***animal* journal**

**Invited review: Bioinformatic methods to discover the likely causal variant of a new autosomal recessive genetic condition using genome-wide data**

G. E. Pollott

**Supplementary Table S1.** *A summary of 34 papers which reported mapping the site of a new variant using single-nucleotide polymorphism methods; first phase methods*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Method | Software | Reference | Species | Number of cases/controls | Number of SNP | Target region length (Mb) | Comments and OMIA hyperlink |
| Homozygosity mapping, Multipoint parametric linkage analysis | PLINK, MERLIN | Agerholm *et al*., 2016 | Cattle | 3/8 | 777 962/ 532 965 | 21.54 | Two stages; Second method confirmed the first<http://omia.org/OMIA002022/9913/> |
| Chi squared (Recessive), homozygosity mapping | MERLIN, PLINK | Bauer *et al*., 2017 | Horses | 3/4 | 670 796/ 210 556 | 17 | Two stages; looked for overlapping regions<http://omia.org/OMIA002096/9796/> |
| Chi squared (Allelic), homozygosity mapping, parametric linkage analysis | PLINK | Becker *et al*., 2010 | Sheep | 23/23 | 49 034 | 2.4 | Three stages; 10,000 permutations. Successive refinement of region and confirmation<http://omia.org/OMIA000649/9940/> |
| Fisher's Exact test (Genotypic) , haplotype analysis | R, HAPLOVIEW, PHASE, PLINK | Brooks *et al*., 2010 | Horses | 6/30 | 562 054 | 10/1.6 | Two-stage approach<http://omia.org/OMIA001501/9796/> |
| Autozygosity mapping | ASSHOM, ASSIST | Charlier *et al*., 2008 | Cattle | 12/14 | 60 000 | 2.12 | Congenital muscular dystonia 1; permutations used but not quantified<http://omia.org/OMIA001450/9913/> |
| Autozygosity mapping | ASSHOM, ASSIST | Charlier *et al*., 2008 | Cattle | 7/24 | 25 000 | 3.61 | Congenital muscular dystonia 2; permutations used but not quantified<http://omia.org/OMIA001451/9913/> |
| Autozygosity mapping | ASSHOM, ASSIST | Charlier *et al*., 2008 | Cattle | 8/14 | 60 000 | 0.87 | Crooked tail syndrome; permutations used but not quantified<http://omia.org/OMIA001452/9913/> |
| Autozygosity mapping | ASSHOM, ASSIST | Charlier *et al*., 2008 | Cattle | 3/9 | 60 000 | 2.42 | Ichthyosis fetalis; permutations used but not quantified<http://omia.org/OMIA000547/9913/> |
| Autozygosity mapping | ASSHOM, ASSIST | Charlier *et al*., 2008 | Cattle | 6/24 | 60 000 | 11.78 | Renal lipofuscinosis; permutations used but not quantified<http://omia.org/OMIA001407/9913/> |
| Autozygosity mapping | ASSIST | Charlier *et al*., 2012 | Cattle | 6/15 | 50 000 | 2.46 | Permutations used but not quantified<http://omia.org/OMIA000151/9913/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Drogemuller *et al*., 2011 | Cattle | 14/27 | 54 001/34 174 | 25/2.88 | Two-stage approach; 10,000 permutations<http://omia.org/OMIA001106/9913/> |
| Chi squared (Allelic) | PLINK | Finno *et al*., 2015 | Horses | 15/17 | 51 453 | 1.7 | 52 000 permutations<http://omia.org/OMIA001897/9796/> |
| Chi squared (Allelic), haplotype analysis | GRIDQTL, own routine | Flisikowski *et al*., 2010 | Cattle | 13/27 | 15 631 | 13.3 | Two-stage approach<http://omia.org/OMIA001565/9913/> |
| Autozygosity mapping | ASSHOM | Floriot *et al*., 2015 | Cattle | 190/200 | 50 000 | 2.5 | 50 000 permutations<http://omia.org/OMIA001502/9913/> |
| Chi squared (Allelic), haplotype analysis | PLINK | Fox-Clipsham *et al*, 2011 | Ponies | 18/31 | 54 602/42 0536 | 2.6 | Two-stage approach<http://omia.org/OMIA001578/9796/> |
| Autozygosity mapping, linkage analysis | ASSHOM, ASSIST | Hirano *et al*., 2013 | Cattle | 13/30 | 13 208 | 4.04 | Two-stage approach<http://omia.org/OMIA001817/9913/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Hollmann *et al*., 2017 | Cattle | 26/88 | 46 075 | 4.7 | Two-stage approach<http://omia.org/OMIA002111/9913/> |
| Linear mixed model, haplotype analysis | GEMMA, BEAGLE | Jung *et al*., 2014 | Cattle | 8/20 | 777 962/644 450 | 18.19/1.02 | Two-stage approach<http://omia.org/OMIA001935/9913/> |
| GWAS in GCTA, haplotype analysis | GCTA, BEAGLE | Kipp *et al*., 2015 | Cattle | 23/11,177 | 45 163 | ?/2.5 | Two-stage approach<http://omia.org/OMIA001965/9913/> |
| Homozygosity analysis, mixed model | BEAGLE, ASREML | Kunz *et al*., 2016 | Cattle | 43/117 | 1 958 | 1.91 | Only Chromosome 4<http://omia.org/OMIA000827/9913/> |
| Chi squared (Recessive), mixed model | Golden Helix | Mack *et al*., 2017 | Horses | 14/10 | 41 820 | 3.5 | Second stage used because of population structure<http://omia.org/OMIA001704/9796/> |
| Homozygosity mapping | PLINK | Menoud *et al*., 2012 | Cattle | 3/10 | 777 962 + 54 001 | 18.6 | <http://omia.org/OMIA000341/9913/> |
| Mixed model, homozygosity mapping | GenAbel, PLINK | Murgiano *et al*., 2014 | Cattle | 4/56 | 777 961/ 549 341 | 6.73 | Two stages; overlapping regions found<http://omia.org/OMIA001936/9913/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Myers *et al*., 2010 | Cattle | 7/9 | 54 000 | 4/3.4 | Two stages; overlapping regions found<http://omia.org/OMIA000755/9913/> |
| Homozygosity mapping  | PLINK | Pausch *et al*., 2016 | Cattle | 3/18 | 46 035 | 1.13/8.42 | Two regions found<http://omia.org/OMIA001334/9913/> |
| GWAS (model not reported)  |  | Rafati *et al*., 2016 | Horses | 14/58 | 54 000 | None | Didn’t work so resorted to NGS method<http://omia.org/OMIA002013/9796/> |
| Haplotype-based association mapping | GLASCOW | Sartelet *et al*., 2015 | Cattle | 15/275 | 34 368 | 2.2 | <http://omia.org/OMIA001953/9913/> |
| Autozygosity mapping, haplotype mapping | ASSIST, BEAGLE | Sasaki *et al*., 2016 | Cattle | 6/17 | 26 151 | 3.5 | Two-stage approach<http://omia.org/OMIA002053/9913/> |
| Autozygosity mapping, haplotype mapping | ASSHOM, BEAGLE  | Seichter *et al*., 2011 | Cattle | 60/50 | 44 473 | 0.93 | Two-stage approach<http://omia.org/OMIA001541/9913/> |
| Chi squared (Allelic), homozygosity mapping | PLINK, ASSHOM | Shariflou *et al*., 2012 | Sheep | 10/27 | 40 899 | 4.5/1.1 | Two-stage approach<http://omia.org/OMIA001595/9940/> |
| Chi squared (Recessive), LD mapping | PLINK, HAPLOVIEW | Sironen *et al*., 2011 | Pigs | 9/21 | 27 510 | 10/2 | Two-stage approach<http://omia.org/OMIA001673/9823/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Suarez-Vega *et al*., 2013 | Sheep | 7/33 | 47 864 | 4.8 | 100 000 permutations<http://omia.org/OMIA001867/9940/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Suarez-Vega *et al*., 2015 | Sheep | 20/76 | 44 785 | 4/0.87 | Two-stage approach, 1 million permutations<http://omia.org/OMIA001948/9940/> |
| Chi squared (Allelic), homozygosity mapping | PLINK | Testoni *et al*., 2012 | Cattle | 65/57 | 536 171 | ?/1.23 | Two-stage approach, 0.5 million permutations<http://omia.org/OMIA001722/9913/> |
| Fishers exact test with sliding window | Own method? | Venhoranta *et al*., 2014 | Cattle | 9/38 | 623 881 | 0.61 | <http://omia.org/OMIA001934/9913/> |
| Dfam, haplotype analysis | PLINK, PHASE | Waide *et al*., 2015 | Pigs | 20/152 | >60 000 | 5.6/1 | Two-stage approach after failing with homozygosity mapping (ASSHOM) [http://omia.org/OMIA 001986/9823/](http://omia.org/OMIA%20001986/9823/) |
| Novel method |  | Wells *et al*., 2012 | Chicken | 86/120 | 60 000 | 1.25 | <http://omia.org/OMIA000889/9031/> |
| Paper authors’ own routine |  | Zhao *et al*., 2011 | Sheep | 17/73 | 54 241 | 5.95 | <http://omia.org/OMIA001542/9940/> |

SNP = Single-nucleotide polymorphism; Mb = megabase; GWAS = genome-wide association study; LD = linkage disequilibrium.

**Supplementary Table S2.***A summary of 34 papers which reported mapping the site of a new variant using SNP methods; follow-on methods summary*

|  |  |
| --- | --- |
| Reference and OMIA ID | Follow up analysis |
| Agerholm *et al*., 2016<http://omia.org/OMIA002022/9913/> | WGS of 3 affected calves, SnpEff, Resequencing of candidate genes |
| Bauer *et al*., 2017<http://omia.org/OMIA002096/9796/> | WGS 2 cases and 2 controls; Sanger sequencing for specific gene |
| Becker *et al*., 2010<http://omia.org/OMIA000649/9940/> | Homology to humans suggested candidate genes;  |
| Brooks *et al*., 2010<http://omia.org/OMIA001501/9796/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2008<http://omia.org/OMIA001450/9913/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2008<http://omia.org/OMIA001451/9913/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2008<http://omia.org/OMIA001452/9913/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2008<http://omia.org/OMIA000547/9913/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2008<http://omia.org/OMIA001407/9913/> | Candidate gene identified within region, resequenced this gene |
| Charlier *et al*., 2012<http://omia.org/OMIA000151/9913/> | Candidate gene identified within region, resequenced this gene |
| Drogemuller *et al*., 2011<http://omia.org/OMIA001106/9913/> | Candidate gene identified within region, resequenced this gene |
| Finno *et al*., 2015<http://omia.org/OMIA001897/9796/> | WGS 2 cases and 2 controls; SnpEff |
| Flisikowski *et al*., 2010<http://omia.org/OMIA001565/9913/> | Candidate gene identified within region, resequenced this gene |
| Floriot *et al*., 2015<http://omia.org/OMIA001502/9913/> | Candidate gene identified within region, resequenced this gene |
| Fox-Clipsham *et al*, 2011<http://omia.org/OMIA001578/9796/> | Resequencing of highlighted region |
| Hirano *et al*., 2013<http://omia.org/OMIA001817/9913/> | Exome sequencing in highlighted area |
| Hollmann *et al*., 2017<http://omia.org/OMIA002111/9913/> | WGS on 2 cases and 4 controls (trios) |
| Jung *et al*., 2014<http://omia.org/OMIA001935/9913/> | WGS on 43 animals |
| Kipp *et al*., 2015<http://omia.org/OMIA001965/9913/> | WGS but not yet conclusive |
| Kunz *et al*., 2016<http://omia.org/OMIA000827/9913/> | WGS from 1 000 bulls project |
| Mack *et al*., 2017<http://omia.org/OMIA001704/9796/> | Candidate gene identified within region, resequenced this gene |
| Menoud *et al*., 2012<http://omia.org/OMIA000341/9913/> | Candidate gene identified within region, resequenced this gene |
| Murgiano *et al*., 2014<http://omia.org/OMIA001936/9913/> | WGS and resequencing |
| Myers *et al*., 2010<http://omia.org/OMIA000755/9913/> | Candidate gene identified within region, resequenced this gene |
| Pausch *et al*., 2016<http://omia.org/OMIA001334/9913/> | WGS and resequencing |
| Rafati *et al*., 2016<http://omia.org/OMIA002013/9796/> | WGS data |
| Sartelet *et al*., 2015<http://omia.org/OMIA001953/9913/> | Resequencing of highlighted region |
| Sasaki *et al*., 2016<http://omia.org/OMIA002053/9913/> | Exome sequencing in identified region |
| Seichter *et al*., 2011<http://omia.org/OMIA001541/9913/> | Looked for candidate genes in the identified region |
| Shariflou *et al*., 2012<http://omia.org/OMIA001595/9940/> | Future work suggested |
| Sironen *et al*., 2011<http://omia.org/OMIA001673/9823/> | Candidate gene identified within region, resequenced this gene |
| Suarez-Vega *et al*., 2013<http://omia.org/OMIA001867/9940/> | Candidate gene identified within region, resequenced this gene |
| Suarez-Vega *et al*., 2015<http://omia.org/OMIA001948/9940/> | Candidate gene identified within region, resequenced this gene |
| Testoni *et al*., 2012<http://omia.org/OMIA001722/9913/> | Candidate gene identified within region, resequenced this gene |
| Venhoranta *et al*., 2014<http://omia.org/OMIA001934/9913/> | WGS and resequencing |
| Waide *et al*., 2015[http://omia.org/OMIA 001986/9823/](http://omia.org/OMIA%20001986/9823/) | Reverse transcription in identified region |
| Wells *et al*., 2012<http://omia.org/OMIA000889/9031/> | Candidate gene expression levels |
| Zhao *et al*., 2011<http://omia.org/OMIA001542/9940/> | Candidate gene identified within region, resequenced this gene |

SNP = Single-nucleotide polymorphism; WGS = whole genome sequencing.

**Supplementary Material S1**

**References used in Supplementary Tables S1 and S2**

Agerholm JS, McEvoy FJ, Menzi F, Jagannathan V and Drögemüller C 2016. A *CHRNB1* frameshift mutation is associated with familial arthrogryposis multiplex congenita in Red dairy cattle. BMC Genomics 17, 479. doi: 10.1186/s12864-016-2832-x.

Bauer A, Hiemesch T, Jagannathan V, NeuditschkoM, Bachmann I, Rieder S, Mikko S, Penedo MC, Tarasova N, Vitková M, Sirtori N, Roccabianca P, Leeb T and Welle MM 2017. A Nonsense Variant in the *ST14* Gene in Akhal-Teke Horses with Naked Foal Syndrome. Genes, Genome, Genetics 7, 1315-1321. doi: <https://doi.org/10.1534/g3.117.039511>.

Becker D, Tetens J, Brunner A, Burstel D, Ganter M, Kijas J for the International Sheep Genomics Consortium and Drögemüller C 2010. Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like Homeodomain 3 (*PITX3*) gene. PLoS ONE 5, e8689. doi:10.1371/journal.pone.0008689.

Brooks SA, Gabreski N, Miller D, Brisbin A, Brown HE, Streeter C, Mezey J, Cook D and Antczak DF 2010. Whole-genome SNP association in the horse: identification of a deletion in *Myosin Va* responsible for Lavender Foal Syndrome. PLoS Genetics 6, e1000909. doi:10.1371/journal.pgen.1000909.

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet J-C, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie1 H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ and Georges M 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genetics 40, 449-454. doi:10.1038/ng.96.

Charlier C, Agerholm JS, Coppieters W, Karlskov-Mortensen P, Li W, de Jong G, Fasquelle C, Karim L, Cirera S, Cambisano N, Ahariz N, Mullaart E, Georges M and Fredholm M 2012. A deletion in the bovine *FANCI* Gene compromises fertility by causing fetal death and brachyspina. PLoS ONE 7, e43085. doi:10.1371/journal.pone.0043085.

Drögemüller C, Reichart U, Seuberlich T, Oevermann A, Baumgartner M, Kuhni Boghenbor K, Stoffel MH, Syring C, Meylan M, Muller S, Muller M, Gredler B, Solkner J and Leeb T 2011. An unusual splice defect in the Mitofusin 2 Gene (*MFN2*) is associated with degenerative axonopathy in Tyrolean Grey Cattle. PLoS ONE 6, e18931. doi:10.1371/journal.pone.0018931.

Finno CJ, Stevens C, Young A, Affolter V, Joshi NA, Ramsay S and Bannasch DL 2015. *SERPINB11* frameshift variant associated with novel hoof specific phenotype in Connemara Ponies. PLoS Genetics 11, e1005122. doi:10.1371/journal.pgen.1005122.

Flisikowski K, Venhoranta H, Nowacka-Woszuk J, McKay SD, Flyckt A, Taponen J, Schnabel R, Schwarzenbacher H, Szczerbal I, Lohi H, Fries R, Taylor JF, Switonski M and Andersson M 2010. A novel mutation in the maternally imprinted PEG3 domain results in a loss of *MIMT1* expression and causes abortions and stillbirths in cattle (*Bos taurus*). PLoS ONE 5, e15116. doi:10.1371/journal.pone.0015116.

Floriot S, Vesque C, Rodriguez S, Bourgain-Guglielmetti F, Karaiskou A, Gautier M, Duchesne A, Barbey S, Fritz S, Vasilescu A, Bertaud M, Moudjou M, Halliez S, Cormier-Daire V, Hokayem JEL, Nigg EA, Manciaux L, GuatteoR, Cesbron N, Toutirais G, Eggen A, Schneider-Maunoury S, Boichard D, Sobczak-Thepot J and Schibler l 2015. *C-Nap1* mutation affects centriole cohesion and is associated with a Seckel-like syndrome in cattle. Nature Communications 6, 6894. doi: 10.1038/ncomms7894.

Fox-Clipsham LY, Carter SD, Goodhead I, Hall N, Knottenbelt DC, May PDF, Ollier WE and Swinburne JE 2011. Identification of a mutation associated with fatal foal immunodeficiency syndrome in the Fell and Dales Pony. PLoS Genetics 7, e1002133. doi:10.1371/journal.pgen.1002133.

Hirano T, Kobayashi N, Matsuhashi T, Watanabe D, Watanabe T, Takasuga A, Sugimoto M and Sugimoto Y 2013. Mapping and exome sequencing identifies a mutation in the *IARS* gene as the cause of hereditary perinatal weak calf syndrome. PLoS ONE 8, e64036. doi:10.1371/journal.pone.0064036.

Hollmann AK, Dammann I, Wemheuer WM, Wemheuer WE, Chilla A, Tipold A, Schulz-Schaeffer WJ, Beck J, Schutz E and Brenig B 2017. Morgagnian cataract resulting from a naturally occurring nonsense mutation elucidates a role of *CPAMD8* in mammalian lens development. PLoS ONE 12, e0180665. <https://doi.org/10.1371/journal.pone.0180665>.

Jung S, Pausch H, Langenmayer MC, Schwarzenbacher H, Majzoub-Altweck M, Gollnick NS and Fries R 2014. A nonsense mutation in *PLD4* is associated with a zinc deficiency-like syndrome in Fleckvieh cattle. BMC Genomics 15, 623. <http://www.biomedcentral.com/1471-2164/15/623>.

Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, Pausch H, Fries R, Thaller G, Tetens J, Pott J, Piechotta M and Grünberg W 2015. A new Holstein haplotype affecting calf survival. Interbull annual meeting, Orlando, FL, July 11, 2015; Interbull Bulletin No. 49, 49-53.

Kunz E, Rothammer S, Pausch H, Schwarzenbacher H, Seefried FR, Matiasek K, Seichter D, Russ I, Fries R, and Medugorac, I 2016. Confirmation of a non-synonymous SNP in *PNPLA8* as a candidate causal mutation for Weaver syndrome in Brown Swiss cattle. Genetics Selection and Evolution 48, 21. doi: 10.1186/s12711-016-0201-5.

Mack M, Kowalski E, Grahn R, Bras D, Penedo MCT and Bellone R 2017. Two variants in *SLC24A5* are associated with "Tiger-Eye" iris pigmentation in Puerto Rican Paso Fino horses. Genes, Genomes Genetics 7, 2799-2806. doi: 10.1534/g3.117.043786.

Menoud A, Welle M, Tetens J, Lichtner P and Drögemüller C 2012. A *COL7A1* mutation causes dystrophic epidermolysis bullosa in Rotes Hohenvieh cattle. PLoS ONE 7, e38823. doi:10.1371/journal.pone.0038823.

Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O'Toole D, O'Connell JR, Beever JE, Sonstegard TS and Smith TP 2010. A deletion mutation in bovine *SLC4A2* is associated with osteopetrosis in Red Angus cattle. BMC Genomics 11, 337. doi: 10.1186/1471-2164-11-337.

Murgiano L, Jagannathan V, Calderoni V, Joechler M, Gentile A and Drögemüller C 2014. Looking the cow in the eye: deletion in the *NID1* gene is associated with recessive inherited cataract in Romagnola cattle. PLoS ONE 9, e110628. doi:10.1371/journal.pone.0110628.

Pausch H, Venhoranta H, Wurmser C, Hakala K, Iso-Touru T, Sironen A, Vingborg RK, Lohi H, Söderquist L, Fries R and Andersson M 2016. A frameshift mutation in *ARMC3* is associated with a tail stump sperm defect in Swedish Red (*Bos taurus*) cattle. BMC Genetics 17, 49. doi: 10.1186/s12863-016-0356-7.

Rafati N, Andersson LS, Mikko S, Feng C, Raudsepp T, Pettersson J, Janecka J, Wattle O, Ameur A, Thyreen G, Eberth J, Huddleston J, Malig M, Bailey E, Eichler EE, Dalin G, Chowdary B, Anderssson L, Lindgren G and Rubin CJ 2016. Large deletions at the *SHOX* locus in the pseudoautosomal region are associated with skeletal atavism in Shetland Ponies. Genes, Genomes Genetics 6, 2213-2223. doi: 10.1534/g3.116.029645.

Sartelet A, Li W, Pailhoux E, Richard C, Tamma N, Karim L, Fasquelle C, Druet T, Coppieters W, Georges M and Charlier C 2015. Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchor biosynthesis class H gene (*PIGH*) and causes arthrogryposis in Belgian Blue cattle. BMC Genomics 16, 316. doi: 10.1186/s12864-015-1528-y.

Sasaki S, Hasegawa K, Higashi T, Suzuki Y, Sugano S, Yasuda Y and Sugimoto Y 2016. A missense mutation in solute carrier family 12, member 1 (*SLC12A1*) causes hydrallantois in Japanese Black cattle. BMC Genomics 17, 724. doi: 10.1186/s12864-016-3035-1.

Seichter D, Russ I, Förster M and Medugorac I 2011. SNP-based association mapping of Arachnomelia in Fleckvieh cattle. Animal Genetics 42, 544-547. doi: 10.1111/j.1365-2052.2010.02167.x.

Shariflou MR, Wade CM, Kijas J, McCulloch R, Windsor PA, Tammen I and Nicholas FW 2013. Brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in Merino sheep maps to a 1.1-megabase region on ovine chromosome OAR2. Animal Genetics 44, 231-233. doi: 10.1111/j.1365-2052.2012.02392.x.

Sironen A, Uimari P, Venhoranta H, Andersson M and Vilkki J 2011. An exonic insertion within *TEX14* gene causes spermatogenic arrest in pigs. BMC Genomics 12, 591. doi: 10.1186/1471-2164-12-591.

Suárez-Vega A, Gutiérrez-Gil B, Cuchillo-Ibáñez I, Sáez-Valero J, Pérez V, García-Gámez E, Benavides J and Arranz JJ 2013. Identification of a 31-bp deletion in the *RELN* Gene causing lissencephaly with cerebellar hypoplasia in sheep. PLoS ONE 8, e81072. doi:10.1371/journal.pone.0081072.

Suárez-Vega A, Gutiérrez-Gil B, Benavides J, Perez V, Tosser-Klopp G, Klopp C, Keennel SK and Arranz JJ 2015. Combining GWAS and RNA-Seq Approaches for detection of the causal mutation for hereditary junctional epidermolysis bullosa in sheep. PLoS ONE 10, e0126416. doi:10.1371/journal.pone.0126416.

Testoni S, Bartolone E, Rossi M, Patrignani A, Bruggmann R, Lichtner P, Tetens J, Gentile A and Drögemüller C 2012. KDM2B is implicated in bovine lethal multi-organic developmental dysplasia. PLoS ONE 7, e45634. doi:10.1371/journal.pone.0045634.

Venhoranta H, Pausch H, Flisikowski K, Wurmser C, Taponen J, Rautala H, Kind A, Schnieke A, Fries R, Lohi H and Andersson M 2014. In frame exon skipping in *UBE3B* is associated with developmental disorders and increased mortality in cattle. BMC Genomics 15, 890. doi: 10.1186/1471-2164-15-890.

Waide EH, Dekkers JC, Ross JW, Rowland RR, Wyatt CR, Ewen CL, Evans AB, Thekkoot DM, Boddicker NJ, Serão NV, Ellinwood NM and Tuggle CK 2015. Not all SCID pigs are created equally: two independent mutations in the Artemis gene cause SCID in pigs. Journal of Immunology 195, 3171-3179. doi: 10.4049/jimmunol.1501132.

Wells KL, Hadad Y, Ben-Avraham D, Hillel J, Cahaner A and Headon DJ 2012. Genome-wide SNP scan of pooled DNA reveals nonsense mutation in *FGF20* in the Scaleless line of featherless chickens. BMC Genomics 13, 257. doi: 10.1186/1471-2164-13-257.

Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF and Garrick DJ 2011. A novel nonsense mutation in the *DMP1* gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. PLoS ONE 6, e21739. doi:10.1371/journal.pone.0021739.

**Supplementary Table S3**

**Methods and software references plus websites accessed on 25/10/17**

|  |  |
| --- | --- |
| ASREML | Gilmour AR, Gogel BJ, Cullis BR, Welham SJ and Thompson R 2015. ASReml User Guide Release 4.1 Functional Specication, VSN International Ltd, Hemel Hempstead, HP1 1ES, UKhttps://www.vsni.co.uk/software/asreml/ |
| ASSHOM and ASSIST | Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet J-C, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie1 H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ and Georges M 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genetics 40, 449-454. doi:10.1038/ng.96. |
| BEAGLE | Browning SR and Browning BL 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81,1084-1097. doi:10.1086/521987https://faculty.washington.edu/browning/beagle/beagle.html |
| GATK | https://software.broadinstitute.org/gatk/documentation/ |
| GCTA | Yang J, Lee SH, Goddard ME and Visscher PM. 2011. GCTA: a tool for Genome-wide Complex Trait Analysis. American Journal of Human Genetics 88, 76-82.http://cnsgenomics.com/software/gcta |
| GEMMA | Zhou X and Stephens M 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821-824.http://www.xzlab.org/software.html |
| GenAbel | Aulchenko YS, de Koning D and Haley C 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177, 577–585.http://www.genabel.org/ |
| GERMLINE | Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM and Pe’er I 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Research 19, 318–326. <http://doi.org/10.1101/> gr.081398.108http://www.cs.columbia.edu/~gusev/germline/ |
| Golden Helix | http://goldenhelix.com/ |
| GridQTL | Seaton G, Hernandez J, Grunchec JA, White I, Allen J, De Koning DJ, Wei W, Berry D, Haley C and Knott S. 2006. GridQTL: A grid portal for QTL mapping of compute intensive datasets. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, August 13-18, 2006. Belo Horizonte, Brazil.http://gridqt1.cap.ed.ac.uk/gridqtl\_project.htm |
| HAPLOVIEW | Barrett JC, Fry B, Maller B and Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 263–265. https://doi.org/10.1093/bioinformatics/bth457https://www.broadinstitute.org/haploview/haploview |
| MERLIN | Abecasis GR, Cherny SS, Cookson WO and Cardon LR 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nature Genetics 30, 97-101http://csg.sph.umich.edu/abecasis/MERLIN/download/ |
| PHASE | Stephens M, Smith N and Donnelly P 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics, 68, 978--989.http://stephenslab.uchicago.edu/software.html |
| PLINK 1.7 | Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.http://zzz.bwh.harvard.edu/plink/ |
| PLINK 1.9 | Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM and Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4.https://www.cog-genomics.org/plink2 |
| R | R Core Team 2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. |
| SAMTOOLS | Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078-9http://www.htslib.org/doc/samtools.html |
| SnpEff | Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X and Ruden DM 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.". Fly (Austin) 6, 80-92.http://snpeff.sourceforge.net/ |
| VCF files | http://www.internationalgenome.org/wiki/Analysis/vcf4.0/ |

**Supplementary Material S2**

**Models of χ2 tests in PLINK**

Assuming the Lavender Foal Syndrome data used in the paper (Brooks *et al*., 2010) and an example that could be in the homozygous region on horse chromosome 1 (ECA1), these are the four models available in PLINK. For the recessive and dominant models grouping is carried out by minor allele frequency (MAF). In the example below, T is the minor allele (29/72 compared to 43/72 for A). Note that this may or may not be the target allele found as homozygous in all cases in the target area and so genotypic distributions may change from 0/0/6 to 6/0/0 in the same region of the genotype depending on the **MAF** at the single-nucleotide polymorphism (SNP) in question. This may be more noticeable in the allelic, dominance or recessive tests.

--model gen (Genotypic model)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AA | AT | TT | Total |
| Cases | 6 | 0 | 0 | 6 |
| Controls | 8 | 15 | 7 | 30 |
| Total | 14 | 15 | 7 | 36 |

χ2 table in Hardy-Weinberg Equilibrium (HWE) with an allele frequency of the target allele (A) = 0.5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Observed*** | AA | AT | TT | Total |
| Cases | 6 | 0 | 0 | 6 |
| Controls | 8 | 15 | 7 | 30 |
| Total | 14 | 15 | 7 | 36 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Expected*** | AA | AT | TT | Contribution to χ2 value1 |
| Cases | 2.25 | 2.5 | 1.25 | 10 (0.83) |
| Controls | 11.25 | 12.5 | 6.25 | 2 (0.17) |
| χ2 value | 12 | P < 0.01 |  |  |

1 Proportion shown in parentheses

--assoc/--model (Allelic model)

|  |  |  |  |
| --- | --- | --- | --- |
|  | A | T | Total |
| Cases | 12 | 0 | 12 |
| Controls | 31 | 29 | 60 |
| Total | 43 | 29 | 72 |

N.B. T is the minor allele in this example

--model dom (Dominant model)

|  |  |  |  |
| --- | --- | --- | --- |
|  | AA | TT/AT | Total |
| Cases | 6 | 0 | 6 |
| Controls | 8 | 22 | 30 |
| Total | 14 | 22 | 36 |

--model rec (Recessive model)

|  |  |  |  |
| --- | --- | --- | --- |
|  | AA/AT | TT | Total |
| Cases | 6 | 0 | 6 |
| Controls | 23 | 7 | 30 |
| Total | 29 | 7 | 72 |

Results from running the tests shown using the Lavender Foal Syndrome (LFS) dataset in PLINK based on EquCab2.0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Top SNP position | Genotype/allele frequencies | Region start | Region end |
| χ2 Genotypic  | ECA1:133 508 742 | 5/0/1 0/22/8 | ECA1:129 228 091 | ECA1:139 718 117 |
| χ2 Allelic  | ECA1:135 938 654 | 0/12 35/25 | ECA1:129 228 091 | ECA1:139 718 117 |
| χ2 Recessive  | ECA1:133 508 742 | 5/1 0/30 | ECA1:130 449 869 | ECA1:139 718 117 |
| χ2 Dominance  | ECA1:135 938 654 | 0/6 28/2 | ECA1:126 619 234 | ECA1:138 722 881 |
| Final paper result | Mutation at ECA1:138 235 715 |  | Haplotype analysis ECA1:136 812 666 | Haplotype analysis ECA1:138 375 254 |

The four different models all found the region containing the new autosomal recessive mutation in the **LFS** dataset but this varied in length from 9.27 megabases (Mb) (Recessive model) to 12.1 **Mb** (Dominance model). Notice that the **SNP** with the highest χ2 value for the genotypic and recessive models did not have completely homozygous cases and so the result was more dependent on the genotype (allele) frequencies in controls to achieve significance. None of the ‘top SNP’ were in the region found by haplotype analysis or the various runs of homozygosity (ROH) methods reported in this paper.

 

Figures A (Left-hand plot) and B (Right-hand plot) showing the chi-squared values and probabilities for allele frequencies from 0.1 to 0.9 for the LFS dataset in HWE

 Plotting the χ2 values for the LFS data for all frequencies of the target allele A in controls, assumed to be in HWE, from 0.1 to 0.9 produces Figure A, with the probabilities for a single χ2 and Fisher’s Exact Test plotted in Figure B. Using this simulation as a guide, then it will not be possible to pick up a significant χ2 value unless the target allele has a frequency of less than ~0.7 in the controls (Figure B). Thus any SNP which is monomorphic in controls for the target allele will not be found to have a significant χ2 test. In fact there will need to be at least half of the controls with the alternate allele (in heterozygotes or the other homozygote) in order to see a significant association of that SNP. Secondly, should the mutation occur in a monomorphic region of the chromosome, but nearby SNP have a target allele frequency < 0.7 then these adjacent SNP will appear significant if they are in the same ROH containing the target SNP but not directly next to the mutation. Thus the widely used χ2 test will only be effective when one adjacent SNP allele segregates with the mutation and one does not, to a large degree. This was not the case in the LFS data and the region only appeared as significant because nearby SNP had more genotypes containing ‘the alternate allele’. In fact, the SNP with the lowest *P* value was not in the region found to be homozygous in all cases but had a genotypic distribution (say TT/AT/AA) of 1/0/5 in cases and 8/22/0 in controls. It was not even homozygous for the ‘target’ SNP in all cases and had no homozygotes for the target SNP in controls; the allele frequency was 36/60 (0.6) for the ‘unaffected’ allele in controls; a highly significant result as shown in Figure 2B. If a correction for multiple testing were applied to these data then a region found significant would have to have an even higher χ2 value than discussed here and would require very low allele frequencies of the target allele in controls. Even the permutation methods reported in Table 4 failed to find the target SNP.

**Supplementary Material S3**

**Details of the Autozygosity-By-Difference Method**

1. Genotype and allele frequencies are calculated for each single-nucleotide polymorphism (SNP) in cases and the commonest homozygous genotype in cases (CHG) identified at each **SNP**. This should facilitate identifying the homozygous candidate SNP genotype since all cases should be homozygous for the same alleles at the site of the new mutation.
2. Each SNP is coded 1 or 0 in all cases and controls; 1 is allocated when the SNP genotype is that of the **CHG** for a given SNP. The code 0 is allocated to SNP of the other homozygote and heterozygotes. Missing SNP genotypes are coded as 1, after deleting all monomorphic SNP from the dataset. This allows for the possibility that the mutation may result in an unrecognisable sequence at the candidate SNP.
3. Each SNP in each animal (cases and controls) is scored if it is located in a run of homozygosity. A run of homozygosity (ROH) is identified as being adjacent SNP of the same homozygous genotype as the CHG in cases (given a value of 1 in the previous step). All SNP in such runs are given the same score equal to its length in base pairs. This is calculated to include half the distance between the first (and last) SNP in a **ROH** and the next SNP on the chromosome.
4. At each SNP the mean score in both cases and controls, and also the difference between the means of cases and controls, are calculated. This ensures that the effect of any ROH that are breed or population specific is ‘removed’ from the calculation. It also allows for any effects of ‘incomplete penetrance’ or late-onset conditions to be accounted for and inspected.
5. The location of the mutation is likely to be in the region with the highest autozygosity-by-difference (ABD) scores but taking due account of the likely long ROH scores in cases. Care should be taken to inspect the scores, best done graphically using a Manhattan plot, to see if there is evidence of incomplete penetrance and interpret the cases’ and controls’ scores appropriately. The highest scores should be in a region of two or more monomorphic SNP, in cases, but this should be verified by inspection of the original genotypes. The new mutation should be situated between two SNP, monomorphic in cases. The method can also be used to store the ROH scores for each subject individually and the results inspected in an appropriate spreadsheet program.
6. Significance of the identified scores, either case ROH or **ABD** scores, is calculated by permutation. This is achieved by repeatedly rescoring the dataset with phenotypes randomly allocated to animals and including the original cases and controls in the permutations (after Charlier *et al*., 2008), say 1 000 times (N). The probability of each score is the proportion of occasions when that score, and all those greater than it, was achieved, out of N times the number of SNP tested.



**Supplementary Figure S1** Results of analysing the Lavender Foal Syndrome dataset using a genotypic model (3x2) with Fisher’s Exact Test. -Log10 probabilities shown for all 36 651 autosomal single-nucleotide polymorphism with a minor allele frequency > 0.05. Bonferroni correction value *P* = 5.85. (Based on the EquCab2.0 build of the horse genome).

 **Supplementary Figure S2** Results of calculating ASSHOM scores (Charlier et al., 2008) for the Lavender Foal Syndrome dataset cases. (Based on the EquCab2.0 build of the horse genome).

 **Supplementary Figure S3** Results of calculating ASSIST scores (Charlier et al., 2008) for the Lavender Foal Syndrome dataset cases. (Based on the EquCab2.0 build of the horse genome).



**Supplementary Figure S4** Results of calculating mean run of homozygosity (ROH) scores for controls in the Lavender Foal Syndrome dataset using the Autozygosity-By-Difference method. (Based on the EquCab2.0 build of the horse genome).