Short Report

Haplotype sharing test maps genes for familial cardiomyopathies


Identifying a mutation in a heterogeneous disease such as inherited cardiomyopathy is a challenge because classical methods, like linkage analysis, can often not be applied as there are too few meioses between affected individuals. However, if affected individuals share the same causal mutation, they will also share a genomic region surrounding it. High-density genotyping arrays are able to identify such regions shared among affected individuals. We hypothesize that the longest shared haplotype is most likely to contain the disease-causing mutation. We applied this method to two pedigrees: one with arrhythmogenic right ventricular cardiomyopathy (ARVC) and one with dilated cardiomyopathy (DCM), using high-density genome-wide SNP arrays. In the ARVC pedigree, the largest haplotype was on chromosome 12 and contained a causative \textit{PKP2} mutation. In the DCM pedigree, a causative \textit{MYH7} mutation was present on a large shared haplotype on chromosome 14. We calculated that a pedigree containing at least seven meioses has a high chance of correctly detecting the mutation-containing haplotype as the largest. Our data show that haplotype sharing analysis can assist in identifying causative genes in families with low penetrance Mendelian diseases, in which standard tools cannot be used due to lack of sufficient pedigree information.

PA van der Zwaag\textsuperscript{a},
JP van Tintelen\textsuperscript{a}, F Gerbens\textsuperscript{b*},
JHD Jongbloed\textsuperscript{a}, LG Boven\textsuperscript{a},
JJ van der Smagt\textsuperscript{a},
WP van der Roest\textsuperscript{a},
IM van Langen\textsuperscript{c},
H Bikker\textsuperscript{c}, RNW Hauer\textsuperscript{d},
MP van den Berg\textsuperscript{a}, RMW Hofstra\textsuperscript{a} and GJ te Meerman\textsuperscript{a}

\textsuperscript{a}Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, 
\textsuperscript{b}Department of Medical Genetics, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands, 
\textsuperscript{c}Department of Clinical Genetics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, 
\textsuperscript{d}Department of Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands, and 
\textsuperscript{e}Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

\textsuperscript{*}Deceased 15 May 2008.

\textsuperscript{†}We dedicate this paper to Frans Gerbens, who contributed greatly to this work. To our sorrow he passed away during the course of these investigations.

Key words: ARVC – chromosome mapping – DCM – genome – haplotypes – linkage – SNP array

Corresponding author: Dr Gerard J te Meerman, Department of Genetics, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands.
Tel.: +31 503617100;
fax: +31 503617230;
e-mail: g.j.te.meerman@med.umcg.nl

Received 15 March 2010, revised and accepted for publication 17 May 2010

459
Identifying a mutation in a heterogeneous disease is a challenge with potentially important clinical consequences, because it may direct treatment and facilitate identification of family members at risk. Cardiomyopathies are a clear example of such a disease. These are associated with mechanical and/or electrical dysfunction that usually exhibit inappropriate ventricular hypertrophy or dilatation and are frequently genetic (1). Around 30, 20 and 10 genes, respectively, have been described underlying dilated cardiomyopathy (DCM [MIM#115200]), hypertrophic cardiomyopathy (HCM [MIM#192600]), and arrhythmogenic right ventricular cardiomyopathy (ARVC [MIM#107970]) (2–6). The importance of identifying a mutation is illustrated by the fact that sudden death is more common in HCM patients carrying a troponin T gene (TNNT2) mutation (7, 8), and in DCM patients carrying a lamin A/C gene (LMNA) mutation (9–12).

The detection of a mutation also has important implications for family members. Cascade screening will identify mutation carriers, and early interventions, such as lifestyle modifications, use of medications, and implanting an implantable cardioverter defibrillator (ICD), will reduce morbidity and mortality. Excluding a mutation in family members will allow them to be dismissed from regular cardiological follow-up (13, 14). In cardiomyopathies, identifying a mutation is challenging and time-consuming, and the yield from screening is small for most genes (15).

In low penetrance Mendelian diseases, founder mutations are likely to occur and these are co-inherited with adjacent chromosomal regions which are identical-by-descent (IBD). IBD genomic regions can be found by high-density genome-wide single nucleotide polymorphism (SNP) arrays. However, many shared haplotypes will be found, some of which will be identical-by-state (IBS) while others may be IBD. Such IBS haplotypes are commonly present in the general population and not associated with disease. In general, short haplotypes are common and either represent ancient IBD stretches of DNA or are IBS. Larger shared haplotypes are younger and have a high probability of being IBD. Haplotypes with 15 or more polymorphic markers covering about 100,000 base pairs already have a substantial probability of being IBD (16). Mutant genes tend to have a common ancestor from a geographic subpopulation and will show extended haplotype sharing surrounding the mutant gene. It is this expected difference in length between IBS and IBD haplotypes that we use to identify disease loci.

To identify disease-associated haplotypes, we designed the haplotype sharing test (HST) using high-density SNP array data. As a proof-of-principle, we applied the HST to two pedigrees: one consisting of four putative, remotely related families with autosomal dominant (AD) ARVC (Fig. 1a), and one with AD DCM (Fig. 1b).

Materials and methods

Clinical evaluation

Four small families with AD ARVC were investigated (Fig. 1a). Three families could be traced back to two ancestral founder couples. Family 4 could not be linked to the pedigree, but their ancestors lived in the same geographical region suggesting a possible founder mutation. DNA and clinical information was available for 16 individuals.

From a four-generation family with AD DCM (Fig. 1b), DNA and clinical information were available for 19 family members.

The studies were approved by the ethics committees of the UMC Utrecht and UMC Groningen and informed consent was obtained from all participants.

Genotype data

In the ARVC pedigree, genome-wide genotyping with the genechip® Mapping 10K SNP array (Xba131; Affymetrix, Santa Clara, CA, USA) was performed according to manufacturer’s protocols. Data from the arrays were converted to genotypes using genechip® DNA analysis software 2.0 (Affymetrix, Santa Clara, CA, USA).

In the DCM pedigree, genome-wide genotyping with the human 610-quadt beadchip® 610K SNP array (Illumina, San Diego, CA, USA) was performed according to manufacturer’s protocols. Data from the arrays were converted to genotypes using genomestudio® data analysis software (Illumina, San Diego, USA).

Haplotype sharing test

Using microsoft® Office Excel 2007 (Microsoft, Redmond, WA) software, the HST was developed to search for shared haplotypes. An algorithm checks for consistency of genotypes with shared risk haplotypes using data from affected individuals and obligate carriers only. The algorithm starts at the p-telomere of a chromosome and compares all the genotypes of the first SNP. When all affected individuals share an allele for the
SNP, the comparison continues with the next SNP. A run is terminated when a real inconsistency (i.e., opposite homozygosity) in an SNP occurs. After a run terminates, the algorithm computes the total number of SNPs and the length of the haplotype in megabase (Mb) and centiMorgan (cM), based on the deCODE genetic map (17). A new run starts with the following SNP. The algorithm scans all the SNPs on each chromosome and computes the lengths of all the shared haplotypes.

**Results**

**Haplotype sharing test**

In all four families of the ARVC pedigree, the HST revealed a largest haplotype run of 93 SNPs.
Fig. 2. (a) Haplotype length distribution across the genome after haplotype sharing analysis using a 10K single nucleotide polymorphism (SNP) array in the ARVC families 1–4. The largest region of 24.1 cM or 93 SNPs, flanked on either side by the SNP markers rs724903 and rs408435, pointed to chromosome 12p12.1-q13.13, where *PKP2* is localized. (b) Haplotype length distribution across the genome after haplotype sharing analysis using a 610K SNP array in the DCM pedigree. The largest region of 20.8 cM or 2707 SNPs, flanked on either side by the SNP markers rs11621148 and rs17101350, pointed to chromosome 14q11.2-q13.1, where *MYH7* is localized.

on chromosome 12, spanning 24.1 cM (Fig. 2a and Table 1). This haplotype run was substantially larger than other areas which were most probably shared due to hidden distant ancestry or IBS (average shared haplotype length: 4.0 SNPs and 2.1 cM).

In the DCM pedigree, the HST revealed a largest haplotype run of 2707 SNPs on chromosome
**Haplotype sharing test in ARVC and DCM**

Table 1. The largest shared haplotype across some affected individuals from the ARVC pedigrees (in gray)

<table>
<thead>
<tr>
<th>SNPid</th>
<th>Position</th>
<th>Mb</th>
<th>cM</th>
<th>hap1</th>
<th>hap2</th>
<th>hap1</th>
<th>hap2</th>
<th>hap1</th>
<th>hap2</th>
<th>hap1</th>
<th>hap2</th>
<th>hap1</th>
<th>hap2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs724903</td>
<td>23.2</td>
<td>42.8</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs763853</td>
<td>23.2</td>
<td>42.8</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#15 SNPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs951821</td>
<td>28.7</td>
<td>51.3</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1601981</td>
<td>29.0</td>
<td>51.5</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs986570</td>
<td>29.2</td>
<td>51.7</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs986569</td>
<td>29.2</td>
<td>51.7</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs958478</td>
<td>29.2</td>
<td>51.8</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3862404</td>
<td>30.3</td>
<td>53.3</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1429630</td>
<td>30.3</td>
<td>53.4</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1320925</td>
<td>30.3</td>
<td>53.4</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2114903</td>
<td>30.8</td>
<td>53.9</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs720480</td>
<td>31.0</td>
<td>54.0</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs724417</td>
<td>31.2</td>
<td>54.3</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1844986</td>
<td>31.3</td>
<td>54.4</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.2489+4A&gt;C</td>
<td>32.8</td>
<td>55.4</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#50 SNPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2133675</td>
<td>33.1</td>
<td>55.5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs955648</td>
<td>33.1</td>
<td>55.5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1392339</td>
<td>33.1</td>
<td>55.5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1457677</td>
<td>33.3</td>
<td>55.6</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2389173</td>
<td>33.3</td>
<td>55.6</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1386934</td>
<td>33.7</td>
<td>55.6</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1386937</td>
<td>33.8</td>
<td>55.7</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1525885</td>
<td>33.9</td>
<td>55.7</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8186671</td>
<td>36.5</td>
<td>56.1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8186802</td>
<td>36.5</td>
<td>56.1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4088478</td>
<td>36.5</td>
<td>56.1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8186742</td>
<td>36.5</td>
<td>56.1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* The phase of heterozygous SNPs is inferred from the parental haplotypes (not listed) resulting in termination of the run after 93 SNPs (see VI:8). The identified mutation c.2489+4A>C is present on the shared haplotype.

14, spanning 20.8 cM (Fig. 2b). This haplotype run was also substantially larger than other areas (average shared haplotype length: 8.1 SNPs and 0.06 cM).

**Gene finding**

We looked for known disease-associated genes in the largest haplotypes. In the ARVC pedigree, the largest shared haplotype stretched over a 24.1 cM region (28.0 Mb), containing over 200 genes. Considering ARVC, the most likely candidate gene was the plakophilin 2 (*PKP2* [MIM#602861]) gene. This gene was identified as an highly prevalent ARVC-associated gene during the course of this study (18, 19).

The largest shared haplotype in the DCM pedigree stretched over a 20.8 cM region (11.5 Mb), also containing over 200 genes. Considering DCM, the most likely candidate gene in this region was the cardiac beta myosin heavy chain (*MYH7* [MIM#160760]) gene. *MYH7* (missense) mutations are found in 4–10% of patients with idiopathic DCM (20–25).

**Mutation identification**

*PKP2* mutation screening in patients from the ARVC pedigree revealed a pathogenic splice-site mutation c.2489+4A>C (26), segregating with the disease. This mutation was identified as clinically relevant by reverse-transcription...
polymerase chain reaction (RT-PCR) in lymphocytes showing the absence of exon 12 in the PKP2 transcript sequence in all affected subjects (data not shown). The absence of exon 12 leads to a frameshift in the coding sequence, resulting in an aberrant protein of 848 amino acids opposed to 863 amino acids for wild-type PKP2 isoform 2b. 

MYH7 mutation screening in the proband from the DCM pedigree revealed a c.2710C>T missense mutation which changes a highly conserved hydrophilic and polar arginine residue into a hydrophobic and nonpolar cysteine residue in the corresponding MYH7 protein (p.Arg904Cys). All available affected family members also carried this mutation, whereas unaffected family members did not.

Both mutations were absent in over 150 ethnically matched controls (300 alleles).

Step-by-step analysis

Tables 2 and 3 show the results of step-by-step haplotype sharing analysis in both pedigrees. An analysis of only five individuals in each pedigree indicated the mutation-containing haplotype was indeed the largest. Additional individuals narrowed down the size of the shared haplotype, thereby decreasing the total number of genes to be considered. In the DCM pedigree (Table 3), we predicted the size of the unrecombined area surrounding the mutation by using the formula: 200 cM/number of meioses (27). The size of the haplotypes was comparable to the expected size.

Requirement for HST

To establish the minimal number of affected members within a pedigree to successfully apply the HST, we calculated all possible combinations of different numbers of meioses in the DCM pedigree. The results are shown in Table 4, illustrating that five or six meioses revealed the MYH7 haplotype as the largest in 4/19 (21%) or 9/15 (60%) combinations, respectively. With seven meioses, the MYH7 haplotype is the largest in 8/9 (89%) combinations.

Discussion

Rare Mendelian disorders have a mutational spectrum which is characterized by the presence of very few major mutations but a large number of very rare mutations. It is therefore reasonable to search for the presence of common mutant genes in patients with the same disease phenotype. It has been shown that haplotypes identical over 0.1 cM have a substantial probability of originating from a common founder (16). Such haplotypes can easily be detected with current high-density genotyping methods.

If two individuals have a common ancestor, they may share a haplotype that has been shortened by

---

Table 2. In the step-by-step haplotype analysis of the ARVC pedigree, family 1 was analyzed first, with subsequent analysis of families 2–4

<table>
<thead>
<tr>
<th>IDs</th>
<th>Average haplotype</th>
<th>Largest haplotype</th>
<th>Rank PKP2 haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNPs cM</td>
<td>SNPs cM</td>
<td></td>
</tr>
<tr>
<td>Family 1</td>
<td>V:2/V:4/V:1/V:2</td>
<td>13.6 5.3</td>
<td>349 82.0 3</td>
</tr>
<tr>
<td>+ Family 2</td>
<td>+V:3</td>
<td>7.3 3.1</td>
<td>217 52.9 1</td>
</tr>
<tr>
<td>+ Family 3</td>
<td>+V:5/V:3</td>
<td>4.9 2.3</td>
<td>120 31.6 1</td>
</tr>
<tr>
<td>+ Family 4</td>
<td>+V:10/V:8</td>
<td>4.0 2.1</td>
<td>93 24.1 1</td>
</tr>
</tbody>
</table>

*aIncluding haplotypes >1 SNP, ranked according to their length in cM.

Table 3. In the step-by-step haplotype analysis of the DCM pedigree, four members of generation II were analyzed first, with subsequent analysis of their offspring

<table>
<thead>
<tr>
<th>IDs</th>
<th>Average haplotype</th>
<th>Largest haplotype</th>
<th>Calculated haplotype size</th>
<th>Rank MYH7 haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNPs cM</td>
<td>SNPs cM</td>
<td>Meioses cM</td>
<td></td>
</tr>
<tr>
<td>Generation II</td>
<td>II:5/II:6/II:8/II:11</td>
<td>32.3 0.21</td>
<td>11,129 78.4</td>
<td>4 50.0 9</td>
</tr>
<tr>
<td>+ offspring II:1</td>
<td>+III:1</td>
<td>17.5 0.12</td>
<td>3589 27.0</td>
<td>6 33.3 1</td>
</tr>
<tr>
<td>+ offspring II:5</td>
<td>+III:3</td>
<td>12.5 0.09</td>
<td>3089 27.0</td>
<td>7 28.6 1</td>
</tr>
<tr>
<td>+ offspring II:11</td>
<td>+III:6/IV:2</td>
<td>8.1 0.06</td>
<td>2707 20.8</td>
<td>9 22.2 1</td>
</tr>
</tbody>
</table>

*aIncluding haplotypes >1 SNP, ranked according to their length in cM.

*bBased on the formula: 200 cM/number of meioses = size unrecombined area surrounding a mutation (27).

This haplotype is 16.5 cM.
recombination, but that 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. Of course, for many genomic areas, a 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: first investigate the largest areas shared between patients or obligate carriers for mutant genes. The carrier frequency of such areas in the population must be fairly low, making the association between a rare variant and a rare marker haplotype more likely under the hypothesis that the shared region harbors a mutant gene that has led to ascertainment of the individuals who carry them. Occasionally, however, recombinations will take place near 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 even be as small as the mapping interval used for linkage analysis with a sufficient number of meioses.

The results of our analyses confirm our hypothesis. Figure 2 shows the shared haplotypes identified in the pedigrees that were analyzed using a 10K and a 610K SNP array. We identified many short IBS or perhaps IBD haplotypes with an average length of 4.0 SNPs (2.1 cM) and 8.1 SNPs (0.06 cM), respectively, and in both pedigrees, we found one larger haplotype of 24.1 cM (93 SNPs) and 20.8 cM (2707 SNPs), respectively. A common gene underlying ARVC, namely PKP2, was present in the largest haplotype of the ARVC pedigree. Screening of the gene revealed a pathogenic splice-site mutation (26).

In the DCM pedigree, the largest haplotype encompassed MYH7, a common gene underlying DCM. Screening revealed a missense mutation of

---

### Table 4. Results of the HST for all possible combinations of different numbers of meioses in the DCM pedigree

<table>
<thead>
<tr>
<th>Five meioses</th>
<th>Rank</th>
<th>MYH7 haplotype</th>
<th>Largest haplotype (cM)</th>
<th>Six meioses</th>
<th>Rank</th>
<th>MYH7 haplotype</th>
<th>Largest haplotype (cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>II:5/II:6/II:8 + III:1</td>
<td>1</td>
<td>27.0</td>
<td>II:5/II:6/II:8 + IV:2</td>
<td>1</td>
<td>20.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:11 + III:1/III:3</td>
<td>1</td>
<td>35.1</td>
<td>II:5 + III:1 + IV:2</td>
<td>1</td>
<td>20.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:8 + III:1/III:3</td>
<td>1</td>
<td>47.4</td>
<td>II:6 + III:3 + IV:2</td>
<td>1</td>
<td>21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:6/II:8 + IV:2</td>
<td>1</td>
<td>68.1</td>
<td>II:5/II:6/II:8 + III:1 + III:3</td>
<td>1</td>
<td>27.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III:3 + IV:2</td>
<td>2</td>
<td>31.6</td>
<td>II:6/II:8 + III:1/III:3</td>
<td>1</td>
<td>27.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:8 + III:1/III:3</td>
<td>2</td>
<td>37.5</td>
<td>II:8/II:11 + III:1/III:3</td>
<td>1</td>
<td>27.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:6/II:8/II:11 + III:1</td>
<td>2</td>
<td>46.0</td>
<td>II:6/II:11 + III:1/III:3</td>
<td>1</td>
<td>35.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:5/II:6 + IV:2</td>
<td>2</td>
<td>47.0</td>
<td>II:8 + III:1 + IV:2</td>
<td>1</td>
<td>47.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:6 + III:1/III:3</td>
<td>2</td>
<td>59.0</td>
<td>II:6/II:8 + III:1/III:3</td>
<td>1</td>
<td>47.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III:1 + IV:2</td>
<td>2</td>
<td>74.1</td>
<td>II:8 + III:3 + IV:2</td>
<td>2</td>
<td>21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:6 + III:1/III:3</td>
<td>2</td>
<td>82.4</td>
<td>III:1/III:3/III:6</td>
<td>2</td>
<td>21.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:5/II:6/II:11 + III:1</td>
<td>3</td>
<td>44.7</td>
<td>II:6 + III:1 + IV:2</td>
<td>2</td>
<td>52.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:8 + III:3/III:6</td>
<td>5</td>
<td>47.7</td>
<td>Ranked 1st/average size 9/15</td>
<td>28.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:6/II:8/II:11 + III:3</td>
<td>5</td>
<td>56.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:5/II:6/II:8 + III:6</td>
<td>7</td>
<td>37.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II:5 + III:1/III:6</td>
<td>7</td>
<td>47.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranked 1st/average size</td>
<td>4/19</td>
<td>48.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the DCM pedigree, the largest haplotype encompassed MYH7, a common gene underlying DCM. Screening revealed a missense mutation of PKP2, which is present in the largest haplotype of the ARVC pedigree.
a highly conserved residue (p.Arg904Cys), with large physicochemical differences between the amino acids, which was shown to segregate with the disease in the pedigree. Furthermore, missense mutations in the surrounding residues 901, 905, 906, and 908 have been identified previously in HCM (28–31).

Based on the formula 200 cM/number of meioses, we expected the mutation-containing haplotype to be around 22.2 cM, which corresponds well with the observed haplotype of 20.8 cM. When performing classic linkage in this pedigree, the maximal logarithm of odds (LOD) score could not exceed 2.4 \([(\text{no. of meioses} - 1) \times \log 2]\) illustrating that haplotype sharing analysis can be a valuable tool in pedigrees too small for linkage analysis.

Our results show that haplotype sharing analysis is a useful tool for narrowing down the area in which to search for a mutated gene. This might be particularly useful in genetically heterogeneous disorders. Moreover, we prove that the theoretical model proposed by Miyazawa et al. to use homozygosity mapping through haplotype analysis to identify shared segments is indeed effective (32). More recently, others have described statistics-based methods that calculate the size of shared haplotypes to map genes in genotyped pedigrees (33–35). The advantage of these methods is that they extend the principle of homozygosity mapping for recessive inherited diseases to cases of AD diseases.

The results of the HST for different numbers of meioses in the DCM pedigree suggest that seven meioses have a high chance of correctly detecting the mutation-containing haplotype (Table 4). Using fewer meioses is however possible, but the risk of detecting a false-positive largest haplotype (i.e. not containing the mutated gene) is substantial. We therefore recommend to apply the HST in pedigrees with at least seven meioses.

Identifying a large shared haplotype is relatively easy, because current molecular biology tools, such as SNP genotyping technology, easily yield hundreds of thousands of reliable genotypes. Subsequent mutational analysis can be performed by sequencing, and current technology is powerful enough to investigate a few candidate regions and confirm the results, thereby contributing significantly to the clinical diagnosis of heritable diseases.

We found genes known to be associated with ARVC and DCM in the largest haplotypes of the studied pedigrees. Mutations in these genes could have been identified by a candidate gene approach, because mutations in \textit{PKP2} and \textit{MYH7} are frequently identified in these diseases (18–25). When a candidate gene approach fails to identify a mutation, a novel disease gene could be identified by screening the largest haplotype. With the decline of costs for SNP arrays, the HST could prove to be more cost-effective than the screening of selected candidate genes.

We recommend haplotype sharing analysis as a tool to assist in identifying genes in those low penetrance Mendelian diseases in which standard tools cannot be used due to lack of substantial pedigree information or in which the mutation rate is low enough to expect founder mutations to be shared among patients. Identifying the largest shared haplotype can be used as a pre-screening method to identify candidate genes.

Acknowledgements

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

Funding

Netherlands Heart Foundation (2003B062 and 2007B132).

Competing interest

The authors declare that they have no conflicting interests.

References


