ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12296

On the use of the transmission disequilibrium test to detect pseudo-autosomal variants affecting traits with sex-limited expression

Mahmoud Elansary*, Anneleen Stinckens†, Naima Ahariz*, Nadine Cambisano*, Wouter Coppieters*, Eli Grindflek‡, Maren van Son‡, Nadine Buys† and Michel Georges*

*Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Bd de Colonster B43, 4000 Liège, Belgium. †Unit of Livestock Genetics, Division of Gene Technology, Department of Biosystems, KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium. ‡(The Norwegian Pig Breeders Association) NORSVIN, Hamar, Norway.

Summary

We herein describe the realization of a genome-wide association study for scrotal hernia and cryptorchidism in Norwegian and Belgian commercial pig populations. We have used the transmission disequilibrium test to avoid spurious associations due to population stratification. By doing so, we obtained genome-wide significant signals for both diseases with SNPs located in the pseudo-autosomal region in the vicinity of the pseudo-autosomal boundary. By further analyzing these signals, we demonstrate that the observed transmission disequilibria are artifactual. We determine that transmission bias at pseudo-autosomal markers will occur (i) when analyzing traits with sex-limited expression and (ii) when the allelic frequencies at the marker locus differ between X and Y chromosomes. We show that the bias is due to the fact that (i) sires will preferentially transmit the allele enriched on the Y (respectively X) chromosome to affected sons (respectively daughters) and (ii) dams will appear to preferentially transmit the allele enriched on the Y (respectively X) to affected sons (respectively daughters), as offspring inheriting the other allele are more likely to be non-informative. We define the conditions to mitigate these issues, namely by (i) extracting information from maternal meiosis only and (ii) ignoring trios for which sire and dam have the same heterozygous genotype. We show that by applying these rules to scrotal hernia and cryptorchidism, the pseudo-autosomal signals disappear, confirming their spurious nature.

Keywords pig, quantitative trait loci, SNP, transmission disequilibrium test

Introduction

With the availability of SNP arrays covering the entire genome for a growing number of species including domestic animals, genome-wide association studies (GWASs) have become widely used to identify loci influencing developmentally, medically or agronomically important phenotypes. A major concern when performing such a GWAS is to properly account for stratification, which is pervasive in many populations, particularly in domestic animals. A variety of approaches have been proposed to that effect including genomic control, structured association, inclusion of principal components or a random polygenic effect in the model, or 'simultaneous search' (i.e., fitting all SNPs simultaneously in the model using, for instance, Bayesian methods). An alternative approach is to simultaneously test for association and linkage using the so-called transmission disequilibrium test (TDT) (Spielman *et al.* 1993). In its most common implementation, the TDT relies on genotyped sire–dam–offspring trios. When studying a binary trait (such as a disease), the offspring typically expresses the phenotype, whereas its parents may or may not. To test for association and linkage between the trait of interest and a SNP, one identifies trios for which one or both parents are heterozygous (say for alleles '1' and '2') and, within those, determines the ratio of transmissions of the 1 vs. the 2 allele from parents to affected offspring. A significant departure from the expected 50–50% ratio strongly suggests that the preferentially transmitted SNP allele is

Address for correspondence

M. Georges, Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000 Liège, Belgium.
E-mail: michel.georges@ulg.ac.be

Accepted for publication 15 March 2015

associated with a closely linked risk allele. This elegant test is known to effectively protect against population stratification. It is not that commonly implemented, as the identification of trios is sometimes difficult and the TDT is less cost-effective than is regular association (e.g., case–control design) in terms of added information per genotyped individual.

Congenital defects are a major economic and welfare issue in livestock. As several of these are at least in part genetically determined, identifying risk loci by a GWAS may provide useful information to lower their incidence by marker-assisted selection. Cryptorchidism (CR) and scrotal hernia (SH) are two such defects that are common in pigs. Reported incidence of CR and SH ranges from 1% to 1.4% (Knap 1986; Rothschild *et al.* 1988) and 1.7% to 6.7% (Thaller *et al.* 1996) respectively, with corresponding heritability estimates (on underlying liability scale) ranging from 0.06 to 0.33 (Knap 1986) and 0.21 to 0.86 (Mikami & Fredeen 1979; Althoff *et al.* 1988; Thaller *et al.* 1996) respectively.

We herein report the use of the TDT to perform a genome scan for CR and SH in a European commercial pig population. It led to the identification of a genome-wide significant signal immediately adjacent to the pseudo-autosomal boundary (PAB) for both CR and SH. We subsequently demonstrated that these signals are artifactual and result from (i) the sex-limited expression of the study's phenotypes and (ii) the linkage disequilibrium (LD) between the corresponding SNPs and the sex-specific portions of the gonosomes. We propose a set of rules that will properly correct for the corresponding artifact.

Materials and methods

Pedigree material

Belgium

Blood samples were collected from a total of 708 animals corresponding to 167 trios with SH offspring and 295 trios with CR offspring. Animals were sampled from five commercial lines developed from Pietrain, Large White and Landrace stock. Samples were obtained from the nucleus farms of RaSe Genetics. All piglets born in the farms were visually inspected in the farrowing crates, 2 days after birth.

Norway

Blood samples were collected from a total of 453 animals corresponding to 144 trios with SH offspring and 106 trios with CR offspring. Animals were of the Norwegian Landrace breed and collected in 35 different Norwegian nucleus pig farms. Putative SH and CR cases were reported by farmers, and diagnoses confirmed by experienced Norvin breeding consultants following Grindflek *et al.* (2006).

SNP genotyping

Genomic DNA was isolated from venous blood using standard procedures. Genotyping was carried out using the Illumina PorcineSNP60 v2 BeadChip, using procedures recommended by the manufacturer. SNP calling was performed using GENOMESTUDIO version 2011.1 (Illumina). SNPs with a call rate <0.98 and Hardy–Weinberg deviation with a *P*-value $<3 \times 10^{-4}$ were eliminated. Animals with a call rate <0.95 and gonosome genotypes inconsistent with their sex were eliminated. The genotyping and quality control procedures applied to the Norwegian samples were previously described in Grindflek *et al.* (2011). After quality control, 552 of genotyped Belgian animals (i.e., 78%) and 366 genotyped Norwegian animals (i.e., 81%) were retained for further analyses. The number of SNPs retained for further analyses in the different cohorts and when combining cohorts are given in Table S1.

Transmission Disequilibrium Test (TDT)

The TDT was conducted using custom-made scripts. SNPs were analyzed one by one. We identified fully genotyped trios for which at least one of the parents was heterozygous. We then determined how often allele 1 vs. allele 2 was transmitted by heterozygous parents to their affected offspring. For trios in which all individuals are 12, allele 1 and allele 2 are both transmitted once although the parent of origin remains undetermined. We then computed whether the observed transmission ratio departed significantly from expectation using a chi-squared test. To perform analyses across cohorts (whether across diseases or across countries), we either treated the merged cohort as a single population and performed the TDT as described above or we summed the population-specific chi-squared test values with corresponding degrees of freedom to yield an 'across-population' chi-squared statistic and hence a *P*-value. Resulting *P*-values were Bonferroni corrected for the number of analyzed SNPs. Following Lander & Kruglyak (1995), corrected *P*-values ≤0.05 (expected by chance alone once per 20 genome scans) were considered to be genome-wide significant, whereas corrected *P*-values ≤0.63 (expected by chance alone once per genome scan; see also Harmegnies *et al.* 2006) were considered to be genome-wide suggestive.

Haplotype-based TDT

Haplotyping was conducted using BEAGLE software version 3.3.2 (Browning & Browning 2009; http://faculty.washington.edu/browning/beagle/beagle.html) for the four pseudo-autosomal SNPs closest to the PAB (*ALGA0111396, ALGA0109503, ALGA0103468* and *MARC0045707*). When using haplotyped data, the TDT was conducted separately for each haplotype. Thus, we

identified trios for which at least one of the parents was heterozygous for the considered haplotype and then counted the numbers of times that haplotype was transmitted vs. untransmitted to affected offspring. Haplotype-based analyses were also conducted using information from maternal meioses only, that is, counting only the number of times heterozygous dams transmitted the considered haplotype to their affected offspring. In these maternal-only analyses, we either included or excluded ambiguous matings, that is, matings for which sire and dam had the same heterozygous genotype. For each one of these analyses, we computed haplotype-specific *P*-values for the observed transmitted vs. untransmitted ratio using a chi-squared test. To appropriately correct the ensuing *P*-values for the analysis of multiple haplotypes, while accounting for the correlation between haplotype-specific transmission ratios, we used simulations. We randomly sampled two haplotypes for sires and dams based on their real population frequency and then sampled one haplotype at random from each parent to generate the offspring. We then performed the haplotype-based TDTs as described above and stored the lowest *P*-value (obtained for the best haplotype) for each simulation. We performed a total of 1000 simulations. Corrected *P*-values for the real data were then computed as the proportion of simulations with a *P*-value as high or higher than the real *P*-value. We further applied a Bonferroni correction for the realization of three tests, as we analyzed SH and CR separately and then jointly.

Results

A GWAS by means of the regular TDT reveals a genome-wide significant risk locus for both CR and SH in the pseudo-autosomal region

We obtained samples from a total of 311 SH trios and 401 CR trios from breeding companies in Belgium and Norway. Genotyping the corresponding DNAs with the Illumina Porcine SNP60 array followed by stringent quality control resulted in 280 CR and 262 SH trios usable for the GWAS. By applying the TDT, as described in Materials and methods, we obtained a genome-wide significant ($P < 1.2 \times 10^{-6}$) pseudo-autosomal signal in the immediate vicinity of the PAB for both CR and SH. In these analyses, Norwegian and Belgian animals were treated jointly as a single population. At first glance, this finding suggested that the same variant(s) affecting a gene residing on the pseudo-autosomal region (PAR) might confer risk to both CR and SH in the Belgian as well as Norwegian populations. We

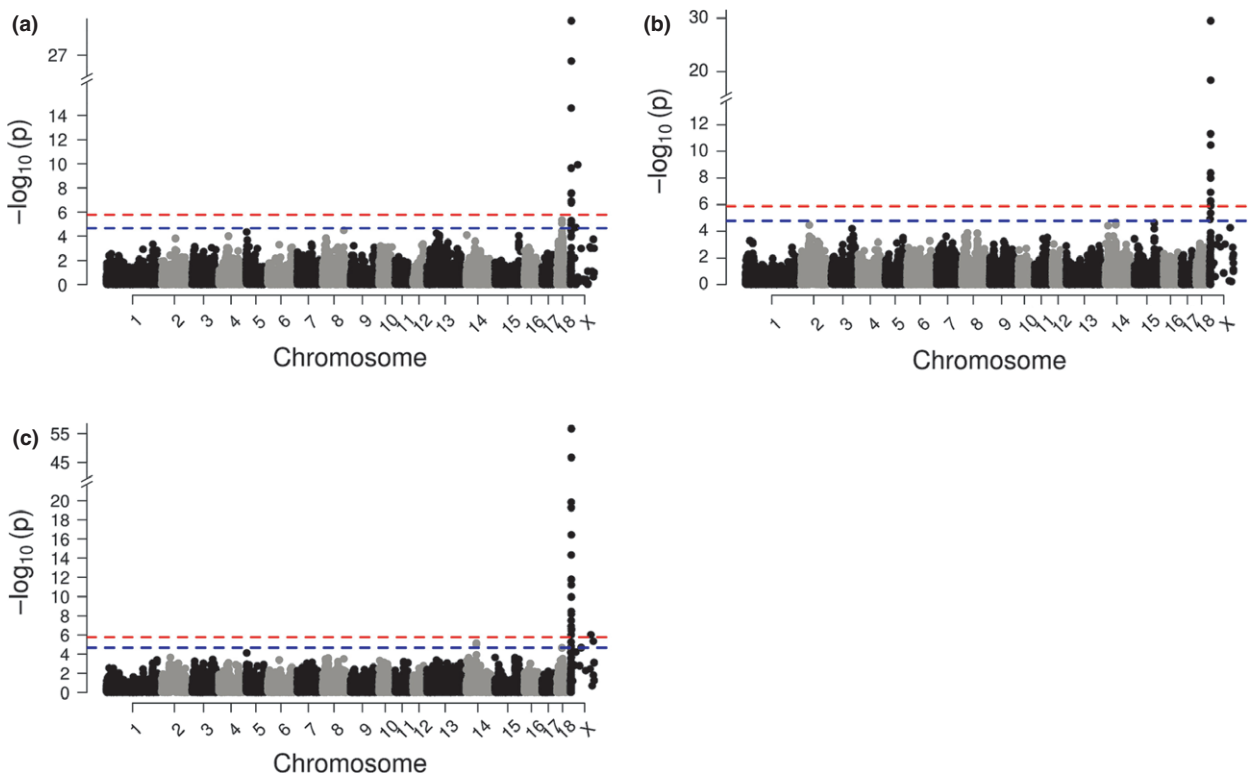

Figure 1 Results of a transmission disequilibrium test-based genome-wide association study extracting information from paternal and maternal meiosis for cryptorchidism (a), scrotal hernia (b) and the two diseases combined by summing the chi-squared tests with corresponding degrees of freedom (c). The genome-wide significant and suggestive thresholds correspond to the red and blue horizontal lines respectively.

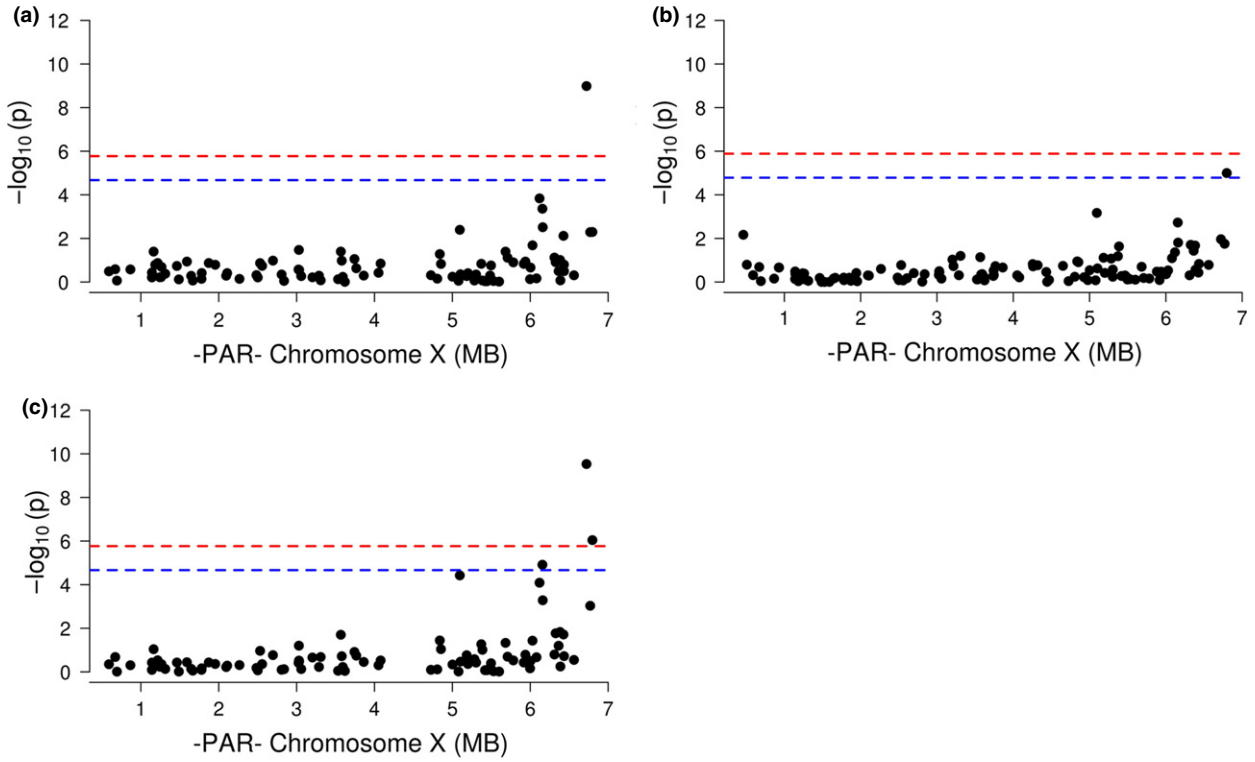

Figure 2 Results of a transmission disequilibrium test-based association study of the porcine pseudo-autosomal region (PAR) extracting information from maternal meiosis only for cryptorchidism (a), scrotal hernia (b) and the two diseases combined by summing the chi-squared tests with corresponding degrees of freedom (c). The genome-wide significant and suggestive thresholds correspond to the red and blue horizontal lines respectively.

therefore performed a joint analysis of all animals together (both diseases and countries), and this increased log $(1/P)$ values to 57 in the PAR.

In addition, we obtained one signal exceeding the genome-wide suggestive threshold for CR (chromosome 18, position 30 341 177 bp, $P = 4.7 \times 10^{-6}$) and one affecting both defects combined (chromosome 14, position 60 757 776 bp, $P = 7.2 \times 10^{-6}$). The latter was driven primarily by SH (chromosome 14, position 60 757 776 bp, $P = 2.2 \times 10^{-5}$) (Fig. 1). Both results were obtained by treating animals from Norway and Belgian as being part of the same population.

The pseudo-autosomal association remains present when restricting the analysis to female meioses, but not when performing haplotype-based analyses

As both CR and SH affect only males, we realized that a spurious TDT signal would be generated if the disease-associated SNPs were also in LD with the gonosome-specific sections of the sex chromosomes, that is, that the allelic frequencies would be different on the Y and X chromosomes. In such cases, the allele that is preferentially associated with the Y chromosome would obviously be preferentially transmitted by fathers to their affected sons. As the disease-associated SNPs map very close to the PAB, this hypothesis seemed very plausible. As a matter of fact, the SNPs that were significantly associated with the diseases were indeed in LD with sex in both the Belgian and Norwegian pig populations (Table S2).

To overcome this obvious issue, we decided to extract information from maternal meioses only: Are mothers preferentially transmitting one allele to their affected offspring? Dams carry two X chromosomes and can, *a priori*, transmit either one to their sons. We reanalyzed the PAR region using this approach, that is, we counted only how often dams transmitted either the 1 or 2 allele and ignored the alleles transmitted by the sires. To our surprise, the same markers yielded genome-wide significant association signals when analyzing CR ($P = 1 \times 10^{-9}$) or both defects jointly ($P = 2.9 \times 10^{-10}$) and a genome-wide suggestive association when analyzing SH ($P = 1 \times 10^{-5}$) (Fig. 2). Again this suggested, at first glance, that variants increasing the risk for both CR and SR mapped in the vicinity of the PAB.

To further the characterization of the corresponding locus, we decided to perform a haplotype-based analysis hoping to identify a haplotype that would be in stronger LD with the putative causative variant than either SNP considered alone, hence increasing the power of the TDT. To that end, we phased the genotype data corresponding to the four pseudo-autosomal SNPs closest to the PAB

 doi: 10.1111/age.12296

Table 1 Haplotype-based TDT. The number of times the corresponding haplotype was transmitted (T) or untransmitted (UnTr) by a heterozygous dam to her affected offspring.

	All matings			Unambiguous matings		
Haplotype	Tr	UnTr	*P*-value	Tr	UnTr	*P*-value
SH						
1313	10	13	0.999	10	13	0.999
1321	38	24	0.827	38	24	0.827
3111	14	14	1	14	14	1
3113	7	0	0.206	7	0	0.206
3121	44	46	1	44	46	1
3313	27	29	1	27	29	1
3321	32	48	0.827	32	48	0.827
CR						
1313	19	20	1	18	20	1
1321	30	35	0.999	30	33	1
3111	24	16	0.990	24	16	0.993
3113	7	3	0.990	5	3	0.999
3121	66	53	0.991	66	53	0.994
3313	21	40	0.299	21	40	0.198
3321	38	38	1	38	38	1
ALL						
1313	29	33	1	28	33	0.999
1321	68	59	0.999	68	57	0.999
3111	38	30	0.999	38	30	0.999
3113	14	3	0.123	12	3	0.355
3121	110	99	0.999	110	99	0.999
3313	48	69	0.646	48	69	0.606
3321	70	86	0.987	70	86	0.993

All matings consider all dams; unambiguous matings consider only matings for which sires and dams have distinct diplotypes. Haplotype: haplotypes for *ALGA0111396*, *ALGA0109503*, *ALGA0103468* and *MARC0045707* in that order: 1 = T, 2 = G and 3 = C. SH, scrotal hernia; CR, cryptorchidism. ALL, both diseases combined. *P*-values were computed empirically by simulation and then Bonferroni corrected for the realization of three tests.

(*ALGA0111396*, *ALGA0109503*, *ALGA0103468* and *MARC0045707*; spanning 237 kb and showing the strongest association with CR and SH) using BEAGLE. Seven of the 16 possible haplotypes accounted for >95% of the chromosomes in the dataset. These were subjected to a TDT using maternal meioses only. Contrary to expectations, none of transmission ratios departed significantly from Mendelian 1:1 proportion when accounting for the realization of multiple tests (Table 1).

Bias of the TDT when analyzing association between pseudo-autosomal markers and traits with sex-limited expression

These findings made us suspect that even the signal obtained on the basis of maternal meioses only (i.e., when analyzing SNPs one by one rather than when using haplotypes) might be artifactual. To address this, we modeled the outcome of a TDT for pseudo-autosomal markers located at θ recombination units from the PAB, for which one allele (say, allele 1) was observed at frequency p_x on X chromosomes and $p_y = p_x + \delta(1-p_x)$ on Y chromosomes. The SNP was assumed not to have any effect (i.e., we modeled the null hypothesis) on a trait with sex-limited expression. Thus, the actual transmission ratio from heterozygous (1x: where x is any of the other alleles) parents was 1:1.

When considering paternal and maternal meioses, the TDT generated 'artifactual' transmission distortion, as expected (Fig. 3a). The allele associated with the Y chromosome was overtransmitted to affected sons (trait with male-limited expression) or undertransmitted to affected daughters (trait with female-limited expression). The transmission distortion increased with increasing LD, with decreasing distance from the PAB and with decreasing frequency of the Y-associated allele. It was (nearly) equally severe in both unambiguous (i.e., different parental genotypes) and ambiguous (i.e., same parental heterozygous genotypes, shared with the non-informative half of the offspring) matings. It was not improved when using multi-allelic (>2) marker systems (Fig. S1A).

Against our initial expectation, the TDT also generated artifactual transmission distortion when extracting information from maternal meioses only (Fig. 3b). As above, the allele associated with the Y chromosome was overtransmitted to affected sons or undertransmitted to affected daughters. The effect of LD, distance from the PAB and frequency of the Y-associated allele was also the same as above. However, the distortion was observed only in ambiguous matings (i.e., same parental heterozygous genotypes, shared with the non-informative half of the offspring) and not in unambiguous matings (i.e., different parental genotypes). The transmission distortion is due to the fact that offspring inheriting the allele that is not associated with the Y from a mother that carries the Y-associated allele have a higher probability to be non-informative than those inheriting the Y-associated allele. The degree of transmission distortion was reduced when using multi-allelic (>2) markers (Fig. S1B).

Adapting the TDT to search for associations between pseudo-autosomal markers and traits with sex-limited expression

The analysis described above indicates that one should not apply the regular TDT (extracting information from paternal and maternal meiosis) to test for association between a sex-limited trait and pseudo-autosomal markers that are in LD with sex. However, the same analysis indicates that the TDT can be applied in this scenario, as long as (i) one extracts information only from maternal meioses and (ii) one considers only non-ambiguous trios in the analysis.

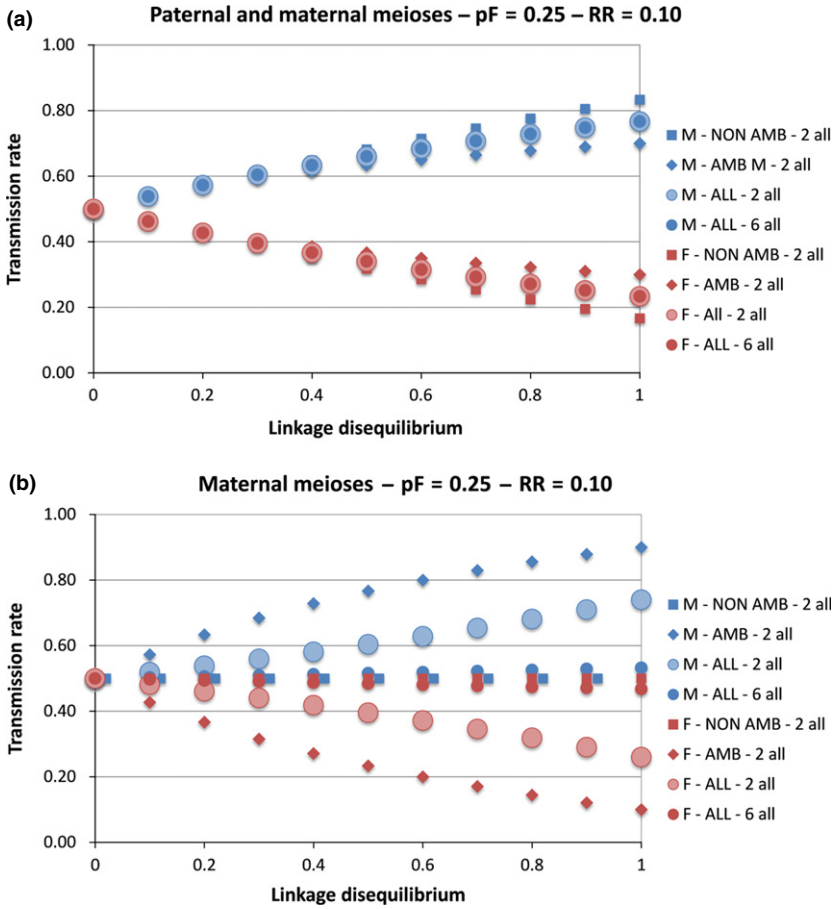

Figure 3 Expected transmission rates (*Y*-axis) for a Y-associated marker allele from heterozygous parents to affected sons (male-limited trait, in blue) or affected daughters (female-limited trait, in red) as a function of the linkage disequilibrium between the tested allele and the Y chromosome (*X*-axis), the frequency of the tested allele in females (pF = 0.25) and the recombination rate between the marker and the pseudo-autosomal boundary (RR = 0.1). (a) transmission disequilibrium test (TDT) extracting information from paternal and maternal meioses. (b) TDT extracting information from maternal meioses only. M – NON AMB – 2 all: male-limited trait, non-ambiguous matings (i.e., different parental genotypes), biallelic marker; M – AMB – 2 all: male-limited trait, ambiguous matings (i.e., same parental genotypes), biallelic marker; M – ALL – 2 all: male-limited trait, all matings, biallelic marker; M – ALL – 6 all: male-limited trait, all matings, multi-allelic (6) marker; F – NON AMB – 2 all: female-limited trait, non-ambiguous matings, biallelic marker; F – AMB – 2 all: female-limited trait, ambiguous matings, biallelic marker; F – ALL – 2 all: female-limited trait, all matings, biallelic marker; F – ALL – 6 all: female-limited trait, all matings, multi-allelic (6) marker.

Under these conditions, the transmission ratio expected under the null hypothesis is the Mendelian 1:1 ratio. However, if the LD between the SNP and sex is high, the proportion of unambiguous yet informative matings may be low, and this will considerably reduce the power to detect an association if it exists. One approach to overcome this limitation is to use multi-allelic (>2) marker systems. This can be performed, for instance, using microsatellite markers, or more conveniently given the widespread use of SNP genotyping arrays, using haplotype information.

Accordingly, we reanalyzed the available haplotype data for CR and SH, yet restricted ourselves to unambiguous matings. None of the haplotypes exhibited a significant transmission distortion when applying these rules (Table 1). This allowed us to conclude that the associations that were initially observed between the pseudo-autosomal SNPs and CR and SH were spurious and due to the bias introduced by the TDT as initially applied.

Discussion

We herein describe a previously unrecognized issue when applying the TDT to detect association between pseudo-autosomal markers and sex-limited traits. As their name implies, pseudo-autosomal markers are typically handled as autosomal ones in association studies. The problem will arise in the case of LD between the corresponding markers and the sex-specific segments of the gonosomes, that is, if the allelic frequencies differ between X and Y chromosomes. This is most likely for markers that are located close to the PAB. We show that the problem can be overcome by

 doi: 10.1111/age.12296

applying simple rules, namely (i) limiting the analysis to maternal meioses and (ii) ignoring ambiguous matings (i.e., matings for which sire and dam have the same heterozygous genotype). LD between pseudo-autosomal markers and sex would also generate problems in regular association studies for traits with sex-limited expression if the controls would include individuals from both sexes. Thus, cases and controls need to be matched for sex when performing regular association analyses with pseudo-autosomal markers.

Our analysis led us to the conclusion that the TDT signals detected in the PAR were spurious. We detected no other genome-wide significant signals anywhere in the genome. A suggestive signal was observed for CR on chromosome 18 (30 341 177 bp). The orthologous region in the human genome is chr7: 117 210 649–118 210 649. To the best of our knowledge, this region is not known to contain any gene previously implicated in cryptorchidism. Cryptorchidism has been associated with mutations in *INSL3* and *LGR8* in mice and humans (respective positions: chr19: 17 816 513 bp and chr13: 31 739 542 bp) (Ferlin *et al.* 2003). Another suggestive signal was obtained on chromosome 14 (60 757 776 bp) when analyzing CR and SH jointly. It was driven primarily by associations with SH. The corresponding orthologous region in the human genome is chr1: 233 891 313–234 891 313. It has not been implicated in SH before. Neither the chromosome 14 nor the chromosome 18 regions correspond to regions that were listed as most promising in a previously reported microsatellite linkage scan of the complete porcine genome (Grindflek *et al.* 2006). Note that the number of analyzed animals and the marker density were higher in the present study. Additionally, these regions do not coincide with the *HOXA10*, *ZFPM2* or *MMP2* genes previously associated with SH (Zhao *et al.* 2009).

The chromosome 14 and chromosome 18 associations reported above were obtained by considering Norwegian and Belgian animals as belonging to one population. Analyses were also performed separately on Belgian and Norwegian samples or by summing chi-squared values obtained within country. This did not reveal more significant signals than the ones reported in Fig. 1 (data not shown). Performing a joint analysis for CR and SH was first conducted as a result of the coincident association signals observed on the PAR, which suggested that both diseases might involve common genetic determinants. This was not devoid of biological meaning. As mentioned by Grindflek *et al.* (2006), the genitofemoral nerve, the caudal ligament gubernaculum and controllers of testicular descent may influence closure of the processus vaginalis. Moreover, a genetic correlation of 0.2 has been previously reported between both diseases (Mikami & Fredeen 1979).

In conclusion, our study highlights a previously unrecognized issue related to the use of the TDT with pseudo-autosomal markers. Moreover, it does not reveal obvious genome-wide significant association signals for either CR or SH, despite the use of relatively large cohorts. This indicates that, although heritable, the genetic architecture of these disorders is likely to be polygenic without the involvement of major gene effects.

Acknowledgements

This work was funded by the European FP7 PIGENDEF project. We would like to thank Norsvin and the Research Council of Norway for funding the Norwegian part of the data collection, the breeding technicians in Norsvin for collecting blood samples and recording data from the nucleus herds in Norway and Hanne Hamland in Norsvin for handling samples and genotyping. We acknowledge RaSe Genetics for funding the Belgian part of the data collection, collecting blood samples and recording data from the nucleus herds in Belgium.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AS, NB, and EG collected and organized sampling; EG, MVS, WC, NA and NC generated SNP genotypes and performed quality control; NB and MG designed experiments; ME and MG analyzed data; and ME, AS, EG, NB and MG wrote the manuscript.

References

Althoff W., Mayer M., Richter L. & Simon D. (1988) Zur erblichen Abhängigkeit der Geburtsfehler Brüche und Binnenhodigkeit beim Schwein. *Züchtungskunde* **60**, 319–29.

Browning B.L. & Browning S.R. (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics* **84**, 210–23.

Ferlin A., Simonato M., Bartoloni L., Rizzo G., Bettella A., Dottorini T., Dallapiccola B. & Foresta C. (2003) The INSL3-LGR8/GREAT ligand-receptor pair in human cryptorchidism. *The Journal of Clinical Endocrinology and Metabolism* **88**, 4273–9.

Grindflek E., Moe M., Taubert H., Simianer H., Lien S. & Moen T. (2006) Genome-wide linkage analysis of inguinal hernia in pigs using affected sib pairs. *BMC Genetics* **7**, 25.

Grindflek E., Lien S., Hamland H., Hansen M.H., Kent M., van S.M. & Meuwissen T.H. (2011) Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. *BMC genomics* **12**, 362.

Harmegnies N., Davin F., De Smet S., Buys N., Georges M. & Coppieters W. (2006) Results of a whole-genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross. *Animal Genetics* **37**, 543–53.

Knap P.W. (1986) Congenital defects inheritance of AI boars: genetic parameters and breeding value estimation procedures. *Livestock Production Science* **15**, 337–52.

 doi: 10.1111/age.12296

Lander E. & Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–7.

Mikami H. & Fredeen H.T. (1979) A genetic study of cryptorchidism and scrotal hernia in pigs. *Canadian Journal of Genetics and Cytology* **21**, 9–19.

Rothschild M.F., Christian L.L. & Blanchard W. (1988) Evidence for multigene control of cryptorchidism in swine. *Journal of Heredity* **79**, 313–4.

Spielman R.S., McGinnis R.E. & Ewens W.J. (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *The American Journal of Human Genetics* **52**, 506–16.

Thaller G., Dempfle L. & Hoeschele I. (1996) Maximum likelihood analysis of rare binary traits under different modes of inheritance. *Genetics* **143**, 1819–29.

Zhao X., Du Z.Q., Vukasinovic N., Rodriguez F., Clutter A.C. & Rothschild M.F. (2009) Association of *HOXA10, ZFPM2,* and *MMP2* genes with scrotal hernias evaluated via biological candidate gene analyses in pigs. *American Journal of Veterinary Research* **70**, 1006–12.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Expected transmission rates (*Y*-axis) for a Y-associated marker allele from heterozygous parents to affected sons (male-limited trait; in blue) or affected daughters (female-limited trait; in red), as a function of the linkage disequilibrium between the allele and the Y chromosome (*X*-axis), the frequency of the Y-associated allele (0.25 – 0.5 – 0.75; rows), and the recombination rate between the marker and the pseudo-autosomal boundary (0.0 – 0.1; columns).

Table S1 Number of quality controlled SNPs retained for analyses performed either in specific cohorts or by combining multiple cohorts across countries and/or disease.

Table S2 Linkage disequilibrium between four pseudo-autosomal SNPs mapping closest to the PAB and sex in the parents of the CR and SH trios.