**Original Scientific paper**
10.7251/AGRENG2001077M
UDC 757.113.2:636.1:636.1

# GENOME-WIDE DISTRIBUTION OF AUTOZYGOSITY ISLANDS IN SLOVAK WARMBLOOD HORSE

Nina MORAVČÍKOVÁ[1*], Radovan KASARDA[1], Marko HALO[2], Ondrej KADLEČÍK[1], Anna TRAKOVICKÁ[1], Kristína LEHOCKÁ[1], Barbora OLŠANSKÁ[1], Juraj CANDRÁK[1]

[1]Department of Animal Genetics and Breeding Biology, Slovak University of Agriculture in Nitra, Slovakia
[2]Department of Animal Husbandry, Slovak University of Agriculture in Nitra, Slovakia
*Corresponding author: nina.moravcikova1@gmail.com

## ABSTRACT

The objective of this study was to estimate the distribution of autozygosity islands represented by homozygous segments (ROHs) in the genome of the Slovak Warmblood horse. The Slovak Warmblood is a very efficient breed with the excellent characteristics of a sport horse. The study included 37 animals that were genotyped by GGP Equine70k chip (71,947 SNPs). Only animals (36) and autosomal SNPs (62,439) with call rate >90% and minor allele frequency >1% were included in subsequent analyses. The homozygous segments were defined as stretches with minimum 15 consecutive homozygous SNPs of >500 kb with minimum density 1 SNP per 100 kb and maximum gap between markers of 1,000 kb. The heterozygous or missing calls were not accepted. The analysis indicated in total of 8,501 autozygosity islands in the genome of the Slovak Warmblood horse. The majority of identified segments (85.42%) were most likely derived from the remote ancestors in the past. Only 0.15% of detected segments resulted from the recent selection events affecting the genetic structure of studied population. The proportion of segments varied across chromosomes. The major fraction of autosome residing in ROH was found on ECA1 (8.30%), while ECA31 showed the lowest ROH coverage. The scan for overlapping homozygous segments shared by more than 50 % of animals demonstrated that the ECA6 autosome may be under strong selection pressure. Inside those selection signals, several genes were identified including them associated with immunity and reproduction.

**Keywords**: *footprints of selection, horse, genomic data, runs of homozygosity.*

## INTRODUCTION

The Slovak Warmblood horse is one of the national breeds that is a descendant of Austro-Hungarian warmblood breeds and horses which were ennobled from Arabians, English purebreds and halfbloods. Compared to Czech Warmblood, it has more blood of the Hungarian halfbred, Furioso, Przedswit, Gidran and Nonius

and little of the Oldenburg and East Fresian horse. Recently, the breed was improved through the use of Trakenher and Hanoverian blood. Since 1961, the Slovak Warmblood is bred especially for the purpose of being sport and work horses. It belongs to the medium sized horses with well-developed skeleton and rectangular body frame. The Slovak Warmblood is very efficient breed with excellent movement, good character and lively temperament. The main goal of its breeding is to fix the type and breed in purity, with emphasis on the massive type (Hendricks, 2007; National Stud Farm Topoľčianky, 2019).

Modern horse breeds represent heterogeneous populations selected for specific appearance and performance traits. In general, genomic regions under strong selection pressure due to specific breeding for traits of interest display low genetic variability resulting in increase of proportion of continuous homozygous segments that are common in individuals of target population or breed (Metzger *et al.*, 2015). The development and implementation of new genomic tools, including high-density SNP genotyping arrays, make it possible to analyse such genomic regions and overall genetic background of populations at a deeper level. Previous studies revealed that the application of SNP arrays allow for better resolution of the genome-wide determination of diversity parameters, including selection signatures, genetic bottleneck and trend of inbreeding compared to the most commonly used pedigree analysis. To conserve the genetic resources of native horse breed, several studies highlighted the importance of genome-wide analyses focusing on the distribution of long consecutive homozygous genotype segments referred to as runs of homozygosity (ROH) (Metzger *et al.*, 2015; Kamiński *et al.*, 2017; Druml *et al.*, 2018; Grilz-Seger *et al.*, 2018). The ROHs are a result of parents transmitting identical haplotypes, which can be used to estimate autozygosity (Grilz-Seger *et al.*, 2018). It was demonstrated that the distribution of ROH segments in the genome give insight into a complex population history, genetic events in the past, demographic evolution of a population over time and genetic relatedness among individuals (Peripolli *et al.*, 2017). In addition, the ROHs can be used to assess the impact of selection on the genome trough identification of autozygosity islands. Provided that the frequency of certain alleles increases due to positive selection the selection signatures can be valuable resource for mapping of causative mutations (Nolte *et al.*, 2019).

The aim of this study was to analyse the distribution of autozygosity islands represented by runs of homozygosity (ROHs) in the genome of the Slovak Warmblood horse and to identify genomic region under strong selection pressure due to breeding for specific traits of interest.

## MATERIAL AND METHODS

The genomic information for 37 animals were used to determine the distribution of autozygosity islands in the genome of the Slovak Warmblood horse. Animals were genotyped for 71,947 SNPs by using GGP Equine70k genotyping array. The quality of genotyping data was checked by R package plinkQC (Anderson *et al.*, 2010) and subsequent SNP pruning was performed using PLINK v1.9 (Chang *et*

*al.*, 2015) to filtered out i) SNP markers with uknown chromosomal position as well as SNPs located on sex chromosomes, ii) animals and autosomal SNPs with call rate < 90 % and iii) SNP markers with minor allele frequency < 1 % across animals. The Hardy-Weinberg equilibrium (HWE) limit was set to 0.00001.

The autozygosity islands were defined as ROH segments that contain minimum 15 consecutive homozygous SNPs of > 500 kb with minimum density 1 SNP per 100 kb, maximum gap between markers of 1,000 kb and no missing or heterozygous SNPs in a run, following the study of Nolte et al. (2019). The distribution of ROHs in the genome was scanned by R package detectRUNS (Biscarini *et al.*, 2018). To identify signals of selection the runs incidence per each SNP was calculated. The genome-wide occurrence of SNPs in ROH was expressed as the proportion of overlapping ROH shared among animals. The genomic regions significantly affected by positive selection were recognized based on the SNPs in ROHs shared by more than 50 % of the entire sample. For functional analysis, genomic regions covering signals of positive selection were scanned for annotated genes in the equine reference assembly EquCab2.0 by web based tool Biomart from Ensembl database (https://www.ensembl.org/).

## RESULTS AND DISCUSSION

From totally 71,947 SNP markers, up to 68,214 loci were positioned across equine autosomes (ECA). Due to low level of genotyping call rate one animal and 1,655 SNPs were removed from the database. The MAF limit did not nt meet 4,964 loci. After SNP pruning overall 36 animals and 62,439 SNP markers covering 2,240,031.34 kb of the autosomal genome were retained for the detection of autozygosity islands. Among adjacent SNP markers the average spacing at level 35.89 ± 34.66 was found. The obtained genotyping rate across samples and SNPs markers (98.26 %) was in agreement with previous studies in horses (Kamiński *et al.*, 2017; Druml *et al.*, 2018; Griltz-Seger *et al.*, 2019).

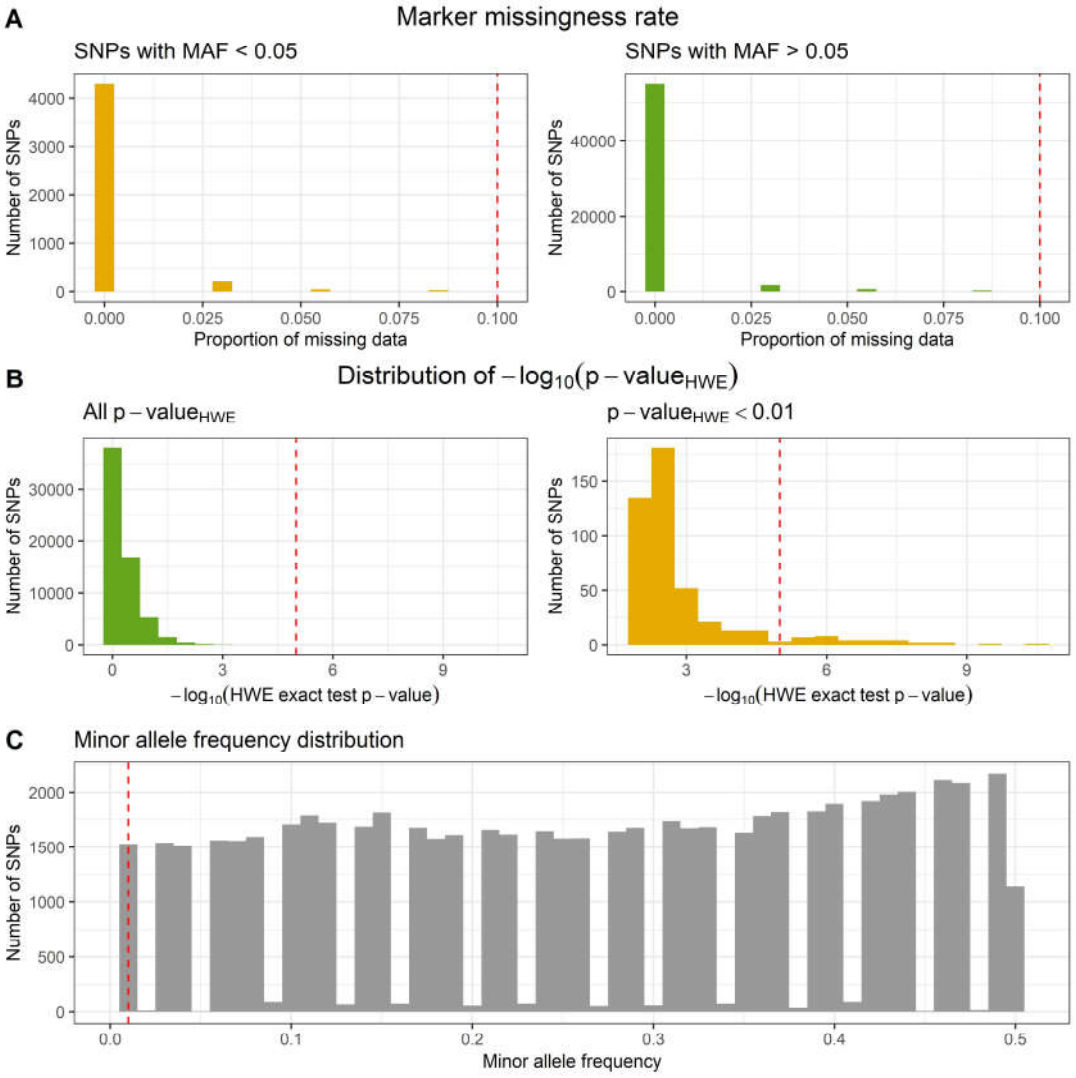


Figure 1. Quality of genotyping data in sample of Slovak Warmblood horse per SNP markers

In the sample of the Slovak Warmblood horses totally 8,501 ROH segments covering in average 8.45 % (189.41 Mb) of the autosomal genome were identified. The similar proportion of ROHs was found e.g. in the genome of Slovak and Hungarian Lipizzan horses (Grilz-Seger *et al.*, 2019) or Posjave horse breed (Grilz-Seger *et al.*, 2018). The number and mean length of the ROH segments was not uniform and varied across autosomes. The highest ROHs proportion was found on ECA1 (8.3 %), while the lowest number of ROHs was detected on ECA31 (0.98 %). The majority of identified segments (85.42%) were most likely derived from

the remote ancestors in the past. Only 0.15 % of detected segments resulted from the recent selection events affecting the genetic structure of studied population. The distribution and proportion of autosomes residing in ROH are in accordance with previous studies in population of Polish Konik horse (Kamiński *et al.*, 2017), Austrian Noriker (Grilz-Seger *et al.*, 2019) or Haflinger horse breed (Druml *et al.*, 2018). Assuming that the ROHs with extreme frequency are most likely results of intensive breeding for traits of interest during the grading-up process of the breed, in the genome of Slovak Warmblood horse totally 8 regions across 7 autosomes (1, 2, 6, 9, 11, 15 and 16) significantly affected by positive selection were identified (Fig. 1). The longest region was found on ECA6, while the shortest was localized on ECA9. The scan for overlapping homozygous segments shared by more than 50 % of animals indicated that mainly the ECA6 autosome showed strong impact of selection. Inside detected selection signals overall 80 protein-coding genes were identified (Table 1).

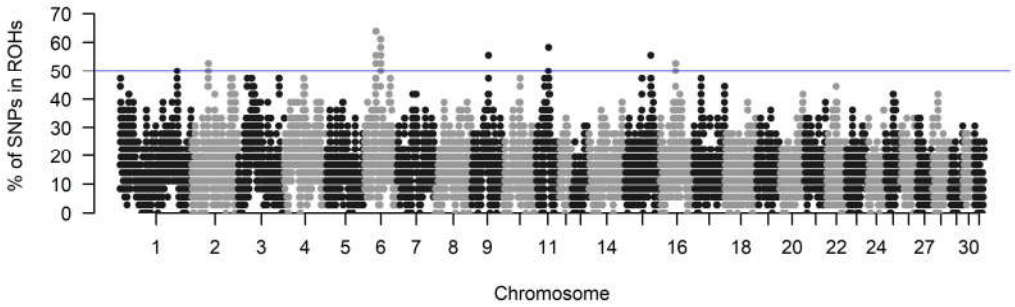


Figure 2. Proportion of overlapping ROH segments shared among analysed Slovak Warmblood horses

Table 1. Signal of selection in genome of Slovak Warmblood horse expressed by ROH segments with extreme frequency

| ECA | Start position (Mb) | End position (Mb) | Region size (Mb) | Protein-coding genes |
|---|---|---|---|---|
| 1 | 147.61 | 148.06 | 0.44 | STARD9, TTBK2, CDAN1, HAUS2, LRRC57, SNAP23, ZNF106, CAPN3, GANC, VPS39 |
| 2 | 42.74 | 43.12 | 0.38 | RERE, SLC45A1, ERRFI1, PARK7, TNFRSF9 |
| 6 | 29.36 | 30.02 | 0.66 | CACNA2D4, DCP1B, LRTM2, ADIPOR2 |
| | 41.18 | 42.71 | 1.54 | LRP6, MANSC1, BORCS5, DUSP16, CREBL2, GRP19, CDKN1B, APOLD1, DDX47, GRPC5A, GRPC5D, HEBP1, FAM234B, GSG1, EMP1, GRIN2B |
| 9 | 44.25 | 44.53 | 0.27 | UQCRB, MTERF3, PTDSS1, SDC2 |
| 11 | 32.32 | 33.58 | 1.26 | MSI2, CCDC182, MRPS23, CUEDC1, VEZF1, SRSF1, DYNLL2, EPX, MKS1, LPO, MPO, TSPOAP1, MIR142, RNF43, SUPT4H1, HSF5, MTMR4, TEX14, RAD51C PPM1E, TRIM37, SKA2 |
| 15 | 67.26 | 67.83 | 0.57 | LBH, YPELP |
| 16 | 39.90 | 40.62 | 0.71 | SLC26A6, TMEM89, UQCRC1, MIR711, PFKFB4, SHISA5, TREX1, ATRIP, CCDC51, PLXNB1, FBXW12, SPINK8, NME6, ECATH-3, ECATH-2, CDC25A, MAP4 |

Inside the genomic region on ECA1 the STARD9, TTBK2 and CAPN3 genes are located. The STARD9 gene was in humans associated with various defects, including epilepsy, acquired microcephaly, and blindness (Okamoto *et al.*, 2017). The TTBK2 gene (Tau tubulin kinase 2) is essential for initiating the assembly of primary cilia in the embryo (Goetz *et al.*, 2012). The CAPN3 gene (Calpain 3) is a member of the calpain family that are intracellular calcium-dependent cysteine proteases found in most eukaryotes. It is well known that calpains are involved in the proteolysis of functionally relevant structural proteins such as the myofibrillar proteins and cytoskeletal anchorage complexes (Bhat *et al.*, 2018). The ADIPOR2, LRP6, GRPC5A, and EMP1 genes were found within the regions showing selection signatures on ECA6. The biological function of these genes was mostly studied in humans. It was suggested that the ADIPOR2 gene could be a determinant for atherosclerosis independent of insulin resistance status, possibly by affecting ADIPOR2 protein levels (Halvatsiotis *et al.*, 2010). The mutations in the LRP6 gene was associated with many complex human diseases, including metabolic syndrome, cancer, Alzheimer's disease and osteoporosis (Wang *et al.*, 2018). The GRPC5A gene plays a role in spontaneous and environmentally induced lung carcinogenesis (Wang *et al.*, 2016). Ahmat Amin *et al.* (2018) revealed that the epithelial membrane protein 1 (EMP1) gene has elevated expression in the cancer cells. Within the region on ECA9 two biologically important genes were identified; the UQCRB gene that is important for mitochondrial complex III stability, electron transport, cellular oxygen sensing and angiogenesis and the MTERF3 gene which is a negative regulator of mtDNA transcription initiation (Park *et al.*, 2007; Kim *et al.*, 2017). The CUEDC1, VEZF1 and SKA2 genes that are located within the identified selection signals on ECA11 are involved in estrogen pathway (CUEDC1), development of blood vascular and lymphatic system (VEZF1) and cell cycle regulation (SKA2) (Gowher *et al.*, 2008; Lopes *et al.*, 2018; Xie and Bu, 2018). Several genes positioned within the detected region on ECA16 were found to be involved in the response to oxidative stress and apoptosis (TREX1, CDC25A), innate immunity (ECATH2, ECATH3) and regulation of cell cycle (CDC25A) (Scocchi *et al.*, 1999; Shen and Huang, 2012; Barizzone *et al.*, 2013).

## CONCLUSION

As expected, due to genetic background of Slovak Warmblood horse the proportion of genome residing in ROH was comparable to other horse populations in Europe. The study showed that the ROH segments covered in average 8.45 % of the autosomal genome expressed by SNPs on the chip. The majority of identified autozygosity islands were most likely derived from the remote ancestors. Only 0.15% of the genome was affected by the recent mating of relatives. Scan for signals of selection revealed seven genomic regions that was significantly affected by positive selection during the grading-up process of Slovak Warmblood horse breed. From biological point of view, genes identified directly in the regions under selection pressure are mostly involved in the genetic control of cell cycle regulation, immunity and reproduction.

**ACKNOWLEDGEMENT**

This study was supported by the Slovak Research and Development Agency (APVV-14-0054and APVV-17-0060) and VEGA (1/0742/17).

**REFERENCES**

Ahmat Amin M.K.B., Shimizu A., Zankov D.P., Sato A., Kurita S., Ito M., Maeda T., Yoshida T., Sakaue T., Higashiyama S., Kawauchi A., Ogita H. (2018). Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1. Oncogene, 37, 5416-5434.

Anderson C.A., Pettersson F.H., Clarke G.M., Cardon L.R., Morris A.P., Zondervan K.T. (2010). Data quality control in genetic case-control association studies. Nature Protocols, 5, 1564-1573.

Barizzone N., Monti S., Mellone S., Godi M., Marchini M., Scorza R., Danieli M.G., D'Alfonso S. (2013). Rare variants in the TREX1 gene and susceptibility to autoimmune diseases. BioMed Research International, 471703.

Bhat Z.F., Mortona J.D., Masona S.L., El-Din A., Bekhit A. (2018). Role of calpain system in meat tenderness: review. Food Science and Human Wellness, 7, 196-204.

Biscarini F., Cozzi P., Gaspa G., Marras G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes, CRAN (The Comprehensive R Archive Network).

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4, 7.

Druml T., Neuditschko M., Grilz-Seger G., Horna M., Ricard A., Mesaric M., Cotman M., Pausch H., Brem G. (2018). Population Networks Associated with Runs of Homozygosity Reveal New Insights into the Breeding History of the Haflinger Horse. Journal of Heredity, 109, 384-392.

Goetz S.C., Liem K.F. Jr., Anderson K.V. (2012). The spinocerebellar ataxia-associated gene Tau tubulin kinase 2 controls the initiation of ciliogenesis. Cell, 151, 847-858.

Gowher H., Stuhlmann H., Felsenfeld G. (2008). Vezf1 regulates genomic DNA methylation through its effects on expression of DNA methyltransferase Dnmt3b. Genes and Development, 22, 2075-2084.

Grilz-Seger G., Druml T., Neuditschko M., Mesarič M., Cotman M., Brem G. (2019). Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. Animal Genetics, 50, 334-346.

Grilz-Seger G., Mesarič M., Cotman M., Neuditschko M., Druml T., Brem G. (2018). Runs of Homozygosity and Population History of Three Horse Breeds With Small Population Size. Journal of Equine Veterinary Science, 71, 27-34.

Halvatsiotis I., Tsiotra P.C., Ikonomidis I., Kollias A., Mitrou P., Maratou E., Boutati E., Lekakis J., Dimitriadis G., Economopoulos T., Kremastinos D.T., Raptis S.A. (2010). Genetic variation in the adiponectin receptor 2 (ADIPOR2) gene is associated with coronary artery disease and increased ADIPOR2 expression in peripheral monocytes. Cardiovascular Diabetology, 9, 10.

Hendricks B. (2007). International Encyclopedia of Horse Breeds. University of Oklahoma Press, Norman.

Kamiński S., Hering D.M., Jaworski Z., Zabolewicz T., Ruść A. (2017). Assessment of genomic inbreeding in Polish Konik horses. Polish Journal of Veterinary, 20, 603-605.

Kim H.C., Chang J., Lee H.S., Kwon H.J. (2017). Mitochondrial UQCRB as a new molecular prognostic biomarker of human colorectal cancer. Experimental and Molecular Medicine, 49, e391.

Lopes R., Korkmaz G., Revilla S.A., van Vliet R., Nagel R., Custers L., Kim Y., van Breugel PC., Zwart W., Moumbeini B., Manber Z., Elkon R., Agami R. (2018). CUEDC1 is a primary target of ERα essential for the growth of breast cancer cells. Cancer Letters, 436, 87-95.

Metzger J., Karwath M., Tonda R., Beltran S., Águeda L., Gut M., Gut I.G., Distl O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics, 16, 764.

National Stud Farm Topoľčianky (2019). http://nztopolcianky.sk/index.php/en/breeding/breeds/811-the-slovak-warmblood. Accessed on 09/07/2019.

Nolte W., Thaller G., Kuehn Ch. (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. PLoS One, 14, e0215913.

Okamoto N., Tsuchiya Y., Miya F., Tsunoda T., Yamashita K., Boroevich K.A., Kato M., Saitoh S., Yamasaki M., Kanemura Y., Kosaki K., Kitagawa D. (2017). A novel genetic syndrome with STARD9 mutation and abnormal spindle morphology. American Journal of Medical Genetics Part A, 173, 2690-2696.

Park C.B., Asin-Cayuela J., Cámara Y., Shi Y., Pellegrini M., Gaspari M., Wibom R., Hultenby K., Erdjument-Bromage H., Tempst P., Falkenberg M., Gustafsson C.M., Larsson N.G. (2007). MTERF3 is a negative regulator of mammalian mtDNA transcription. Cell, 130, 273-285.

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., Baldi F. (2017). Runs of homozygosity: current knowledge and applications in livestock. Animal Genetics, 48, 255-271.

Scocchi M., Bontempo D., Boscolo S., Tomasinsig L., Giulotto E., Zanetti M. (1999). Novel cathelicidins in horse leukocytes. FEBS Letters, 457, 459-464.

Shen T., Huang S. (2012). The role of Cdc25A in the regulation of cell proliferation and apoptosis. Anti-Cancer Agents in Medicinal Chemistry, 12, 631–639.

Wang J., Farris A.B., Xu K., Wang P., Zhang X., Duong D.M., Yi H., Shu H.K., Sun S.Y., Wang Y. (2016). GPRC5A suppresses protein synthesis at the endoplasmic reticulum to prevent radiation-induced lung tumorigenesis. Nature Communications, 7, 11795.

Wang Z.M., Luo J.Q., Xu L.Y., Zhou H.H., Zhang W. (2018). Harnessing low-density lipoprotein receptor protein 6 (LRP6) genetic variation and Wnt signaling for innovative diagnostics in complex diseases. Pharmacogenomics Journal, 18, 351-358.

Xie M., Bu Y. (2018). SKA2/FAM33A: A novel gene implicated in cell cycle, tumorigenesis, and psychiatric disorders. Genes and Diseases, 6, 25-30.