Consanguinity and Polygenic Diseases: A Model for Antibody Deficiencies

Valentina Di Pierro a, Roberta Zuntini a, Caterina Cancrini b, Andrea Finocchi b, Giulia Angelino b, Paolo Rossi b, Simona Ferrari a

Medical Genetics Unit, S.Orsola-Malpighi University Hospital, Bologna, and Department of Pediatrics, University of Rome ‘Tor Vergata’ and Children’s Hospital Bambino Gesù, Rome, Italy

Key Words
Antibody deficiencies · Common variable immunodeficiency · Consanguinity

Abstract
Primary immunodeficiencies represent a heterogeneous group of disorders of the immune system, predisposing to various types of infections. Among them, common variable immunodeficiency is the most common symptomatic antibody deficiency. It includes several different forms characterized by defects in the terminal stage of B lymphocyte differentiation, leading to markedly reduced immunoglobulin serum levels and increased susceptibility to bacterial infections. The clinical phenotype is complex, including autoimmunity, granulomatous inflammation, lymphoproliferative disorders and malignancies. Rare autosomal recessive mutations in a number of single genes have recently been reported. However, the underlying genetic defects remain unknown in the majority of cases. In order to seek new genes responsible for the disease, we studied a consanguineous Italian family through exome sequencing combined with homozygosity mapping. Six missense homozygous variants passed our filtering selection and at least two of them were associated with some aspects of the pathological phenotype. Our data remark the complexity of immune system disorders and emphasize the difficulty to understand the significance of genetic results and their correlation with the disease phenotype.

Introduction

Primary immunodeficiencies (PIDs) represent a heterogeneous group of disorders of the immune system, predisposing to various types of infections. Many patients suffer from other clinical manifestations, including autoimmunity, inflammation and malignancies. So far, more than 180 PID disease-related genes have been identified [1]. This number is rapidly growing and 19 new genes have been identified after the last update release [2]. It is difficult to establish the reliable prevalence rate of PIDs worldwide. Most of the available PID statistics are based on data from patient registries, which represent a powerful tool in order to assess the natural history and the underlying incidence or prevalence rate of the diseases [3]. Furthermore, a revision of the definition of PIDs is required due to the continuously growing variety of new phenotypes associated
with inborn errors of the immune system, potentially increasing estimates of their impact on the population [4].

It has recently been suggested that PIDs are not as rare as usually thought [5]. The differences observed between country registries can be the consequence of an underestimation of the real number of PID cases due to the limited resources available for the correct diagnosis of PIDs, particularly in developing countries, or the diagnostic delay. This issue is particularly remarkable for PIDs with a mild phenotype or for adult-onset PIDs, which are strongly underrepresented in the registries.

The most frequent symptomatic primary antibody deficiency in adults is common variable immunodeficiency disease (CVID; MIM No. 240500), whose prevalence is estimated to be around 1/150,000 among populations of European origin, based on the data of the ESID 2011 registry [6]. It includes a heterogeneous group of disorders characterized by markedly reduced immunoglobulin serum levels, lack of antibody responses and increased susceptibility to bacterial infections [7]. Lower respiratory tract infections caused by encapsulated bacteria are the most important clinical manifestations leading to chronic lung disease. Autoimmunity, chronic gastrointestinal diseases, splenomegaly, granulomatous inflammation, lymphoproliferative disorders and malignancies occur in CVID patients with variable frequencies [8–12]. The heterogeneity of the CVID phenotypes suggests a complex etiology, which includes different monogenic defects as well as the combined effects of several susceptibility alleles together with environmental factors. In spite of several years of investigation, the genetic defects underlying CVID are still largely unknown. To date, heterozygous and homozygous mutations in the TNFRSF13B gene, encoding the TACI protein, have been identified in approximately 10% of CVID patients [13, 14]. However, the association between the presence of TACI genetic defects and the CVID clinical phenotype has been clearly established only in few cases.

Candidate gene studies have allowed to successfully identifying the molecular basis of autosomal recessive forms of CVID in single patients, suggesting that at least a proportion of CVID cases may be caused by single gene defects, and confirming the presence of genetic heterogeneity. In 2003, a homozygous deletion in the ICOS gene was identified in nine CVID cases [15]. Then, homozygous mutations in the CD19 [16], BAFF-R [17], CD20 [18], CD81 [19] and LRBA genes [20] have been described in few patients born to consanguineous parents. More recently, novel genes involved in CVID phenotype have been discovered using next-generation sequencing approaches, which led to the identification of mutations in the PRKCD [21, 22], and NFkB2 genes [23].

Parental consanguinity may represent a risk factor involved in the CVID pathogenesis. A recent paper published by the DEFI study group [24], focusing on the comparison between CVID patients with or without demonstrated parental consanguinity, pointed out that the clinical phenotype of CVID patients born to consanguineous parents is associated to a higher incidence of splenomegaly, granulomatous disease and lymphoma, suggesting an increased risk for lymphoproliferative diseases. Moreover, one third of the affected offspring of consanguineous unions developed opportunistic infections usually associated with severe T-cell defects, while only 5% of the other group of CVID patients showed such complications. Moreover, the cases with parental consanguinity are expected to present higher frequency of early-onset disease, as reported in a study on a group of CVID Iranian subjects of consanguineous marriage where the rate of parental consanguinity was significantly higher in affected children than in adults (47 vs. 7%, p = 0.002) [25].

The identification of an increasing number of single-gene defects is configuring CVID as a constellation of different diseases, clearly emphasizing its heterogeneity. This perspective opens a number of questions. How did the presence of specific genetic defects shape the evolution of innate and adaptive immunity over time? Why do some genetic defects of the immune system manifest late in life? Can early-onset CVID-like phenotypes really be classified as CVID? The answers to these questions require an integrated approach, involving a systematic clinical and immunological classification of disease subgroups [26, 27] and population-genetic studies, in order to understand the evolutionary dynamics of the immune system and the relationship between the genetic variations in the human genome – single nucleotide polymorphisms (SNPs), copy number variants (CNVs) or mobile elements insertions – and the disease manifestations [28–30].

Whole exome sequencing is providing new insights into the genetics underlying Mendelian traits, and it is a promising tool for identifying new gene defects in CVID, as already reported for some CVID-like disorders [21–23]. The critical point of this technology remains to understand which variants are really relevant for the phenotype. Using whole genome sequencing data from 185 individuals generated as part of the pilot phase of the 1000 Genomes Project [31], MacArthur et al. [32] estimated that, depending on the ethnic background, each individual genome carries 103–121 variants leading to gene loss,
as nonsense, frameshift or large deletions or insertions and mutations that disrupt splice sites, including about 20 homozygous variants. This suggests that some loss-of-function alterations are well tolerated in healthy individuals either because they are benign variations in redundant genes or because they are variants that do not seriously disrupt the gene function. As expected, the number of homozygous variants is higher in consanguineous individuals, increasing the probability to carry severe recessive disease alleles or alleles that have an impact on phenotype and disease risk.

In this paper, we present original data obtained through an integrated approach, which combines homozygosity mapping and whole exome sequencing analysis, applied to a CVID patient born to consanguineous parents.

**Case Report: A Consanguineous Family with a Severe Form of Antibody Deficiency**

We studied a consanguineous Italian family (fig. 1), consisting of two siblings from a first-cousin marriage, affected by an early-onset form of antibody deficiency. Of the two affected sisters, one died of respiratory failure; the other one – the index case – showed a severe immunological phenotype (table 1), characterized by low to undetectable levels of serum immunoglobulins, a decreased number of peripheral B cells, a poor response to vaccines and an increased susceptibility to infections, especially respiratory and digestive tract infections due to *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Candida albicans*, *Clostridium beijerinckii*, *Escherichia coli*, *Enterococcus faecalis* and *Salmonella* types B and C. Furthermore, she had a sepsis due to *Campylobacter jejuni*. She also suffered from other symptoms including bronchiectasis and chronic lung disease, chronic sinusitis, osteopenia, adverse reactions to some medications, chronic diarrhea and duodenal mucosal atrophy. Moreover, the patient experienced several episodes of autoimmune hemolytic anemia, thrombocytopenia and elevated liver transaminase levels, not related to hepatitis, CMV or EBV infections and liver lesions.

Since exome sequencing has been swiftly integrated with homozygosity mapping to accelerate the investigation of recessive disorders in consanguineous families [33, 34], we combined homozygosity mapping and whole exome sequencing in order to identify the gene(s) responsible for the complex phenotype. A two-step approach has been applied: the first step used the genome-wide SNP genotyping to identify autozygous regions to narrow down the search space for possible loci; the second step examined exome sequences to detect genetic variations at basepair resolution. Genome-wide SNP genotyping and homozygosity mapping were performed in both affected sisters using Affymetrix Genome-Wide Human SNP Array 6.0 Nsp/Sty Assay (Santa Clara, Calif., USA) and Runs of Homozygosity utility implemented in PLINK [35], defining 8 homozygosity regions >2 Mb on chromosomes 2, 5, 6, 8, 11 and 22.

![Table 1. Immunological findings in our CVID patient](image)

<table>
<thead>
<tr>
<th>Age</th>
<th>27 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>At the time of the study</td>
<td>childhood</td>
</tr>
<tr>
<td>At the onset of infections</td>
<td>childhood</td>
</tr>
<tr>
<td>At the diagnosis</td>
<td>11 years</td>
</tr>
</tbody>
</table>

**Lymphocyte subsets**

| CD3+, % of total lymphocytes     | 82%      |
| CD3+CD4+                         | 43%      |
| CD3+CD8+                         | 42%      |
| CD4/CD8                          | 1.02%    |
| CD4+CD45RA+, % of CD3            | 36%      |
| CD4+CD45RO+, % of CD3            | 64%      |
| CD8+CD45RA+, % of CD3            | 34%      |
| CD8+CD45RO+, % of CD3            | 66%      |
| CD19+, % of total lymphocytes    | 1%       |
| CD16+CD56+, % of total lymphocytes| 11%     |

**Serum immunoglobulin levels at the onset**

| IgG                              | 183 mg/dl |
| IgA                              | <5 mg/dl  |
| IgM                              | 7 mg/dl   |

**Vaccine responses (one month after immunization)**

| Tetanus toxoid IgG               | 0.3 IU/ml* |
| Autoantibodies present           | none       |

* Reference value of the lab >0.5 IU/ml.

---

Di Pierro/Zuntini/Cancrini/Finocchi/ Angelino/ Rossi/Ferrari

Hum Hered 2014;77:144–149
DOI: 10.1159/000362364
Considering the results of a recent genome-wide study, conducted in a large group of CVID individuals, where the authors detected a high burden of CNVs among which numerous were unique or rare exonic deletions and duplications [36], we performed a CNV analysis on the Affymetrix data, using Partek® Genomics Suite (Partek Inc., St. Louis, Mo., USA). After having excluded the presence of a rare CNV in both patients, genomic DNA of one affected sibling (S2) was evaluated by exome sequencing in order to identify the causative variant. The sample was processed by in solution hybridization using NimbleGen SeqCap EZ Exome Library v3.0 (Roche). Exon-enriched DNA was sequenced as 101-bp paired-end reads by the Illumina HISeq2000 platform following the manufacturer's instruction. Raw image files were processed by the Illumina CASAVA pipeline for base calling and generating the reads set. The reads were aligned with BWA [37] against the human genome reference sequence (hg19 build). PCR duplicate reads were removed with the Picardtools MarkDuplicates utility (picardtools.sourceforge.net), and a GATK package was used for variant identification and filtering. Identified variants were annotated with Annovar based on the NCBI RefGene (www.ncbi.nlm.nih.gov) and UCSC KnownGene (genome.ucsc.edu) databases.

A mean coverage of 100.1× was achieved. We excluded from further analyses variants covered by ≤5 reads, considering a coverage higher than 5× sufficient to identify novel homozygous variants with high specificity, as has already been reported in previous studies [38, 39]. In particular, 98.6% of the targeted bases were covered >5× and about 95% of targeted bases showed a coverage >20×, sufficient to identify novel homozygous variants with high specificity. A total of 19,279 exomic variations from the reference sequence were found, including 34 homozygous novel variants. The subsequent comparison with the homozygosity mapping results identified a total of 178 variants, all single nucleotide nonsynonymous substitutions, in the candidate regions. Among them, after filtering out variants already reported in SNP databases at a frequency >0.005 and variants with uncorrected segregation pattern in the family, 6 missense substitutions passed the selection (table 2).

Initially, we focused our attention on the previously unreported missense variant in C7 gene, a component of the complement system, which provides the initial response to prevent infections by pathogenic microorganisms. To date, more than 20 molecular defects associated with complete or subtotal C7 deficiencies have been reported, leading to an increased risk of recurrent infections caused by Neisseria meningitis [40]. Neisseria infections have never been reported in our patients. Furthermore, the C7 mutations alone do not explain the severe agammaglobulinemia and the low number of B lymphocytes, so they might not be responsible for the increased susceptibility to the large spectrum of infections recurrent in the patient. For these reasons, this case cannot be classified as a classical C7 deficiency.

It is rather likely that the complexity of the general diagnosis reflects an equivalent complexity at the genetic level. In this model, each variant predisposes to some aspects of the final disease, but only when it is associated with other alterations, it plays a pathological role. According to this hypothesis, we investigated a novel homozygous variant in DNAH8, encoding a heavy chain of an axonemal dynein and predicted to be damaging by PolyPhen 2 [41]. Mutations in other genes of the same family have been reported in primary ciliary dyskinesia (MIM No. 244400), a genetically heterogeneous recessive disorder of motile cilia that leads to chronic oto-sino-pulmo-

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Effect</th>
<th>EVS frequency</th>
<th>1000G frequency</th>
<th>dbSNP ID</th>
<th>PolyPhen 2 prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>HJURP</td>
<td>Chr2: 234,749,399</td>
<td>missense p.H676R</td>
<td></td>
<td></td>
<td></td>
<td>benign</td>
</tr>
<tr>
<td>C7</td>
<td>Chr5: 40,977,888</td>
<td>missense p.S308W</td>
<td></td>
<td></td>
<td></td>
<td>damaging</td>
</tr>
<tr>
<td>DNAH8</td>
<td>Chr6: 38,783,340</td>
<td>missense p.Q1144E</td>
<td></td>
<td></td>
<td></td>
<td>damaging</td>
</tr>
<tr>
<td>DAAM2</td>
<td>Chr6: 39,865,007</td>
<td>missense p.I856T</td>
<td>0.0041</td>
<td>0.0014</td>
<td>rs61748650</td>
<td>benign</td>
</tr>
<tr>
<td>CUL7</td>
<td>Chr6: 43,021,587</td>
<td>missense p.G4S</td>
<td>0.0003</td>
<td></td>
<td></td>
<td>benign</td>
</tr>
<tr>
<td>MICAL3</td>
<td>Chr22: 18,364,065</td>
<td>missense p.R749Q</td>
<td>0.0004</td>
<td></td>
<td></td>
<td>benign</td>
</tr>
</tbody>
</table>

EVS = Exome Variant Server, NHLBI Exome Sequencing Project; 1000G = 1000 Genomes.
nary diseases [42]. Due to its role in respiratory cilia motility, DNAH8 could be associated with chronic sinusitis and respiratory failure, one of the main symptoms of the index case and her sister’s cause of death. We also found a substitution in MICAL3 that, even if it is predicted to be benign by PolyPhen 2, could influence the concentration of liver enzymes in plasma, as hypothesized in a recent genome-wide association study [43], and explain the high levels of liver transaminases measured in the index case. The role of the other detected variants is not likewise clear in defining the examined phenotype. For example, defects in CUL7 have been described in 3M syndrome type 1 (MIM No. 273750), an autosomal recessive disorder characterized by distinctive facial features and severe prenatal and postnatal growth retardation not diagnosed in either of the affected siblings.

As CVID is a complex and heterogeneous disease, it is possible that in some cases it is not inherited as a single gene Mendelian disorder, but rather fits a model in which mutations in a small number of genes may interact genetically to produce the phenotype. Accordingly, evidences for digenic inheritance have been found in some genetic diseases such as facioscapulohumeral muscular dystrophy type 2 [44] and thrombocytopenia with absent radius syndrome [45].

We are aware that none of the mutations identified in our patients have a conclusive role in the disease pathogenesis. In particular, we cannot exclude the possibility that the clinical phenotype might be caused by a homozygous change in regulatory or evolutionary conserved sequences outside the coding regions.

A multidisciplinary approach, combining whole genome sequencing with the analysis of gene and protein expression profiling, is required to eventually clarify the genetic causes of CVID in this family and may help us to understand the physiological and pathogenetic significance of the identified genetic variants. This is particularly important for a heterogeneous disease such as CVID, which may include different disorders involving immune and autoimmune dysregulation, defects of innate immunity, lymphoproliferative and autoimmune-lymphoproliferative syndrome.

Our data highlight, one more time, the genetic complexity of antibody disorders, emphasizing the difficulty to translate the results of better knowledge to the bedside.

Acknowledgements

This research has been supported by a grant from Telethon Foundation (GGP12052 to S.F.). We would like to thank Prof. Isabella Quinti (Sapienza University of Rome) and the members of the Medical Genetics Unit for their helpful advices and discussion.

References


8 Chapel H, Cunningham-Rundles C: Update in understanding common variable immunodeficiency disorders (CVIDs) and the management of patients with these conditions. Br J Haematol 2009;145:709–727.


10 Yong PF, Thaventhiran JE, Grimbacher B: ‘A rose is a rose is a rose’, but CVID is not CVID: common variable immune deficiency (CVID), what do we know in 2011? Adv Immunol 2011;111:47–107.


