

VU Research Portal

Addressing reproductive risk in consanguineous couples

Teeuw, M.E.

2015

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Teeuw, M. E. (2015). *Addressing reproductive risk in consanguineous couples*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

CHAPTER 3

Genomic kinship coefficient does not seem helpful in distinguishing between consanguineous parents with a high and with a low risk for adverse reproductive outcome

Kelmemi W/ Teeuw ME, Bochdanovits Z, Ouburg S, Jonker MA, Alkuraya F, Hashem M, Kayserili H, Van Haeringen A, Sheridan E, Masri A, Al Ammouril A, Cobben JM, Rizzu P, Kostense PJ, Dommering CJ, Henneman L, Chaabouni H, Heutink P, Ten Kate LP, Cornel MC.
Submitted

ABSTRACT

The offspring of consanguineous couples are at increased risk of congenital disorders. The risk increases as parents are more closely related. Individuals that have the same degree of relatedness according to their pedigree, show variable genomic kinship coefficients. To investigate whether we can differentiate between couples with high- and low risk for offspring with congenital disorders, we have compared the genomic kinship coefficient of consanguineous parents with a child affected with an autosomal recessive disorder with that of consanguineous parents with only healthy children, corrected for the degree of pedigree relatedness.

151 consanguineous couples (73 cases and 78 controls) from 10 different ethnic backgrounds were genotyped on the Affymetrix platform and passed quality control checks. After pruning SNPs in linkage disequilibrium, 57,358 SNPs remained. Kinship coefficients were calculated using three different toolsets: PLINK, King and IBDelphi, yielding five different estimates (IBDelphi, PLINK (all), PLINK (by population), King robust (all) and King homo (by population)). We performed a one-sided Mann Whitney test to investigate whether the median relative difference regarding observed and expected kinship coefficients is bigger for cases than for controls. Furthermore, we fitted a mixed effects linear model to correct for a possible population effect.

Although the estimated degrees of genomic relatedness with the different toolsets show substantial variability, correlation measures between the different estimators demonstrated moderate to strong correlations. Controls have higher point estimates for genomic kinship coefficients. The one-sided Mann Whitney test did not show any evidence for a higher median relative difference for cases compared to controls. Neither did the regression analysis exhibit a positive association between case-control status and genomic kinship coefficient.

In this case-control setting, in which we compared consanguineous couples corrected for degree of pedigree relatedness, a higher degree of genomic relatedness was not significantly associated with a higher likelihood of having an affected child. Further translational research should focus on which parts of the genome and which pathogenic mutations couples are sharing. Looking at relatedness coefficients by determining genome-wide SNPs does not seem to be an effective measure for prospective risk assessment in consanguineous parents.

INTRODUCTION

Consanguineous marriages – unions between relatives up to the fifth degree – occur in many different parts of the world and are the preferred type of marriage within several populations due to cultural and socioeconomic advantages (Hamamy et al., 2011). Consanguinity is associated with an increase in congenital/hereditary disorders in offspring, in particular autosomal recessive (AR) disorders. An estimated excess of 1.7-2.8% in morbidity and 3.5% in mortality has been observed among the offspring of first cousins compared to the offspring of non-consanguineous parents (Bennett et al., 2002). The risk is dependent on the inbreeding coefficient and the presence of a pathogenic mutation passed on by a common ancestor. Of all first-cousin couples, a proportion of less than four times the excess in risk (as consanguineous parents can be carrier for more than one AR disorder) are estimated to be a carrier couple for an AR disorder because of their consanguinity. Consequently, there is a chance of 25% (or more if they are a carrier couple for more than one AR disorder) that each of their children will be affected by the associated disorder (Teeuw et al., 2010; Hamamy et al., 2011). The risk of being a carrier couple rises substantially with evidence of a family history for a given genetic disorder (Bennett et al., 2002), but often no family history is known and yet the risk is increased. This makes personalized risk assessment difficult, and discriminating between high-risk (25% or more) and low-risk couples (comparable to the risk for the general population) often impossible.

In order to provide the best possible risk information and genetic counselling for consanguineous couples before pregnancy, health care providers generally aim to estimate the total risk by both calculating the inbreeding coefficient based on the pedigree and by taking a family history. As described earlier, calculating the exact inbreeding coefficient is difficult, due to complex undocumented genealogies and the fact that DNA identical-by-descent (IBD) between consanguineous partners is subject to variation because of stochastic events during meiosis (Teeuw et al., 2010).

In theory, an increase in the proportion of DNA IBD sharing will increase the chance of the presence of pathogenic alleles in both parents and consequently the risk of having children affected by autosomal recessive disorders. In the present study, whose methodology was published earlier (Teeuw et al., 2010), the following hypothesis was tested: in a case-control setting, consanguineous parents with a child affected by an autosomal recessive disorder and without a family history of AR disorder have more DNA IBD than similarly related consanguineous parents with only healthy children. If the hypothesis can be confirmed, knowing the genomic proportion of DNA IBD may be helpful to assess risk more precisely in consanguineous couples.

METHODS

Participants

The aim was to include 100 consanguineous couples (cases) with one or more children affected by an autosomal recessive disorder and 100 consanguineous couples (controls) with a family relation comparable to the cases and with only healthy children (at least three). Information on the identity of all first-degree to third-degree family members of both partners of the couples was obtained as far as this was known to the couple. Cases were excluded if another individual in the family affected by the same disorder was known. Further details on the methods of ascertainment have been described earlier (Teeuw et al., 2010). Considered for inclusion in our study were 263 consanguineous couples. For 168 of these couples (82 case and 86 control couples), reliable pedigree information and blood, saliva or DNA samples suitable for whole genome analysis (or genotypic information) were available. It was attempted to have equal numbers of case and control couples of every ethnic background. Moreover, similar distributions of pedigree relatedness among both the case couples and the control couples were aimed for. There were 10 cases where we had a case and control couple from the same family. The 168 couples originated from 10 different populations (Tunisia, Saudi Arabia, Turkey, Jordan, Morocco, Pakistan, Iraq, Iran, Afghanistan and the Netherlands).

Sample preparation, genotyping and quality control

For the couples other than the Saudi Arabian couples, DNA was extracted from the individual's saliva samples (15 case couples and 72 control couples) according to standard procedures; for 54 case couples, DNA extracted from blood was available. Three case couples and two control couples were excluded due to insufficient DNA quality of one of the partners. A total of 66 and 70 non-Saudi case and control couples respectively were genotyped according to manufacturer's protocol using Affymetrix 6.0K SNP arrays. The Saudi Arabian couples (13 case couples and 14 control couples) were subjected to genotyping using Affymetrix 250K arrays (6 individuals) and Affymetrix Axiom arrays (48 individuals), after DNA extraction from whole blood (See Table S1). Genotypes of 79 case and 84 control couples in total were available.

PLINK was used to perform post-genotyping quality control (Purcell et al., 2007). Individuals with a genotyping rate of <95% were excluded. All genotype data were merged for overall analysis; Tunisian, Saudi, and Turkish couples were also merged into separate files for further analysis for each population. Duplicated SNPs were removed and only SNPs were included with a genotype call rate of >95%. Further quality control included removal of SNPs with a minor allele frequency of less than 5%. After quality control, 12 couples were excluded and 302 individuals (73 case couples and 78 control couples) genotyped for 143,512 markers were available for the analysis (see Table 1).

To obtain a pruned subset of markers with low linkage disequilibrium, the PLINK-indep-pairwise option was used with parameters `50 5 1.5`. (57,358 SNPs remaining). Multidimensional scaling was performed to analyse the population substructuring by using the PLINK MDS plot option. Results were entered in the statistical package R version 3.0.1 (<http://www.r-project.org/>) and SPSS version 20 for Windows (IBM Corp., 2011). The MDS plot was inspected for population clustering and case/control matching.

Analysis of pairwise relatedness

A kinship coefficient based on the pedigree reported for each couple was calculated according to the method as described by Wright (1922).

Although pairwise coefficients of relatedness in genomic data can be calculated based on known allele frequencies in a population, these allele frequencies are frequently not known and often calculated from the sample by estimators of relatedness (Wang, 2002). Lack of homogeneity of the sample or sampling errors can lead to false estimates (Wang, 2002; Manichaikul et al., 2010). Since in some populations individuals are genetically more similar than in other countries, to estimate the genomic pairwise relatedness of our sample we used three different estimators (PLINK, King, IBDelphi) to account for population stratification as well as for the inbreeding in our sample. Moreover, we estimated the relationship coefficients from the whole set of samples (overall analysis), as well as from separate sets containing only couples from one population (population subgroup analysis). This latter analysis was performed only for the Tunisian, Saudi and Turkish couples, given the small sample sizes from the other populations.

PLINK uses a method-of-moments approach where the probability of sharing 0, 1 or 2 SNPs IBD is calculated. The total proportion of SNPs IBD is calculated based on the estimated allele frequency of all SNPs and assumes homogeneity (Purcell et al., 2007). King uses the same approach, and offers two different methods: King homo, which assumes homogeneity of the sample, and King robust, which provides robust relationship inference allowing for heterogeneity of the sample by a *robust* approach that accounts for population stratification (Manichaikul et al., 2010). Finally, IBDelphi is an algorithm that analyses raw data of high-density SNP genotypes from a consanguineous couple by looking for homozygous regions of over 0.5 Mb in both genomes that lack SNPs that exclude IBD (Carr et al., 2011).

In PLINK, pairwise relatedness between partners of each couple was calculated with the `--genome --rel-check` command in PLINK. In King, the pruned subset of SNPs was used to calculate pairwise IBD through the kinship parameter (for the overall analysis) and homo parameter (for the population subgroup analysis). Finally, individual genotype files were entered pairwise in IBDelphi, producing IBD measures.

All estimates of pairwise relatedness (pedigree, PLINK, King and IBDelphi) were entered in the statistical package R and SPSS version 20. Pearson's correlation coefficients were calculated for correlations between the different estimates. R_{gen} represents the relatedness as derived from the genotype, while R_{ped} was calculated based on the pedigree information reported.

The ratio $R = (R_{gen} - R_{ped}) / R_{ped}$ was used as a measure of the degree of similarity between R_{gen} and R_{ped} , with R_{gen} being the observed measure of pairwise relatedness (resulting from our analyses by the four different approaches) and R_{ped} the kinship coefficient between the parents of a child based on the pedigree. If, for a couple, R_{ped} is higher (lower) than R_{gen} , R is negative (positive). By dividing the difference by R_{ped} , we consider the relative differences. The possible influence of population was ignored first, and the alternative hypothesis was tested that the median of the distribution of ratio R of cases (couples with affected children) is higher than the median of the distribution of controls (couples with only healthy children) with the one-sided non-parametric Mann Whitney test at level 0.05. Since most (96 of the 151 couples) couples come from Tunisia, they were subsequently selected to filter out a possible population effect, and the same test was performed based on these selected data. Next, a mixed effects linear model was assumed. The outcome variable in the model is equal to ratio R , the covariates consist of an intercept, the fixed effect 0-1 variable "whether the couple has an affected child (covariate equals 1) or not (covariate equals 0)" (i.e. case or control), and a random effect "population". The population effect on the association between the outcome variable R and the case-control status was investigated and the one-sided alternative hypothesis was tested regarding whether the regression parameter for case-control status was positive, corrected for a possible population effect if "population" is a confounder.

RESULTS

Seventy-three case and 78 control couples from 10 different populations were analysed (Table 1; Table S1). Sixty-five of the 73 case couples had a child with an autosomal recessive disorder that was diagnosed by molecular testing, two case couples had a child that was diagnosed by biochemical testing, and the remaining six case couples had a child with a clinical diagnosis for an AR disorder. The respective AR disorders are presented as Supplementary Information (Table S2). Control couples had on average 4.2 children ($SD=1.6$), ranging from three to 11.

Table 1. Case and control couples included for analysis

Country	Case	Control
Tunisia	47	49
Saudi Arabia	11	12
Turkey	6	10
Morocco	3	1
Jordan	2	3
Pakistan	1	1
The Netherlands	1	1
Afghanistan	1	0
Iraq	0	1
Iran	1	0
Total	73	78

Substructuring in the populations

In Figure 1 two components resulting from within-group multidimensional scaling analysis of all individuals are plotted against each other in a scatter plot. As expected, evidence of substructuring into the different populations is present, with each population distinguishable from the others. Saudi individuals are observed to cluster separately from the other populations in the lower right corner of the figure as a result of using a different marker set for the genotyping. To explore comparability of cases and controls, they were plotted separately (Figure 2). This shows consistent clustering, with an apparent overlap of cases and controls.

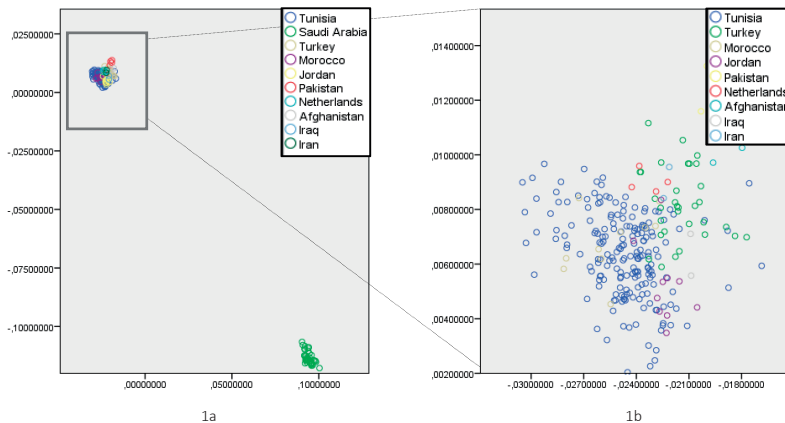


Figure 1a+1b. 1a shows an MDS two-dimensional plot of all individuals in the sample, showing separate clustering of Saudi Arabian individuals (lower right corner). 1b is an enhancement of upper left corner of Figure 1a.

Relationship inference estimates

The estimated degrees of relatedness show a substantial variability between the different estimators (Table 2). PLINK (all) kinship coefficients estimates resulted in a mean for case couples of 0.151 (SD= 0.06), the mean for control couples was 0.164 (SD= 0.05). King robust (all) estimates for all samples are lowest (cases: mean = 0.039 (SD= 0.35) vs. controls: mean = 0.050 (SD=0.031)). In contrast to our hypothesis, controls show higher point estimates for degree of genomic relatedness. Despite the differences in level of the estimate, correlation measures between the different estimators show moderate to strong correlations. Pearson's correlation coefficients were greater or equal to $r = 0.739$, $p < 0.001$ (Table 3).

Genomic estimates compared to pedigree-based estimates

The results of the genomic estimates of the estimators were plotted against the relatedness coefficient based on the pedigree (Figure 3). As expected, higher pedigree-based estimates, on average, correlate with higher genomic estimates. In a proportion of couples, genomic relatedness is much increased compared to the expected relatedness based on the pedigree, resulting in a skewed distribution of relatedness towards higher values, which can be explained by hidden consanguineous loops.

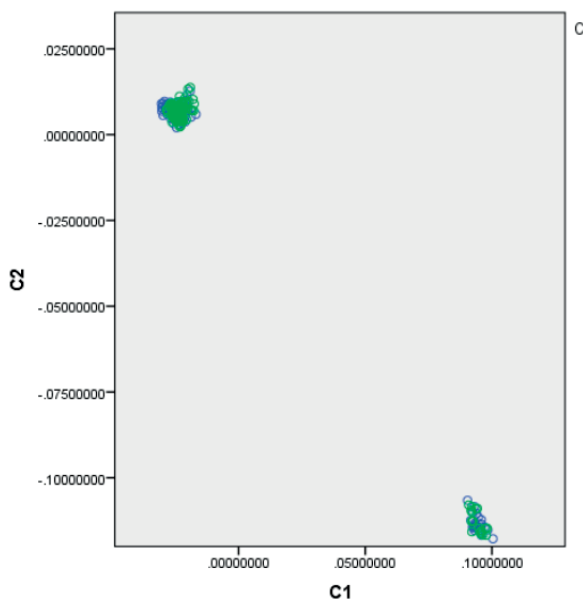


Figure 2. MDS two-dimensional plot of all individuals in the sample. Cases indicated with green dots, controls with blue dots.

Table 2. Mean and standard deviation of pedigree-based relatedness coefficient and relatedness inference estimates for different estimators

	Mean (standard deviation)	
	Case	Control
Pedigree	0.110 (0.058)	0.108 (0.049)
PLINK (all)	0.151 (0.064)	0.164 (0.054)
King robust (all)	0.039 (0.035)	0.050 (0.031)
PLINK (by population)	0.126 (0.063)	0.141 (0.051)
King homo. (by population)	0.054 (0.034)	0.062 (0.030)
IBDelphi	0.121 (0.059)	0.133 (0.053)

all = whole sample; by population = separate estimates for Tunisian, Saudi and Turkish couples

Table 3. Pearson's correlation coefficient between estimators

	IBDelphi	PLINK (all)	King robust (all)	PLINK (by pop.)	King homo (by pop.)
PLINK (all)	.947	1	.880	.963	.962
King robust (all)	.739	.880	1	.814	.831
PLINK (by pop.)	.926	.963	.814	1	.964
King homo (by pop.)	.919	.962	.831	.964	1
IBDelphi	1	.947	.739	.926	.919

Comparing estimates of IBD in cases to IBD in controls

The PLINK (all) mean and median values of the genetic difference ratio R in the sample of couples with affected children are equal to 0.276 and 0.313, respectively, whereas these values unexpectedly increase to 0.347 and 0.338 in the sample of couples with unaffected children. Our hypothesis was thus not confirmed. The considerable difference between the means and the medians are caused by a few couples with high R-values. For 10 case couples, a control couple from the same family was available. The association between their R-values was computed and considered by making scatter plots. Since no association was seen, the R-values of case couples and control couples within a family are assumed to be independent in the further analysis. The one-sided Mann Whitney test based on the whole data set gave a p-value of 0.907. The other three methods for estimating DNA IBD yielded similar p-values (Table 4). Based on this test, it cannot be concluded that couples with affected children are genetically more similar than expected, based on their kinship relatedness, than couples without affected

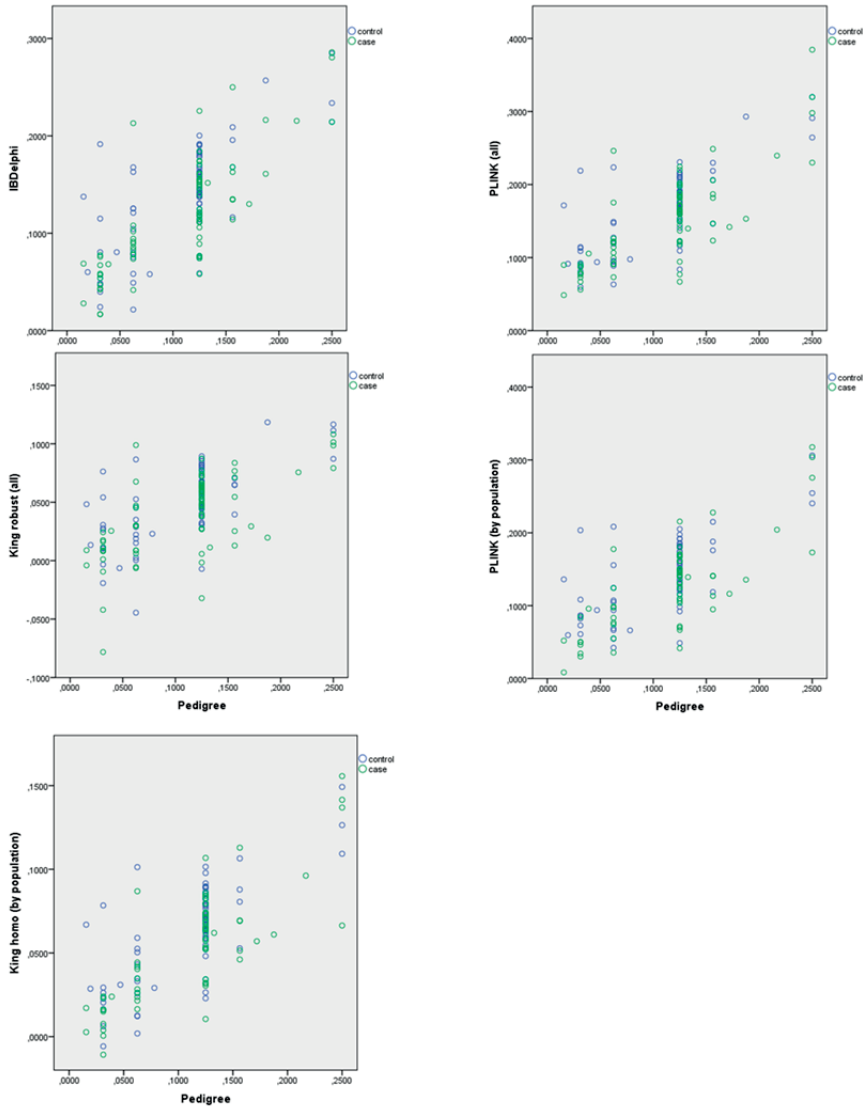


Figure 3. Pedigree kinship coefficient plotted against kinship coefficient estimates of different estimators. Cases indicated with green dots, controls with blue dots.

children. Next, the parameters in the mixed effects linear model were estimated. In all models, “population” seems not to be a confounder and was left out of the model (the estimated standard deviation was very small compared to the estimated standard deviation for the measurement error in the model). Also including the covariate “Rgen” into the linear model gave a better fit, but the association between the variable “case-control status” and the outcome variable hardly changed. This was expected due to the

design of the study. It was tested whether the regression parameter for the covariate case-control status was significantly bigger than zero in the linear model. In the latter analysis, several influence points were left out. With PLINK (all) measures, the regression parameter was estimated as -0.028 and the alternative hypothesis that the regression parameter is positive was not rejected.

Estimates and p-values for the other methods can be found in Table 4. The conclusion still holds true: it cannot be concluded that couples with affected children have more DNA IBD than couples with only healthy children.

Table 4. Results of Mann Whitney test and mixed effects linear model for ratio R

	Mean / Median		Mann Whitney test	Mixed effects model	
	case	control	p-value	β	p-value
PLINK (all)	0.276 / 0.313	0.347 / 0.338	0.907	-0.028	0.625
King robust (all)	-0.615 / -0.595	-0.462 / -0.534	0.990	-0.152	0.992
PLINK (by population)	0.209 / 0.145	0.594 / 0.233	0.991	-0.157	0.965
King homo (by population)	-0.506 / -0.490	-0.341 / -0.428	0.996	-0.086	0.980
IBDelphi	0.278 / 0.192	0.479 / 0.252	0.854	-0.085	0.834

Another way of accounting for a possible population effect is by selecting individuals from a single population and performing the Mann Whitney test based on this subset only. For the Tunisian data only (96 couples), PLINK (by population), King homo (by population) and IBDelphi, a Mann Whitney test was performed separately. For PLINK (by population), the mean and the median of ratio R equal respectively 0.132 and 0.073 for the couples with an affected child and 0.510 and 0.182 for those without an affected child. The couples without an affected child show, on average, a higher genomic kinship relatedness than expected compared to the couples with an affected child. The p-value of the one-sided Mann Whitney test now equals 0.990. For the results of the other estimators, see Table 5. So, by restricting the analysis to Tunisian couples, it cannot be concluded that couples from Tunisia with an affected child have more DNA IBD than expected based on their kinship relatedness than couples from Tunisia without affected children. The values ratio R mean and median rather show the reverse of what was expected. The three other methods for estimating IBD all lead to the same conclusion.

Table 5. Results of Mann Whitney test Tunisian subpopulation for ratio R*

	Mean / Median		Mann Whitney test
	case	control	p-value
PLINK (by population)	0.132 / 0.073	0.510 / 0.182	0.990
King homo (by population)	-0.529 / -0.520	-0.372 / -0.440	0.982
IBDelphi	0.275 / 0.165	0.426 / 0.165	0.815

* Only measures by population are presented given the relative homogeneity of the sample

DISCUSSION

The hypothesis that consanguineous couples with an affected child have, on average, a higher amount of DNA IBD than consanguineous couples with only healthy children was not confirmed by this study. For clinical purposes determining the amount of DNA IBD can, therefore, not differentiate between consanguineous couples at high risk vs. those at low risk. Results from PLINK and IBDelphi show that both case couples and control couples share on average more DNA IBD than expected based on the pedigree. This is a phenomenon more often found especially in populations with a tradition of consanguinity, where the proportion of DNA IBD is larger due to hidden consanguineous loops in the extended family (Woods et al., 2006). Moreover, extended tracts of homozygosity have been found more generally in populations not even associated with recent consanguinity (Stevens et al., 2012; Curtis and Vine, 2010; Kirin et al., 2010; McQuillan et al., 2008).

In this study we have ascertained cases and controls from a population at increased risk, namely consanguineous couples. The exact magnitude of DNA IBD did not seem to make any difference between the couples who did have an affected child - and were thus revealed as carriers of the same pathogenic mutation - and those who had only healthy children. Chance plays a role in the passing of parental mutations to the next generation, and thus in having two pathogenic mutations in a child. It is possible that our selection of couples contained control couples who, in reality, are carrier couples for the same pathogenic mutation but have had only healthy children. However, the chance of being a couple *without* a similar pathogenic mutation is *a priori* already high (>90% for first cousins, >95% for second cousins). Only couples were selected with three children at least (increasing these chances even more), leading to a small differentiation in *a posteriori* chances between the different numbers of healthy children. It was therefore decided not to apply weighting according to the number of children.

In order to answer our research question, we performed one-sided tests rather than two-sided tests, since it was hypothesized that case couples have more DNA IBD than control couples. From the values of the means and median as well as from the estimate of the regression parameter in the (mixed) regression model, we see an opposite effect in our data: the control couples have more DNA IBD than expected based on their kinship coefficient than the couples with an affected child. This could be due to recall bias where control couples may have reported less consanguineous relationships in the family than case couples who were counselled in clinical genetic centres because of their affected child or children. Case couples may have made more effort earlier when confronted with the autosomal recessive disorder of their child to report the family history and family loops, resulting in more awareness regarding consanguinity in their family. This may have led to a relatively lower R_{ped} for controls and thus a higher ratio R . Another explanation could be the fact that control couples originated from more

endogamous subpopulations, but this would only serve as a possible explanation in, for example, the Tunisian population, and not in the Saudi population (where most control couples were from the same family). If the trend towards controls sharing at least as much DNA IBD as cases is in fact a real finding, more research is needed to look into possible explanations, like differences in the nature of the shared segments. Genes are unequally distributed throughout the genome (Zoubak et al., 1996; Cooper et al., 2010) and consanguineous couples can as a result inherit more or less 'quiet' regions.

This study has several limitations. Although the initial goal was to ascertain an intrafamilial control for every case couple, the lack of available and suitable candidates eventually resulted in only 10 cases for which this was achieved. The goal was to ascertain 100 cases and 100 controls, counting a possible loss of 15%. However, ascertainment was difficult and the sample turned out rather heterogeneous. It can be expected, nevertheless, that including more couples will eventually not lead to a considerable effect size and a possible clinical application of estimating DNA IBD in consanguineous couples.

Estimating relatedness coefficients is known to be complicated, especially in the case of population stratification and/or inbreeding (Manichaikul et al., 2010; Wang, 2011). Our sample contains a variety of ethnicities and individuals from populations with a tradition of consanguinity. The heterogeneity of the sample might be causing inflated estimates of kinship coefficients due to population stratification. However, this principle applies to both the case and the control group. Moreover, the three different estimators were consistent in *not* showing that cases share more DNA IBD than controls.

In this study, consanguineous partners who have a child with an AR condition and are therefore both carriers of the same AR disorder do not share more DNA IBD than similarly related consanguineous couples who do not show evidence of carrier status of an AR disorder. In order to assess the risk for a consanguineous couple in clinical practice in the preconception phase, measuring the amount of DNA IBD does not seem to be an effective approach. Further research on other genomic techniques to improve risk assessment for consanguineous couples is needed as current approaches are often unable to identify couples at high risk of having affected children (Teeuw et al., 2010). A promising possibility may be the application of exome sequencing in the preconception phase, although there are still many challenges to face, like dealing with variants of unknown significance (Alkuraya, 2013).

Table S1. In- and exclusion of couples

Population		Case couples	Case individuals	Control couples	Control individuals
TUN	Considered eligible	50		102	
	Reliable pedigree and suitable sample:	50		52	
	Blood		100		
	Saliva				104
	Genotype				
	Remaining after DNA quality control	50		52	
	Remaining after data quality control	47		49	
SAU	Considered eligible	14		14	
	Reliable pedigree and suitable sample:	13		14	
	Blood				
	Saliva				
	Genotype		26		28
	Remaining after DNA quality control	13		14	
	Remaining after data quality control	11		12	
TUR	Considered eligible	9		12	
	Reliable pedigree and suitable sample:	9		12	
	Blood		4		
	Saliva		14		24
	Genotype				
	Remaining after DNA quality control	7		10	
	Remaining after data quality control	6		10	
MOR	Considered eligible	4		3	
	Reliable pedigree and suitable sample:	3		2	
	Blood				
	Saliva		6		4
	Genotype				
	Remaining after DNA quality control	3		2	
	Remaining after data quality control	3		1	
JOR	Considered eligible	7		7	
	Reliable pedigree and suitable sample:	2		3	
	Blood				
	Saliva		4		6
	Genotype				
	Remaining after DNA quality control	2		3	
	Remaining after data quality control	2		3	
PAK	Considered eligible	16		16	
	Reliable pedigree and suitable sample:	1		1	

Population		Case couples	Case individuals	Control couples	Control individuals
	Blood				
	Saliva		2		2
	Genotype				
	Remaining after DNA quality control	1		1	
	Remaining after data quality control	1		1	
NLD	Considered eligible	1		2	
	Reliable pedigree and suitable sample:	1		1	
	Blood				
	Saliva		2		2
	Genotype				
	Remaining after DNA quality control	1		1	
	Remaining after data quality control	1		1	
AFG	Considered eligible	1			
	Reliable pedigree and suitable sample:	1			
	Blood				
	Saliva		2		
	Genotype				
	Remaining after DNA quality control	1			
	Remaining after data quality control	1			
IRQ	Considered eligible	2		2	
	Reliable pedigree and suitable sample:	1		1	
	Blood				
	Saliva		2		2
	Genotype				
	Remaining after DNA quality control	0		1	
	Remaining after data quality control	0		1	
IRN	Considered eligible	1			
	Reliable pedigree and suitable sample:	1			
	Blood		2		
	Saliva				
	Genotype				
	Remaining after DNA quality control	1			
	Remaining after data quality control	1			
TOTAL	Considered eligible	105		158	
	Reliable pedigree and suitable sample	82		86	
	Remaining after DNA quality control	79		84	
	Remaining after data quality control	73		78	

TUN= Tunisia, SAU= Saudi Arabia, TUR= Turkey, MOR= Morocco, JOR= Jordan, PAK= Pakistan, NLD= the Netherlands, AFG= Afghanistan, IRQ= Iraq, IRN= Iran

Table S2A. Specification of disorders in children of case couples, n=73

Case	POP	R	AR disorder in child (MIM code)	Diagnosis
1	TUN	0.1250	AR deafness 12 (MIM#601386)	Molecular
2	TUN	0.1563	AR deafness 7 (MIM#600974)	Molecular
3	TUN	0.0625	AR deafness 2 (MIM#600060)	Molecular
4	TUN	0.1250	AR deafness 2 (MIM#600060)	Molecular
5	TUN	0.0625	AR deafness 3 (MIM#600316)	Molecular
6	TUN	0.0313	AR deafness 31 (MIM#607084)	Molecular
7	TUN	0.1250	AR deafness 6 (MIM#600971)	Molecular
8	TUN	0.0625	AR deafness 7 (MIM#600974)	Molecular
9	TUN	0.1250	Bardet-Biedl Syndrome 4 (MIM#209900)	Molecular
10	TUN	0.1250	Bardet-Biedl Syndrome1 (MIM#209900)	Clinical
11	TUN	0.0156	Bardet-Biedl Syndrome1 (MIM#209900)	Molecular
12	TUN	0.0313	Bardet-Biedl Syndrome1 (MIM#209900)	Molecular
13	TUN	0.1563	Bardet-Biedl Syndrome12 (MIM#209900)	Molecular
14	TUN	0.1250	Bardet-Biedl Syndrome2 (MIM#209900)	Molecular
15	TUN	0.1250	Bardet-Biedl Syndrome2 (MIM#209900)	Molecular
16	TUN	0.1719	Bardet-Biedl Syndrome5 (MIM#209900)	Molecular
17	TUN	0.1328	Bardet-Biedl Syndrome7 (MIM#209900)	Molecular
18	TUN	0.1250	Bardet-Biedl Syndrome8 (MIM#209900)	Molecular
19	TUN	0.1250	Bardet-Biedl Syndrome8 (MIM#209900)	Molecular
20	TUN	0.1250	Congenital Adrenal Hyperplasia due to 11-Beta-Hydroxylase deficiency (MIM#202010)	Molecular
21	TUN	0.2500	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
22	TUN	0.1563	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
23	TUN	0.1250	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
24	TUN	0.1875	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
25	TUN	0.0313	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
26	TUN	0.1250	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
27	TUN	0.0156	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
28	TUN	0.1250	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
29	TUN	0.0313	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
30	TUN	0.1250	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular

Case	POP	R	AR disorder in child (MIM code)	Diagnosis
31	TUN	0.0625	Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency (MIM#201910)	Molecular
32	TUN	0.1250	Criggler Najjar Type 1 (MIM#218800)	Molecular
33	TUN	0.1250	Criggler Najjar Type 1 (MIM#218800)	Molecular
34	TUN	0.1250	Criggler Najjar Type 1 (MIM#218800)	Molecular
35	TUN	0.2500	Criggler Najjar Type 1 (MIM#218800)	Molecular
36	TUN	0.1250	Familial Mediterranean fever (MIM#249100)	Molecular
37	TUN	0.0313	Familial Mediterranean fever (MIM#249100)	Molecular
38	TUN	0.1250	Familial Mediterranean fever (MIM#249100)	Molecular
39	TUN	0.1563	Familial Mediterranean fever (MIM#249100)	Molecular
40	TUN	0.0313	Familial Mediterranean fever (MIM#249100)	Molecular
41	TUN	0.1250	Familial Mediterranean fever (MIM#249100)	Molecular
42	TUN	0.0625	Familial Mediterranean fever (MIM#249100)	Molecular
43	TUN	0.1250	Familial Mediterranean fever (MIM#249100)	Molecular
44	TUN	0.0625	Familial Mediterranean fever (MIM#249100)	Molecular
45	TUN	0.0625	Spinal Muscular Atrophy type I (MIM#253300)	Molecular
46	TUN	0.1563	Spinal Muscular Atrophy type I (MIM#253300)	Molecular
47	TUN	0.1250	Spinal Muscular Atrophy type I (MIM#253300)	Molecular
48	SAU	0.1250	3-M syndrome 1 (MIM#273750)	Molecular
49	SAU	0.0313	Bardet-Biedl Syndrome 1 (MIM#209900)	Molecular
50	SAU	0.2500	Bardet-Biedl Syndrome 10 (MIM#209900)	Molecular
51	SAU	0.1250	D-bifunctional protein deficiency (MIM#261515)	Molecular
52	SAU	0.1250	Hyaline fibromatosis syndrome (MIM#228600)	Molecular
53	SAU	0.0625	Junctional epidermolysis bullosa. Herlitz type (MIM#226700)	Molecular
54	SAU	0.1875	Microcephaly 2, primary autosomal recessive (MIM#604317)	Molecular
55	SAU	0.0313	Retinitis Pigmentosa (MIM#268000)	Clinical, known X-l + AD genes excluded
56	SAU	0.1250	Retinitis Pigmentosa (MIM#268000)	Clinical, known X-l + AD genes excluded
57	SAU	0.1250	Retinitis Pigmentosa (MIM#268000)	Clinical, known X-l + AD genes excluded
58	SAU	0.2500	Spastic paraplegia 18, autosomal recessive (MIM#611225)	Molecular
59	TUR	0.0391	Ataxia-telangiectasia (MIM#208900)	Molecular
60	TUR	0.1250	Citrullinemia (MIM#215700)	Molecular
61	TUR	0.0313	Familial Mediterranean fever, AR (MIM#249100)	Molecular
62	TUR	0.2168	Glycogen storage disease Ib (MIM#232220)	Molecular
63	TUR	0.0625	Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type A, 3 (MIM#253280)	Molecular

Case	POP	R	AR disorder in child (MIM code)	Diagnosis
64	TUR	0.0625	Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type A, 2 (MIM#613150)	Molecular
65	MOR	0.0625	3-Methylcrotonyl-CoA carboxylase deficiency (no subtyping)	Biochemical
66	MOR	0.0313	Joubert syndrome (no subtyping)	Clinical
67	MOR	0.1250	Lipoprotein lipase deficiency (MIM#238600)	Molecular
68	JOR	0.1250	Metachromatic leukodystrophy (MIM#250100)	Clinical
69	JOR	0.1250	Methylmalonic aciduria (no subtyping)	Biochemical
70	PAK	0.1250	Lamellar ichthyosis (MIM#242300)	Molecular
71	NLD	0.0625	Lamellar ichthyosis (MIM#242300)	Molecular
72	AFG	0.1563	Usher syndrome, type 1B (MIM#276900)	Molecular
73	IRN	0.0313	Polycystic kidney disease, infantile, type 1 (MIM#263200)	Molecular

POP= population, R = kinship coefficient based on pedigree, AR = autosomal recessive, MIM= Mendelian inheritance in man (www.omim.org).TUN= Tunisia, SAU= Saudi Arabia, TUR= Turkey, MOR= Morocco, JOR= Jordan, PAK= Pakistan, NLD= the Netherlands, AFG= Afghanistan, IRN= Iran

Table S2B. Specification of control couples

Control	POP	R	Number of healthy children
1	TUN	0.125	3
2	TUN	0.0313	5
3	TUN	0.125	3
4	TUN	0.0625	4
5	TUN	0.125	5
6	TUN	0.125	5
7	TUN	0.125	3
8	TUN	0.0313	4
9	TUN	0.125	3
10	TUN	0.125	5
11	TUN	0.0625	3
12	TUN	0.0625	5
13	TUN	0.0625	3
14	TUN	0.0781	5
15	TUN	0.125	5
16	TUN	0.125	4
17	TUN	0.125	4
18	TUN	0.125	3
19	TUN	0.125	5
20	TUN	0.125	4
21	TUN	0.125	3
22	TUN	0.125	3
23	TUN	0.125	3
24	TUN	0.125	4
25	TUN	0.0625	4
26	TUN	0.125	3

Control	POP	R	Number of healthy children
27	TUN	0.25	4
28	TUN	0.125	4
29	TUN	0.125	4
30	TUN	0.0625	5
31	TUN	0.0313	4
32	TUN	0.125	4
33	TUN	0.0625	3
34	TUN	0.0625	7
35	TUN	0.0625	3
36	TUN	0.125	3
37	TUN	0.125	3
38	TUN	0.125	3
39	TUN	0.0313	3
40	TUN	0.0195	3
41	TUN	0.0469	3
42	TUN	0.125	4
43	TUN	0.1563	4
44	TUN	0.125	4
45	TUN	0.0156	3
46	TUN	0.125	3
47	TUN	0.125	5
48	TUN	0.125	4
49	TUN	0.125	4
50	SAU	0.125	4
51	SAU	0.125	7
52	SAU	0.0313	3
53	SAU	0.125	8
54	SAU	0.0625	3
55	SAU	0.125	6
56	SAU	0.125	4
57	SAU	0.125	6
58	SAU	0.125	11
59	SAU	0.25	5
60	SAU	0.125	4
61	SAU	0.25	8
62	TUR	0.125	5
63	TUR	0.125	3
64	TUR	0.0313	3
65	TUR	0.125	4
66	TUR	0.1563	3
67	TUR	0.1563	3
68	TUR	0.0313	5
69	TUR	0.125	3
70	TUR	0.0313	3
71	TUR	0.1563	3

Control	POP	R	Number of healthy children
72	MOR	0.125	3
73	JOR	0.125	9
74	JOR	0.0313	3
75	JOR	0.125	8
76	PAK	0.0625	3
77	NLD	0.125	5
78	IRQ	0.1875	5

*POP = population, R = kinship coefficient based on pedigree,
TUN= Tunisia, SAU= Saudi Arabia, TUR= Turkey, MOR= Morocco,
JOR= Jordan, PAK= Pakistan, NLD= the Netherlands, IRQ= Iraq*