide additional references and phenes. This study was able to update location information for many of the likely causal variants that are listed in OMIA, information that is particularly important in the absence of EVA IDs. In March 2020, OMIA listed EVA IDs for only 3 porcine variants (ear size: rs338733115; coat colour, white belt, KIT-related: rs328592739, and malignant hyperthermia: rs344435545). The inclusion of EVA IDs for 13 additional variants will allow automated updates to location information if new reference genomes become available once OMIA is hyperlinked to EVA as part of a planned update.

Further curation is needed to increase content in OMIA's text entry fields for many of the porcine phenes, particularly in relation to information about clinical signs and pathology.

Table 2. Defects/disease-related phenes - Summary information about Mendelian traits in pigs for which likely causal variants have been listed in OMIA (Online Mendelian Inheritance in Animals, 2021). OMIA ID, phene name, gene, year of publication and PubMed IDs of papers describing the likely causal variants are listed. Updates to variant information are summarised as updates to chromosome (Chr), genomic DNA (g.), coding DNA (c.) protein (p.) locations, addition of EVA IDs (EVA), analysis of data via the Ensembl Variant Effect Predictor (VEP) and verbal description field (text). Location details are available in OMIA (https://omia.org/results/?search_type=advanced&gb_species_id=9823&result_type=variant).

OMIA ID	Phene (variant phenotype)	Gene	Year	PubMed ID	Updated information
000499-9823	Hypercholesterolaemia	LDLR	1998	9556295	g. / VEP / EVA
000576-9823	Knobbed acrosome	BOLL	2020	32975846	
000621-9823	Malignant hyperthermia	RYR1	1991	1862346	g. / VEP
000636-9823	Membranoproliferative glomerulonephritis type II	CFH	2002	12466119	Chr / g.
000683-9823	Muscular hypertrophy (double muscling)	MSTN	2008	18822098	Chr / g. / EVA
000837-9823	Vitamin D-deficiency rickets, type I	CYP27B1	2003	12915218	Chr
		CYP27B1	2003	12915218	Chr
000862-9823	Resistance to oedema disease	FUT1	2000	11132149	Chr /g. / c. / p. / VEP / EVA
001058-9823	Von Willebrand disease III	VWF	2017	29208651	Chr
001085-9823	Meat quality (Rendement Napole)	PRKAG3	2000	10818001	g. / c. / p. / VEP / EVA
		PRKAG3	2001	11729159	g. / c. / p. / VEP / EVA
001128-9823	Pale soft exudative meat	PHKG1	2014	25340394	Chr / g. / text / EVA
001200-9823	Tremor, high-frequency	MYH7	2012	23153285	Chr / g.
001334-9823	Sperm, short tail	SPEF2	2006	16549801	Chr
001401-9823	Waardenburg syndrome, type 2A	MITF	2016	27349893	g. / text
001436-9823	Non-shivering thermiogenesis, absence	UCP1	2006	16933999	Chr
001673-9823	Spermatogenic arrest	TEX14	2011	22136159	Chr
001685-9823	Stress syndrome	DMD	2012	22691118	Chr / g. / c. text / EVA / VEP
001718-9823	Dwarfism, Schmid metaphyseal chondrodysplasia	COL10A1	2000	11130976	Chr / g. / c. / VEP
001952-9823	Microtia	HOXA1	2015	26035869	g. / VEP

001986-9823	Severe combined immunodeficiency disease, autosomal, T cell-negative, B cell-negative, NK cell-positive, with sensitivity to ionizing radiation	DCLRE1C	2015	26320255	g.
		DCLRE1C	2015	26320255	g.
002161-9823	Leg weakness, MSTN-related	MSTN	2019	30699111	
002178-9823	Abortion, BBS9-related	BBS9	2018	30231021	
002180-9823	Abortion due to haplotype DU1	TADA2A	2019	30875370	
002181-9823	Abortion due to haplotype LA1	POLR1B	2019	30875370	
002182-9823	Abortion due to haplotype LA2	URB1	2019	30875370	
002183-9823	Abortion due to haplotype LA3	PNKP	2019	30875370	
002210-9823	Congenital hypothyroidosis	DUOX2	2019	30651277	g. / text
002232-9823	Myopathy, congenital, SPTBN4-related	SPTBN4	2019	31850074	
002268-9823	Vitamin C deficiency	GULO	2004	15112110	Chr / text
002283-9823	Arthrogryposis multiplex congenita	KIF21A	2020	32686171	
002287-9823	Hypopigmentation and deafness	KIT	2020	33042408	g. / text
002306-9823	Fecundity, BMP15-related (Infertility and increased litter size)	BMP15	2021	33413103	

Table 3. Non-disease related phenes - Summary information about Mendelian traits in pigs for which likely causal variants have been listed in OMIA (Online Mendelian Inheritance in Animals, 2021). OMIA ID, phene name, gene, year of publication and PubMed IDs of papers describing the likely causal variants are listed. Updates to variant information are summarised as updates to chromosome (Chr), genomic DNA (g.), coding DNA (c.) protein (p.) locations, addition of EVA IDs (EVA), analysis of data via the Ensembl Variant Effect Predictor (VEP) and verbal description field (text).

OMIA ID	Phene (variant phenotype)	Gene	Year	PubMed ID	Updated information
000209-9823	Coat colour, dominant white	KIT	1996	8875890	Chr
001199-9823	Coat colour, extension (red)	MC1R	1998	9799269	Chr / g. / c. / p. / text / EVA /VEP
	Coat colour, extension (red)	MC1R	1998	9799269	Chr / g. / c. / p. / text / EVA /VEP
	Coat colour, extension (dominant black)	MC1R	1998	9799269	Chr / g. / c. / p. / text / EVA /VEP
	Coat colour, extension (dominant black)	MC1R	1998	9799269	Chr / g. / c. / p. / text / EVA /VEP
	Coat colour, extension (dominant black)	MC1R	1998	9799269	Chr / g. / c. / p. / text / EVA /VEP
	Coat colour, extension (black spotting on red or white background)	MC1R	2001	11404341 28411032	Chr / g. / c. / p. / text / EVA /VEP
001216-9823	Coat colour, roan	KIT	2011	21749430	Chr
001249-9823	Coat colour, brown, TYRP1-related	TYRP1	2011	20978532	Chr / g.
001743-9823	Coat colour, patch	KIT	1998	9724328	Chr
001745-9823	Coat colour, white belt, KIT-related	KIT	2012	23151514	Chr
	Coat colour, white belt, KIT-related	KIT	2016		g. / VEP / text
000319-9823	Ears, folded (Ear size)	MSRB3	2018	30587124	
001579-9823	Ear size (large floppy ears)	PPARD	2011	21573137	g. / VEP
	Ear size (ear size)	WIF1	2019	30815903	
001089-9823	Blood group system ABO	GGTA1	2011	21554350	Chr

CONCLUSIONS

Despite the vast amount of research performed previously to investigate traits and disorders in pigs, there are gaps in the knowledge held and potentially new phenes to be found. In comparison to other companion and livestock species relatively few inherited diseases and traits have been characterised at the molecular level and research should focus on the identification and investigation of emerging inherited conditions in pigs.

This study highlighted that further curation of OMIA data relating to pigs with a focus on updating textual fields is needed. Planned updates to OMIA will reduce the need to manually update

location information for variants that have an EVA ID and researchers are encouraged to submit likely causal variants to EVA to facilitate this process. Easy access to accurate location information of likely causal variants can facilitate the development of diagnostic SNP panels that can be used by industry to genotype animals for desirable or unfavourable alleles and therefore allow for more informed selection decisions; greatly improving the health and welfare of not only individual breeding populations, but the pig population as a whole.

Feedback on current information presented in OMIA and suggestions for additional information can be emailed to the OMIA curators (frank.nicholas@sydney.edu.au or imke.tammen@sydney.edu.au).

ACKNOWLEDGEMENTS

Stephanie Shields received the Australian Pork Research Institute Honours Scholarship award from APRIL.

REFERENCES

Bassols A., Costa C., Eckersall D., Osada J., Sabria J. and Tibau J. (2014) *Proteomics Clinical Applications* **8**: 715.

Ensembl (2021) Variant Effect Predictor. https://asia.ensembl.org/info/docs/tools/vep/index.html

EMBL-EBI (2021) European Variation Archive. https://www.ebi.ac.uk/eva/

Fujii J., Otsu K., Zorzato F., de Leon S., Khanna V.K., Weiler J.E., O'Brien P.J. and MacLennan D.H. (1991) *Science.* **253**: 448.

Nicholas F.W. (2020) *Animal Genetics* **52**: 3.

National Center for Biotechnology Information (2021a) *Genome Data Viewer*:https://www.ncbi.nlm.nih.gov/genome/gdv/

National Center for Biotechnology Information (2021b) *NCBI Genome Remapping Service*: https://www.ncbi.nlm.nih.gov/genome/tools/remap.

Online Mendelian Inheritance in Animals (2021) https://omia.org/

PubMed (2021) PubMed. Bethesda, Maryland. https://pubmed.ncbi.nlm.nih.gov/