Identifying Genes Responsible for Intellectual Disability in Consanguineous Families

Zafar Iqbal\textsuperscript{a} Hans van Bokhoven\textsuperscript{a, b}

\textsuperscript{a}Department of Human Genetics, Nijmegen Center for Molecular Life Sciences, and \textsuperscript{b}Department of Cognitive Neurosciences, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

Key Words
Autosomal recessive intellectual disability · Heterogeneity · Consanguinity · Next-generation sequencing · Exome sequencing

Abstract
Consanguinity is an important determinant of birth defects including intellectual disability (ID). Consanguineous populations have a relatively high prevalence of autosomal recessive forms of intellectual disability (ARID), which constitute a highly heterogeneous group of disorders both in their clinical presentation and in their genetic aetiology. The availability of large cohorts of consanguineous families and the advent of next-generation sequencing techniques is currently accelerating the pace of gene identification in ARID. Because of the extreme heterogeneity, it is anticipated that hundreds of ARID (candidate) genes will be identified in the near future. With this robust progress, the proof of causality of the identified candidate genes is challenging. To this end, genetic recurrence, cellular assays and animal modelling would serve as three non-exclusive strategies, in order to assign causality to a certain gene. Extensive genetic investigations in consanguineous populations will help in reducing the total disease burden through proper genetic counselling. Moreover, such findings will be helpful to elucidate different pathways and further for possible therapeutic interventions.

Intellectual Disability

Intellectual disability (ID) is a common form of cognitive impairment. Two to three percent of the general population can be defined as having ID based on a theoretical assumption when –2 standard deviations are converted into a percentile [1–3]. However, a population-based meta-analysis revealed an ID prevalence of 1% [4]. Affected individuals have an intelligence quotient (IQ) below 70, with an onset before the age of 18 years and significant limitations in two or more adaptive skills, such as self care, communication skills, problem solving and logical thinking [5]. ID is usually further sub-classified according to the IQ level: mild (IQ level: 50–69), moderate (IQ level: 35–49), severe (IQ level: 20–34) and profound ID (IQ level: <20) [1, 6, 7]. The prevalence of mild ID is 85%, that of moderate ID 10%, whereas only 5% of the patients...
Genes Responsible for Intellectual Disability in Consanguineous Families

Hum Hered 2014;77:150–160
DOI: 10.1159/000360539

are classified as severe/profound ID [4, 8]. The term ID can be dissected into syndromic ID and non-syndromic ID. Patients with syndromic ID have other associated abnormalities, such as structural brain and skeletal anomalies, metabolic defects and facial dysmorphisms, while patients with non-syndromic ID only have ID in common without consistent recognizable associated features [9].

**Causes of ID**

ID is an extremely heterogeneous group of disorders. Genetic and non-genetic factors contribute approximately equally to the development and to the prevalence of ID [10]. Non-genetic factors include environmental insults, nutritional factors, toxic exposure during pregnancy and obstetric-related complications. Socio-economic factors, such as level of income, family circumstances and cultural environment, are likely to have a contributory role to the prevalence of ID. Thus, the prevalence of ID is two-to threefold higher in the poor and underdeveloped World as compared to Western countries [11]. In poor countries malnutrition is one of the main risk factors. Prenatal and perinatal complications such as hypoxia, fetal brain damage due to the maternal transmission of infectious diseases (e.g. rubella, toxoplasmosis and cytomegalovirus infection), trauma and low birth weight are also important contributors to the prevalence of ID. In developed countries, especially in the USA, fetal alcohol syndrome has a prevalence of 0.5–2 cases per 1,000 live births and is the most common avoidable cause of ID [11–13].

The genetic causes of ID can be subdivided into chromosomal abnormalities and monogenic/multigenic disorders, including metabolic defects. Chromosomal anomalies are of different types, such as aneuploidies (numerical disorders), submicroscopic deletions and duplications, copy number variations (CNV) and balanced/unbalanced translocations. Despite recent advances, the genetic causes of about 40% of ID remain unexplained [14]. Chromosomal anomalies are a major cause of ID, as 22% of severe ID cases have an underlying chromosomal defect [15]. The most common cause of ID among chromosomal aneuploidies is trisomy 21 (Down syndrome) with a high prevalence rate of 1/1,000 live births [16, 17]. CNVs can cause recurrent conditions of ID usually associated with other features, such as in Williams-Beuren syndrome, Prader-Willi syndrome and Smith-Magenis syndrome [18]. These CNVs can affect several kb to Mb of DNA that give rise to a contiguous gene syndrome involving altered dosage of several genes. However, in some cases it has been shown that haploinsufficiency of a single dosage-sensitive gene can explain the ID, such as EHMT1 in Kleefstra syndrome [19], KANSL1 in Koolen-de Vries syndrome [20], RAI1 in Smith-Magenis syndrome [21], LIS1 in the 17p13.3 microdeletion syndrome [22] and MECP2 in the MECP2 duplication syndrome [23].

Metabolic disorders are often associated with ID. In a large study, screening of the known metabolic conditions revealed that congenital metabolic disorders account for approximately 3% of cases with unexplained ID [24]. The clinical features of some metabolic disorders can be prevented, such as hypothyroidism and phenylketonuria [25]. In developed countries, specific programs have been installed – such as the heel prick/Guthrie card test as a neonatal screening program in the UK and the Netherlands – for the early detection and treatment of such disorders [11]. Unfortunately, in most developing countries, there is no systematic neonatal screening program employed yet [12].

To date, more than 450 genes have been implicated in ID [26]. The monogenic causes of ID can be classified into X-linked, autosomal dominant and autosomal recessive ID.

**Monogenic Forms of ID**

The estimated total number of genes underlying ID (dominant and recessive) is in the range of 2,000–3,000 [26]. Until recently, genetic studies on ID have focused on the X chromosome [27], mainly because of the accessibility of large numbers of families with apparent X-linked inheritance of an ID phenotype (XLID) [28, 29]. As a consequence, a total of 112 genes have been associated with XLID [30; Kalscheuer VM et al., submitted]. Based on these studies, it is now estimated that XLID accounts for 10–12% of all males with ID [31].

As autosomal dominant ID (ADID) mostly results in reduced fecundity, it is very rare to find large pedigrees with autosomal dominant ID that are suitable for traditional genetic studies, such as linkage analysis. An isolated occurrence is most commonly seen in clinical practice in Western countries. This may suggest that de novo DNA rearrangements and variations are a common cause of autosomal dominant ID [32, 33]. Indeed, analyses in various large patient cohorts have reveal that CNVs may account for 15–20% of all ID individuals, depending on the patient selection used [34, 35]. Recently, whole-exome sequencing and whole-genome sequencing approaches
have been applied to identify de novo DNA variants in ID and autism patients [36–41]. These studies estimate that severe ID could be caused by potential de novo variants in approximately 35–45% of isolated patients [36, 42].

The discovery of causative gene mutations in autosomal recessive ID (ARID) has been hampered by the paucity of large families and the great heterogeneity of the condition [43–45]. For this reason, the identification of causative genes in syndromic forms of ARID has been most successful, as it allowed the identification of multiple families with apparently the same condition and hence the pooling of genetic data. Consequently, a considerable number of genes (>250 genes; Dr. T. Kleefstra, pers. commun.) have been identified in syndromic forms of ARID. By contrast, the identification of genes for non-syndromic ARID (nsARID) is lagging far behind [11, 26, 31, 46]. So far, >40 novel loci have been reported for nsARID, which suggests that no frequent genetic defect is to be expected, at least in the studied cohorts [47–52]. Until 2007, only three genes (PRSS12 [43], CRBN [44] and CC2D1A [45]) had been implicated in nsARID [48]. In 2007, Najmabadi et al. [51] started a project exploiting consanguineous marriages in the Iranian population (consanguinity rate: >40%), and identified eight novel loci for nsARID. In subsequent studies, several novel genes have been identified, such as GRIK2 [53], TUSC3 [54], TRAPPC9 [50], ST3GAL3 [55], and ZC3H14 [56], with homozygous mutations associated with nsARID. Subsequently, mutations in several other genes have been linked to nsARID: TECR in a consanguineous Israeli family [57], KIAA1033 in a large consanguineous family from Oman [58], and finally MED23 [59].

With the advent of high throughput techniques in the field of DNA sequencing, many more ARID genes have been identified. Recently, targeted next-generation sequencing (NGS) has been applied to 136 consanguineous Iranian families of population in which mostly a single locus had previously been identified by homozygosity mapping. This study revealed segregating variants in 29 known ARID genes and another 50 variants in genes that had not previously been linked to ARID [46]. Of the latter group, 29 genes were implicated in nsARID. However, the non-syndromic phenotype connected to specific gene mutations may not be permanent, as it has been witnessed that syndromic features may only be recognized when more families and patients will be identified and carefully compared for their phenotypic features. Recently another gene, CRADD, has been associated with nsARID [60]. Taken together, 41 genes have been associated with nsARID to date, of which only 10 genes are found to be recurrently mutated in multiple ARID families (table 1). Thus, the causality of mutations in most nsARID genes is not proven beyond doubt given their single occurrence in ID, and caution is warranted as has recently been demonstrated for previously established XLID genes [61]. Nonetheless, the above study with 50 new candidate ARID

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene</th>
<th>MIM No.</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PRSS12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>606709</td>
<td>[43]</td>
</tr>
<tr>
<td>2</td>
<td>CRBN</td>
<td>609262</td>
<td>[44]</td>
</tr>
<tr>
<td>3</td>
<td>CC2D1A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>610055</td>
<td>[45]</td>
</tr>
<tr>
<td>4</td>
<td>GRIK2</td>
<td>138244</td>
<td>[53]</td>
</tr>
<tr>
<td>5</td>
<td>TUSC3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>601385</td>
<td>[54]</td>
</tr>
<tr>
<td>6</td>
<td>TRAPPC9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>611966</td>
<td>[50]</td>
</tr>
<tr>
<td>7</td>
<td>ST3GAL3</td>
<td>606949</td>
<td>[55]</td>
</tr>
<tr>
<td>8</td>
<td>ZC3H14</td>
<td>613279</td>
<td>[56]</td>
</tr>
<tr>
<td>9</td>
<td>TECR</td>
<td>610057</td>
<td>[57]</td>
</tr>
<tr>
<td>10</td>
<td>KIAA1033</td>
<td>NA</td>
<td>[58]</td>
</tr>
<tr>
<td>11</td>
<td>MED23</td>
<td>605042</td>
<td>[59]</td>
</tr>
<tr>
<td>12</td>
<td>ADR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>102750</td>
<td>[46]</td>
</tr>
<tr>
<td>13</td>
<td>ADRA2B</td>
<td>104260</td>
<td>[46]</td>
</tr>
<tr>
<td>14</td>
<td>ASC3C</td>
<td>614217</td>
<td>[46]</td>
</tr>
<tr>
<td>15</td>
<td>ASCL1</td>
<td>100790</td>
<td>[46]</td>
</tr>
<tr>
<td>16</td>
<td>C11orf46</td>
<td>612295</td>
<td>[46]</td>
</tr>
<tr>
<td>17</td>
<td>C12orf57</td>
<td>NA</td>
<td>[46]</td>
</tr>
<tr>
<td>18</td>
<td>C8orf41</td>
<td>614426</td>
<td>[46]</td>
</tr>
<tr>
<td>19</td>
<td>C9orf86</td>
<td>610615</td>
<td>[46]</td>
</tr>
<tr>
<td>20</td>
<td>CASP2</td>
<td>600639</td>
<td>[46]</td>
</tr>
<tr>
<td>21</td>
<td>CCNA2</td>
<td>123835</td>
<td>[46]</td>
</tr>
<tr>
<td>22</td>
<td>COO5</td>
<td>NA</td>
<td>[46]</td>
</tr>
<tr>
<td>23</td>
<td>EEF1B2</td>
<td>600655</td>
<td>[46]</td>
</tr>
<tr>
<td>24</td>
<td>ELP2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NA</td>
<td>[46]</td>
</tr>
<tr>
<td>25</td>
<td>ENTPD1</td>
<td>601752</td>
<td>[46]</td>
</tr>
<tr>
<td>26</td>
<td>FASN</td>
<td>600212</td>
<td>[46]</td>
</tr>
<tr>
<td>27</td>
<td>HIST3H3</td>
<td>602820</td>
<td>[46]</td>
</tr>
<tr>
<td>28</td>
<td>INPP4A</td>
<td>600916</td>
<td>[46]</td>
</tr>
<tr>
<td>29</td>
<td>MAN1B1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>604346</td>
<td>[9]</td>
</tr>
<tr>
<td>30</td>
<td>NDST1</td>
<td>600853</td>
<td>[46]</td>
</tr>
<tr>
<td>31</td>
<td>PECR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>605843</td>
<td>[46]</td>
</tr>
<tr>
<td>32</td>
<td>PRMT10</td>
<td>307150</td>
<td>[46]</td>
</tr>
<tr>
<td>33</td>
<td>PRRT2</td>
<td>614386</td>
<td>[46]</td>
</tr>
<tr>
<td>34</td>
<td>RALGDS</td>
<td>601619</td>
<td>[46]</td>
</tr>
<tr>
<td>35</td>
<td>RGS7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>602517</td>
<td>[46]</td>
</tr>
<tr>
<td>36</td>
<td>SACER</td>
<td>611611</td>
<td>[46]</td>
</tr>
<tr>
<td>37</td>
<td>TRMT1</td>
<td>611669</td>
<td>[46]</td>
</tr>
<tr>
<td>38</td>
<td>UBR7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>613816</td>
<td>[46]</td>
</tr>
<tr>
<td>39</td>
<td>ZCCHC8</td>
<td>NA</td>
<td>[46]</td>
</tr>
<tr>
<td>40</td>
<td>ZNF526&lt;sup&gt;b&lt;/sup&gt;</td>
<td>614387</td>
<td>[46]</td>
</tr>
<tr>
<td>41</td>
<td>CRADD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>603454</td>
<td>[60]</td>
</tr>
</tbody>
</table>

NA = Not available. *A recurrence has been identified. †This gene has been implicated in ID and ASD. ‡This gene has been implicated in nsARID with associated dysmorphic features.
Genes clearly indicates the power of studying recessive disorders in highly consanguineous populations [46]. It should be noted that the majority of genes implicated in ARID stem from an extensive genetic study [46] in the Iranian population (table 1), which warrants further studies in other (consanguineous) populations.

**Consanguinity: Focus on Pakistan**

Consanguineous marriage is defined as the union of two individuals having a common ancestor [62]. It is estimated that worldwide 10.4% of all marriages occur among relatives [63], but this percentage is much higher in Arab and central Asian countries. Consanguineous marriages are an important determinant of autosomal recessive diseases [48]. Pakistan belongs to that group of countries where the rate of consanguinity is very high (>60%) with 17–38% first-cousin marriages [64, 65]. Consanguinity is one of the main causes of infant mortality and ID in Pakistan [66]. In different consanguineous populations, the risk for birth defects varies from 2.7 to 15.8% [67]. Consanguineous marriages are considered one of the main aetiological factors in the high prevalence of severe ID among non-European immigrants in Sweden [68]. In humans, the high levels of inbreeding are somehow linked to a generally reduced cognitive ability [69]. Likewise, in mice and Drosophila, it has been shown that high rates of inbreeding have adverse effects on cognitive traits along with other fitness-related traits [70, 71].

There are several factors that directly or indirectly play a role in the still increasing rates of consanguinity in Pakistan. Religious, demographic and social factors influence the prevalence of blood marriages [62, 72]. People tend to follow a number of social values. They prefer to get married within the same religion or even within the same sect of a religion (e.g. Sunni, Shia or Wahabi), within the same caste or sub-caste (e.g. Syed (Bukhari or Gilani, etc.), Rajpoot (Rana or Bhatti, etc.) or Jutt (Cheema or Warraich, etc.), within the same province or district and/or with a person speaking the same language. In the Punjab province, people also prefer a partner whose grandparents came from the same district or village of India at the time before the Partition in 1947. Status and welfare are also important factors (fig. 1). Such economic factors seem to be the strongest binding force, as families want to keep the wealth within the circle of relatives. In addition, they also preserve familial wealth by saving large dowries for brides to be paid to non-relatives. Therefore, large age differences between spouses are observed frequently in this society.

The level of complexity of the Pakistani community can also be seen among Pakistanis who have migrated to other countries. In August 2010, a television program (by T. Ahmad, a British news reporter) revealed that a third of all children born with recessive genetic diseases in Britain belong to the Pakistani community [73], whereas Pakistanis constitute only about 1.5% of the total British population [74]. In Norway, the incidence rate of progressive encephalopathy is estimated to be...
seven times higher in the Pakistani community as compared to the endogenous population [75]. Moreover, genetic counselling in the Pakistani population is also found to be more difficult as compared to other ethnic communities for reasons of guilt and shame. Therefore, those who are at risk of being mutation carriers will remain uninformed through such psychosocial behaviour [76], leading to a continuous transmission of pathogenic alleles.

There is an urgent need to increase the general awareness about the adverse effects of consanguinity in societies where blood marriages are very common. To achieve this, Pakistani communities should be informed about the positive effects of genetic analyses. Along these lines, a decline in the frequency of consanguineous marriages has been observed in some Arab countries, such as Jordan [77] and Lebanon [78]. This was achieved by elevating the levels of female education, urbanization, industrialization and an improved economic status of families, much alike the transitions seen in Western countries in the 19th and 20th centuries [79]. These examples provide a clear roadmap to develop strategies in order to address the issues of increasing rates of consanguinity in Pakistan and other countries that have a high consanguinity rate.

**Homozygosity Mapping**

Homozygosity mapping is a powerful tool to map autosomal recessive gene defects. It has proven to be useful in locating the genetic defects in both consanguineous and non-consanguineous (unrelated) unions [80]. In the latter situation, homozygous disease gene mutations may reflect identity by descent/autozygosity, which means that the disease allele, derived from a common ancestor, is segregating through multiple generations to occur homozygously in the affected individual. This is also the case in consanguineous families, but here the parents are by definition more closely related (second cousins or closer) [63, 64]. Homozygosity mapping takes advantage of the fact that not only the mutation but also the surrounding region is identical on both chromosomes (fig. 2). These regions usually encompass multiple variable DNA markers, such as microsatellite markers or single nucleotide polymorphisms (SNPs). By using these DNA markers, runs of homozygosity (ROH) can be determined. Therefore, this technique is called homozygosity mapping, identity-by-descent mapping or autozygosity mapping [81, 82].

The length of ROH is determined by the rate of recombination events over the generations within a family. Thus, the ROH will be much shorter in children of distantly related individuals as compared to the progeny of first or second cousins. It is estimated that offspring of first-cousin unions contain on average 11% homozygosity genome-wide with 20 homozygous segments >3 Mb. The size of the homozygous region containing a recessive disease gene is estimated to be 26 Mb on average [83]. The percentage of homozygosity decreases after successive recombination events in each generation, resulting in an average length of ROH reduced to a few Mb or less [80, 83].

Families with members that have ID and consanguineous marriages can contribute to the mapping of recessive genetic defects. As compared to outbred (non-related) families, studying consanguineous families is advantageous as they often have multiple sibships and large regions of homozygosity. Although homozygous muta-
tion are a common cause of disease in consanguineous families, it also happens regularly that consanguinity is unrelated to the disorder. Thus, compound heterozygosity can be causative for recessive genetic disorders and one should be aware of this, particularly when studying outbred families or families with affected individuals residing in a single sibship [84, 85].

**Modern DNA Techniques for Identifying Genetic Defects**

Several modern DNA techniques that have emerged in the last few years have accelerated genetic studies, both in the mapping of genetic defect by high-resolution genotyping as well as in the identification of the causative mutation by massively parallel sequencing [11, 26, 29].

**High-Resolution Genotyping Techniques**

The mapping of genetic defects requires genomic markers to pinpoint the position of the defect in the DNA [86]. Until a decade ago, the use of microsatellite markers (highly polymorphic with di-, tri- and tetra-nucleotide repeats) was a standard procedure to genotype genomes. Panels of 300–400 microsatellite markers, equally distributed across the genome with a resolution of approximately 10 Mb (average spacing) genome-wide, were used widely [87]. This method has been successful in large consanguineous families to identify ROH that encompassed the respective genetic defect [88]. However, the resolution is too low to study small non-consanguineous families [88]. Only some ten years ago, high throughput and robust microarray-based SNP genotyping was introduced, which has the additional advantage that they are also highly powerful in the detection of CNVs [89]. Present SNP arrays can detect deletions and duplications up to 10 and 20 kb resolution, respectively, and can also identify ROHs.

**Next-Generation Sequencing**

For over three decades, Sanger sequencing has been the dominant approach for DNA sequencing [90]. This classical method is still the gold standard procedure in genomic diagnostic laboratories. There are, however, several limitations to this technique, regarding the length of individual reads, the number of strands that can be sequenced in parallel and the relative high cost per nucleotide [91, 92]. Recently, massively parallel sequencing, also called NGS or second-generation sequencing, approaches have become widely implemented. A considerable number of new discoveries have since been documented in the field of genetics because of robust NGS technologies, which can massively generate hundreds of Mb to Gb of DNA sequence in a single run depending on the type of platform [90, 92].

NGS can be used to sequence the entire genome of an individual, referred to as whole-genome sequencing, as well as only the exons of the genome, which is referred to as whole exome sequencing or exome sequencing. Currently, exome sequencing is widely used for the elucidation of the molecular basis of (heterogeneous) monogenic disorders. The enriched exomes cover nearly all protein-coding genes, which constitute approximately 1.5% of the genome [93]. About 85% of all Mendelian disease-causing mutations are located within the protein-coding regions [94, 95]. Exome sequencing became popular because of the cheaper costs, improvements in the bioinformatics tools, relatively easy interpretation of the data and less time consumption of the experiments as compared to whole-genome sequencing.

NGS technologies have been very successful in identifying disease-causing genes regardless of the types of DNA variations, phenotypic or genetic heterogeneity and inheritance patterns [96]. The combination of homozygosity mapping with targeted NGS has been powerful for finding autosomal recessive defects [46, 97, 98].

Thus, exome sequencing appears to be a highly attractive ‘all-in-one’ technique that is able to reveal ROH, CNVs and pathogenic variants. As a proof-of-principle, in 2011, Pippucci et al. [99] developed a strategy denoted EX-HOM (Exome HOmozygosity). They studied two affected siblings born to first-cousin parents for a recessive disease. Using exome sequencing, the data revealed several homozygous regions (>1 Mb), accounting for 290 Mb in total. The causative gene search in these homozygous regions identified only three candidate variants, among which a FA2H mutation was the most plausible one. To overcome the problems of genetic heterogeneity and the need of large pedigrees, EX-HOM combines homozygosity mapping and sequencing as a single-step approach to identify recessive genetic defects. Recently, Seelow and Schuelke [100] have offered a new module integrated in the HomozygosityMapper program (http://www.homozygositymapper.org/), in which the raw NGS data can be uploaded to identify ROH. The analysis of NGS data by such programs obviates the prior performance of costly SNP genotyping.

In short, SNP-based homozygosity mapping will become obsolete and it will be more efficient and cost-effective to use available family members for exome se-
quencing. The detection of CNVs by using NGS is still in its early phase, as it critically relies on the coverage efficiency (especially in a single mate pair) and there are many pitfalls to consider [101]. The problem of coverage is less problematic for the detection of homozygous deletions, and microdeletions within exons or spanning several exons can be successfully identified by (targeted) exome sequencing [12]. For heterozygous rearrangements, the detection problem can be circumvented by the use of paired-end/mate pair sequencing, which is not yet commonly used in routine laboratory settings, and by increasing the depth-of-coverage of the reads [102]. Paired-end sequencing can generate sequences of up to 500 bp, whereas mate-pair sequencing can sequence the ends of 2–5-kb DNA fragments [96]. These strategies are very useful to investigate chromosomal rearrangements, such as small CNVs, translocations and inversions. Recently, Talkowski et al. [38] have demonstrated that mate-pair sequencing can efficiently sequence the breakpoints of balanced chromosomal translocations in patients with neurodevelopmental disorders.

The application of NGS for the detection of recessive mutations in consanguineous families from inbred populations can result in multiple variants that map to large homozygous regions shared by the affected individuals of the respective families. This was not an issue in the past, when classical mapping and subsequent Sanger sequencing of a selected number of candidate genes was performed. In our studies, we have identified several families in which mutations affecting multiple genes segregated with the phenotype [98]. The identification of recessive mutations in non-consanguineous families – both in inbred and outbred populations – appears to be highly efficient, as usually less homozygous variants need to be confirmed and validated by Sanger sequencing, which may out-weight the disadvantage of small family size.

There are pros and cons of using NGS over traditional sequencing methods. Because of the uniformly optimized experiments and high depth of NGS, it is less likely to miss a causative variant as compared to conventional methods. On the other hand, using NGS increases the chances of finding multiple variants, which complicates the conclusive selection of the true causative mutation. In addition, NGS has the potential to uncover incidental variants that are not relevant to the disorder under investigation, but which might reveal important information regarding future health concerns of the patient.

What Makes DNA Change, a Pathogenic Variant?

The identification of multiple candidate variants complicates the assignment of the truly causative mutation for the phenotype under investigation. In our studies, multiple DNA variants have identified segregating with the phenotype in single consanguineous families [98]. Similarly, proof of causality is also needed when targeted NGS or exome sequencing reveals only a single segregating variant in a family. In general, there are three non-exclusive strategies to determine which of such variants is causative:

Recurrent Genetic Findings

The identification of recurrent variants – mutations in the same gene in different individuals – is considered to be the most powerful evidence to assign the possible pathogenic role of dysfunction of a certain gene. For this, cohorts containing patients either with non-syndromic ID or with similar syndromic ID phenotypes have to be screened. As ARID is an extremely heterogeneous group of disorders [46], it will be essential to extend the screening to even larger patient groups of different ethnic backgrounds. Collaborations among research groups and diagnostic centres worldwide could be a rich source to find recurrences. In addition, the development and continuous curation of DNA variant databases analogous to publicly available CNV databases, such as DECIPHER and ECARUCA [103, 104], are in progress and should promote the identification of matching gene mutations in patients with an overlapping phenotype. Likewise, the presence of comparable gene variants in control cohorts at frequencies >1% should be used as an exclusion criterion. For example, a high occurrence of truncating gene mutations in the exome variant server of Washington University (http://evs.gs.washington.edu/EVS/) would diminish the likelihood of this gene being involved in a severe disorder such as ID.

Cellular Modelling (in vitro Studies)

In vitro studies might also be helpful in supporting the pathogenic role of the identified variant. A wide variety of biological assays can be used, either in patient-derived cells or in transfected cell approaches [105]. For studying causality of ID candidate genes, electrophysiological studies using rodent primary neuronal cell lines transfected with mutant transgenes or shRNAs appear to be particularly powerful [106, 107]. A promising addition to the repertoire of cellular assays is provided by the use of human induced pluripotent stem cells (hiPSC) [108–112].
Animal Modelling (in vivo Studies)

In vivo studies can also provide additional proof about the causality of a candidate gene. The choice of the animal model depends on the properties of the gene, conservation, suitability to obtain the desired phenotype, time consumption and the budget. In case of inactivating mutations, knockdown/knockout animal models, such as mouse, Drosophila or zebrafish, are a good source to observe phenotypic consequences.

In conclusion, the extreme heterogeneity in ID can be a major complication to provide proof of causality for very rare gene variants. Nevertheless, if resources are available, the latter two strategies might be helpful to predict the biological consequences of such rare gene variants.

Final Remarks

ID is one of the most serious health concerns in the world. For the past five years, the pace of identifying genetic defects in neurodevelopmental conditions has significantly accelerated due to the introduction of massively sequencing techniques. This will increase our knowledge about the molecular and genetic causes of hitherto enigmatic entities and will help filling in the pieces of puzzles in respective pathways that are disrupted in disease and subsequently will pave the way towards therapeutic interventions in a more efficient way. The field of ID therapeutics is in a development phase and hopefully some therapeutic options will emerge in the near future. At present, this large group is best served by improved medical care, which can be optimized by offering conclusive genetic diagnosis. Knowledge of the genetic basis of the ID disorder will allow optimal management of the disease, genetic counselling and prognostic testing. In general, the early detection of the genetic cause of disease will provide the best opportunities for this kind of personalized medicine. The benefits of early detection are evident by comparing medical genetic practice between countries. In the Pakistani population, recessive disorders are common, and especially some metabolic disorders have a high prevalence. The implementation of a national neonatal screening program would help in decreasing the disease load in the population. As more families would be diagnosed with frequent metabolic disorders leading to ID, they could immediately benefit from available therapies.

The high consanguinity rate in combination with the lack of awareness of the associated risks for the recurrence of recessive genetic disorders is another threat for the Pakistani as well as other consanguineous populations. Genetic counselling of the affected families would be essential in order to stop disease propagation. In December 2000, Israel passed a legislation concerning genetic testing, genetic counselling and the protection of privacy rights of patients regarding genetic findings [113]. Such actions taken by other countries, especially where the consanguinity rate is high, will help to educate families regarding the spread of disease alleles to the next generations and to reduce the overall disease burden in the society. Extensive genetic studies will definitely help in elucidating the genetic causes of ID, which in turn will help affected families in terms of genetic counselling. Genetic studies in the highly consanguineous Iranian population have contributed a considerable proportion of causes of nsARID. Investigations of other consanguineous populations at a similar scale would help in reducing the total disease burden through proper genetic counselling. An intensified continuation of such investigations is highly warranted to stop the disease propagation in consanguineous populations.

Acknowledgements

H.v.B. was supported by the European Union’s Seventh Framework Program under grant agreement number 241995, project GENCODYS. Z.I. was supported by the Higher Education Commission (HEC), Islamabad, Pakistan.

References

Genes Responsible for Intellectual Disability in Consanguineous Families


Ahmad T: Pakistan: inbreeding themselves to death. Channel 4, August 2010.


Shaw A, Hurst JA: ‘I don’t see any point in telling them’: attitudes to sharing genetic information with relatives among British Pakistani adults referred to a genetics clinic. Ethn Health 2009;14:205–224.


