university of groningen

University Medical Center Groningen

University of Groningen

Inherited cardiomyopathies

Tintelen, Johannes Peter van

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Tintelen, J. P. V. (2008). *Inherited cardiomyopathies*. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

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.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 21-07-2024

Chapter 9

Haplotype Sharing Test Maps a Gene for Familial Cardiomyopathy

Frans Gerbens[a]
J. Peter van Tintelen[a]
Paul A. van der Zwaag[a]
Ludolf Boven[a]
Jasper J. van der Smagt[b]
Irene M. van Langen[c]
Hennie Bikker[c]
Richard N.W. Hauer[d]
Robert M.W. Hofstra[a]
Gerard J. te Meerman[a]

From the department of Genetics[a], University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; departments of Clinical Genetics[b], and Cardiology[d], University Medical Center Utrecht, Utrecht, the Netherlands; and the department of Clinical Genetics[c], Academic Medical Center, Amsterdam, the Netherlands.

submitted

Abstract
Genomic regions in distantly related family members with an identical genetic disease, can either be identical-by-descent (IBD) or identical-by-state (IBS). The probability for shared haplotypes to be IBD versus IBS increases with increasing length. We hypothesize that the longest shared haplotype is most likely to contain the disease-causing mutation. Therefore we studied 4 small, possibly related families with Arrhythmogenic Right Ventricular Cardiomyopathy using a high-density, genome-wide SNP array. The data were analyzed using the haplotype sharing test (HST). The HST ultimately revealed a single, large haplotype run of 91 SNP markers on chromosome 12 corresponding with a 27.7 Mb genomic area. This haplotype run was substantially longer than all the other shared regions. A causative Plakophilin-2 mutation was identified in this largest shared genomic region.

Our extended genome-wide SNP analysis in combination with HST revealed a large haplotype containing the disease-causing gene. We therefore conclude that our method is a powerful tool in families in which there are insufficient numbers of meioses present to perform classical linkage analysis and where linkage analysis using multimarker analysis over many unknown ancestors is computationally not feasible.

Introduction
In low penetrance Mendelian diseases, disease associated founder mutations are co-inherited with adjacent chromosomal regions which are identical-by-descent (IBD). The region containing the mutation can be found by multimarker haplotype sharing methods. The finding of shared haplotypes has been greatly facilitated by the high-density (10-500K) genome-wide SNP arrays now available. However, at high resolution many shared haplotypes will be found, some of which are identical-by- state (IBS) and others may be IBD. Such IBS haplotypes are commonly present in the general population and are not associated with a disease. The short haplotypes are common and either represent ancient IBD stretches of DNA or are IBS. Longer shared haplotypes are more often IBD. Models for estimating the probability of IBD versus IBS show that haplotypes with 12 or more identical alleles in a region covering about 100,000 base pairs or less already have a substantial probability of being IBD even in unrelated individuals from the same population.[1] It is this expected difference in length between IBS and IBD haplotypes that we now use to identify disease loci.

To identify these longer haplotypes we designed the haplotype sharing

test (HST), which uses isolated patients and parent-offspring data to identify the longest possibly shared haplotypes. We applied the HST to a pedigree consisting of four putative, remotely related families (Figure 1) with autosomal dominantly inherited arrhythmogenic right ventricular cardiomyopathy (ARVC [MIM#107970]) a disease known for its low penetrance.[2,3]

Materials and Methods

Patients

The genealogy of four small nuclear families each containing at least one affected individual with (low penetrance) autosomal dominantly inherited ARVC was investigated (over approximately 6 generations). Three of the families could be traced back to two ancestral founder couples (Figure 1: I-1/I-2 and I-3/I-4/I-5) from the same small village although genealogical investigations could not formally confirm paternity of IV-1. A fourth family (Figure 1; Family 4) could not be linked to the pedigree, but their ancestors lived in the same geographical region. These data suggest segregation of a possible founder DNA in these families. At the time of initial analysis of families 1-3, DNA and clinical information was available for 5 patients or obligate mutation carriers and 6 related individuals. The study was approved by the medical and ethical committee of the UMC Utrecht and written informed consent was obtained from all participants.

Genotype Data

Genome-wide high-density genotyping with GeneChip® Mapping 10K SNP arrays (Xba131; Affymetrix, Santa Clara, California, USA) was performed by Service XS (Leiden, the Netherlands) according to the manufacturer's protocols. The resulting data from the arrays were converted to genotypes by the GeneChip® DNA analysis software 2.0 (GDAS 2.0; Affymetrix).

Haplotype Sharing Analysis

A Delphi 5.0 program that processes trio and isolated individual data was used to search for shared haplotypes. The program checks for consistency of genotypes with extended shared risk haplotypes, using "affected-only" data. To account for genotyping errors a maximum of 1 in 20 SNPs with an inconsistent genotype was allowed.

We first identified all the SNP alleles present in a first trio, including an obligate carrier parent and an affected child (VI-1 and parents) (Figure 1). Then a second trio was analyzed (VII-1 and parents) and alleles shared by all

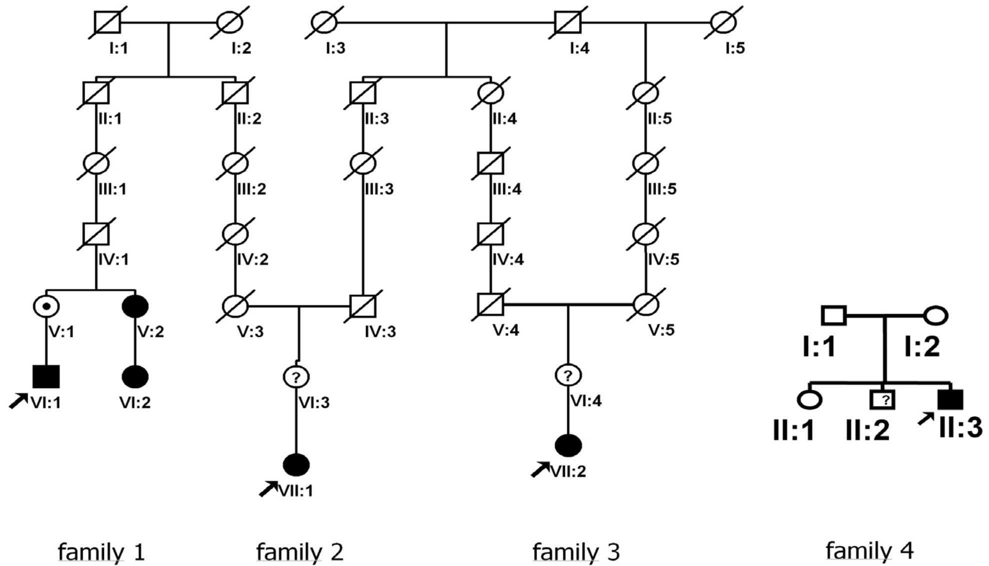

Figure 1: Pedigrees of the families studied with autosomal dominantly inherited ARVC. Circles indicates females; squares indicates males; black symbols represent affected persons; ? represents unknown clinical status because the person has not been investigated or the clinical signs are insufficient to fulfill the formal clinical criteria; line through symbol means deceased; clear symbol in living subjects means that person has no signs of the disease; symbol with dot means obligate carrier without clinical signs.

3 patients were subsequently checked for consistency with the genotypes in all branches of the pedigree (families 3 and subsequently family 4) (Figure 1).

The algorithm starts at the telomeric end of a chromosome. When shared alleles for genotypes as described above are consistent, the comparison continues with the next SNP locus. A run is terminated when a real inconsistency in the observed SNP alleles occurs. After a run terminates, a new one will start with the following SNP. In this way the algorithm scans all the markers on each chromosome.

Probability Scores

When a series of common SNP's are identified that are consistent in all affected individuals, the probability that an allele is shared by all individuals is calculated. This is based on the allele frequencies in the general population, assuming linkage disequilibrium (LD) and assuming independent probabilities

for two loci to be homozygous for the opposite alleles: $p^2 \times (1\text{-}p)^2$, or when a risk allele has been determined with certainty: p^2 or $(1\text{-}p)^2$ depending on the allele frequency of the certain allele. This model neglects the dependency of allelic markers caused by linkage disequilibrium. This will therefore generally result in a lower probability for the shared haplotype compared to the case where LD would have been taken into full account.

For the full shared haplotype these probabilities are multiplied and the $-^{10}\log$ was calculated. If no allele frequency information was available, the allele frequency was set to 0.5. Appropriately formatted allele frequencies for most current genotyping arrays can be generated from the manufacturer's data files. Examples for some platforms (Affymetrix and Illumina) including the program to perform the calculations are provided upon request.

Results

Haplotype Sharing

The Haplotype Sharing Test in families 1-3 revealed a haplotype run of 118 SNP markers on chromosome 12 flanked on either side by the SNP markers rs1163969 and rs1902765 (see Figure 2 and Table 1). This haplotype run was substantially longer than other areas which are most likely shared due to random effects (the average shared haplotype length on autosomes is only 3.9 SNPs (including haplotypes >1 SNP). After we included the data from family 4, the haplotype run covered 91 SNP markers.

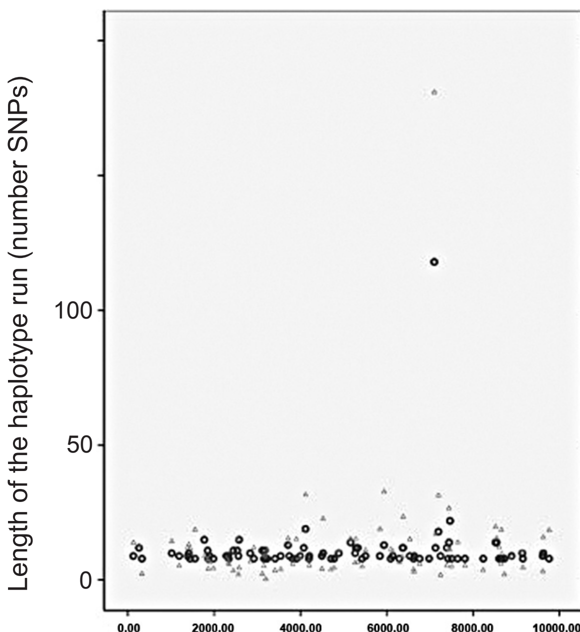

Figure 2. Haplotype length distribution across the genome after haplotype sharing analysis in families 1-3. The circles indicate the length of the haplotype runs. The largest region and most significant region of 118 SNPs pointed to chromosome 12p12.3-q13.13.

Table 1

SNP id	Genomic position	Family 1							Family2				Family3			
		husband of V-1	V-1	VI-1		V-2	VI-2		husband of VI-3	VI-3	VII-1		Husband of VI-4	VI-4	VII-2	
				haplo1	haplo2		haplo1	haplo2				haplo2			haplo1	haplo2
rs1163969	17425668	BB	AA	B	A	AB	A	A	BB	AB	B	B	AA	AB	A	A
rs4129180	18216340	AA	AB	A	A	AA	A	A	AA	AA	A	A	AA	AA	A	A
#42 SNP's																
rs951821	28710773	AB	AB	A	A	AA	B	A	AA	AA	A	A	AA	AA	A	A
rs1601981	28961846	AA	AB	A	B	AB	B	B	NoCall	BB	AB	B	AA	AB	A	B
rs986570	29157485	AB	AB	AB	AB	AA	B	A	AA	AA	A	A	AA	AB	A	A
rs986569	29157743	AB	AB	AB	AB	AA	B	A	AA	AA	A	A	AA	AB	A	A
rs958478	29246991	BB	BB	B	B	BB	B	B	BB	AB	B	B	BB	BB	B	B
rs3862404	30276999	AB	AB	AB	AB	BB	B	B	AB	BB	B	B	BB	BB	B	B
rs1429630	30314116	BB	BB	B	BB	AB	B	B	BB	BB	B	B	BB	BB	B	B
rs1320925	30353247	BB	BB	B	BB	AB	B	B	BB	BB	B	B	BB	BB	B	B
rs2114903	30859954	AB	AA	B	A	AA	B	A	AB	AB	AB	AB	AB	AB	A	A
rs720480	31023510	BB	AA	B	A	AB	AB	AB	AB	AB	A	A	AB	AB	AB	AB
rs724417	31264161	BB	AB	B	A	AB	AB	AB	BB	AB	B	A	BB	AB	B	A
rs1844986	31366772	AB	AA	A	A	AA	B	A	AA	AB	A	A	AB	AA	B	A
c2489+72_73delinsA	32840238	GT/GT	GT/A	GT	A	GT/A	nd	nd	AA	GT/A	nd	nd	nd	AA	nd	nd
c.2489+13_2489+14insC	32840297	TT	T/CT	T	T	TT	T	T	TT	T/CT	T	T	CT/CT	TT	T	T
c.2489+4A>C	32840306	AA	AC	A	C	AC	A	C	AA	AC	A	C	AA	AC	A	C
rs2133675	33086481	AA	AA	A	A	AA	A	A	AA	AA	A	A	AA	AA	A	A
rs955648	33105418	AA	BB	A	B	NoCall	AB	AB	AA	BB	A	B	AB	AB	AB	AB
rs1392339	33143421	AA	AB	A	B	AB	A	B	AA	BB	A	B	AB	BB	AB	B
rs1457677	33306174	BB	AB	B	A	AB	AB	AB	BB	AA	B	A	AB	AB	AB	AB
rs2389173	33321150	BB	BB	B	B	BB	B	B	BB	BB	B	B	BB	BB	B	B
rs1386934	33674567	AB	BB	A	B	BB	B	B	NoCall	AB	AB	AB	AB	AB	AB	AB
rs1386937	33761564	AB	AA	B	A	AA	A	A	AB	AA	B	A	AB	AA	B	A
rs1525895	33857763	AA	AB	A	B	BB	B	B	AA	BB	A	B	AB	BB	A	B
rs8186671	36491877	BB	BB	B	B	BB	B	B	BB	BB	B	B	BB	BB	B	B
rs8186802	36506044	AA	AA	A	AA	AA	A	A	AA	AA	A	A	AA	AA	A	A
rs4088478	36517000	AA	AA	A	AA	AA	A	A	AA	AA	A	A	AA	AA	A	A
rs8186742	36545390	AA	AA	A	AA	AA	A	A	AA	AA	A	A	AA	AA	A	A
rs8186744	36545505	AA	AB	A	B	BB	B	B	AA	BB	A	B	AB	BB	B	B
rs8186798	36694832	BB	NoCall	B	A	AA	AA		BB	AA	B	A	NoCall	AA	A	A
#48SNPs																
rs1316607	49329157	AA	AB	A	B	BB	A	B	BB	BB	B	B	AB	BB	A	B
rs1902765	50938088	BB	AB	B	A	AB	AB	AB	AB	AA	AB	AB	AB	AB	B	B

Table 1: The initial 118 SNPs shared haplotype across affected individuals from families 1-3. Genotypes and inferred haplotypes from SNPs flanking the central PKP2 splice site mutation are shown as well as the SNPs surrounding the breakpoints on either side of the shared haplotype.

Gene Finding

The longest shared haplotype stretched over a 27.7Mb region. This genomic area is located on chromosome 12 (region 12p12.3 to 12q13.13) and contains more than 200 known and predicted genes. Considering the ARVC phenotype, the most likely candidate gene in this region was the plakophilin-2 (*PKP2*) gene. This gene was identified during the course of this study, and it is frequently the underlying factor in familial forms of this disease.[4,5]

Mutation Identification

Patients from this pedigree were screened for *PKP2* mutations in a routine diagnostic setting; this revealed a pathogenic splice-site mutation (NM_004572 c.2489+4 A>C) that was segregating with the disease. This splice-site mutation was identified as clinically relevant by reverse-transcription (RT)-PCR in lymphocytes showing the absence of exon 12 in the respective PKP2 transcript sequence (NM_004572 encoding PKP2 protein isoform 2b) in all affected subjects. The deletion of exon 12 leads to a frame shift in the protein coding sequence resulting in an aberrant protein of 848 amino acids as opposed to 863 amino acids for wild type PKP2 isoform 2b (J.J. van der Smagt et al., manuscript in preparation).

Discussion

Rare Mendelian disorders have a mutational spectrum which is characterized by the presence of very few major mutations but larger numbers of very rare variants. It is therefore reasonable to search for the presence of common mutant genes in patients with the same disease phenotype. Nolte and te Meerman have shown that under coalescence assumptions that should be approximately correct for the European population, haplotypes with > 12 consecutive marker alleles identical over 0.1 cM of recombination distance have a substantial probability of originating from a common founder.[1] Such haplotypes can easily be detected with current high density genotyping methods.

If two individuals have a common ancestor they may share a stretch of DNA (a haplotype) that has been shortened by recombination but is still of detectable size. The current density of whole genome screens is so high that shared haplotypes of a few hundred thousand base pairs will contain well over 30 SNPs. If rare Mendelian disorders are investigated, a common disease mutation may coalesce to a common ancestor well within 100 generations, leading to expected genomic sharing of several centimorgans,

with hundreds of markers on the haplotype. Of course, for many genomic areas random coalescence to recent ancestors may occur as well. Due to adverse selection most descent lines with deleterious mutations will gradually be eliminated leading to shorter coalescence times to a common ancestor. The expectation is therefore that shared areas surrounding disease mutations will be larger in size than shared areas due to a common ancestor for a neutral genomic area. This leads to a heuristic principle: investigate the largest areas shared between patients or obligate carriers for mutant genes first. The carrier frequency of such areas in the population must be rather low, making the association between a rare variant and a rare marker haplotype more likely under the hypothesis that the shared region harbours a mutant gene that has led to ascertainment of the individuals that carry them. Occasionally however, recombinations will take place near to the mutant position and the corresponding shared haplotype may become small. Fortunately the total size of shared haplotypes that could contain a mutant gene is often not so large, and may be even as small as the mapping interval of linkage analysis with a sufficient number of meioses, to have certainty of linkage.

The results of our analysis confirm our hypothesis. Figure 2 shows the shared haplotypes identified in the patients from families 1-3 that were analyzed. We identified many short IBS or perhaps IBD haplotypes with an average length of 12 SNPs and one longer haplotype of 92 SNPs. A recently identified gene underlying ARVC, namely PKP2, was also present in the longest haplotype, which contained more than 200 genes. Subsequent screening of the gene revealed a splice mutation leading to an abnormal RNA fragment on RT-PCR (J.J. van der Smagt et al., manuscript in preparation).

The pedigrees suggest that the there are at least 23 meioses separating the carriers of the mutation. The size of the unrecombined area surrounding a common mutation originating from a common founder decreases inversely with the number of independent meioses according to the formula 200 cM / number of meioses distance. The age of the mutation since the common founder is therefore estimated to be more than 8 generations ago, on the basis of the pedigree, and likely not much more recently present in the common ancestor. As we cannot see the complete history of this mutation, its exact age is difficult to determine and could be well over 8 generations as the selection against this mutant is rather weak. Our results show that haplotype sharing analysis is a useful tool for narrowing down the area in which to search for the gene rather than providing conclusive evidence. This might be particularly useful in low penetrance,

genetically heterogeneous disorders.

Moreover we prove that the theoretical model proposed by Miyazawa et al. is indeed effective.[6] They showed the use of homozygosity mapping through haplotype analysis as a tool to directly compare autosomes among several patients in order to identify shared segments. The great advantage of our analysis is that it extends the well known principle of homozygosity mapping to cases where the disease mutation is present on a single chromosome in affected individuals. Identifying a large shared haplotype nowadays is not a problem since current genotyping technology easily yields hundreds of thousands of reliable SNP genotypes in a genome. Subsequent mutational analysis can be done by sequencing, and current technology is powerful enough to investigate a few candidate regions and confirm the results. We propose that haplotype sharing analysis will assist in gene identification in those low penetrance Mendelian diseases where standard tools cannot be used due to lack of substantial pedigree information or where the mutation rate is low enough to expect founder mutations to be shared among patients.

Acknowledgements

We thank Eric Hennekam, genealogist, for the genealogical investigations and Jackie Senior for her help in preparing this manuscript. We are also indebted to Arthur Wilde, cardiologist, who performed the cardiological investigations in family 4 and Chantal Conrath MD, who helped collecting data from the families.

Funding: the Netherlands Heart Foundation (2003B062 and 2007B132).

References

1. Nolte IM, te Meerman GJ. The probability that similar haplotypes are identical by descent. Ann Hum Genet. 2002; 66:195-209.
2. Dalal D, James C, Devanagondi R, Tichnell C, Tucker A, Prakasa K, Spevak PJ, Bluemke DA, Abraham T, Russell SD, Calkins H, Judge DP. Penetrance of mutations in plakophilin-2 among families with arrhythmogenic right ventricular dysplasia/cardiomyopathy. J Am Coll Cardiol. 2006; 48:1416-1424.
3. Syrris P, Ward D, Asimaki A, Sen-Chowdhry S, Ebrahim HY, Evans A, Hitomi N, Norman M, Pantazis A, Shaw AL, Elliott PM, McKenna WJ. Clinical expression of plakophilin-2 mutations in familial arrhythmogenic right ventricular cardiomyopathy. Circulation. 2006; 113:356-364.
4. Gerull B, Heuser A, Wichter T, Paul M, Basson CT, McDermott DA, Lerman BB, Markowitz SM, Ellinor PT, MacRae CA, Peters S, Grossmann KS, Drenckhahn J, Michely B, Sasse-Klaassen S, Birchmeier W, Dietz R, Breithardt G, Schulze-Bahr E, Thierfelder L. Mutations in the desmosomal protein plakophilin-2 are common

in arrhythmogenic right ventricular cardiomyopathy. Nat Genet. 2004; 36:1162-1164.

5. van Tintelen JP, Entius MM, Bhuiyan ZA, Jongbloed R, Wiesfeld AC, Wilde AA, van der Smagt J, Boven LG, Mannens MM, van Langen IM, Hofstra RM, Otterspoor LC, Doevendans PA, Rodriguez LM, van Gelder IC, Hauer RN. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation. 2006; 113:1650-1658.
6. Miyazawa H, Kato M, Awata T, Kohda M, Iwasa H, Koyama N, Tanaka T, Huqun, Kyo S, Okazaki Y, Hagiwara K. Homozygosity haplotype allows a genomewide search for the autosomal segments shared among patients. Am J Hum Genet. 2007; 80:1090-1102