STRING data mining of GWAS data in canine hereditary pigment-associated deafness

Maria Kelly-Smith, George M. Strain*

Comparative Biomedical Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 70803 USA

ABSTRACT

Most canine deafness is linked to white pigmentation caused by the piebald locus, shown to be the gene *MITF* (*melanocyte inducing transcription factor*), but studies have failed to identify a deafness cause. The coding regions of *MITF* have not been shown to be mutated in deaf dogs, leading us to pursue genes acting on or controlled by *MITF*. We have genotyped DNA from 502 deaf and hearing Australian cattle dogs, Dalmatians, and English setters, breeds with a high deafness prevalence. Genome-wide significance was not attained in any of our analyses, but we did identify several suggestive associations. Genome-wide association studies (GWAS) in complex hereditary disorders frequently fail to identify causative gene variants, so advanced bioinformatics data mining techniques are needed to extract information to guide future studies.

STRING diagrams are graphical representations of known and predicted networks of protein-protein interactions, identifying documented relationships between gene proteins based on the scientific literature, to identify functional gene groupings to pursue for further scrutiny. The STRING program predicts associations at a preset confidence level and suggests biological functions based on the identified genes.

Starting with (1) genes within 500 kb of GWAS-suggested SNPs, (2) known pigmentation genes, (3) known human deafness genes, and (4) genes identified from proteomic analysis of the cochlea, we generated STRING diagrams that included these genes. We then reduced the number of genes by excluding genes with no relationship to auditory function, pigmentation, or relevant structures, and identified clusters of genes that warrant further investigation.

Introduction

Most deafness in dogs is hereditary, congenital, and associated with white pigmentation. It is the genetic disorder with highest prevalence in dogs, affecting up to 30% in some breeds and reported in over 100 breeds (Strain, 2011; 2012; 2015; 2020). The mechanism of inheritance does not appear to be simple Mendelian. Genome wide association studies (GWAS) and other molecular genetic studies of deafness by our lab and others have failed to identify loci significantly associated with deafness (Cargill, 2004; Kluth and Distl, 2013; Stritzel et al., 2009; Sommerlad et al., 2010; Hayward et al., 2020), a common outcome with complex disorders (Altshuler et al., 2008; Rudy, 2012). Numerous specific candidate genes have been ruled out: *EDNRB* and *KIT* (Metallinos and Rine, 2000), *MYO15A* (Rak et al., 2002), *PAX3* (Brenig et al., 2003), *TMC1* and *TMIE* (Mieskes and Distl, 2006), *SILV* (Stritzel et al., 2007), and *ESPN, MYO3A, SLC26A5,* and *USH1C* (Mieskes and Distl, 2007). These unsuccessful outcomes indicate the need for different bioinformatics data mining approaches.

Deafness is clearly linked to white pigmentation in numerous species and breeds, which in the dog is usually caused by the classical gene locus piebald (*S*), identified as the gene *MITF* (*melanocyte inducing transcription factor*), located on canine chromosome CFA20 (Strain, 2015). Deafness of this type results from primary degeneration of the stria vascularis of the cochlear duct from loss or absence of melanocytes, followed by secondary degeneration of cochlear hair cells at 3 to 4 postnatal weeks of age (Strain, 2012).

MITF has been identified and sequenced in the reference boxer genome, but no alterations in *MITF* have yet been shown to be associated with deafness in dogs, although variants cause deafness in other species: human Waardenburg Syndrome type IIa and Tietz Syndrome (NCBI, 2020), horse (Blatter et al., 2013), cow (Philipp et al., 2011), pig (Chen L et al., 2016), mink (Markakis et al., 2014), and mouse ((Tachibana et al., 2003)Tachibana et al., 2003). *MITF* is described as the master transcriptional regulator of pigmentation, and more than 125 distinct genes are known to directly or indirectly regulate pigmentation. These actions are mediated through (1) transcription factors

---

Table 1

Partial listing of biological processes identified with linked genes and the probability of being a false discovery. Processes of ear and inner ear development, developmental pigmentation, and inner ear morphogenesis are highlighted by the indicated color code used in Fig. 2.

Biological Process (GO) Functional Enrichments in the STRING Network				
GO term[a]	Description	Count in gene set[b]	False discovery rate[c]	Color code[d]
GO:0007605	sensory perception of sound	56 of 144	1.41e-47	
GO:0007600	sensory perception	71 of 901	1.56e-21	
GO:0050877	nervous system process	81 of 1271	1.39e-19	
GO:0043583	ear development	32 of 204	2.97e-16	green
GO:0048839	inner ear development	30 of 177	5.06e-16	blue
GO:0060113	inner ear receptor cell differentiation	13 of 49	2.47e-08	
GO:0048066	developmental pigmentation	12 of 38	2.47e-08	red
GO:0042471	ear morphogenesis	16 of 112	3.69e-07	
GO:0042472	inner ear morphogenesis	14 of 92	1.35e-06	yellow
GO:0060122	inner ear receptor cell sterocilium organization	9 of 26	1.57e-06	

[a] Biological process gene groups based on the bioinformatics Gene Ontology (GO) initiative (Jackson Laboratory, 2020)
[b] number of GO term genes present in this network
[c] probability of a false association
[d] color of gene symbol fills in Fig. 2.

such as *PAX3, SOX10, TCF, LEF-1*, and *CREB*; (2) through upstream signaling factors such as *WNT* acting on *FZD* and then β catenin, *αMSH* acting on *MC1R* and then cAMP, and *EDN3* acting on *EDNRB* and *KITLG* acting on *KIT*, both then acting (3) through the posttranslational signaling pathway factor *MAPK* (Strain, 2015). Mutations of any of the genes for these proteins can potentially produce hereditary deafness.

Mutations of more than 100 genes have been shown to cause deafness in humans (Van Camp and Smith, 2020), but many have been eliminated as causative in canine deafness (reviewed in Strain, 2015). Recent proteomic studies of mouse cochlear inner hair cells and rat stria vascularis tissue have identified genes expressed therein as potential deafness candidates (Hickox et al., 2017; Uetsuka et al., 2015).

One advanced data mining technique for complex hereditary disorders is assembly of STRING diagrams that identify known and predicted physical and functional relationship networks among the proteins of candidate genes (Hu et al., 2018; Doncheva, Morris, Gorodkin, & Jensen, 2019; Szklarczyk et al., 2019). This provides a rational approach for identifying genes for consideration as causative in hereditary disorders. The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)[1] open access database and web resource were developed and continue to be expanded by a consortium of leading academic institutions. The relationships identified by the program are based on a broad variety of interaction information sources (databases, experiment literature, etc.), with the database covering nearly 25,000,000 proteins from over 5,000 organisms. Interactions between and among candidate genes can be predicted at different confidence levels with the program. The confidence level represents the approximate probability that a predicted link exists between two gene proteins or among a group of gene proteins. Available preset confidence limits are 15%, 40%, 70%, and 90% (lowest to highest). Low confidence limits display more interactions but also more false positives. Biological processes (based on established gene networks) can be projected for the diagram genes with a calculated probability of false discovery. Examples of such biological processes relevant to this study are shown in Table 1, where ear and inner ear development, developmental pigmentation, and inner ear morphogenesis processes were seen as highly relevant. Gene sets associated with the biological processes were developed based on the bioinformatics Gene Ontology initiative (Jackson Laboratory, 2020). Assembled genes in the diagram, including program-suggested additional genes, can be winnowed down by review of the functions of the identified genes/proteins plus knowledge of disorder pathology and increasing the stringency of required confidence levels. This process is subjective to a degree, but is information-based. Resulting clusters of

candidate genes can subsequently be investigated using whole genome and targeted NGS sequencing of DNA from affected and unaffected subjects.

The objectives of the current study were to perform STRING analysis of relevant assembled genes from several resources to further direct genomic studies of pigment-associated deafness in dogs, and to demonstrate the utility of this data mining technique in canine molecular genetic studies.

Methods

Foundational data for the study came from a GWAS of 502 deaf and hearing dogs (Australian cattle dogs, Dalmatians, English setters) (Hayward et al., 2020). An initial candidate list of 400 genes was assembled from (1) genes located within 500 kb of single nucleotide polymorphisms (SNPs) from our previous GWAS that had approached significance (235 genes), (2) pigmentation genes (31 genes) (UniProt Consortium, 2019), (3) genes identified in humans as responsible for non-syndromic autosomal recessive and dominant deafness (113 genes) (Van Camp and Smith, 2020), and (4) genes identified from proteomic studies of mouse cochlear inner hair cells and rat stria vascularis tissues that were also located near our GWAS SNPs (21 genes) (Hickox et al., 2017; Uetsuka et al., 2015). These sources were chosen to cast a wide net in identifying potential candidate genes that had not been identified in GWAS studies.

The STRING program was implemented to identify groupings from among the candidate genes by relevant biological processes and associated p values. Identified gene clusters were subsequently evaluated for potential exclusion based on relevance to deafness, ear development, and/or pigmentation, since this form of deafness is pigment-associated and is an early postnatal developmental disorder. Identified clusters were candidates to then pursue in subsequent directed gene sequencing studies.

Results

The STRING diagram for the initial cohort of 400 gene proteins is shown in Fig. 1 at a confidence level of 40%. Gene proteins show connection lines for known interactions identified by the Gene Ontology initiative (Jackson Laboratory, 2020) and the confidence level represents the approximate probability that a predicted link exists between two genes. Genes with gray halos (31) were pigmentation genes, genes with brown halos (113) were human deafness genes, and genes with turquoise halos (21) were from hair cell and stria vascularis proteomics studies; the remaining genes (235) were derived from our GWAS study. Circle fill colors are randomly assigned by the program in
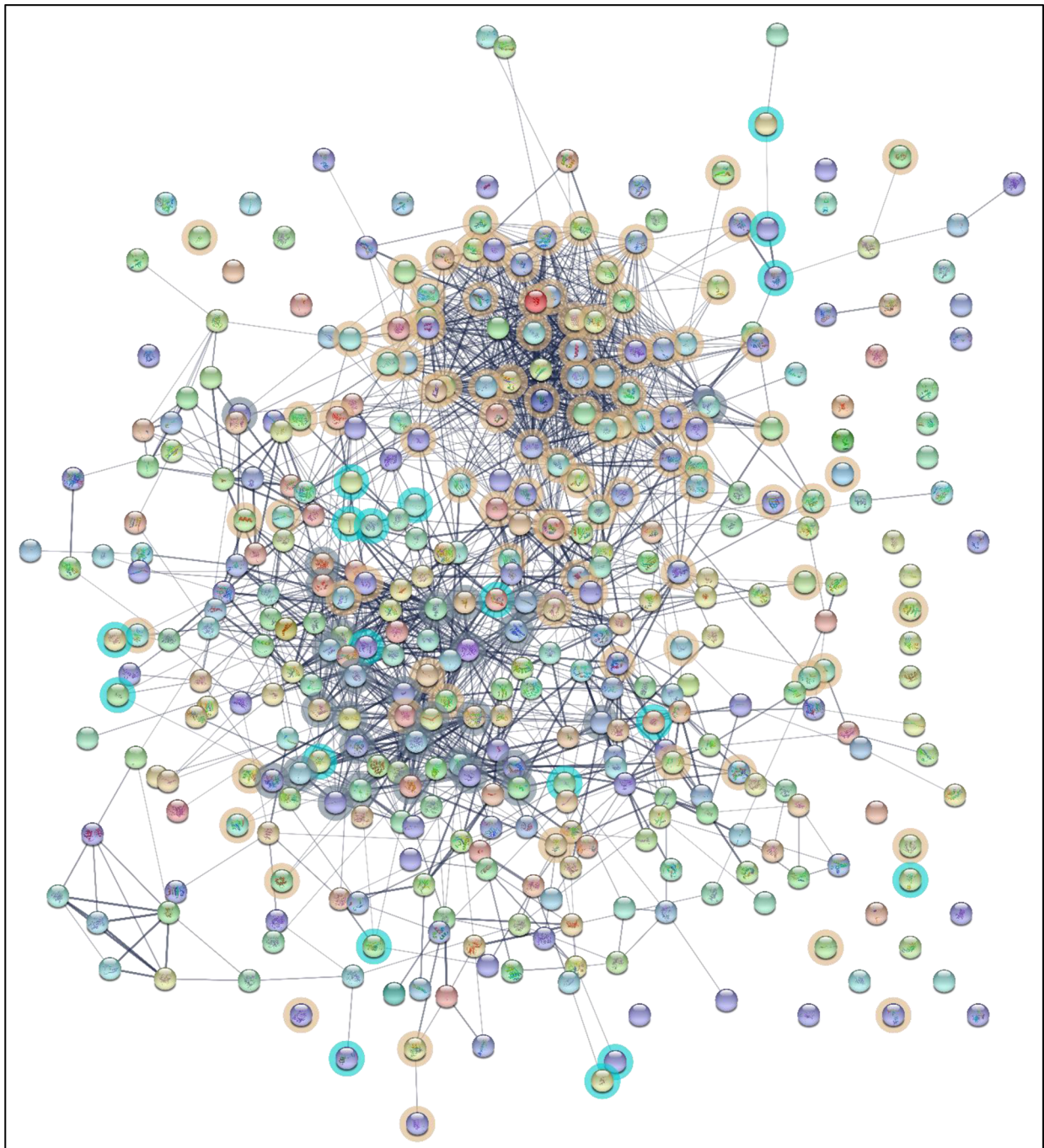
Fig. 1. Initial cohort of 400 candidate genes shown at 40% confidence level. Genes without halos (235) are from our earlier GWAS studies, identified as genes near SNPs approaching significance. Genes with gray halos (31) are pigmentation genes. Genes with brown halos (113) are human deafness genes. Genes with turquoise halos (21) are from hair cell and stria vascularis proteomics studies.

this figure. A listing of the candidate genes and their sources (GWAS, pigment, etc.) is available in Supplementary Table 1.

Following the elimination of most singleton genes and functionally non-relevant doublet and triplet gene protein pairings, and increasing association stringency by assigning a 90% confidence level, the pool was reduced to 115 genes (Fig. 2; the 115 genes are indicated in Supplementary Table 1 with a gray table cell background). Biological processes relevant to audition and inner ear development were suggested by the STRING program from this pool using gene ontology (partial listing in Table 1), with calculated probabilities against a false association ranging from a high of 1.41e-47 to a low of 1.57e-06 (Table 1). Circle fill colors in Fig. 2 represent the selected biological

processes indicated in Table 1, and include multiple colors for some genes that are involved with multiple biological processes, with halo colors as in Fig. 1. Icons within circles represent protein structure information when available. The selected biological processes color coded in Fig. 2 highlight gene clusters with highly significant associations. No genes suggested by the STRING program beyond the initial 400 were included in subsequent analyses.

Gene clusters were then evaluated for potential relevance to the known deafness phenotype and mechanism by asking: could the known function(s) of the gene at all be construed to contribute to the disorder pathology? Several potential clusters are indicated in Fig. 2 by the letters A to I, but numerous other clusters are possible. As an example of
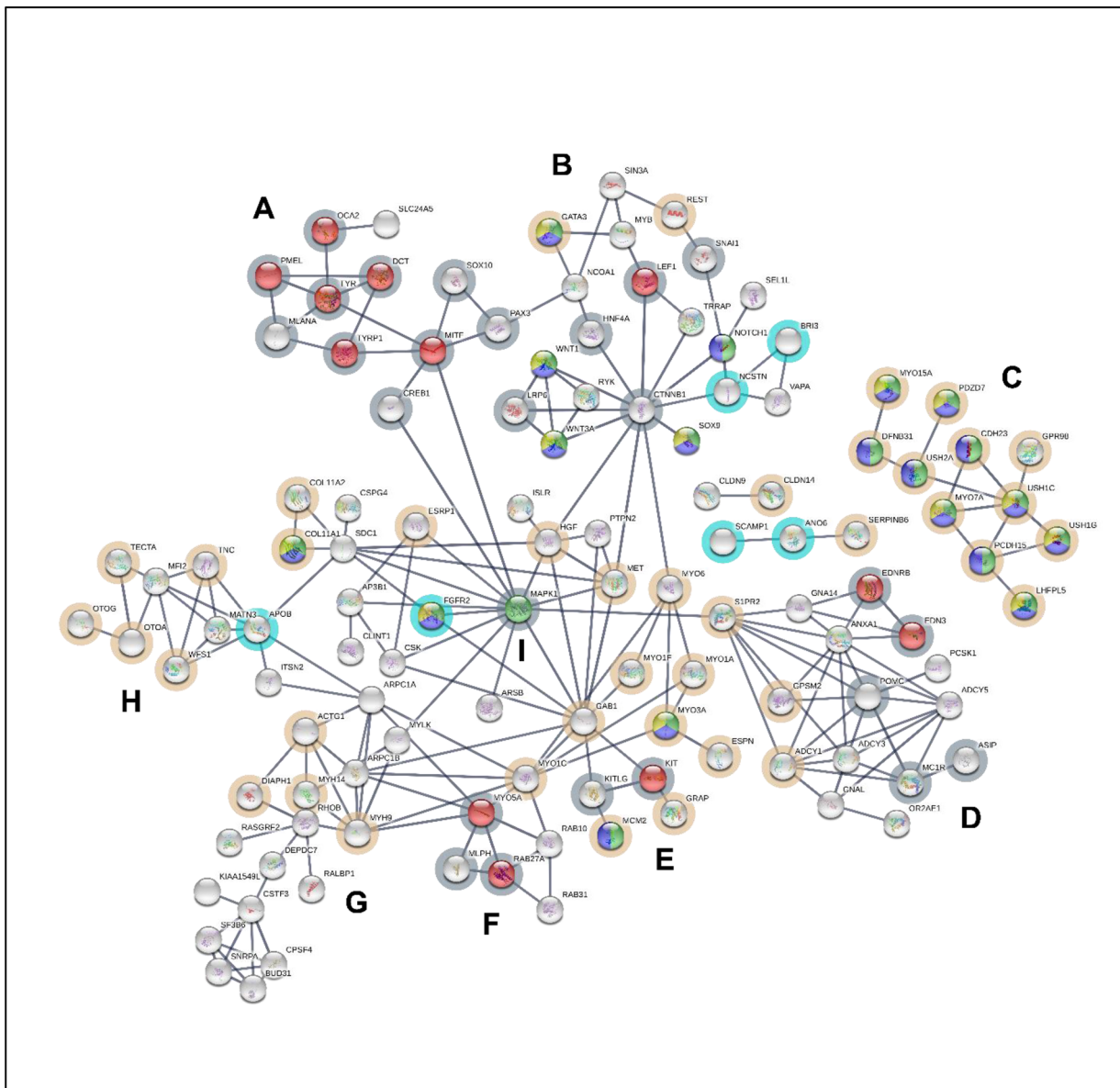
Fig. 2. Candidate pool reduced to 115 gene proteins by increasing the confidence level to 90% and elimination of singleton and functionally non-relevant gene protein pairs and triplets. Fill colors of blue, green, red, or yellow refer to biological processes identified in Table 1. Labels A through I indicate possible gene clusters for further study.
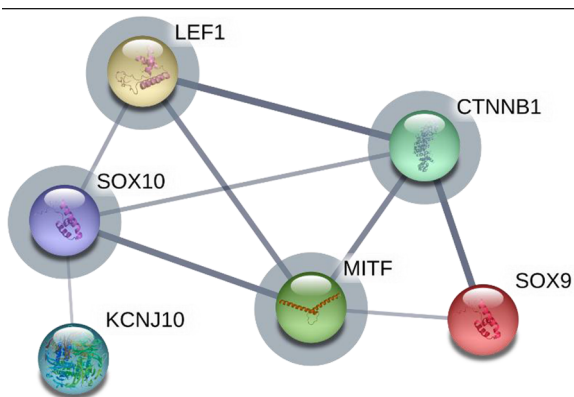

Fig. 3. Cluster of candidate genes currently undergoing further evaluation as the cause of deafness. Confidence level set to 40%, where the line thickness indicates the relative degree of confidence prediction of the interaction.

the elimination process, cluster C was eliminated because it had no connection to any of the GWAS SNPs, even though all are proteins from deafness genes.

Fig. 3 shows one cluster of candidate genes currently undergoing further evaluation as the cause of canine deafness, drawn from clusters labeled A and B in Fig. 2. Inclusion of genes in this cluster resulted from *a priori* knowledge (*MITF*) as well as suggested pairings from the STRING analysis. Genes *MITF, SOX10, LEF1*, and *KCNJ10* were chosen because of their centrality to both pigmentation and having been demonstrated to cause deafness when mutated (Van Camp and Smith, 2020). Gene *CTNNB1* (*catenin beta 1*) is a pigment gene that is also a part of the Wnt signaling pathway (Saito et al., 2003), and *SOX9* was chosen because it is involved in otic formation (Trowe et al., 2010) and is also one of the SOXE transcription factor genes (*SOX8, SOX9,* and *SOX10*) (Haseeb and Lefebvre, 2019) which might potentially substitute functionally for *SOX10* in canines.

Discussion

GWAS of complex hereditary disorders may fail to identify causative gene variants. This has been true for pigment-associated congenital deafness in dogs, where GWAS identified promising deafness-associated SNPs but significance levels were not met. In our GWAS, the results were further complicated by the possibility that differences in genetic mechanisms may exist across the three included dog breeds. Using the more holistic approach of STRING permits identification of gene relationships that might not be evident by traditional GWAS, with a focus on biological processes known to be relevant to cochlear structure and pigmentation. The results provide clusters of interacting genes and gene products, such as those in Fig. 3, for further investigation by WGS and NGS methods.

Our initial cohort of 400 genes was reduced to 115 by applying a more stringent association confidence level and removing gene singletons and doublet and triplicate gene pairings without functional association with audition and/or pigmentation. This was knowledge based, but still had a subjective aspect, and more than one set of outcomes could have resulted. The analysis outcomes do not provide a definitive target pathway, but instead guide subsequent studies by including information not previously evident. A direct or indirect relationship to MITF/pigmentation was a foundational starting point in this study. The STRING diagrams made evident gene-gene interactions that might not otherwise have been identified, but also reduced the noise of the multitude of potential involved genes and relationships: the potential interactions among our initial 400 candidate genes required an advanced bioinformatics data mining tool to progress in identifying a causative variant.

Prior to these analyses, the gene *CTNNB1* had not been under consideration, but it repeatedly appeared as STRING-identified relationships were considered. The cluster labeled C in Fig. 2 was clearly deafness-related (i.e., Usher Syndrome) but did not have a pigmentation association nor location near SNP-identified genes, and so was eliminated. A combination of clusters A and B resulted in a still large group, but evaluation of individual included genes for potential functional association with ear development and/or pigmentation enabled reduction of included genes until the cluster of Fig. 3 resulted. All of the genes in Fig. 3 have interactions as indicated by the connection lines, but other genes might still legitimately have been included. In the study described herein, the 400 initial genes exceeded the power of a human brain to extract critical, relevant gene relationships that might be informative in pursuit of an answer. STRING is a useful data mining tool to implement when GWAS alone does not provide an answer.

Conclusions and future work

STRING analysis provides a bioinformatics data mining technique to extract potential gene targets for analysis of complex genetic disorders. We are currently pursuing examination of genes identified from this analysis that did not on their own appear to be significantly associated with deafness from GWAS, opening further avenues for establishing the molecular genetic cause or causes of hereditary deafness in dogs.

Ethics statement

No animals were used in this study. DNA samples used were collected under previous studies in accordance with Louisiana State University Institutional Animal Care and Use Committee guidelines (LSU IACUC 15-104).

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.vas.2020.100118.

References

Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science, 322*, 881–888.

Blatter, M., Haase, B., Gerber, V., Poncet, P. A., Leeb, T., Rieder, S., Henke, D., Janett, F., & Burger, D. (2013). Klinische und genetische Befunde bei einem männlichen Freiberger mit der neuen Fellfarbe Macchiato [Clinical evaluation of the new coat colour macchiato in a male Franches-Montagnes horse]. [Article in German]. *Schweizer Archiv für Tierheilkunde, 155*, 229–232.

Brenig, B., Pfeiffer, I., Jaggy, A., Kathmann, I., Balzari, M., Gaillard, C., & Dolf, G. (2003). Analysis of the 5′ region of the canine PAX3 gene and exclusion as a candidate for Dalmatian deafness. *Animal Genetics, 34*, 47–50.

Cargill, E. J. (2004). *Development of a multiplexing strategy for whole genome scans of the domestic dog and analysis of hereditary deafness in the Dalmatian*Texas A&M University87.

Chen, L., Guo, W., Ren, L., Yang, M., Zhao, Y., Guo, Z., Yi, H., Li, M., Hu, Y., Long, X., et al. (2016). A de novo silencer causes elimination of MITF-M expression and profound hearing loss in pigs. *BMC Biology, 14*, 52.

Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network analysis and visualization of proteomics data. *Journal of Proteome Research, 18*, 623–632.

Haseeb, A., & Lefebvre, V. (2019). The SOXE transcription factors—SOX8, SOX9 and SOX10—share a bi-partite transactivation mechanism. *Nucleic Acids Research, 47*, 6917–6931.

Hayward, J. J., Kelly-Smith, M., Boyko, A. R., Burmeister, L., De Risio, L., Mellersh, C., ... Strain, G. M. (2020). Genome-wide association study of deafness in three canine breeds. *PLoS One, 15*(5), e0232900. https://doi.org/10.1371/journal.pone.0232900.

Hickox, A. E., Wong, A. C., Pak, K., Strojny, C., Ramirez, M., Yates, J. R., 3rd, ... Savas, J. N. (2017). Global analysis of protein expression of inner ear hair cells. *Journal of Neuroscience, 37*, 1320–1339.

Hu, W. P., Zeng, Y. Y., Zuo, Y. H., & Zhang, J. (2018). Identification of novel candidate genes involved in the progression of emphysema by bioinformatic methods. *International Journal of Chronic Obstructive Pulmonary Diseases, 13*, 3733–3747.

Jackson Laboratory Mouse Genome Informatics. 2020. Gene Ontology Browser. http://www.informatics.jax.org/vocab/gene_ontology(accessed 27 January 2020).

Kluth, S., & Distl, O. (2013). Congenital sensorineural deafness in Dalmatian dogs associated with quantitative trait loci. *PLoS One, 8*, e80642.

Markakis, M. N., Soedring, V. E., Dantzer, V., Christensen, K., & Anistoroaei, R. (2014). Association of MITF gene with hearing and pigmentation phenotype in Hedlund white American mink (*Neovison vison*). *Journal of Genetics, 93*, 477–481.

Metallinos, D., & Rine, J. (2000). Exclusion of EDNRB and KIT as the basis for white spotting in Border Collies. *Genome Biology, 1* RESEARCH0004.

Mieskes, K., & Distl, O. (2006). Elimination of TMC1 and TMIE as candidates for hereditary non-syndromic deafness in Dalmatian dogs. *Animal Genetics, 37*, 519–521.

Mieskes, K., & Distl, O. (2007). Evaluation of ESPN, MYO3A, SLC26A5 and USH1C as candidates for hereditary non-syndromic deafness (congenital sensorineural deafness) in Dalmatian dogs. *Animal Genetics, 38*, 533–537.

NCBI. 2020. MITF melanocyte inducing transcription factor [*Homo sapiens* (human)]. https://www.ncbi.nlm.nih.gov/gene/4286(accessed 27 January 2020).

Philipp, U., Lupp, B., Mömke, S., Stein, V., Tipold, A., Eule, J. C., Rehage, J., & Distl, O. (2011). A MITF mutation associated with a dominant white phenotype and bilateral deafness in German Fleckvieh cattle. *PLoS One, 6*, e28857.

Rak, S. G., Drögemüller, C., Kuiper, H., Leeb, T., Quignon, P., André, C., & Distl, O. (2002). Cloning and chromosomal localization of MYO15A to chromosome 5 of the dog (*Canis familiaris*). *Chromosome Research, 10*, 407–410.

Rudy, G.2012. Have we wasted 7 years and $100 million dollars on GWAS studies?https://blog.goldenhelix.com/have-we-wasted-7-years-and-100-million-dollars-on-gwas-studies/(accessed 27 January 2020).

Saito, H., Yasumoto, K., Takeda, K., Takahashi, K., Yamamoto, H., & Shibahara, S. (2003). Microphthalmia-associated transcription factor in the Wnt signaling pathway. *Pigment Cell Research, 16*, 261–265.

Sommerlad, S. F., McRae, A. F., McDonald, B., Johnstone, I., Cuttell, L., Seddon, J. M., & O'Leary, C. A. (2010). Congenital sensorineural deafness in Australian stumpy-tail cattle dogs is an autosomal recessive trait that maps to CFA10. *PLoS One, 5*, e13364.

Strain, G. M. (2011). *Deafness in Dogs and Cats.* Wallingford, UK: CAB International150.

Strain, G. M. (2012). Canine deafness. *Veterinary Clinics of North America: Small Animal Practice, 42*, 1209–1224.

Strain, G. M. (2015). The genetics of deafness in domestic animals. *Frontiers in Veterinary Science, 2.* https://doi.org/10.3389/fvets.2015.00029 29 (20 pages).

Strain, G.M.2020. Deafness in Dogs and Cats. http://www.lsu.edu/deafness/deaf.htm (accessed 27 January 2020).

Stritzel, S., Wöhlke, A., & Distl, O. (2007). Elimination of SILV as a candidate for congenital sensorineural deafness in Dalmatian dogs. *Animal Genetics, 38*, 662–663.

Stritzel, S., Wöhlke, A., & Distl, O. (2009). A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. *Journal of Animal Breeding and Genetics, 126*, 59–62.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research, 47*, D607–D613.

Tachibana, M., Kobayashi, Y., & Matsushima, Y. (2003). Mouse models for four types of Waardenburg syndrome. *Pigment Cell Research, 16*, 448–454.

Trowe, M.-O., Shah, S., Petry, M., Airik, R., Schuster-Gossler, K., Kist, R., & Kispert, A. (2010). Loss of Sox9 in the periotic mesenchyme affects mesenchymal expansion and differentiation, and epithelial morphogenesis during cochlea development in the mouse. *Developmental Biology, 342*, 51–62.

Uetsuka, S., Ogata, G., Nagamori, S., Isozumi, N., Nin, F., Yoshida, T., ... Inohara, H., et al. (2015). Molecular architecture of the stria vascularis membrane transport system, which is essential for physiological functions of the mammalian cochlea. *European Journal of Neuroscience, 42*, 1984–2002.

UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research, 47*, D506–D515.

Van Camp, G., Smith, R.J.H.2020. Hereditary hearing loss homepage. http://hereditaryhearingloss.org/(accessed 27 January 2020).