Detecting genetic regions associated with height in the native ponies of the British Isles by using high density SNP genotyping

Skujina, I.; Winton, C. L.; Hegarty, M. J.; McMahon, R.; Nash, D. M.; Davies-Morel, M. C. G.; McEwan, N. R.

Published in:
Genome

DOI:
10.1139/gen-2018-0006

Publication date:
2018

Citation for published version (APA):
Skujina, I., Winton, C. L., Hegarty, M. J., McMahon, R., Nash, D. M., Davies-Morel, M. C. G., & McEwan, N. R. (2018). Detecting genetic regions associated with height in the native ponies of the British Isles by using high density SNP genotyping. *Genome*, *61*(10), 767-770. https://doi.org/10.1139/gen-2018-0006

Detecting Genetic Regions Associated With Height in the Native Ponies of the British Isles by Using High Density SNP Genotyping

I. Skujina [1], C. L. Winton [1], M. J. Hegarty [1], R. McMahon [1, 2], D. M. Nash [1], M. C. G. Davies-Morel [1] and N. R. McEwan [1, 3]

[1] **Institute of Biological Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Wales, SY23 3DA**

[2] **Molecular Haematology, Haematology Laboratory, Level 2, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, Scotland EH16 4SA.**

[3] **School of Pharmacy and Life Sciences, Robert Gordon University, Garthdee Road, Aberdeen, Scotland, AB10 7GJ**

Address for communications: n.mcewan@rgu.ac.uk

Abstract

Height is an important characteristic in the equine industry although little is known about its genetic control in native British breeds of ponies. This study aimed to map QTL data with the withers height in 4 pony breeds native to the British Isles, including 2 different sections within Welsh Cobs. In this study, a genome-wide analysis approach using the *Illumina EquineSNP50 Infinium BeadChip* was applied to 105 ponies and cobs. Analysis identified 222 highly significant height-associated SNPs ($P \leq 10^{-5}$), among which three SNPs on ECA9 have also been previously reported elsewhere. The highest number of significant SNPs associated to height in the native British horses were located on ECA1, ECA8 and ECA16.

Running header:

Height related SNPs in native British ponies

Keywords Height, Horse, Single Nucleotide Polymorphisms, Genome Wide Association, High Density Genotyping , Quantitative Trait Loci.

Introduction

Height in humans is an intensively investigated polygenic trait (Hirschhorn & Lettre, 2009). Its analysis can be traced back over a century to proposals that the normal distribution of human height can be explained as the interaction of many inherited factors with individually small effects (Galton, 1886; Fisher, 1918). Since height is an easily and accurately measured highly heritable trait it has served as a model for other quantitative traits (Palmert & Hirschhorn, 2003). Successful cross-species identification of genes contributing to body size regulation should provide insights into mechanisms of growth and development (Lettre *et al*., 2008), and also provide guidance in making selection decisions in breeding programmes. Size, as measured by height at the withers, is an important equine character as it impacts directly on a horse's performance and maintenance costs. Barrey *et al*. (2002) noted that taller horses have slower stride frequencies and longer stride lengths, leading to increased speed, which is essential in many competitive horse sports, and height at withers in dressage is important because of its relationship with kinematic variables at trot (Sánchez *et al*., 2013; Sánchez-Guerrero *et al*., 2016). Height is a key parameter that determines whether an animal is classified as a horse or a pony and in some breeds may be the main defining characteristic of the breed or sub-grouping, *e.g*., sections A, B, C and D in Welsh Mountain ponies and cobs.

Although genetic variation explains up to 90% of the height in some human populations (Silventoinen *et al*., 2000), most single nucleotide polymorphisms (SNPs) identified for height individually account for a very small proportion of observed variance (Visscher, 2008). In order to identify SNPs contributing to height variance, large cohorts of many thousands of individuals and extremely high density microarrays using millions of SNPs are necessary. In a study of 183,727 individuals, the GIANT consortium identified 180 loci significantly associated with adult height based on genome-wide analysis (GWA) that together could only explain around 10% of observed variation. Allen *et al*. (2010) estimated a sample size of 500,000 is required to identify 99·6% of these loci as genome-wide significant. Since height has been actively selected in equine breed development, both between breeds and within breeds, it is likely that rare height-associated mutations of moderate effect will have occurred and have been selected in some breeds and not others. Studying associations across breeds may therefore detect loci fixed by selection which may differ from those associated with height variation within contemporaneous breeds. Comparison of GWA in horses between and within breeds might therefore be a powerful tool to provide insight into loci that are potential candidates in other species as well, and hence good targets for further investigations.

Methods

DNA samples were obtained from either hair root or cheek swab samples. In total, 120 adult (4+ years old) ponies of 4 different breeds native to the British Isles were sampled and genotyped. Genome-wide SNP genotypes of these individuals were obtained using *Illumina EquineSNP50 Infinium BeadChip* (San Diego, CA; Fan *et al*., 2010) at Central Biotechnology Services (Henry Wellcome Building,

School of Medicine, Cardiff University, Heath Park, Cardiff, Wales) following the manufacturer's instructions. Of these original 120 animals, 107 had the required phenotype information and 2 of these horses were excluded because of too high an autosomal heterozygosity value (FDR <1%; mean autosomal heterozygosity of the sample = 0.29; SEM = 0.07; heterozygosity of the removed samples ≥ 0.61), leaving 105 horses that passed all initial quality metrics and their SNP genotyping results were subsequently compared to phenotypic data using *GenABEL* software (Aulchenko *et al*., 2007) in R.

Initially, genomic relatedness analysis was performed using classical multidimensional scaling (MDS) in R. The two groups of Welsh ponies – Welsh Mountain ponies (Section As) and Welsh Cobs (Section Ds) – showed a close relationship clustering as a single (A&D) group containing all 43 individuals from the two sections A (n=23) and D (n=20), with the other breeds - Highland ponies (n = 17), Fell ponies (n = 23) and Connemara ponies (n = 22) - forming three separate more distant clusters (Supplement S1).

Two separate GWAS runs were performed 1) using all 4 breeds (n=105), i.e., "105 group" and 2) in Welsh ponies and cobs only (n=43), i.e., "A&D group". The first analysis was used to show common polymorphisms associated with height in different breeds, whilst possibly containing "false-positives" as polymorphisms may not be controlling the height but show significance due to the interbreed differences and the history of each breed. The second GWAS analysis ("A&D" group) was therefore performed within individuals from a single breed, with extreme height variation: Sections As must not exceed 12.0 hands (122 cm) and Section Ds must be taller than 13.2 hands (137 cm). In a way, this served as a control, similar to occasional inclusion of half-siblings into affected groups in case-control association studies. This then reduces the chance of population substructures producing spurious associations when a whole population is analysed. It also means that the SNPs with the top significance values if appearing in both groups are particularly strong candidates for a genuine association. The different Welsh sections originated from Welsh Mountain ponies, but have subsequently been bred for different purposes. In total 18,950 SNPs for all 105 horses and 14,832 specifically for the A&D group (i.e. not all SNPs identified across the bigger dataset were identified in the smaller dataset) were obtained with sufficient quality scores to permit analysis.

Results

Out of the 18,950 SNPs that passed the quality control in the 105 group and 14,832 SNPs in the A&D group, 402 (2.12%) and 108 (0.73%) showed significant associations ($P \leq 1x10^{-5}$). The full list of SNPs associated with height ($P \leq 1x10^{-5}$) in both groups is shown in Tables S2 and S3. In the 105 group, 19 SNPs showed very strong evidence of association with height (Figure 1). The uncorrected p-values for these SNPs were $<10x10^{-9}$ and two of them – SNPs BIEC2-382246 (ECA17) and BIEC2-1054436 (ECA8) were significant at $P=3.30x10^{-12}$ and $P=7.48x10^{-12}$ respectively (Figure 1a). These still achieved statistical significance ($1.8x10^{-4}$ and $2.3x10^{-4}$ respectively) after the Bonferroni correction.

The peak P-values in the A&D group reached a significance level of $P \leq 1x10^{-8}$ (BIEC2-1124358, $P=5.78x10^{-8,}$ ECAX; BIEC2-258538, $P=5.80x10^{-8}$, ECA14; BIEC2-1050295 $P=7.47x10^{-8}$, ECA8; BIEC2-1052645, $P=8.7x10^{-8}$, ECA8) and another 16 SNPs had P-values $\leq 1x10^{-7}$ (Figure 1b). Two of the five most significant SNPs in the group of all 105 ponies occurred within 706 bp on Chromosome 26 (positions 14008370 and 14009076) and the highest density of significant height associations were mapped on ECA1, ECA10, ECA8, ECA16 and ECA26, whereas in the A&D group the ECA1, ECA3, ECA8 and ECA14 had most of the significantly associated SNPs (Figure 1). In total, 222 SNPs were significantly associated ($P \leq 1x10^{-5}$) in both groups (Table S4). A few larger regions appeared particularly saturated with significant SNPs: 3.2 Mb on chromosome 1 (positions 86760461-89963952); 1.2Mb on chromosome 3 (positions 58379387- 59602599); 0.75 Mb on chromosome 9 (positions 34937791- 35685534); 4.8Mb on chromosome 11 (positions 42468715- 47262501); 4.8Mb on chromosome 16 (positions 31105067- 35930428) and 4.1Mb on chromosome 17 (positions 62419312- 66476932). The last of these included the most significant SNP (BIEC2-382246) in the 105 group and two other highly significant ($<1x10^{-8}$) SNPs. Similarly, the average number of significant SNPs per chromosome (Figures 1) was highest in both the 105 group and the A&D group on chromosomes 1, 8 and 16. Table S5 shows the complete map of all regions. Therefore these regions are good candidates for height regulating loci, whereas the loci with height significance in the 105 group only (e.g. the most significant SNPs on ECA26 and ECA10) may be aberrant positives resulting from breed specificity.

Discussion / Conclusions

Three of the SNPs identified here (BIEC2-1105377, BIEC2-1105370 and BIEC2-1105372) located on chromosome 9 had been reported previously (Signer-Hasler *et al*., 2012) in Franches-Montagnes horses, as being significant for height in ~140bp vicinity form *zinc finger and AT hook domain containing* (ZFAT) gene, which has been associated with height in multiple human populations (Takeuchi *et al*., 2009; Allen *et al*., 2010). An intergenic association in a gene-sparse region 410 kbp upstream of the ZFAT was reported by Makvandi-Nejad *et al*. (2012) as one out of the four loci explaining 83% of the variation in several different horse breeds. Conversely, no significant QTLs were found near LCORL/NCAPG on ECA3, which had explained ~18% of height variance in German Warmbloods (Tetens *et al*., 2013). Similarly, Makvandi-Nejad *et al*. (2012) reported 4 QTLs accounting for 83% of the variation in 16 various horse breeds, none of which was significantly associated with height in these native British ponies. However significant SNPs that mapped on ECA3, ECA6, ECA9 and ECA11 were detected here and were identified in a previous analysis which included Welsh Mountain ponies and Welsh ponies (Makvandi-Nejad *et al*., 2012) although the frequency of alleles for these loci did not follow the same pattern, indicating potential differences in the native breeds here. Genes identified in association with genetic markers included those associated with myosin synthesis.

The different Welsh sections originated from Welsh Mountain ponies, but have subsequently been bred for different purposes. Section As were initially used as pack ponies and for riding and later as pit ponies. In contrast, the Section Ds (Cobs) have been selected to serve as draft horses, and to carry or pull substantial weights. Consequently, the SNPs that are significant in both groups have a strong probability to be genuinely height associated polymorphisms in native British pony breeds.

In conclusion, this work presents how combined inter-breed and intra-breed GWAS data can be used to identify SNPs associated with height and therefore exclude SNPs that appear to be observed due to selection for breed-specific characteristics rather than to the actual trait studied. The novel height-associated SNPs and genomic regions thereby enhance our insight into loci associated with height-related traits in general.

Acknowledgements

CLW was funded by the award of an IBERS Ph.D. studentship

Ethics statement

All sampling was done in accordance with the recommendations and permissions of the Aberystwyth University's Animal Welfare and Ethical Review Body (AWERB) and no Home Office licence was required for non-invasive sampling according to the Animals (Scientific Procedures) Act 1986 (ASPA). All samples were collected from animals with the permission of the owners.

References

Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467:** 832-838.

Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23:** 1294-1296.

Barrey E, Desliens F, Poirel D, Biau S, Lemaire S, Rivero J L, Langlois B. 2002. Early evaluation of dressage ability in different breeds. *Equine Veterinary Journal* **34:** 319-324.

Fan B, Du ZQ, Gorbach DM, Rothschild MF. 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian-Australasian Journal of Animal Sciences* **23:** 833-847.

Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh* **52:** 399–433.

Galton F. 1886. Hereditary stature. *Nature* **33:** 317.

Hirschhorn JN, Lettre G. 2009. Progress in genome-wide association studies of human height. *Hormone Research in Paediatrics* **71:** 5-13.

Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T. 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature genetics* **40:** 584-591.

Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredo AI, Bellone RR, Mezey JG, Brooks SA, Sutter NB. 2012. Four loci explain 83% of size variation in the horse. *PLoS ONE* **7:** e39929.

Palmert MR, Hirschhorn JN. 2003. Genetic approaches to stature, pubertal timing, and other complex traits. *Molecular Genetics and Metabolism* **80:** 1-10.

Sánchez MJ, Gómez MD, Peña F, Monterde JG, Morales JL, Molin A, Valera M. 2013. Relationship between conformation trait and gait characteristics in Spanish Purebred horses. *Archiv Tierzucht* **13:** 137-148.

Sánchez-Guerrero MJ, Molina A, Gómez MD, Peña F, Valera M. 2016. Relationship between morphology and performance: Signature of mass-selection in Pura Raza Español horse. *Livestock Science* **185:**148-155.

Silventoinen K, Kaprio J, Lahelma E, Koskenvuo M. 2000. Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *American Journal of Public Health* **90:** 627–630.

Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, Rieder S. 2012. A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS ONE* **7:** e37282.

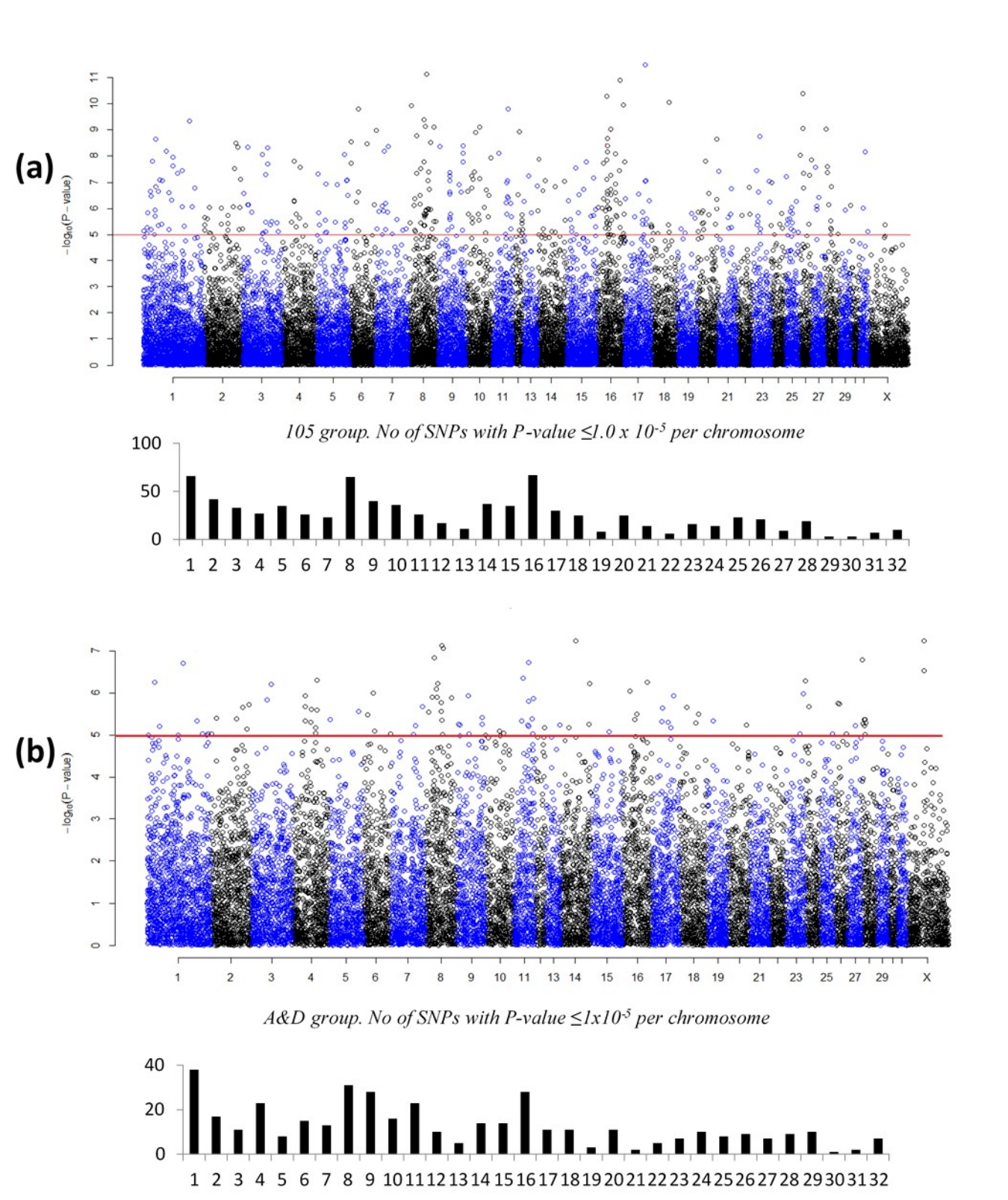
Takeuchi F, Nabika T, Isono M, Katsuya T, Sugiyama T, Yamaguchi S, Kobayashi S, Yamori Y, Ogihara T, Kato N. 2009. Evaluation of genetic loci influencing adult height in the Japanese population. *Journal of human genetics* **54(12):749**.

Tetens J, Widmann P, Kühn C, Thaller G. 2013. A genome-wide association study indicates LCORL/ NCAPG as a candidate locus for withers height in German Warmblood horses. *Animal Genetics* **44:** 467-471.

Visscher PM. 2008. Sizing up human height variation. *Nature Genetics* **40:** 489–490.

Figure 1. **SNP association analysis.**
(a) Upper panel: Results from genome-wide association analysis for genotype-height GWAS in all 105 ponies. The y-axis plots -$\log_{10}$ (P-values) and the x-axis plots the physical position of the SNPs sorted by chromosome number and chromosome position. The strength of association between each SNP and trait is calculated on the basis of the prevalence of each SNP in horses with different heights. The most significant SNP was found on chromosome 17 (BIEC2-417495). **Lower panel:** the number of SNPs ($P \leq 10^{-5}$) per chromosome in the 105 group.
(b) Upper panel: results from genome-wide association analysis for genotype-height GWA in the A&D group of 43 animals; Welsh section A ponies and section D cobs. The most significant ($P < 5.78 \times 10^{-8}$) SNP (BIEC2-1124358) was found on the X chromosome. **Lower plot**: the number of SNPs ($P \leq 10^{-5}$) per chromosome in the A&D group. Chromosome No 32 is the X chromosome

Supporting information
Additional supporting information may be found in the online version of this article.
Figure S1. Plot of scores for multidimensional scaling. The groups of dots identify genetic outliers. The slightly larger cluster (circled) represents Section the A&D group
Figure S2.The full list of SNPs associated with height ($P \leq 1 \times 10^{-5}$) in the A&D group.
Table S3. The full list of SNPs associated with height ($P \leq 1 \times 10^{-5}$) in the A&D group.
Table S4. The list of SNPs associated with height ($P \leq 1 \times 10^{-5}$) present in both groups, i.e. overlapping SNPs.
Table S5. Genomic regions, with highest number of SNPs associated to height ($P \leq 1 \times 10^{-5}$).


(a)
−log10(P−value)
0 1 2 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 21 23 25 27 29 X
105 group. No of SNPs with P-value ≤1.0 x 10-5 per chromosome
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
(b)
−log10(P−value)
0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 21 23 25 27 29 X
A&D group. No of SNPs with P-value ≤1x10-5 per chromosome
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32