Aus dem Institut für Medizinische Genetik und Humangenetik
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: The WASH complex member SWIP

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Fabienne Ropers

aus Freiburg im Breisgau

Gutachter/in: 1. Prof. Dr. K. Sperling

2. Prof. Dr. rer. nat. G. Rappold

3. Prof. Dr. med. E. Schwinger

Datum der Promotion: 07.09.2012

Abstract

High throughput sequencing has greatly facilitated the elucidation of genetic disorders, but compared to recessive X-linked diseases, the search for genetic defects underlying autosomal recessive disorders remains a challenge. In a large consanguineous family with autosomal recessive intellectual disability, we have combined homozygosity mapping, targeted exon enrichment and high throughput sequencing to identify the underlying gene defect.

Using homozygosity mapping we identified a 13 Mb region on chromosome 12, which could be narrowed down to 12.5 Mb using microsatellite finemapping. This region contains 159 genes. After targeted exon enrichment of the specified region, high throughput sequencing was performed.

Only two molecular changes remained after appropriate SNP filtering, including a non-synonymous sequence change in the SWIP (Strumpellin and WASH-Interacting Protein) gene. SWIP is a member of the recently discovered WASH complex, which is involved in actin polymerization and multiple endosomal transport processes. Based on high pathogenicity and evolutionary conservation scores, as well as functional considerations, this gene defect is considered to be the cause for ID in this family. In line with this assumption, we could show that this mutation leads to significantly reduced SWIP levels and to destabilization of the entire WASH complex. Thus, our findings suggest that SWIP is a novel gene for autosomal recessive ID.

Zusammenfassung

Neue Sequenzierungsverfahren (*engl*. high throughput sequencing) haben die Aufklärung genetischer Defekte vereinfacht. Im Vergleich zu rezessiven X-gebundenen Erkrankungen, stellt die Suche nach Genen für autosomal rezessive Krankheiten noch eine Herausforderung dar. Um den ursächlichen Gendefekt in einer grossen konsanguinen Familie mit autosomal rezessiver mentaler Retardierung zu identifizieren, haben wir eine Kombination von Homoyzygotiekartierung, gezielter Exonanreicherung und Hoch-Durchfluss-Sequenzierung angewandt.

Die Homozygotiekartierung ergab eine 13 Mb grosse Region auf Chromosom 12. Diese Region konnte durch Feinkartierung mit Mikrosattelitenmarker auf 12.5 Mb eingegrenzt werden. Die identifizierte Region enthielt 159 Gene. Nach gezielter Exonanreicherung dieser Gene wurde eine Hoch-Durchfluss Sequenzierung durchgeführt.

Im Anschluß an die kritische Durchsich der SNP Varianten blieben zwei molekulare Veränderungen übrig. Eine davon war eine nicht synonyme Sequenzveränderung in dem SWIP Gen (Strumpellin and WASH-Interacting Protein). SWIP ist ein Bestandteil des vor kurzem beschriebenen WASH-Komplexes, welcher eine Rolle in der Aktin Polymerisierung und mehreren endosomalen Transportprozessen spielt.

Aufgrund der evolutionären Konserviertheit der betroffenen Aminosäure sowie funktioneller Erwägungen wurde dieser Gendefekt als ursächlich für die Erkrankung angesehen. Übereinstimmend mit dieser Annahme konnten wir zeigen, dass die Mutation zu einer deutlichen Reduktion der SWIP Expression und zu einer Destabilisierung des gesamten WASH Komplexes führt. Zusammenfassend belegen unsere Ergebnisse, dass wir mit SWIP ein neues Krankheitsgen für autosomal rezessive mentale Retardierung gefunden haben.

Table of Contents

Introduction:

Intellectual disability

The definition of mental retardation involves three key criteria: (i) low general intellectual functioning as measured by IQ score, (ii) difficulties in adaptive behaviour as expressed in conceptual, social and practical adaptive skills and (iii) manifestation before 18 years of age. Conceptual skills refer to the understanding of e.g. language and literacy, money and self-direction. Social skills include interpersonal interaction, self-esteem and gullibility, and practical adaptive skills involve daily living skills, occupational skills, travel, use of money etc. [1].

In recent years, organisations such as the American Association of Mental Retardation (AAMR) proposed a change in terminology and replaced the term “mental retardation” with “intellectual disability” (ID). Consequently, their name changed from AAMR to American Association on Intellectual and Developmental Disabilities (AAIDD). The term “intellectual disability” is perceived to better reflect the nature of the disorder and convey more dignity and respect to the subjects. It is emphasized that the social-ecological concept of disability better exemplifies the interaction between the individual and its environment. Rather than being an absolute, invariant trait of a person, it focuses on the role that individualized support can play in enhancing individual functioning [2].

Prevalence

The prevalence of ID is dependent on the diagnostic criteria used. There has been an ongoing debate about them. To categorize ID into severe or moderate forms, intelligence quotients are used. Generally, an IQ lower than 70-75 is interpreted as sub-average intellectual functioning.
A binary classification of severity ID is often used, with an IQ<50 being considered as severe ID, although others use a lower cut-off level of IQ < 40 [3]. Here, it is important to note that standard IQ tests are based on normally distributed intelligence levels in the population, and were not designed to reliably estimate individual capacities and skills on the lower end of the scale. Furthermore, an IQ test is a measure not of innate intellectual functioning but of performance on a set of skills defined by test instruments. Most tests are dependent on language skills and designed against a specific cultural background.

The scores obtained by mere IQ testing can be of little value from an educational perspective. Thus, the intensity of support needed and the ability to learn and adapt should be included in estimating the measure of disability.

Besides variations in classification systems, the registers used to estimate the overall prevalence of ID also contribute to the considerable variation in prevalence among countries and regions.
Estimates using educational data sources tend to be higher than estimates based solely on IQ measurements. Using an IQ cut off <40 to define severe ID, the prevalence in children of school age is around 1.5/1000, using a cut-off < 50 IQ points 3-4/1000 ([4],[5] reviewed in [3]). Moderate ID defined as IQ 40/50-70, has an estimated prevalence of 5/1000, as observed in a Californian study [6]. In a recent meta-analysis of 52 studies, the prevalence of ID (IQ<70) was 10.37/1000, corresponding to the estimates in the aforementioned studies [7].
A higher prevalence of ID among male children was first noted by Penrose et al. [8]. This difference in prevalence is most prominent in mild ID and can be partly attributed to biological factors, such as X linked-conditions and a higher susceptibility (in males as compared to females) to negative effects of low birth weight and prematurity on IQ. It is also possible that boys are more likely to be referred to special educational services.

In developing countries the impact of ID has been studied to a lesser extent, though the prevalence is higher. Estimates of severe ID from developing countries are consistently above 5/1000, whereas those from developed countries are less than 5/1000 [9].

Etiologic factors that predominantly affect third world countries, such as iodine deficiency, congenital infection, lacking hygiene, malnutrition of mother and child are conditions that bear a possibility for intervention.

Etiology

It can be argued that the milder forms of ID represent the lower end of a normally distributed intelligence curve with a mean at 100, whereas more severe forms of MR (IQ < 40-50) are regarded as consequences of severe disease and birth complications, or genetic defects. This assumption is supported by the higher proportion of clear etiological causes in severe ID compared to moderate ID [10].

There are several classification systems of etiological factors for intellectual disability. One classification divides risk factors and causes of intellectual disability into [11]:

Risk factors and causes of ID		
1. genetic	syndromic and non-syndromic, metabolic diseases	
2. external prenatal disorders	maternal conditions and exposure	
	multiple births	
	high maternal age	
3. perinatal disorders	low birth weight and small head circumference	
	infections	
	birth trauma	
4. postnatal disorders	infections	
	trauma	
	environmental factors, e.g.:	low parental education
		lack of stimulation
		social environment
		malnutrition and dietary deficiency
		other conditions related to poverty
5. unknown etiology	The percentage of cases with unknown etiology varies between different studies but is commonly thought to be around 50-60% in moderate to severe ID, and up to 85% in mild ID (as reviewed in(12)).	

As has been repeatedly shown, low birth weight is the strongest predictor for ID [6, 12]. Socioeconomic factors also seem to play an important role. Many studies have consistently found that the prevalence of moderate ID was strongly associated with low socioeconomic status [13]. More specifically, Stromme *et al.* found that children of parents with lower socioeconomic status had a significantly increased risk of moderate ID, but not of severe ID [14]. When ID is subdivided into isolated ID and ID with other neurological conditions, a strong inverse relationship between maternal education and isolated ID was found [15]. Growing up in a disadvantaged social environment - even more than individual and familiar factors - could contribute to a decline in IQ scores in young school children [16].

Given the heterogeneous etiology of ID, one could assume that in low-income countries factors related to poverty such as infections and malnutrition outweigh genetic causes. However, there is no clear evidence for this assumption. Some genetic diseases associated with ID, are more common in low-income countries, such as haemoglobinopathies. Consanguinity, which is especially prominent in certain populations in the Middle East, is another factor increasing the frequency of genetic causes in ID. Intermarriage is most common between first degree cousins, increasing the likelihood of accumulation of homozygous mutations (reviewed in [9]).

Genetic causes

A differentiation can be made between monomorphic ID and syndromic ID.
Based on accompanying anomalies, a further classification is possible. This classification allows clustering of patients and facilitates the search for mutations. Indeed, for several syndromic forms of ID, the underlying gene defect has been identified. The same holds true for microcephaly or central nervous system malformations associated with ID.

However, the majority of ID cases do not have accompanying anomalies. Until recently, the genetic diagnostic panel was restricted to general methods like karyotyping or exclusion of mutations in the most common genes.
The establishment of comparative genome hybridization and single nucleotide polymorphism microarrays has allowed a high-resolution evaluation of the entire genome for submicroscopic deletions and duplications, as well as for single nucleotide polymorphisms. This molecular karyotyping has major advantages, especially in the genetic evaluation of non-syndromic ID patients, as it does not depend on phenotypic classification.

The overall percentage of genetic causes in the etiology of ID can only be estimated, as the detection rate is highly dependent on the pre-selection of the patient group, the clinical information available, the diagnostic approach, the quality of the sample material, and investigator's experience.

In a study (conducted in 2006) using conventional karyotyping, locus-specific FISH analysis and molecular karyotyping, a group of around 1100 ID patients of unknown etiology with or without other anomalies was examined. In 40% of the patients, a genetic aberration could be identified [17]. However, the study design contained a selection bias, as referral to the genetic center (and thus inclusion) was decided by a clinician. Thus, children with an identifiable cause of ID (such as extreme prematurity with concomitant low birth weight) were less likely to be referred, the proportion of genetic causes in the overall population might be overestimated. Indeed, other studies have described that 25-35% of ID can be attributed to genetic causes, with the percentage increasing proportionally with severity of ID (among others [18-20].

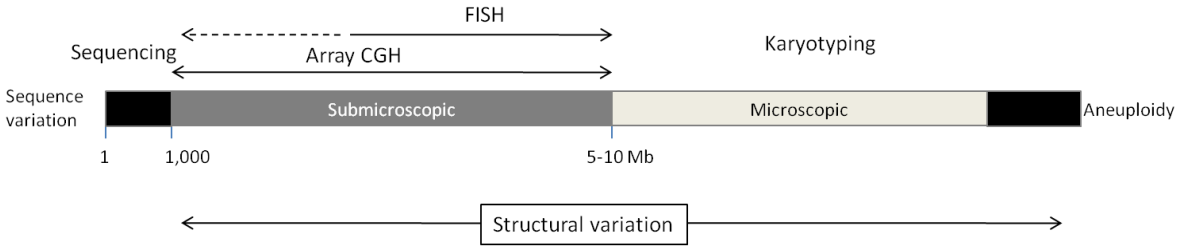


Figure 1. From chromosome to single nucleotide.
Genetic aberrations form a continuum from single base mutations to whole chromosomal aneuploidies. Different technologies are required to interrogate genetic variants at different levels, from sequencing to conventional karyotyping.

Modified from PhD thesis D. Koolen, Department of Human Genetics, Nijmegen, Netherlands

The detected genetic variants range from chromosomal aneuploidy to single nucleotide mutations.
Trisomy 21 is the most common chromosomal aberration, accounting for 9% of ID cases with multiple congenital anomalies [21].
Around 5% of patients with idiopathic developmental delay carry chromosomal aberrations detectable by microscopic analysis [22, 23].
The advent of high-resolution comparative genome hybridization has now enabled the genome-wide detection of copy number variations as small as 1 kb, and high throughput sequencing enables whole genome sequencing. Several studies led to the conclusion that rare, *de novo* submicroscopic Copy Number Variations (CNVs) account for 10(-15)% of cases with idiopathic ID and a normal karyotype [24-27].

Numerous CNVs were found to be associated with specific phenotypes. Recurrent interstitial microdeletions and -duplications, involving at least 19 genomic intervals, have been identified in patients with ID [22, 25, 26, 28].
The vast majority of these have been shown to result from non-allelic homologous recombination (NAHR) between sufficiently large duplicated sequence blocks separated by a stretch of unique DNA. Thus, these studies have greatly expanded the spectrum of known genomic disorders (reviewed in [29]). In many of these disorders and several microdeletion syndromes, dosage-sensitive genes have been identified that are responsible for most of the observed clinical features [28, 30].
However, many CNVs show incomplete penetrance, as several unaffected carriers were found, underscoring the clinical variability. In addition, a large number of clinically neutral CNVs have been detected by Array CGH. Indeed, more than 1000 CNVs can be present in a single individual, and CNVs are now considered to be a common form of structural genomic

variation [31].
In order to advance the clinical interpretation of these CNVs, the continuous collection of CNVs and their associated phenotypes in databases is of major importance [22].

The development of high throughput sequencing (or next generation sequencing- NGS) including the bioinformatic tools to handle the huge amount of data, represent yet another great step in the search for genetic defects. Moreover, since the costs are rapidly dropping, this technique will become widely applicable in the near future. The future, however, has already begun. This has been illustrated in an exceptional study based on 272 consanguineous families with early-onset, autosomal recessive intellectual disability. First, the candidate gene regions were identified by homozygosity mapping. In a next step, all relevant exons were enriched by hybrid capture with oligonucleotide arrays and screened for mutations by NGS. Apart from mutations in 23 known disease genes, 50 novel (candidate) genes for autosomal recessive have been elucidated. Moreover, these findings gave new insights into the complex protein networks underlying brain development and function [32]. It is realistic to assume that NGS will soon be implemented as universal intake test in Clinical Genetics.

X-linked ID

As early as in 1938, a higher rate of ID was observed in males. This observation has been confirmed by several studies, and the male to female ratio averages 1.3:1 [3].
X-linked mental retardation has been extensively studied in recent decades, leading to the identification of over 90 genes (see review [22]). Clinical and genetic observations have shown that X-linked ID is a very heterogeneous set of conditions responsible for a large proportion of inherited mental retardation. X-linked ID is estimated to cause 10% of all inherited cases of ID in males [33]. Around 2/3 of XLID is estimated to be non-syndromic [34].
One third of the currently known X chromosomal genes involved in ID have been found in non-syndromic ID. Half of those have been implicated both in syndromic and non-syndromic MR, and considerable phenotypic variability was observed even within families.
Most of the causative X chromosomal genes were found through cooperation in international consortia. Linkage analysis, breakpoint analysis, screening and sequencing of exonic regions of the X chromosome, and screening for CNVs lead to the elucidation of a genetic cause in around 50% of the families with syndromic and non-syndromic forms of XLID (reviewed in [35, 36]).
The most common gene involved in X linked mental retardation is *FMR* leading to fragile X

syndrome and estimated to cause up to 25% of all X-linked ID cases, followed by mutations in the *ARX* gene. The other genes each account only for a small percentage of all patients [34, 37, 38].

Autosomal ID

In contrast to XLID, only in recent years attention turned to the investigation of autosomal forms of ID. This might seem surprising as in the majority of the patients the disease is transmitted in an autosomal mode of inheritance.
Autosomal-dominant *de novo* mutations are thought to contribute significantly to sporadically occurring ID cases. Many of these are due to balanced translocations, and breakpoint analysis has contributed to the identification of numerous causative genes. The recent detection of *de novo* CNVs by array-based techniques also helped identify new disease genes. Not only the occurrence of *de novo* CNVs in at least 10% of ID cases, but also the recent finding of *de novo* mutations 6/10 individuals with ID using high throughput sequencing [39], are indicative of the relative contribution of dominant mutations to the pathogenesis of ID.

The finding of 50 new genes for autosomal recessive ID making use of high throughput sequencing and exon enrichment can be regarded as a major breakthrough and proof for the wide range of genes that are important for neurological functioning [32].
Previously, the heterogeneity of the disorder and the rather small families in western countries had complicated genetic mapping of the disease locus. The genes elucidated at the time of our study, had all been identified by homozygosity mapping in large consanguineous families. Their function is summarised in Table 2.

Table 2. Genes involved in non syndromic autosomal recessive intellectual disability (Identified till January 2011)

Gene	Function	Reference
PRSS12	Encodes neurotrypsin, a neuronal serine protease that is found in presynaptic nerve endings of cortical synapses of hippocampal, cortical, amygdale and motor neurons. It cleaves agrin. Agrin is involved in the formation of filopodia on neuronal axons and dendrites. Activity-dependent presynaptic exocytosis of neurotrypsin and the resulting proteolytic cleavage of agrin at CNS synapses have been proposed as a mechanism promoting the activity-dependent formation of dendritic filopodia.	MIM 606709 [40], [41-44]
CRBN	Encodes Cereblon, a potassium channel modulator. It binds to the C terminus of large conductance calcium-activated potassium BK channel alpha-subunit (KCNMA1; MIM 600150). Mutations appear to disturb the development of these large-conductance Ca-activated K-channels, which causes increased intracellular Ca sensitivity and results in faster activation, and slower deactivation kinetics Direct association of rat Cereblon with Kcnma1 was confirmed by immunoprecipitation in brain lysate, and the 2 proteins colocalized in cultured rat hippocampal neurons.	MIM 609262 [45-47]
CC2D1A	Encodes Freud-1, which binds to a dual repressor element of the serotonin receptor and an intronic repressor element of the dopamine-D2 receptor, transcriptionally repressing these receptors. Both receptors function as pre-synaptic autoreceptors regulating the neurotransmission of serotonin and dopamine, respectively, and have a role in memory and behaviour.	MIM 610055 [48, 49]
GRIK2	Encodes glutamate receptor, ionotropic, kainate 2 (or GluR6), a subunit of kainate receptor. Human GluR6 containing kainate-preferring receptors are present at pre- and postsynaptic level, and have primary involvement in excitatory synaptic activity. Kainate receptors alter the excitability of mossy fiber axons and have been reported to play a role in the induction of long-term potentiation (LTP) at mossy fiber synapses in the hippocampus.	MIM 138244 [50, 51]
TUSC3	Encodes a subunit of an enzymatic complex catalyzing transfer of oligosaccharide side chain on nascent proteins. Mutations in TUSC3 belong to group of congenital disorders of glycosylation (CDGs).	MIM 601385 [52, 53]
TRAPPC9	Encodes NIBP. It interacts directly with NIK and IKK beta, which results in the activation of the NF-kappa B Pathway. NIBP is involved in axonal outgrowth in vitro, possibly involved in neuronal cell survival.	MIM611966 [54-57]
TECR	Encodes trans-2,3-enoyl-CoA reductase. Synaptic glycoprotein involved in the synthesis of very long chain fatty acids (VLCFA).	[58, 59]

Considering the heterogeneity of ID and the fact that over 30 genes are known to cause X linked non-syndromic ID, and the X chromosome contains ~ 4% of the genes, these few known autosomal NS-ID genes are likely to represent the tip of the iceberg [60, 61].

Patients and aim:

In this study we identified a new disease gene for autosomal recessive intellectual disability in a large consanguineous Omani family with 6 affected children.
The offspring of consanguineous parents have considerable homozygous regions that possibly contain recessive mutations inherited from a common ancestor.
Using homozygosity mapping we identified a single large homozygous region present in affected but not in unaffected subjects. Subsequent exon enrichment and sequencing of the homozygous interval, led to the finding of a new disease gene for ID.

Patients

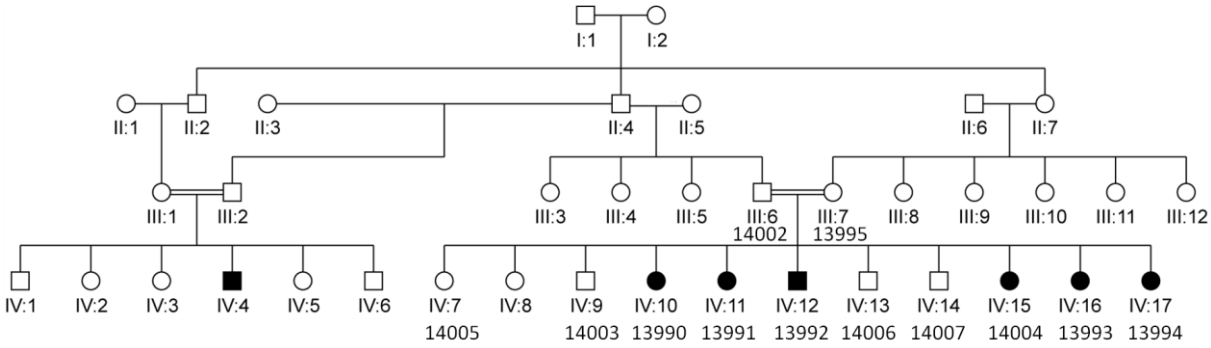


Figure 2. Pedigree

All patients (filled symbols) have moderate to severe ID. The Arabic numbers represent the DNA numbers.

Figure 2 displays the pedigree of the Omani family. III:1 and III:2 and III:6 and III:7 are first degree cousins. In the first branch of the family, one son is affected (IV:4), his 5 siblings are healthy. In the second branch 6 out of 11 children are affected, both male and female.
All affected children have moderate to severe ID/DD (IQ 35-50). They have severe learning impairment, poor language skills and poor adaptive skills. Age of walking was normal, though other (especially fine) motor development was severely delayed. They are of short stature (<3rd centile), with normal head circumference and no dysmorphic features. For more details see table 3. Non-genetic causes for mental retardation could be excluded. Whereas ID was the only clinical abnormality in 6 of the 7 affected children, subject IV:4 also showed signs of spasticity. MRI imaging was normal, but in the absence of perinatal monitoring, a history of perinatal hypoxia could not be ruled out.

Table 3	Clinical features

Intellectual functioning

- delayed acquisition of speech
- poor vocabulary
- recognize around 10 letters of the Arabic alphabet
- Recognize 10 digits, except for mirror images of Arabic 2 and 6
- unable to read and write
- unaware of own age
- know own name and that of sibs/parents
- able to recognize pictures only of simple objects as ball, cat, dog, but not policeman (man)

Conceptual, social, practical adaptive skills

- delayed toilet training
- delay in eating with cutlery
- require supervision with dressing, washing and using toilet
- deficit in conceptual adaptive skills (e.g. unaware of concept money)
- friendly behaviour but no comprehension of social situations

Motor development

- floppiness in infancy
- poor fine motor skills and coordination
- mild dysarthria

Material and methods

Introduction

The work presented in this thesis is the result of a collaboration with the Department of Genetics, Muscat, Sultanate of Oman (identification, examination and counselling of the patients), Cologne Center for Genomics and Institute for Genetics (SNP genome profiling and homozygosity mapping), Max Planck Institute for Molecular Genetics Berlin (exon enrichment, high throughput sequencing, data analysis) and Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, France (creation of SWIP expressing cell lines, Immunoblot experiments). It was published in Human Molecular Genetics 2011 [62].

As part of the research, I visited Oman in February 2009, to see the family of our interest (and multiple others, currently under investigation in our institute). I examined patients IV:12 and IV:17 and their parents (III:6 and III:7).
All linkage calculations were performed in Cologne, but for this and another project, I also performed the calculations in our lab in Berlin.
PCR, all Sanger Sequencing (patients, controls, Irani families), microsatellite finemapping was conducted in our lab.
In this methods and material chapter, I will first describe general principles and then the software and methods applied. The last section will contain information about methods and experiments carried out in the Max Planck Institute Berlin and the Laboratoire d'Enzymologie et Biochimie Structurales.

Section I: general principles

Autozygosity and Homozygosity

The term autozygosity describes homozygosity for markers that are identical by descent. Autozygosity mapping is used for rare autosomal recessive diseases manifesting in consanguineous families, since in those families individuals are likely to be autozygous for markers linked to the disease locus as the mutation together with flanking DNA segments is likely to be inherited from a common ancestor.
If a child is homozygous for a particular marker allele, two identical alleles can be inherited from a common ancestor through both the maternal and paternal line (consanguinity) resulting in autozygosity for that allele (identical by descent), or a second copy of the same allele has entered the family independently leading to homozygosity (identical by state). The

rarer the allele is in the population, the greater the likelihood that homozygosity represents autozygosity [63].

As outlined in the introduction, intellectual disability is not a rare disease in itself, but with a few exceptions, each genetic aberration is only responsible for a small percentage of all cases. A plethora of genetic defects can lead to the same ID-phenotype. Therefore, each ID-carrying pedigree can be regarded as carrying a probably rare mutation, making homozygosity mapping a useful strategy in the case of a recessive mode of inheritance.

The aim of autozygosity/homozygosity mapping is to find homozygous regions identical by descent shared among inbred affected individuals and, as a result, to identify the disease locus.
Homozygosity mapping is usually performed with genome-wide linkage analysis using markers such as microsatellite markers, or predominantly single nucleotide polymorphisms (SNPs). The major advantage of SNPs is their high density throughout the entire genome, providing a high resolution. The major disadvantage is their lower informativity when compared to the polymorphic microsatellite markers.

With linkage analysis in a recessive disease model with a consanguineous pedigree, all homozygous regions segregating with the disease will be detected and heterozygous or non-segregating homozygous regions will be excluded.
The size of these regions depends on several factors:

- The degree of parental consanguinity.
- The number of informative family members.
- The distribution of recombination events.

Whole genome SNP mapping Array

The family was genotyped using the Affymetrix GeneChip® Human Mapping 250 K Array. The mapping was carried out at ATLAS Biolabs GmbH.
The GeneChip® Human Mapping 250K Sty I Array is a powerful SNP genotyping tool for investigation of genetic diseases. 238,000 SNPs are genotyped in one experiment. In a first step DNA is digested by Sty I, followed by adaptor ligation to the Sty-specific base overhang and amplification using adaptor-specific primers. PCR conditions are optimised to preferentially amplify 200-1100 bp fragments. After amplification DNA is fragmented, labelled and hybridised to the Genechip.

Allele-specific hybridization is a way to distinguish allelic variants at the DNA level.
Each SNP is interrogated by 26-40 different 25 bp oligonucleotides. For each SNP, 2 complimentary oligonucleotides and 2 mismatching oligonucleotides are synthesized. The latter have one mismatch in the centre of the complimentary sequence.
The complimentary oligos provide information on the phase of the SNP (AA, AB, BB), the mismatching oligos provide information on the specificity of the hybridization.

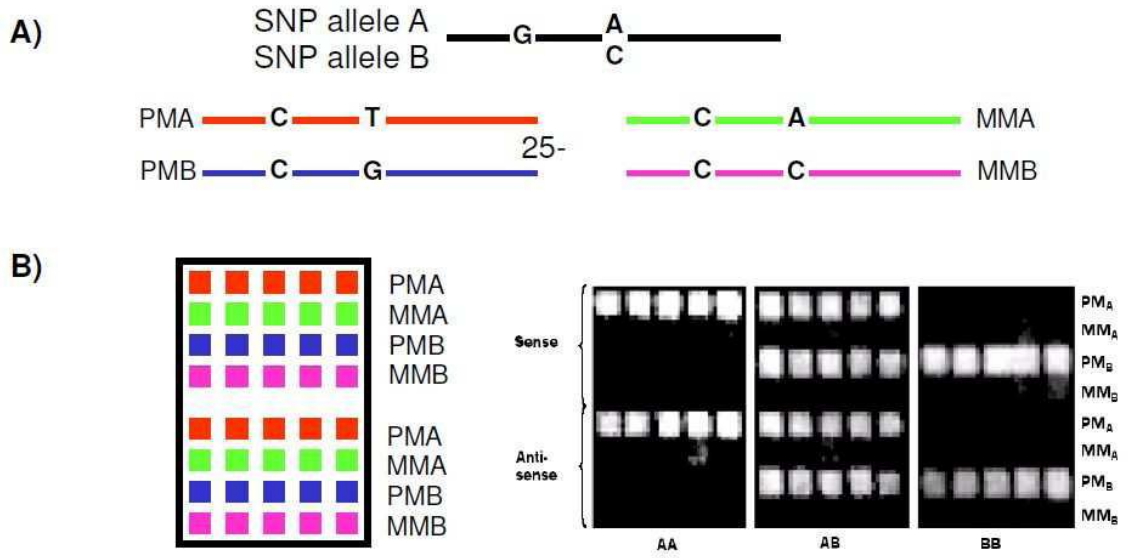


Figure 3. Principle of hybridization. A) Design of mismatch and perfect match oligonucleotides for each SNP. PM: Perfect match (Specific for each allele) MM: Mismatch (one mismatch in the center of PM oligo). **B)** Example of allele-specific yhbridization with 40 probes.

Source: Affymetrix GeneChip® Human Mapping 10K Array manual, Affymetrix website

The allele-specific hybridization signals are measured in intensity. A grid is applied on the resulting image file, and then analysed by provided software. Genotype data are saved in a genotype file. This file can be used for subsequent linkage analysis with appropriate software as Merlin and Genehunter (see below).

Linkage and recombination

To understand the general principle of linkage analysis it is important to discuss the phenomenon recombination first.
If loci lie on the same chromosome one could expect them to segregate together. However, a phenomenon called recombination can interfere with this segregation. In meiosis I the pairs of homologous chromosomes are in an adjacent position and exchange DNA segments, in a process called crossing-over. In crossing-over, only two of four chromatids are involved,

resulting in a recombination fraction of 50%. The recombination fraction never exceeds 50%. Recombination is a function of physical distance, but there are regional differences in the probability of recombination, and one crossing-over event decreases the probability of a second crossing-over event in the vicinity. A genetic map has been created, where distance is defined by the frequency of crossing-over, and described in centiMorgan. By definition, one centiMorgan equals 1% probability of recombination.
Because of these regional differences in frequency of crossing-over events, there are differences between the genetic map and a physical map of the genome.

Linkage analysis

In linkage analysis the probability that loci are linked (i.e. a marker and the disease gene) is calculated. A marker is any polymorphic Mendelian character that can be used to follow a chromosomal segment through a pedigree.
Two genetic loci are linked if they are transmitted together from parent to offspring significantly more often than expected under independent inheritance (0.5). Linkage is a function of the physical and genetic distance between marker and genetic locus: If genetic marker and disease locus are in close proximity, the recombination fraction will be low, and they will be transmitted together. Hence, usually a stretch of adjacent markers will be transmitted *en block*. This cotransmitted stretch of markers is referred to as haplotype.

To define linkage, information is needed regarding the pedigree, disease status of individuals, and marker position. Most importantly, informative meioses are required. A meiosis is informative for linkage when one can identify whether or not the gamete is recombinant [63]. The proportion of recombinant gametes is referred to as the recombination fraction.

Under the assumption that both loci are linked, with a given recombination fraction Theta, the probability of a recombinant meiosis is Theta, and the probability of a non-recombinant meiosis is (1- Theta).
If loci are transmitted independently, the probability of recombination is 50%.
The logarithm of the ratio of the probabilities (i.e. the probability that loci are linked versus the probability that loci are not linked) is called LOD score (Figure 4).

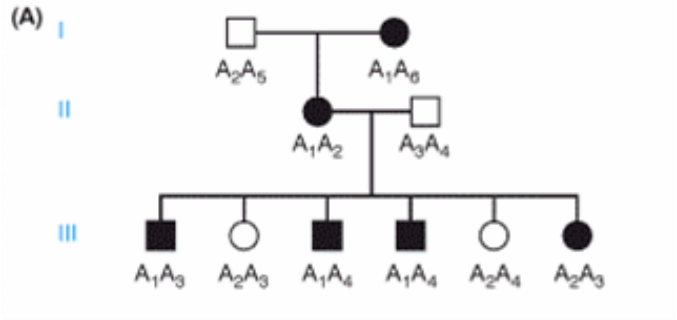


Figure 4. LOD Score Calculation

If marker A1 is linked to the disease locus, there are 5 non-recombinants and 1 recombinant.

Under the assumption of linkage, the likelihood given recombination fraction (Theta), equals $(1 - \theta)^5 \times \theta$.

The likelihood with recombination fraction (½) equals $(½)^{6.}$

The ratio between the 2 likelihoods (hypotheses) is

$$(1 - \theta)^5 \times \theta \quad / \quad (½)^6$$

LOD = logarithm of odds = $^{10}\text{Log}\ ((1 - \theta)^5 \times \theta / (½)^6)$

The maximum LOD score will be found for the most probable recombination being θ = 1/6 with a LOD score of 0.9 [63].

The LOD score is calculated for different values of Theta, the maximum LOD score will be found for the most probable recombination frequency. Large positive scores are evidence for linkage (or cosegregation), and negative scores are evidence against it.

The a priori probability that 2 markers are transmitted together is relatively low, since there are 23 chromosome pairs, and even if markers were located on the same chromosome, the probability of linkage is low. The a priori likelihood of linkage is estimated to be 1:50. When permitting a 5% error rate (95% confidence interval), a LOD score above 3 (log 1000) is regarded as an indication of significant segregation of markers and disease locus.

With increasing amount of markers used in an analysis, the probability of a coincident transmission of 2 markers, that are not truly linked, increases. A stringent procedure would be to increase the cut-off according to the following rule: significant LOD (3) plus Log (N marker). For 1000 markers, the significant LOD score would be >= 6. Since markers on the same chromosome are not transmitted independently, and as markers for which segregation has been excluded increase the chance of other markers to segregate, a LOD of 3 or 3.3 is usually accepted.

When the linkage between 2 markers is evaluated, it is referred to as two-point linkage analysis.

Often genome-wide genetic markers are used, most commonly SNP markers. The high resolution guarantees a dense coverage of the entire genome, but the major disadvantage is the lower informativity of the markers. Multipoint linkage analysis helps to overcome problems caused by the limited informativity, through the simultaneous analysis of multiple (more than two) loci at a time. The relative order of these loci/markers, and their position can be extracted from a genetic map.
The likelihood of the hypothesis linked versus unlinked is tested for any position of the disease gene with a known map of markers.

Section II: Methods used at the Institute of Medical and Human Genetics, Charité, Berlin

Linkage analysis SOFTWARE

As outlined above, linkage analysis can be used to identify autozygous regions. To define linkage, information is needed regarding the pedigree, disease status of individuals, penetrance of the disease, mode of inheritance and marker maps with information on their physical and relative position. There are several mathematical programs designed to perform multipoint linkage analysis. These programs, like Merlin, Allegro and Genehunter, were initially designed for highly informative but smaller microsatellite markers sets (<1000 markers)[64, 65]. Our SNP arrays contained 240,000 markers, exceeding the power of the software.

To overcome this problem, the analysis is performed with subsets of markers using a sliding window mode.

ALOHOMORA software was used for converting the genotyping data into appropriate linkage formats with defined subset sizes [66].

Genotype data obtained from SNP genotyping are introduced and subsets are created.

Coded pedigree information must be introduced, as well as genetic marker map data and allele frequencies. Allele frequencies introduced into the linkage analysis software will affect the LOD score, since linkage disequilibrium is derived or calculated with allele frequency information provided by the user. We used Caucasian frequencies, since Arabic allele frequencies were not available.

A calculation was performed with a 20k subset of the markers, allowing linkage analysis per chromosome without the need of subsets. After identification of a linkage interval on chromosome 12, a detailed analysis with all markers in the area was performed.

Alohomora permits quality control of the introduced pedigree and genotype data in several manners.

Pedigree information control

It is possible to check if the gender information specified in the pedigree file corresponds with the genetic data, by assessing the heterozygous SNPs on the X chromosome. In males, the amount of heterozygous SNPs must theoretically equal zero.

Incorrect pedigree information with regard to kinship can also be identified with Alohomora. Individuals sharing the same degree of relationship will have a similar degree and pattern of

allele sharing. First, pairs are formed according to their relationship, such as siblingpairs, parent-offspring pairs, unrelated individuals etc. Then the mean proportion of allele sharing and its variance are calculated for each pair. Sibpairs are expected to share more alleles than unrelated individuals, as do parent-offspring pairs. However, the variance in sib-pairs will be larger than in parent-offspring pairs, as they can share 2, 1, or zero alleles [64].

Genotype data check

Markers introduced as genotype data can be AA, AB, BB, or of insufficient quality, indicated as no call. Through Alohomora an additional program called Pedcheck can be run, to eliminate Mendelian errors. This software is not only designed to detect Mendelian errors but contains an algorithm to identify and exclude the individual most likely to be in error for that particular SNP.

Another algorithm defines unlikely genotypes. These are equivalent to double recombinations in a short chromosomal segment. Causes for a seemingly double cross-over are genotyping errors and incorrect SNP position in the genetic map.

Parametric versus non-parametric linkage analysis

If linkage analysis is carried out under the assumption of a specific mode of inheritance for the trait locus, the calculation is referred to as a parametric analysis. In contrast, non-parametric linkage analysis does not require a genetic model and is based on shared chromosomal segments in affected subjects. The pedigree of our family indicated an autosomal recessive mode of inheritance, hence this was the selected disease model. We assumed a complete penetrance of the disease.

Ultimately, before starting the true linkage analysis, an allele frequency in the general population must be specified. This is an estimate, derived from the disease frequency and expected allele frequency. In our analysis we assumed an allele frequency of 0.0001.

All calculations were carried out by the ATLAS Biolabs GmbH. However, detailed calculations for this and another research project were carried out in our lab, using all settings and programs mentioned.

Microsatellite fine mapping

Microsatellites are short sequence repeats of 2-6 bases spread throughout the genome. Due to increased DNA polymerase slippage during replication of these repetitive sequences, microsatellites are highly polymorphic resulting in a great interindividual variability.

Before the advent of SNP genotyping, microsatellite markers were used to define

chromosomal segments linked to a disease trait.
With their higher resolution and cheaper and faster methodology SNP genotyping arrays have largely replaced microsatellite genotyping. However, microsatellite markers are still widely used in confirming and fine mapping of loci found by linkage analysis with SNP genotyping data.
In analogy with SNP mapping, for microsatellite marker mapping, the number of informative meioses is crucial: The larger the number, the higher the likelihood of a recombination event. A recombination event in the identified chromosomal segment in an affected individual reduces the segment length shared by all affected subjects, reducing the amount of candidate genes.

With the use of NCBI Map Viewer and SNP Viewer (www.ncbi.nlm.nih.gov/mapview) microsatellite markers that flanked or resided in the chromosomal segment found through SNP linkage analysis were identified. The website also provided information on the expected repeat length, heterogeneity, and suitable primers for PCR.

To the 5' end of the proposed forward primer an M13-sequence was added. Incorporation of this sequence in the PCR product enabled a cost-effective labelling of the PCR product, as in each reaction a universal M13-FAM labelled primer could be used. FAM (6-carboxy-fluorescein) is a fluorescent dye.
For PCR, two forward primers were used in a fixed ratio of 1: 4, with a sequence-specific primer with 5'-end M13 sequence at 2 pmol/reaction, and a FAM- labelled M13 primer at 8 pmol/reaction. The reverse primer was used at 10 pmol/reaction (see figure 5).
After optimisation of PCR conditions, the PCR reaction was conducted with patient samples.

The M13- labelled PCR products were separated according to size and charge using capillary gel electrophoresis.
In order to analyse the labelled PCR product, a mix of 9 µl HiDi-Formamide and 0.2 µl size standard GeneScan 500 HD [ROX] was added to 0.5-1 µl PCR product and denatured for 3 min at 95°C. Subsequently the samples were loaded via electrokinetic injection into the capillary of a 3730 (ABI) DNA-Analyzer and separated electrophoretically. The data obtained were evaluated with the GeneMapper v.3.7, using the size standard to estimate the product length.

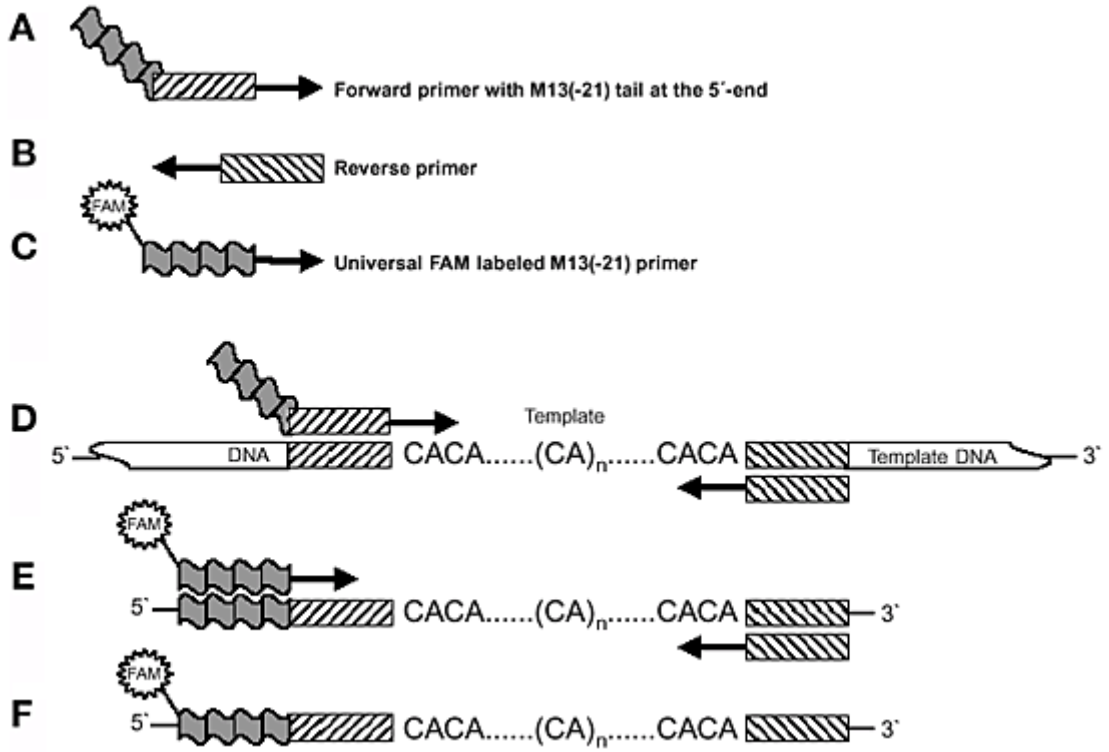


Figure 5. Amplification scheme for the one-tube, single-reaction nested PCR method. (A,B) The hatched boxes indicate the microsatellite-specific primers, (C) the undulating gray box the universal M13(-21) sequence, and the star the fluorescent FAM label. (D) In the first PCR cycles, the forward primer with the M13(-21) tail is incorporated into the PCR products. (E) These products are then the target for the FAM-labelled universal M13(-21) primer, which is incorporated during subsequent cycles at a lower annealing temperature of 53°C. (F) The final labelled product can be analysed on a laser detection system. [67]

Table 4. Microsatellite markers Chromosome 12

Name	Sequence
D12S338-M13-F	CGACGTTGTAAAACGACGGCCAGTCTGGCTGTTGGCTGGA
D12S338-R	AAAATGGGAGGTCACTTCTAAT
D12S1075-M13-F	CGACGTTGTAAAACGACGGCCAGTCCACCCTGTGTAGATGTCCT
D12S1075-R	CTCTTCCCCCAAAGGACTG
D12S1672-M13-F	CGACGTTGTAAAACGACGGCCAGTATCAGGCCACTGCACTC
D12S1672-R	CTGGAAATTCACATCTGCTT
D12S821-M13-F	CGACGTTGTAAAACGACGGCCAGTCATTGCACTTCAGCCTATAC
D12S821-R	GCAACCATAACCCCTATTTAG
D12S1646-M13-F	CGACGTTGTAAAACGACGGCCAGTACCACTCCATTGCTGGC
D12S1646-R	GCTGGGTAAGAACCTCTGC
D12S817-M13-F	CGACGTTGTAAAACGACGGCCAGTCTGTGATGGACATCTTTCCAA
D12S817-R	TCTCTCCTGTCCCACATCC
D12S2079-M13-F	CGACGTTGTAAAACGACGGCCAGTGGGCAGATCACTCGAGGC
D12S2079-R	TGGCCATAACAACCATCTTT
D12S2070-M13-F	CGACGTTGTAAAACGACGGCCAGTGGGTCAGCGAATATTTCCTT
D12S2070-R	TGGCTGACAGAGCCTAAAGT

Amplification of whole genomic DNA

The MPI for molecular genetics holds a database of Iranian families with ARID, that have been SNP genotyped. From this database, 4 families were selected with linkage intervals overlapping the one on chromosome 12, identified in our family.

Due to the limited supply of patients' DNA, whole genomic DNA was amplified using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare*).*

The GenomiPhi kit utilizes bacteriophage Phi29 DNA polymerase to exponentially amplify single- or double-stranded linear DNA templates using random hexamer primers.

For amplification, 0.5 µl of the genomic DNA template is combined with 9 µl sample buffer containing random hexamer primers. The mixture is heat-denatured at 95°C for 3 min and cooled at 4°C to allow random annealing of the hexamers.

Then Phi29 DNA polymerase (1 µl), deoxynucleotide triphosphates and buffer components (9 µl) are added. This reaction mixture is incubated for 90 min at 30ºC (elongation), followed by 10 min at 65°C (denaturation) and cooling to 4°C.

The DNA replication is accurate because of the proof-reading activity of Phi29 DNA polymerase and the amplified genomic DNA can be used as template for any further PCR.

Sequencing

Polymerase chain reaction (PCR) was used to amplify the different exons of the genes of interest. Primers binding to adjacent introns were selected using http://genome.ucsc.edu and ExonPrimer http://ihg.gsf.de/ihg/ExonPrimer.html (table 5). PCR was conducted according to standard procedures using Taq DNA polymerase (Solis Biodyne).

Table 5.

A. *KIAA*1033 primers

Name	Sequence
Kiaa1033_E1F	GTGCTGTGACAGTAGCTGGG
Kiaa1033_E1R	GTCCCTCTCGCAGGAGC
Kiaa1033_E2F	TGCTTTGCGTGACTTTTGAG
Kiaa1033_E2R	CTGCATGGCTTAGAGAAAAGG
Kiaa1033_E3-4F	CAGGATCTAACACAAGTTTAAGTCG
Kiaa1033_E3-4R	AAGTACTCAATGAATGGTAGTCCG
Kiaa1033_E5F	GCAAATGATGTAGCCCTCTG
Kiaa1033_E5R	ACCTGTTTACCAAACATACCAAC
Kiaa1033_E6F	TGCAGAGTAAAAGATGTCCTAATG
Kiaa1033_E6R	CACATCTTTAGTATTTGAAGGTTTCAG
Kiaa1033_E7F	GGTTAACTGGAGGCTTTATTGG
Kiaa1033_E7R	AGTGTTCAAAGAATGCTGAATG
Kiaa1033_E8F	AAATTCCAAGATCATCTGCTCC
Kiaa1033_E8R	TCCAACCATACAAACTCCAAAG

Kiaa1033_E9F	CGTCTACATGAGGCTTAACTTCTAAC
Kiaa1033_E9R	CCTTCTAAATATCAATCCCCAGG
Kiaa1033_E10F	TCAAATGTGTAGAGCCAAACAAG
Kiaa1033_E10R	CCTACTATAATTCTGAGCTTTGTCGTC
Kiaa1033_E11-12F	TTTATACAGGATTATAGATTTGGGTTC
Kiaa1033_E11-12R	CAGTGGCAAAGAAAAGGAAAAG
Kiaa1033_E13F	CAGCTTGTTGTTTAGCCTGC
Kiaa1033_E13R	GGAGATTGCCTGACCCTTAC
Kiaa1033_E14F	AAATAATGTACTGCAATCAGTATCAAG
Kiaa1033_E14R	AAAACATACTCTACCCTTTGTTCAG
Kiaa1033_E15F	GGTAAGTTGAGGAATAGCTGCTG
Kiaa1033_E15R	TTTAAGCCTGGGTCATTTCTC
Kiaa1033_E16F	CCCTTCCCTAGCCTAAGGTC
Kiaa1033_E16R	CCGAGCATATTATGTTTGAAGG
Kiaa1033_E17-18F	TTGAATCATGGAAAAGATAATCAAG
Kiaa1033_E17-18R	AAAATCAATGTTTCTATTAACCCTGTC
Kiaa1033_E19F	CATTCCTTTCATATTGTTTTGGG
Kiaa1033_E19R	AAACTAAAACGGTTTACACAGTGAC
Kiaa1033_E20F	TTCCTTCAGATCGAGGTTTTG
Kiaa1033_E20R	CAGTGAGTCTGCTTCAATCG
Kiaa1033_E21F	AAATGGAGCATAGACATGTAGGG
Kiaa1033_E21R	CCTCAAAGTTTTGTTTAACAGTAGGG
Kiaa1033_E22F	CCCTACTGTTAAACAAAACTTTGAGG
Kiaa1033_E22R	GCTCAATATAAGCAGGACCCC
Kiaa1033_E23F	TTTGTCTTTTAACGTTTGATCG
Kiaa1033_E23R	AAATCAATTCCTTGCCCCTC
Kiaa1033_E24F	TGACCAGTGTTTGCTATGGG
Kiaa1033_E24R	GAGAAAACCACGTTCCAAGC
Kiaa1033_E25F	TTTTCAGGTCCAATATGAAATAGAC
Kiaa1033_E25R	GGGAAAATAAACTAAAACCCC
Kiaa1033_E26F	TGGCATGGTTTATTTGGGAG
Kiaa1033_E26R	AAAATTACATGAGTGGAGTAAAGCTG
Kiaa1033_E27F	GGTTCTTAGGAGTGGAAACCTTG
Kiaa1033_E27R	TGTAATCTTTACATGCATTCATAAACC
Kiaa1033_E28F	CTGAAATGCACAGTTGTTTCC
Kiaa1033_E28R	TGGCTGAAATACTGGACAGC
Kiaa1033_E29F	TTTTCAAAGCCACGCTTC
Kiaa1033_E29R	CAAAGAAAACCAGATCTATTTGTGC
Kiaa1033_E30F	GGGTACCTAAGGCAAATGTTTC
Kiaa1033_E30R	GCAACAGAGCGAGACGAGAC
Kiaa1033_E31F	TGGGAATTATTGGTGAAGGAG
Kiaa1033_E31R	TTCTATATGGCTTCTGCTTTTCAG
Kiaa1033_E32F	GGGCGGGTTACTTATTTCAG
Kiaa1033_E32R	GGTCTGTGTTATTTTCCAAAAGG
Kiaa1033_E33F	GAAGTTTGGCAATACAGCTCTTC
Kiaa1033_E33R	ATCCATTCCACTGGCAAAAG

B. *LHX5* primers

Name	Sequence
LHX5_E1F	CAGCTGCTGAGACAAGAGGC
LHX5_E1R	AGGCTGGGATGGGGATG
LHX5_E2F	CTGGGGAGGGCTGGATG
LHX5_E2R	CCAAGTAGGGAGTGCCAGG
LHX5_E3F	GACTGAAGAGGCGTCTGGC
LHX5_E3R	AGACTCCTCCGAGGCTCC
LHX5_E4F	GGGAAGGGAGAAGAGGGG
LHX5_E4R	GGACCCTTCGCCCTCAG
LHX5_E5F	ACCTTGGGTCGGTGTCG
LHX5_E5R	GTTTTGGTTTCAGGAGGCTG

Primers and surplus dNTPs were digested by Exonuclease I/Antarctic phosphatase (BioLabs). In the next step, the exons were sequenced through the chain-terminator method (Sanger, Nicklen et al. 1977). Here a sequencing primer (identical to our PCR primer) binds to single-stranded template DNA and is elongated by a DNA-polymerase until the synthesis breaks off after incorporation of a dideoxynucleotide (ddNTP, Terminator). The ddNTPs differ from normal dNTPs by the absence of a 3'-OH-group and are labelled with four different fluorescent dyes (BigDye, Applied Biosystems).

Before analyzing a product on the sequencer, a further step of product purification is applied in order to remove the excess residues of ddNTP mix and primers. Unused fluorescently labelled ddNTPs can cause superimposition when evaluating sequences.

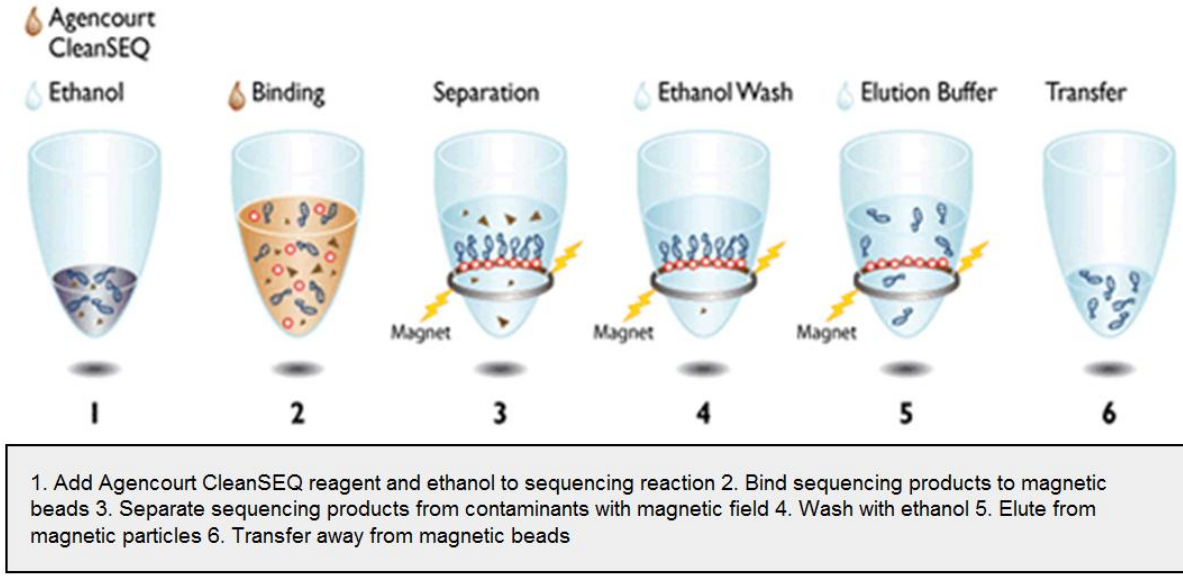


Figure 6. Purification of sequencing products with magnetic beads.

Source: http://www.beckmangenomics.com, Agencourt CleanSEQ

The Agencourt CleanSEQ system in combination with a Biomek NX Multichannel 96 roboter was used for purification of the DNA.
The Agencourt CleanSEQ system is a rapid, high performance dye-terminator removal process based on Solid Phase Reversible Immobilization (SPRI) technology using magnetic beads. The Agencourt CleanSEQ system produces sequences with longer read lengths of high base calling quality, and high signal intensities. Sequencing can be performed with lower quantities of fluorescently labelled ddNTPs/dNTPs (Big Dye, Applied Biosystems), reducing sequencing costs. After purification, the DNA products were loaded into the DNA Sequencer 3730 (ABI). Here the products of the sequencing reaction were fractionated electrophoretically in a capillary gel matrix and analysed by a laser detection system. According to the fluorescent dyes, different signals are obtained and evaluated with Sequence Analysis (ABI) and Sequence Pilot software (JSI medical systems GmbH).

Section III: Methods used by MPI and Laboratoire d'Enzymologie et Biochimie Structurales

Agilent SureSelect DNACapture Array ©
(Experiments performed at the MPI for molecular Genetics, Berlin)
DNA Capture Array is capable of enriching user-defined genomic regions from complex genomes, lowering sequencing costs by focusing only on regions of interest. This array-based target enrichment method was used to selectively amplify a 12.5 Mb region on chromosome 12. The custom-made array contained oligonucleotides that specifically bind the exonic regions of 159 genes located on this chromosomal segment.

The procedure is outlined in the flow scheme (figure 7).
DNA was sheared and purified and the fragment length and quality were analysed.
Sheared DNA was blunt-ended and a 3`-dA overhang was added (adenylation). Adapters were ligated, and the fragments were size selected using gel purification. Fragments in the range of 150-300 bp were selected and purified.
PCR with adapter-specific primers was conducted and a DNA library was created. This library was hybridized on a custom made microarray, displaying oligonucleotides specific for the 1376 exons contained in the identified region on chromosome 12.
After multiple washing steps, the DNA was eluted from the plate. The DNA was again amplified using adapter-specific primers.

Thus, Agilent SureSelect DNA Capture Array © was used to selectively amplify all 1376 exons plus 50 bp flanking sequences of the 159 genes with a total length of 392165 bp, following manufacturer's instructions.

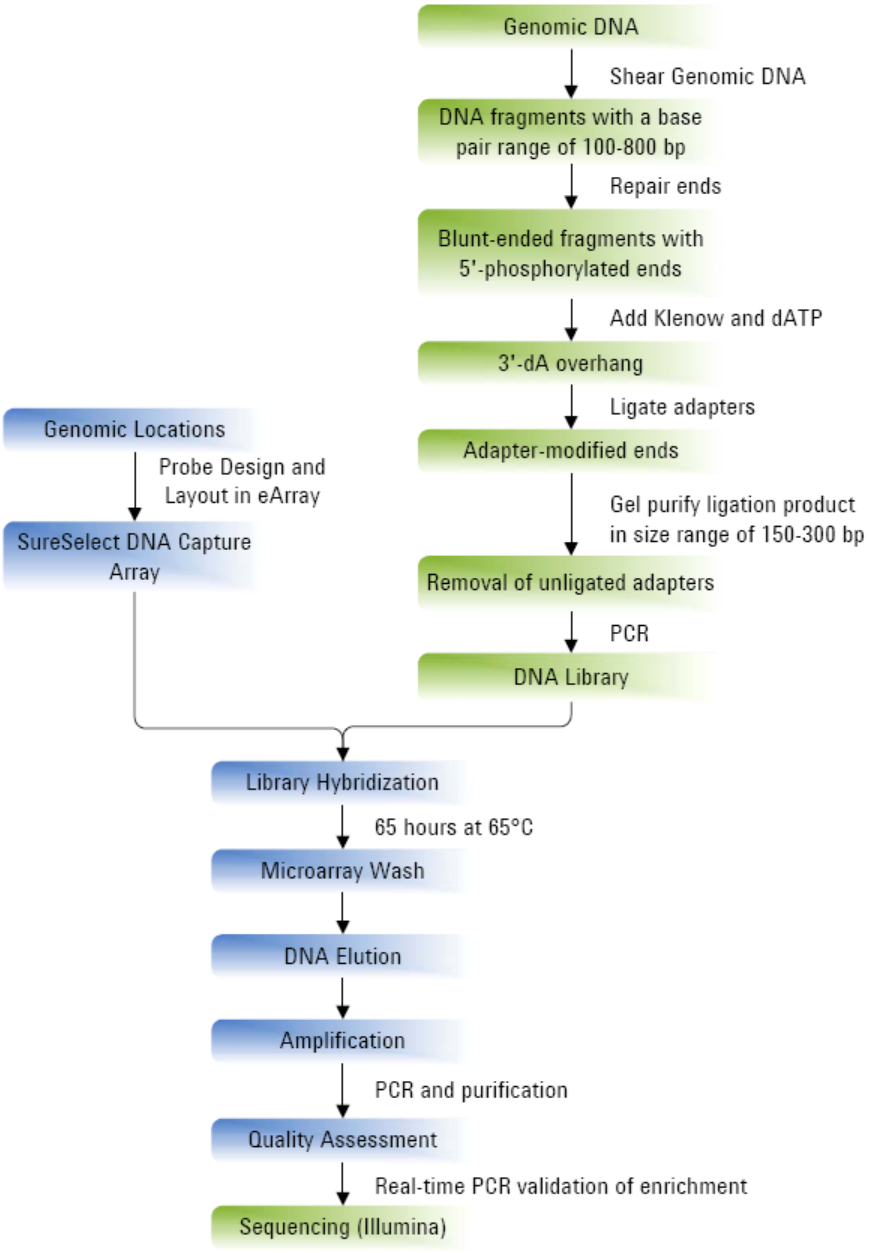


Figure 7. Sure Select DNA Capture Array.
A DNA library is created and hybridized to a custom-made array displaying oligonucleotides specific for the region of interest. In a next step, DNA is eluted and amplified.

Source: http://www.genomics.agilent.com

High throughput sequencing

The sample was analysed through sequencing by synthesis, using Illumina technology and the Illumina Genome Analyzer II GAII at the Max Planck Institute for molecular Genetics,

Berlin.
The 8 channel flow cell has a dense lawn of oligonucleotides that can bind to the adaptors ligated to the DNA fragments. The bound fragments are amplified and the PCR copies are located around the template strain. The reverse strands are cleaved and washed away. The clusters are ready for sequencing.
A mix of primers, DNA polymerase and 4 different fluorescently labelled reversibly terminated nucleotides are added to the plate. All four bases compete with each other to bind to the template. This competition ensures the highest accuracy. After each round of synthesis, the clusters are excited by a laser, and the signal is detected. The fluorescent label and blocking group are then removed, allowing for the addition of the next base.
This method enables single nucleotide sequencing with high accuracy.
Thirty-six base pair single read sequencing of the enriched DNA was performed according to the respective Illumina manual.

Data analysis
Alignment of the DNA fragments and data analysis was carried out by the Max Planck Institute for molecular Genetics.
The total read length encompassed 416 Mb, of which 394 Mb could be mapped to the human reference genome. The total read length in the target region amounts to 23 Mb with a coverage of 99.3%. A mean coverage of 121 fold was reached for the target region.
The 36bp single-end reads were aligned onto human reference genome (hg18) [http://hgdownload.cse.ucsc.edu/goldenPath/hg18/chromosomes/] using SOAP2.20 [68] with parameter settings –a –D –o. Reads that were unambiguously aligned to target regions were used for variant calling. For variant calling, multiple criteria were used: Presence of at least 3 non-identical supporting reads, Phred-like quality score >20 (base calling accuracy > 99%) and allelic percentage >70 (>70% of the reads support variant).

Generation of 3T3 cell lines stably expressing KIAA1033
(Performed by Laboratoire d'Enzymologie et Biochimie Structurales, Centre de Recherche de Gif, CNRS, France)

The DNA fragment encoding full length KIAA1033 was amplified by PCR using the cDNA clone IMAGE:8143997 as a template, then cloned into a modified pCDNA5/FRT/V5-His (Invitrogen) vector in which the CMV promoter was replaced by a EF1α/HTLV chimera promoter amplified from pFUSE FC (Invivogen). This vector tags SWIP at its N-Terminus with $(His)_6$-PC (HHHHHH-EDQVDPRLIDGK) followed by a TEV binding and cleavage site

(DYDIPTTENLYFQG). The c.3056C>G point mutation was introduced using the Quick Change site directed mutagenesis kit (Stratagene).

Flp-In NIH3T3 cells (Invitrogen) were grown in DMEM supplemented with 10% calf serum (Invitrogen). They were transfected with Lipofectamine 2000 (Invitrogen). Stable transfectants obtained by homologous recombination at the FRT site according to the manufacturer's instructions were selected using 100 µg/mL Hygrogold (Invivogen).

Cell extracts and Western Blotting

~ $5x10^6$ cells were washed in PBS then lysed in 400 µl RIPA (50 mM Hepes, 150 mM NaCl, 5 mM EDTA, 1% NP-40 (nonyl phenoxypolyethoxylethanol), 0.5% deoxycholate, 0.1% sodium dodecyl sulfate; pH 7.7) supplemented with Protease Inhibitor Cocktail (1:500, Sigma). Lysate was rocked for 10 min at 4°C then spun at 20000 x *g* for 10 min at 4°C. SDS-PAGE was performed using 4-8% Tris acetate gels [69].

Western blots were revealed using horseradish peroxidase coupled secondary antibodies, Supersignal kit (Pierce) and a Fuji LAS-3000 (Fujifilm).

Polyclonal antibodies targeting WASH were described previously [70]. Anti-tubulin mAb (clone E7) was developed by M. Klymkowsky and obtained from Developmental Studies Hybridoma Bank. Strumpellin pAb (C-14) and PC-tag mAb were from Santa Cruz and Roche, respectively.

Web resources

http://www.ncbi.nlm.nih.gov/mapview

http://www.genedistiller.org

http://genome.ucsc.edu

http://ihg.gsf.de/ihg/ExonPrimer.html

http://hgdownload.cse.ucsc.edu/goldenPath/hg18/chromosomes

http://www.ncbi.nlm.nih.gov/projects/SNP

http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/refGene.txt.gz , July 2010

http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/phyloP44wayAll.txt.gz

Results

We studied an Omani family with non-syndromal autosomal recessive intellectual disability (ARID). Figure 8 shows the pedigree of this family. In its two branches, there are 7 affected patients, 5 females and 2 males, all with non-syndromic ARID. After informed written consent was obtained from the parents, linkage analysis using Affymetrix GeneChip® Human 250 K Sty I Array set was performed on 4 out of 7 affected patients (IV:10, IV:11, IV:16, IV: 17), the parents (III:6, III:7) and 4 unaffected siblings (IV:7, IV:9, IV:13, IV:14). The calculations were carried out in the Cologne Center for Genomics and Institute for Genetics in Cologne, Germany, and corroborated in our institute.

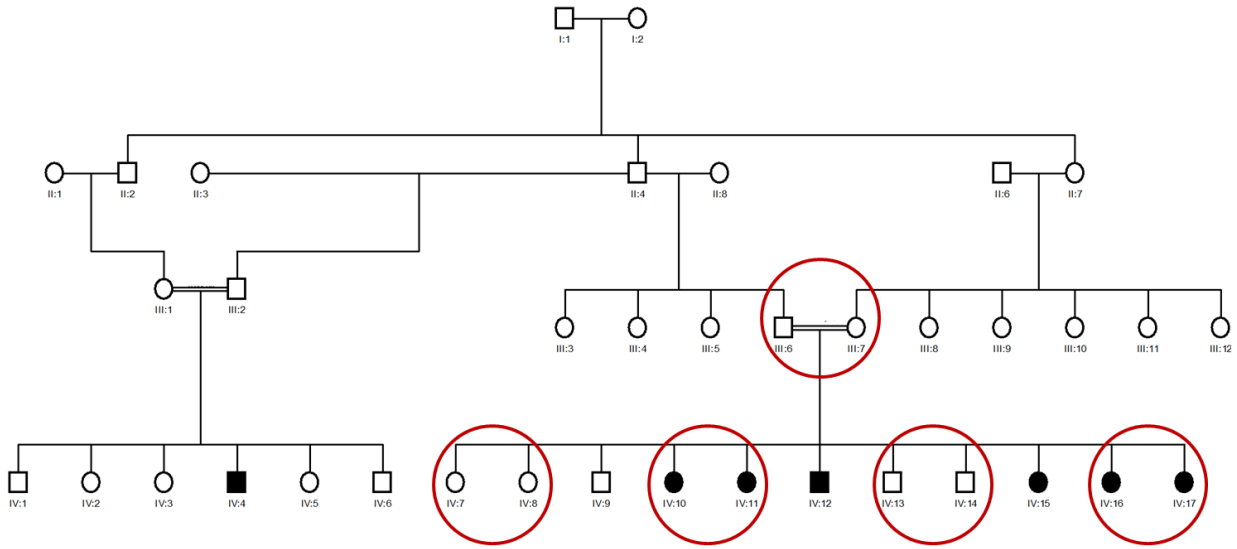
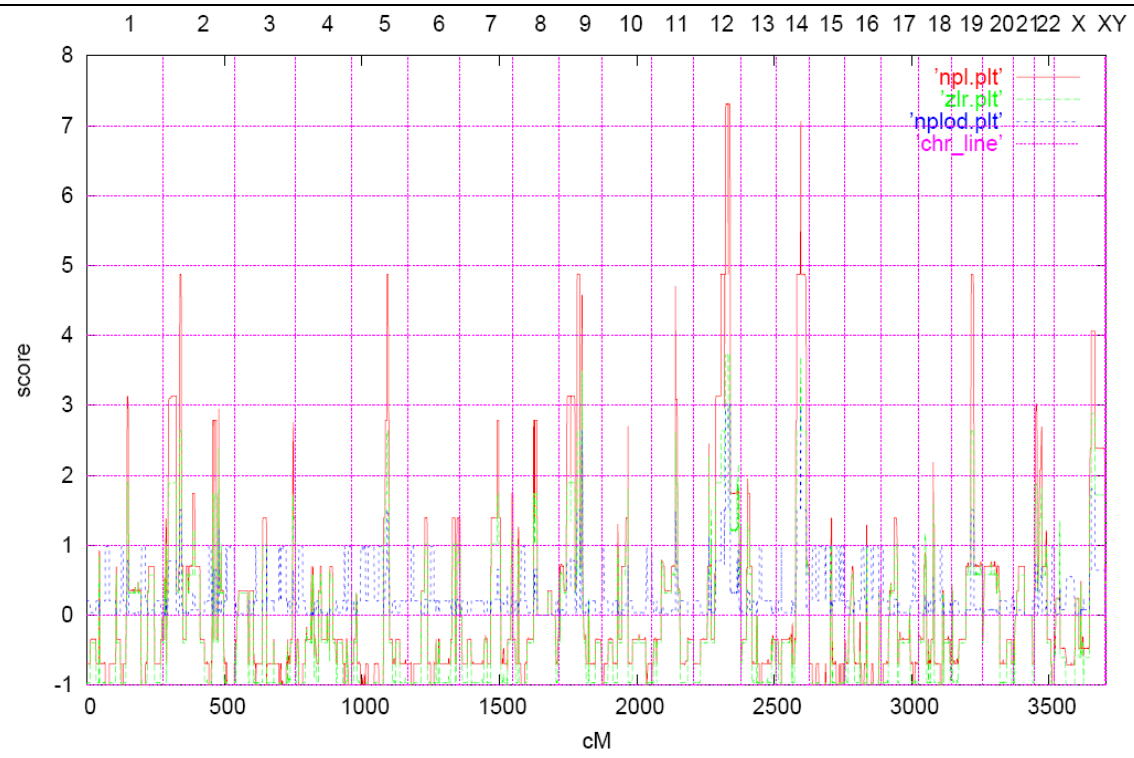


Figure 8. Pedigree. Red circles indicate the subjects included in the SNP analysis. Patients (dark symbols) have moderate to severe intellectual disability.

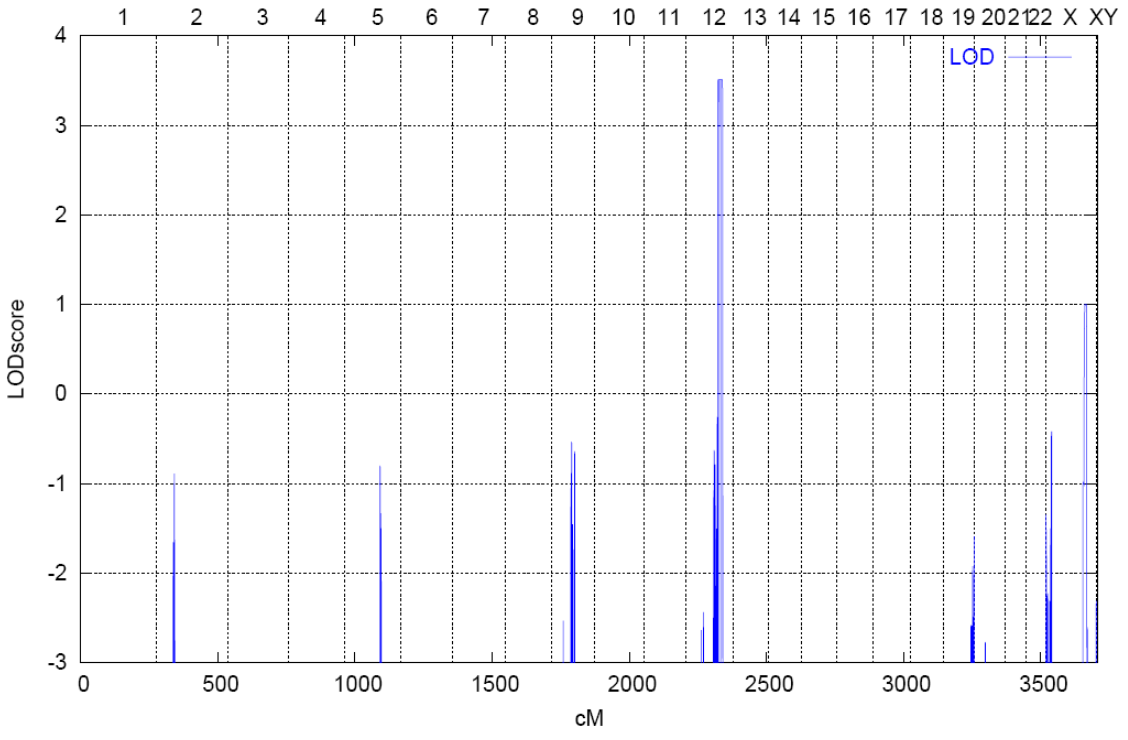
With Alohomora© the kinship relation was confirmed for all samples, and Mendelian errors and unlikely genotypes were excluded.

For linkage calculations Allegro© was used and in a first analysis a reduced panel of 20.000 markers was included. With nonparametric analysis a maximum LOD score was reached at chromosome 12, which could be confirmed in parametric analysis, assuming a recessive model of inheritance, with complete penetrance of the disease and a disease allele frequency in the general population of 0.0001.

A second calculation including all markers on chromosome 12 was performed, and the borders of the chromosomal segment linked to the disease could be defined.



A. Non parametric calculation



B. Parametric calculation

Figure 9. Genome-wide linkage analysis results using a subset of 20.000 markers. In both graphs the bottom x-axis shows the genetic distance in cM and the upper x-axis the chromosome number. The y-axis depicts the LOD score. **A)** Shows the non- parametric LOD score in red peaks while **B)** shows the parametric LOD score in blue peaks.

A single homozygous interval with a maximal LOD score of 3.5 defined by flanking heterozygous SNP markers at positions 102615255 (rs7312283) and 115621170 (rs11068147) (human genome version 18) on chromosome 12q23-24 was identified in a region not previously implicated in autosomal recessive ID.
Allegro constructs haplotypes, and uses those to identify unlikely haplotypes that can result from genotyping errors. In figure 10 the predicted haplotypes are depicted.

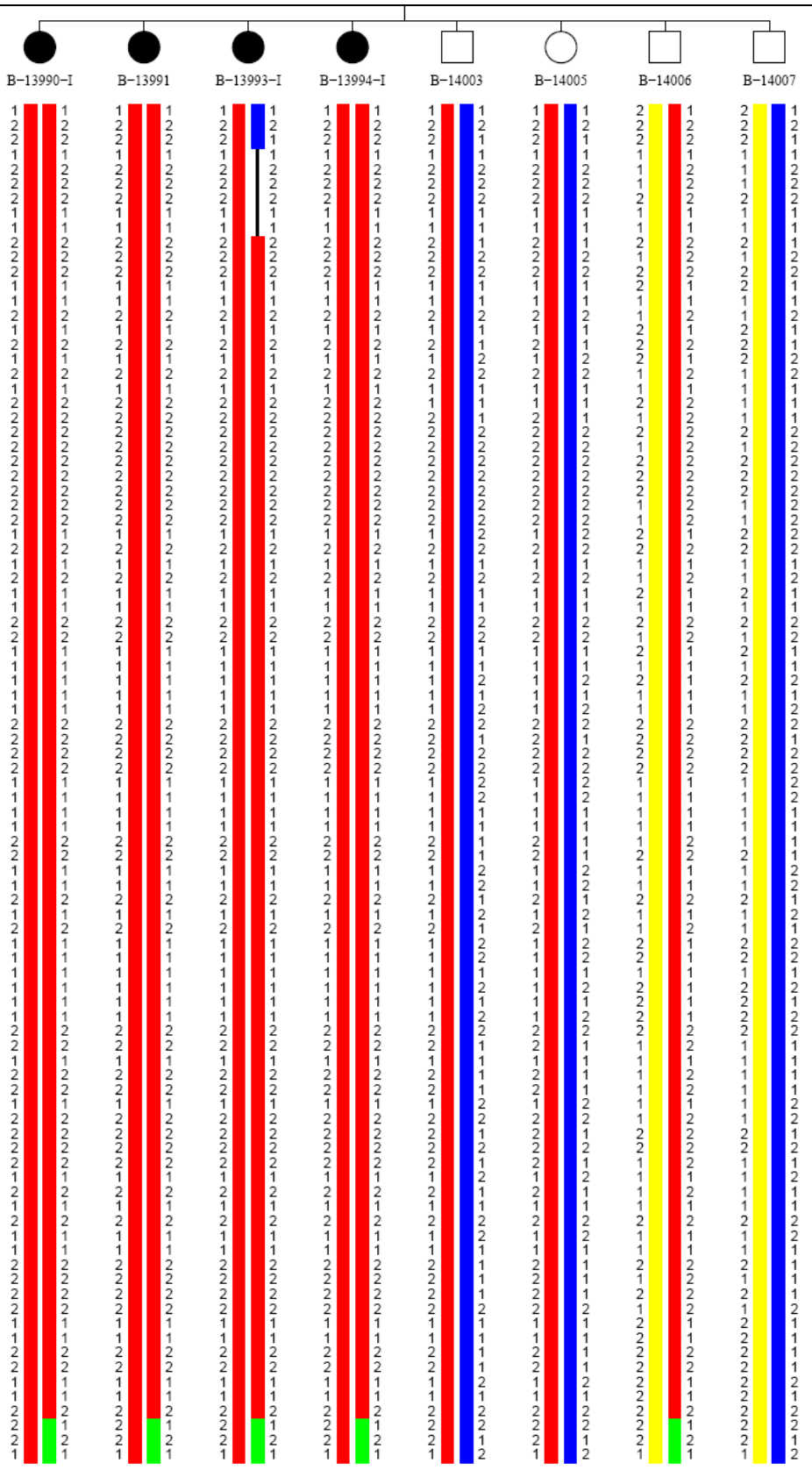


Figure 10. Predicted haplotypes using 10.000 markers. Allegro constructs haplotypes, and uses them to define unlikely phenotypes.
The numbers represent DNA numbers.
13390 IV:10, 13391 IV:11, 13993 IV:16, 13994 IV:17 14003 IV:9, 14005 IV:7 14006 IV:13, 14007 IV:14

These results were refined by genotyping a panel of 19 microsatellite markers in 11 out of 12 siblings and both parents (III:6 and III:7) of the larger family branch. This resulted in an interval of 12.5 Mb between positions 103065571 (D12S388 (last crossing-over)) and 115621170. The distal border could not be narrowed down, as the last crossing over at D12S366 is distal to the border defined by linkage analysis.

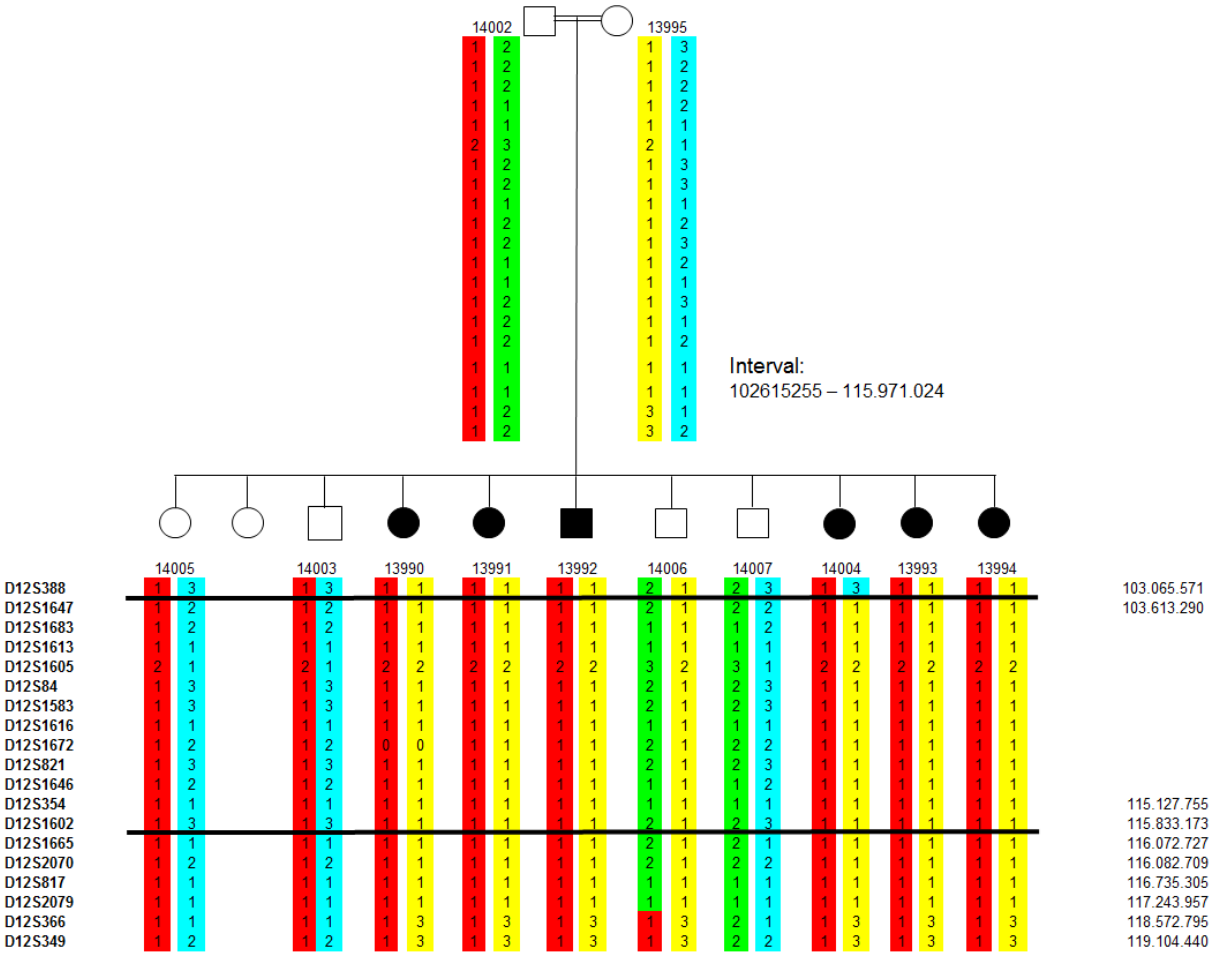
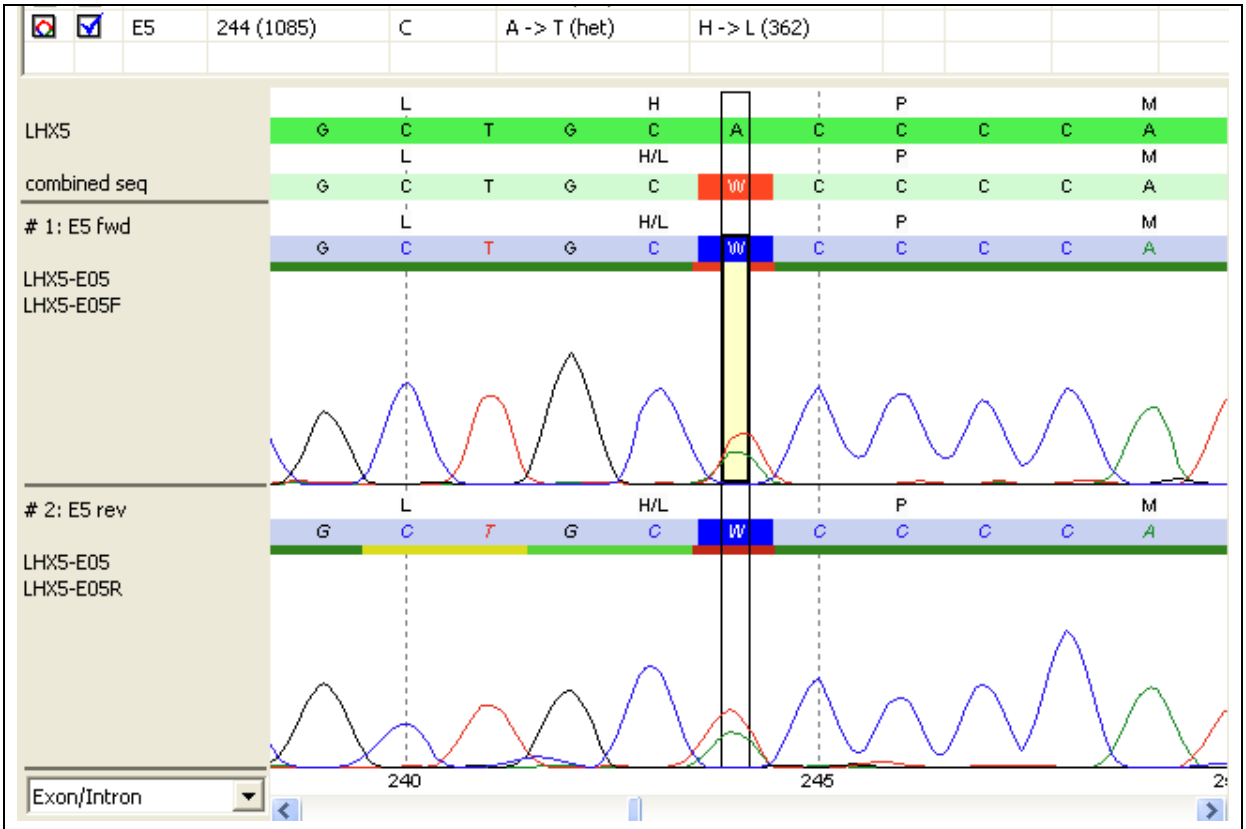


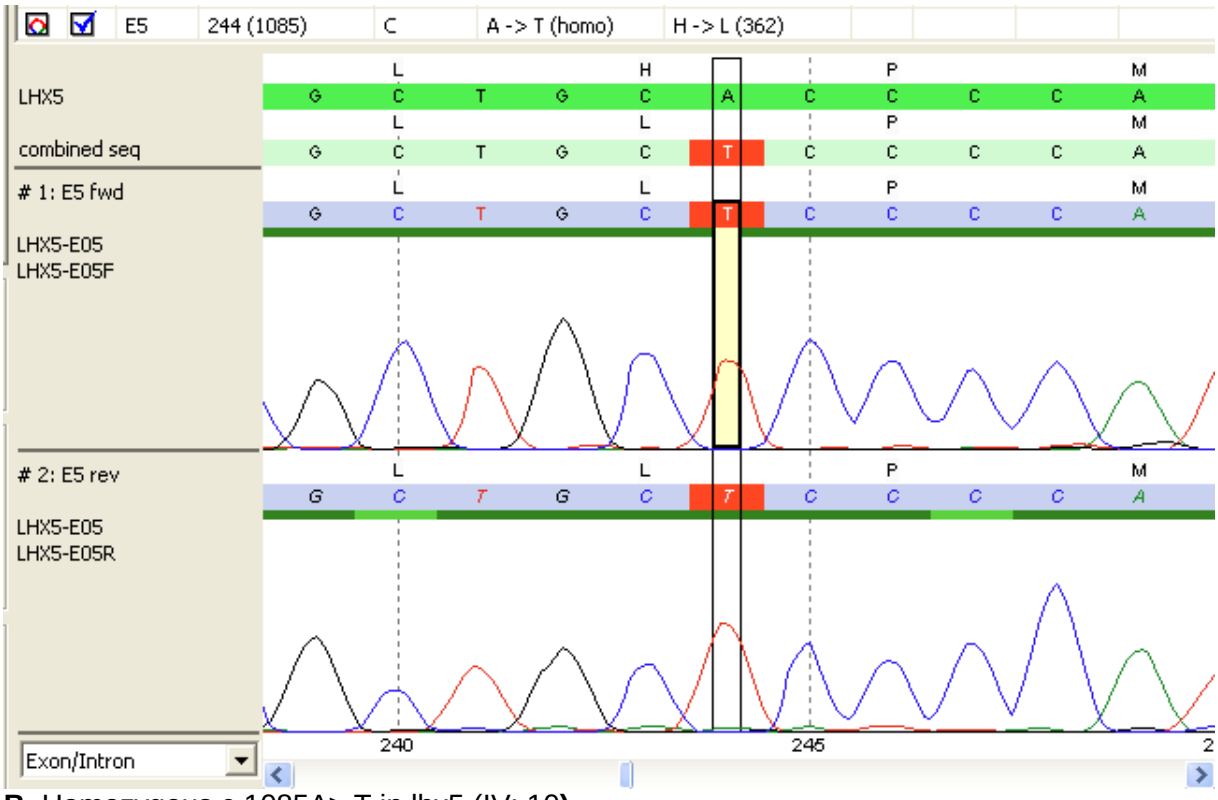
Figure 11. Microsattelite fine mapping of the linkage interval. The proximal border could be narrowed down to 103065571

This segment of 12.5 Mb contains 159 genes. Genedistiller, a new tool designed to facilitate candidate gene selection, was used to define candidate genes. We selected for genes with the word 'brain' in their OMIM report (32 genes) and for genes with a neurological mouse phenotype (12 genes). These genes were prioritised, and *LHX5* was designated as the most likely candidate gene, mainly because of the KO mouse phenotype with learning difficulties and hippocampal malformation.

Sanger sequencing yielded a homozygous mutation c.1085A> T in exon 5 of the gene, resulting in the amino acid substitution H362L.



A. heterozygous c.1085A> T in *LHX5* (III:6)



B. Homozygous c.1085A> T in lhx5 (IV: 10**)**

Figure 12. Sanger sequencing of lhx5 . Subject III:6 carries a heterozygous mutation c.1085A> T in exon 5 of *LHX5* (A), whereas subject IV:10 is homozygous for the mutation (B). Both forward and reverse sequences are shown.

This missense mutation was neither found in 365 Omani controls, nor in 95 Caucasian controls.

In collaboration with the Max Planck Institute for molecular genetics, high throughput sequencing was performed to extend the search for mutations in the remaining 158 genes in this 12.5 Mb homozygous interval.
In subject IV:12 high throughput sequencing was performed with the Illumina Genome Analyser II GAII, after exon enrichment using Agilent SureSelect DNA Capture Array ©. This array- based target enrichment method was used to selectively amplify all 1376 exons of the 159 genes from the homozygous region on chromosome 12, plus 50 bp flanking sequence on each side. Exon enrichment, high throughput sequencing and data analysis were carried out in the Max-Planck-Institute for molecular Genetics, Berlin.

The total read length encompassed 416 Mb, of which 394 Mb could be mapped to the human reference genome. The total read length in the target region amounts to 23 Mb with a coverage of 99.3%. A mean coverage of 121 fold was reached for the target region.

The 36bp single-end reads were aligned onto human reference genome (hg18) [http://hgdownload.cse.ucsc.edu/goldenPath/hg18/chromosomes/] using SOAP2.20 [68]. Reads that were unambiguously aligned to target regions were used for variant calling. For variant calling multiple criteria were used: presence of at least 3 non-identical supporting reads, Phred-like quality score >20 and allelic percentage >70.

High throughput sequencing of all coding sequences in the 12 Mb target region yielded 49 homozygous variants. Forty-six of those were evaluated as presumably neutral polymorphisms by comparison with dbSNP130 [http://www.ncbi.nlm.nih.gov/projects/SNP/], the genomic sequence of 185 individuals available through the 1000-genome project [71] and 200 recently published exomes from Danish individuals [72]. The remaining 3 variants were annotated using RefSeq as basis.
[http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/refGene.txt.gz, July 2010].
Subsequent screening for non-synonymous mutations and changes involving canonical splice sites revealed two potentially disease-causing mutations.
Besides the previously found homozygous mutation c.1085A>T (p.His362Leu) in exon 5 of *LHX5*, we identified another homozygous, probably disease-causing missense mutation, c.3056C > G, in exon 29 of the *KIAA1033* gene. This mutation results in a Pro1019Arg exchange in Strumpellin and WASH-Interacting Protein (SWIP), a member of the WASH complex.

Both changes were validated by Sanger sequencing, and co-segregation with ID was confirmed by re-analysis of the family.

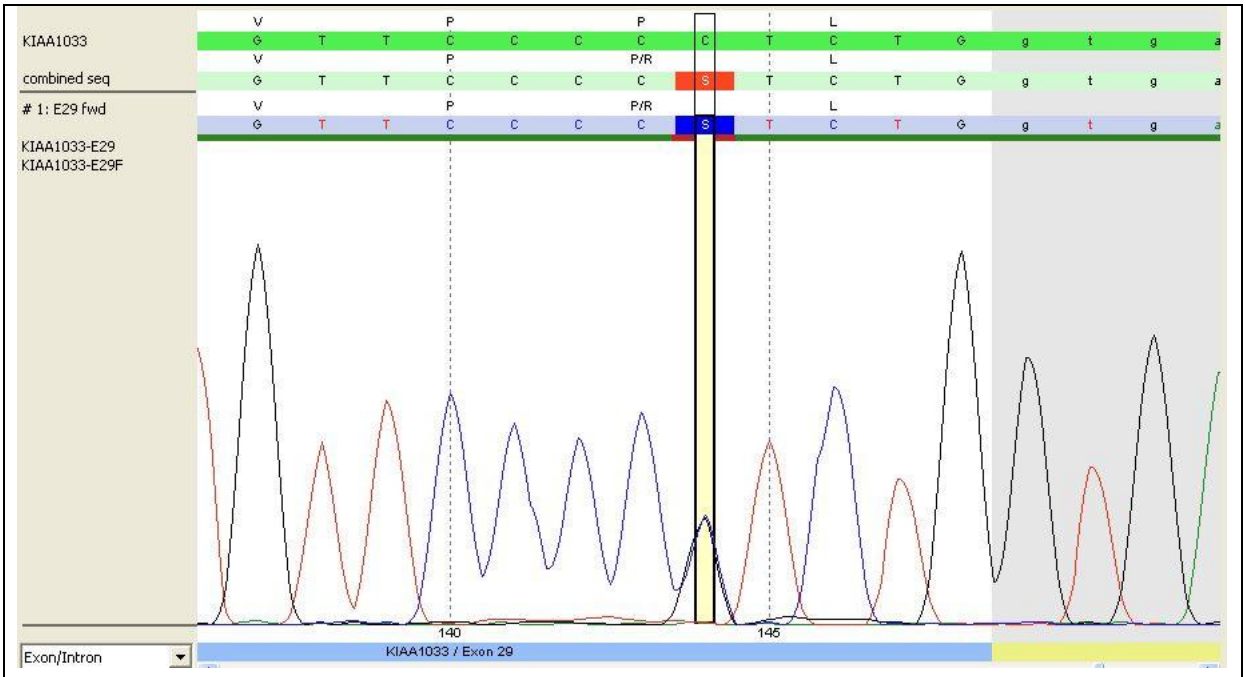
Further evaluation with PhyloP(44vertebrates) *[http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/phyloP44wayAll.txt.gz]* yielded high base conservation scores for both changes. To estimate the pathogenicity of these variants Pmut [73], SIFT [74], PolyPhen2 [75] and MutationTaster [76] were employed. While the *KIAA1033* change was considered as pathogenic throughout, the *LHX5* variant yielded ambiguous results (table 6), which was supported by clinical evidence and functional considerations (see below).

Mutation	**Gene**	**Supporting Reads**	**Allelic percentage**
g.104078052C>G c. 3056C>G p. P1019R	KIAA1033 (NM_015275)	21	100
g. 112385502T>A c. 1085A>T p. H362L	LHX5 (NM_022363)	39	81

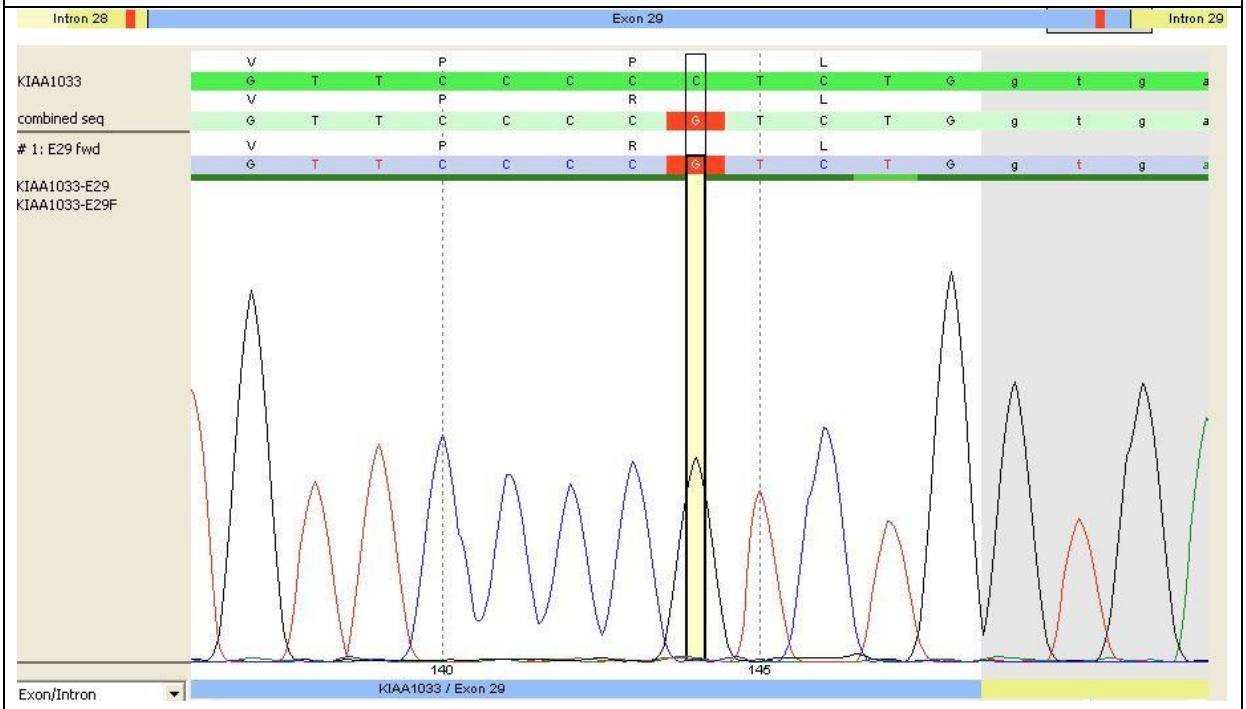
Mutation	**Gene**	**PhyloP (44vertebrate)**	**Pmut**	**SIFT**	**PolyPhen2 (HumVar-trained)**	**Mutation Taster**
g.104078052C>G c. 3056C>G p. P1019R	KIAA 1033	5.37	PATHOL OGICAL	Not scored	Probably damaging	Disease causing
g. 112385502T>A c. 1085A>T p. H362L	LHX5	4.09	PATHOL OGICAL	tolerated	Benign	Disease causing

Table 6.
Pathogenicity predictions for Pro1019Arg in SWIP and His362Leu in LHX5.

Analysis of the 1000-genome database and 200-Danish-exome dataset did not reveal any deleterious mutations in these genes, and this was corroborated by Sanger sequencing of the respective mutated exons in healthy individuals from Oman (365 for *LHX5* (figure 12), 331 controls for *KIAA1033* (figure 13)). Finally, 4 unrelated consanguineous Iranian families with ARID and regions of homozygosity overlapping the relevant linkage interval on chromosome 12q23-24 were screened for mutations in these two genes. No additional mutations were identified.



A. heterozygous c. 3056C>G (IV:13)



B. Homozygous c. 3056C>G (IV:16)

Figure 13. Sanger Sequencing of *KIAA1033*
Subject IV:13 carries a heterozygous mutation c.3056C>G in *KIAA 1033* (A), whereas subject IV:16 is homozygous for the mutation (B).

LHX5 is a member of the LIM-homeobox genes that encode a family of transcription factors that are highly conserved throughout evolution [77, 78]. It contains 2 zinc finger motifs called LIM domains (AA 3-125) located N-terminally of the DNA binding homeodomain (AA180-239)

[79]. The amino acid exchange at position 362 was located outside the known functional domains of *LHX5*.

The gene is expressed solely during the embryonic development, and plays a role in the anatomic formation of the hippocampus [80]. Since inactivation of the *LHX5* gene in mouse results in gross anatomical changes in the hippocampus, we reasoned that pathogenic mutations in humans leading to a severe cognitive disorder would also be likely to induce observable anatomical changes in the brain. However, MRI scans of 5 patients did not reveal structural abnormalities of the hippocampal area or other brain regions (except for a benign midline cyst in subject IV:11), thereby providing another argument against the pathogenicity of the *LHX5* variant in this family.

In figure 14 a coronal MRI section showing normal hippocampal anatomy is depicted.

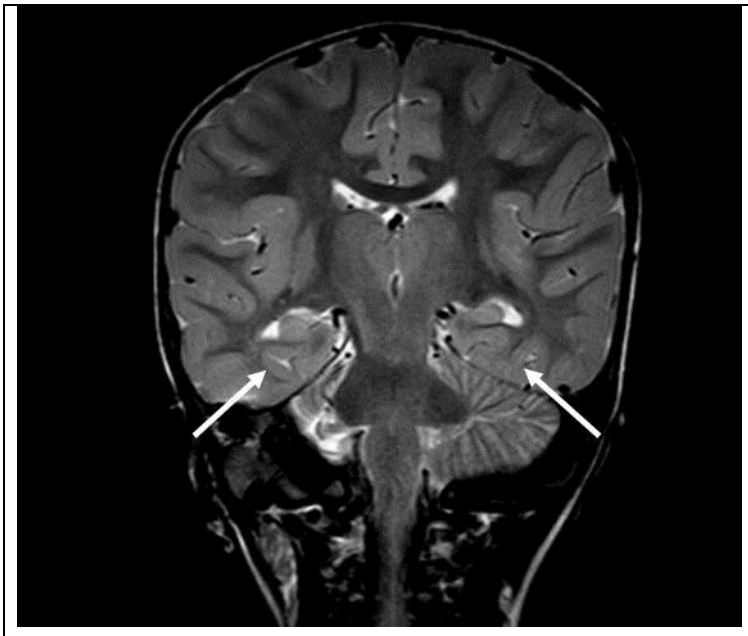


Figure 14.
Coronal MRI section of IV:16.
Arrows indicate the hippocampal area. No anatomic malformations were present.

In an attempt to gain further evidence against a pathogenic role of the *LHX5* mutation in the development of ID in our patients, we sought the cooperation of the group of Rob Willemsen, of the department of.clinical genetics, Erasmus Medical Center, Rotterdam, the Netherlands, specialised in zebrafish experiments. As was shown by Peng *et al*., *lhx5* is required for forebrain development. Injection of *lhx5* morpholino interfering with *lhx5* protein synthesis in zebrafish embryos, resulted in zebrafish with small heads and eyes, missing most of the rostral part of their head [81].

Our aim was to knockdown *LHX5* with Morpholinos in zebrafish, and in the same experiment rescue the phenotype with both wildtype *LHX5* mRNA and *LHX5* mRNA carrying H362L mutation, to prove that both mRNAs would be able to restore the phenotype, and the point mutation introduced in *LHX5* was irrelevant for proper functioning of the transcription factor.

QuickChange Mutagenesis to introduce the mutation in cDNA construct was performed in our lab.
Though in the morpholino knockdown experiment the described phenotype with small head and eyes could be reproduced, the rescue did not succeed, neither with WT *LHX5* mRNA nor with mutated *LHX5* mRNA.
This failure of mRNA rescue of a morpholino knockdown has been described and discussed. The timing and location of expression seem to play a critical role. Misexpression of their protein products seem to preclude rescue of the phenotype in many cases (personal communication R. Willemsen).

LHX5 is expressed neither in peripheral blood, nor in any accessible tissue. Therefore, gene or protein expression profiling was not feasible.
However, normal hippocampal anatomy, contradictory pathogenicity results for lhx5 and a very likely other candidate gene prompted us to focus on the apparently pathogenic mutation of the SWIP gene.

The SWIP protein is a large ubiquitously expressed protein of 1173 amino acids, which has no identifiable domains. However, it has recently been described as a subunit of the WASH (Wasp and Scar homologue) complex [70, 82]. The WASH complex controls the polymerization of actin at the surface of endosomes through the activation of the Arp2/3 complex, a major actin nucleator. WASH-dependent actin polymerization promotes scission of transport intermediates of the different endosomal routes, and affects recycling [70], degradation [83] and retrograde pathways [84, 85].
The stable assembly of the WASH complex involves a core of 5 subunits, namely SWIP, FAM21, Strumpellin, Ccdc53 and WASH, and the recruitment of the heterodimer of Capping Protein [70, 86]. These studies reported that the stability of WASH and Ccdc53 subunits depends on all the other subunits, whereas the stability of a subcomplex formed by SWIP, FAM21 and Strumpellin does not depend on WASH or Ccdc53.
We sought collaboration with the group of Alexis Gautreau at the Laboratoire d'Enzymologie et Biochimie Structurales, at CNRS in Gif-sur-Yvette, France, as this group was the first to show that WASH is bound in a stable multiprotein complex with SWIP as its binding partner.

To examine the effect of the potentially pathogenic mutation of SWIP, Gautreau and his group isolated stable cell lines of 3T3 cells expressing either wild type SWIP or Pro1019Arg mutated SWIP. These two lines were generated by homologous recombination at a defined locus in order to ensure similar expression of the two transgenes. The exogenous proteins encoded by the transgenes are tagged with the Protein C (PC) epitope.

The exogenous wild type SWIP protein was clearly detected by PC western blot, but was at the limit of detection when the PRO1019ARG mutation was present (figure 15). This result suggests that the mutation PRO1019ARG destabilizes the SWIP protein. However, WASH was present at similar levels in both cases, indicating that expression of the mutant SWIP has no dominant negative effect on the assembly or maintenance of the WASH complex in the presence of endogenous wild type SWIP. This finding is consistent with the recessive mode of inheritance of the condition.

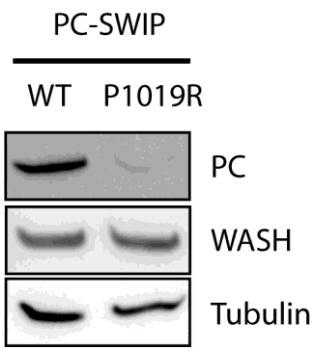


Figure 15. The mutation PRO1019ARG destabilizes SWIP. Stable 3T3 cell lines were established, expressing either PC tagged WT SWIP or Pro1019Arg SWIP from the same locus to ensure similar expression. Cell extracts were analysed by western blot. The level of PRO1019ARG SWIP is strongly reduced compared to WT SWIP. The expression of PRO1019ARG has no effect on the level of WASH complex.

Experiments and graphic by A. Gautreau and E. Derivery.

In our genetic lab lymphoblastoid cell lines from patients, healthy relatives and unrelated healthy subjects were established. Protein extracts from the cell lines were analysed by western blot using the two antibodies available to us for the WASH complex, a home-made WASH antibody [70] and a commercially available Strumpellin antibody. The affected patients, carrying the homozygous Pro1019Arg mutation of the SWIP protein, exhibited reduced levels of both Strumpellin and WASH compared to healthy relatives. Healthy relatives had similar levels of Strumpellin and WASH, independent of their carrier status (figure 16). These results indicate that the Pro1019Arg mutation of SWIP impairs the stability of the WASH complex when mutant SWIP is the sole source of this subunit.

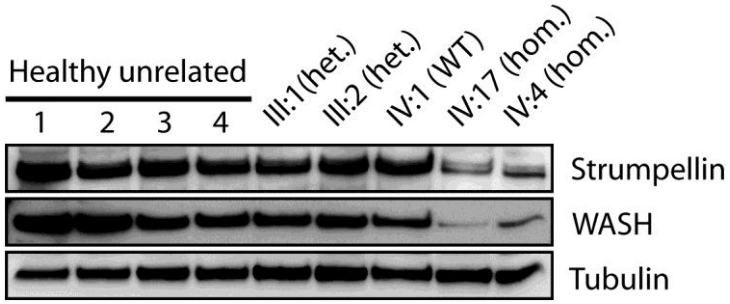


Figure 16. The PRO1019ARG mutation of SWIP destabilizes the WASH complex.
Extracts prepared from different lymphoblastoid cell lines were analysed by western blot with the indicated antibodies. Strumpellin and WASH are subunits of the WASH complex, whereas Tubulin is a loading control. Patients and healthy relatives from the family, whose pedigree is shown in Figure 1, as well as healthy unrelated subjects were analysed. The homozygous patient IV:17 has a severely reduced amount of WASH and Strumpellin, compared to his two parents, III:1 and III:2, which are heterozygous for the PRO1019ARG mutation, and the other controls. A similarly reduced amount of WASH complex is observed in a cousin, patient IV:4, who is also homozygous for the PRO1019ARG mutation of SWIP.

Experiments and graphic by A. Gautreau and E. Derivery..

Thus, we were able to show that the substitution of the highly conserved proline at position 1019 of SWIP leads to markedly reduced concentrations of this protein and the WASH complex it is embedded in.

Discussion

In collaboration between the Genetic Department in Muscat, Oman, and the Institute of Medical and Human Genetics in Berlin, we received genetic material from a large consanguineous Omani family with intellectual disability. The apparent autosomal recessive mode of inheritance and the large number of offspring provided an excellent opportunity for homozygosity mapping. Linkage analysis with both autosomal recessive and nonparametric models revealed a linkage interval on the long arm of chromosome 12. Closer inspection of the haplotypes and microsatellite finemapping could confirm and narrow down the region to 12.5 Mb.
As intellectual disability is an extremely heterogeneous disorder, and a wide spectrum of genes have been implicated in ARID this far, a variety of candidate genes were considered.

Initially *LHX5* was identified as the most likely candidate gene in this interval on the basis of its role in forebrain development, more specifically the hippocampus. The hippocampus plays an important role in memory and learning [87-89]. Furthermore, the existing mouse knock-out model produced a phenotype with combined cognitive and motor impairment.

LHX5 is a member of the LIM-homeobox genes that encode a family of transcription factors that are highly conserved throughout evolution [77, 78]. Current understanding of the manifold function of LHX genes points to a major role in forebrain development, as well as patterning and differentiation of diverse other cell types ([77, 90]. Lhx5 was first cloned in zebrafish, and subsequently in higher animals. Expression was detected in the forebrain, and more specifically in hippocampal precursor cells [80]. It contains 2 zinc finger motifs called LIM domains (AA 3-125) located N-terminally of the DNA binding homeodomain (AA180-239)[79].
Zhao *et al.* created a functional *lhx5* knockout mouse by targeted deletion of exon 2-4 abolishing the second LIM domain and the homeobox domain. The mice were born live, however most of them died within a few days after birth [80].
In embryonic development (at E 18.5 of a total gestation time of 27 days) the hippocampus showed histological defects. The ventricular zone was thicker than normal, corresponding to an increased number of proliferating cells. These cells did migrate, but failed to position themselves properly to form the distinctive structures of the hippocampus. These postmitotic cells acquired certain identities of neurons or glial cells, but did not differentiate further into the various subclasses of hippocampal neurons in the absence of *lhx5* function.

The pyramidal cell layer in the Ammons' horn and the granule cell layer in the dentate gyrus were the hippocampal structures most affected.
Hence, in mice *lhx5* seems to be required for proper differentiation and migration of neurons in the hippocampus, a forebrain structure that is considered essential for cognition, learning, and memory in adult animals [80].

Besides the abnormal neuroanatomy of the hippocampal area, the choroid plexus of both the lateral ventricle and the third ventricle was missing, and the anterior callosal axons failed to cross the midline. There were no detectable morphological defects in other regions of the forebrain, hindbrain, midbrain or spinal cord where *lhx5* is normally expressed.
This is possibly due to a functional compensation by *lhx1*, which is co-expressed with *lhx5* in these areas [80].

Behavioural phenotyping of *lhx5* null mutant mice revealed deficits on learning and memory tasks. General health, neurological reflexes and sensory abilities appeared to be normal in those mice that survived the first postnatal days.
Motor tests showed impaired performance on some measures of coordination, balance and gait as well has increased general motor activity [91].

Sequencing of *LHX5* revealed a non-synonymous mutation resulting in amino acid substitution H362L. Pathogenicity predicting programs yielded ambiguous results regarding the point mutation in *LHX5* located outside the known functional domains.
Since inactivation of the *lhx5* gene in mouse results in gross anatomical changes in the hippocampus, we reasoned that pathogenic mutations in humans leading to a severe cognitive disorder would also induce observable anatomical changes in the brain. However, MRI scans of 5 patients did not reveal structural abnormalities of the hippocampal area, thereby providing another argument against the pathogenicity of the *LHX5* variant in this family.

The homozygous interval contains 158 other genes. In cooperation with the Max Planck Institute for molecular genetics, high throughput sequencing was performed, after exon enrichment of the identified interval.

Sequencing of the homozygous interval revealed 2 non-synonymous mutations. The previously identified mutation in *LHX5* that resulted in amino acid substitution H162L, and a second mutation in *KIAA1033* coding for SWIP resulting in amino acid substitution P1019R. Controls did not carry these mutations. No additional mutations could be identified in 4

unrelated consanguineous Iranian families with ARID and regions of homozygosity overlapping the relevant linkage interval on chromosome 12q23-24.

In contrast to the *LHX5* mutation, pathogenicity of the SWIP mutation was unambiguously predicted by all applied programs. As this gene is part of a multiprotein complex that has a putative role in endosomal fission, and many other genes involved in endosomal trafficking and actin dynamics have been identified in ID families, we focussed on this gene.

KIAA1033 encodes a protein termed SWIP (Strumpellin and WASH interacting protein). SWIP is a member of the WASH complex, a novel type I actin nucleation promoting factor (NPF).
Actin nucleation -the association of actin into polymerisable seeds of dimers and trimers- is driven by actin-binding nucleation factors.
Branched actin filament networks are primarily regulated by the nucleation factor Arp2/3. Actin polymerization by the Arp2/3 complex is known to generate a force that pushes or remodels membranes [92], and Arp 2/3 has been shown to function in a broad spectrum of cell processes, like migration, endocytosis, endosomal trafficking and cell-cell communication [93].
Nucleation by Arp 2/3 complex is activated by a class of proteins called type I nucleation promoting factors, which bind to the Arp2/3 complex and actin monomers [94].

All members share a C-terminal Arp2/3 binding CA-module (connector-acidic) linked N terminally to different numbers of actin monomer binding WH2 domains (figure 17). The N terminal regions are highly divergent, and have several functions in targeting of subcellular compartments, offering binding sites for Rho-GTPases or other regulatory proteins [86, 95, 96].

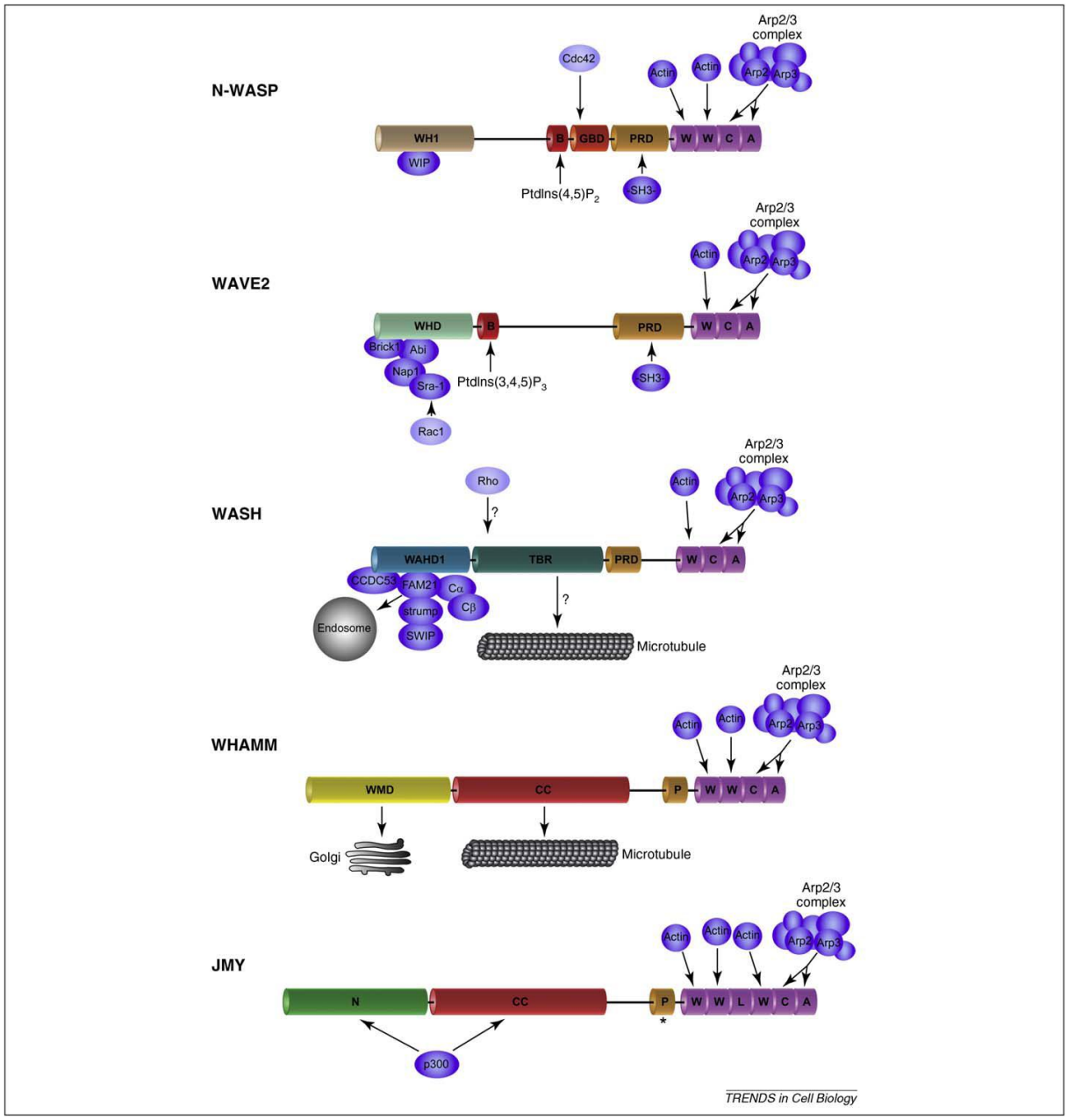


Figure 17. Domain organization of human type I Nucleation Promoting Factors (NPFs). Representative, canonical NPFs are shown, including ubiquitous N-WASP and WAVE2 and the novel family members WASH, WHAMM and JMY. The C-termini of all NPFs shown are characterized by Arp2/3-binding CA-modules linked N-terminally to different numbers of actin monomerbinding WH2 domains (W). The N-termini are highly divergent and are thought to mediate subcellular targeting, Rho-GTPase-binding (Cdc42 in the case of N-WASP[95] or a Rho subfamily member in the case of WASH [86, 97], associations with diverse regulators, or into large protein complexes (i.e. WAVE2 and WASH).

Abbreviations: -SH3-, SH3-domain containing protein; A, acidic region; B, basic domain; C, connector region; Ca, a-subunit of heterodimeric capping protein, Cb, b-subunit of heterodimeric capping protein; CC, coiled-coil region; Ccdc53, coiled-coil domain containing protein 53; GBD, GTPase-binding domain; PRD, proline-rich domain; P,polyproline domain; L, monomer-binding linker (homologous to actin monomer-binding linker in Spir); N, N-terminus; strump, Strumpellin; SWIP, Strumpellin and WASH interacting protein; TBR, tubulin-binding region; W, WH2 domain; WAHD1, WASH homology domain 1; WHD, WAVE homology domain (also known as SHD, Scar homology domain); WH1, WASP homology 1 domain; WIP, WASP interacting protein; WMD, WHAMM membrane interaction domain; * indicates the description of JMY isoforms lacking the polyproline domain [98].

Source [82]

WASP (Wiskott-Aldrich syndrome protein) was the first reported member of this NPF I family, followed by 3 WAVE isoforms. Recently the novel members WASH, WHAMM and JMY were characterized [99-101].
WASP and Neural-WASP are essential for efficient receptor endocytosis [92], while WAVE proteins play a role in lammelipodia formation, migration and cellular adhesion (cellular protrusions). WHAMM controls trafficking from the endoplasmic reticulum to the Golgi [99], the role of JMY is still debated (reviewed by [82]) .

WASH (WASP and Scar homologue) has recently been identified as a member of the WASP family of nucleation promoting factors. The WASH complex is located at the surface of endosomes [70, 83, 84] and it promotes membrane scission of transport intermediates through the branched actin network it induces [70]. This newly discovered complex has already been implicated in numerous endosomal transport pathways, namely the recycling pathway [70], the degradation pathway [83] and the retrograde pathway [84, 85]. These trafficking pathways are especially important in neurons, where cargoes have to travel long distances. Alterations in these pathways are at the origin of several forms of neuropathies [102].
WASH is bound in a multiprotein complex through its N terminal coiled-coil region contained within the WASH homology domain (WAHD1, see figure 17). Members of this complex include CCDC53, Fam 21, Strumpellin and SWIP (Strumpellin and WASH interacting protein) (encoded by *KIAA1033*) [70].

Importantly, we have been able to demonstrate that the mutation of SWIP found in patients with ID has a deleterious effect on the stability of SWIP and the entire WASH complex. In stable multiprotein complexes, like WASH or the analogous WAVE complex, subunits usually depend on each other for their stability [103]. By depleting SWIP expression, it was shown that SWIP is required for the stability of all the other core subunits of the WASH complex [86]. Here we found the same effect in lymphoblastoid cell lines derived from the patients, indicating that the single point mutation Pro1019Arg is able to exert the same effect as the absence of SWIP. In line with the autosomal recessive transmission in our family of ID patients, we have found that the expression of the mutant SWIP subunit does not impair the stability of the WASH complex in a dominant-negative way.
Interestingly, another subunit of the WASH complex has been implicated in a hereditary neurological disease. Mutations in the subunit Strumpellin were identified as a rare cause of autosomal dominant hereditary spastic paraplegia, a disease characterized by the degeneration of upper motor neurons with onset from the 3rd decade [104]. However, these Strumpellin mutations did not lead to destabilization of the WASH complex [86], suggesting

that the mutations affect protein-protein interactions other than the ones required for complex assembly or maintenance.

Multiple other ID genes are involved in actin dynamics and vesicle trafficking. *STXBP1* [105] and *SYP* [106] encode proteins that associate with SNARE complexes involved in membrane fusion. GDI1 and CASK regulate the Rab3 GTPase [107, 108], important for neurotransmitter release and synaptic plasticity. Several ID genes are associated to Rho family of GTPases, that also play a prominent role in actin dynamics. Examples are OPHN1 that acts as a Rho GTPase activating protein (GAP)[109], *ARHGEF6* and *FGD1*, encoding Guanine nucleotide exchange factors (GEFs) for Rac1 and Cdc42 [110-112] or the Rac and Cdc42 effector, PAK3 [110, 111, 113]. Here, it is worth mentioning that WASH has been found to be regulated by Rho in Drosophila [96], even though this does not seem to be the case for the human WASH complex [86]. Whether through interaction with the Rho pathway or interference with other endosomal pathways, we showed that mutated SWIP and its deleterious effect on WASH complex assembly leads to severe neuronal malfunction.

In view of the compelling evidence for a causative role of SWIP, and the arguments against an involvement LHX5 in this disorder, we consider it less likely that the mutations found in the two closely linked and therefore co-segregating genes *LHX5* and *KIAA1033* both contribute to the described phenotype, but at least formally, such a digenic model cannot be excluded.

By combining homozygosity mapping and high throughput sequencing to efficiently analyse homozygous chromosome regions, we were able to identify SWIP as a novel candidate gene for non-syndromic ARID. The spectrum of gene defects that are linked to RHO/RAC/CDC42 pathway, vesicle trafficking and actin polymerization is hereby expanded to the WASP family of nucleation promoting factors.

References

1. Luckasson, R. and A. Reeve, *Naming, defining, and classifying in mental retardation.* Ment Retard, 2001. **39**(1): p. 47-52.
2. Schalock, R.L., et al., *The renaming of mental retardation: understanding the change to the term intellectual disability.* Intellect Dev Disabil, 2007. **45**(2): p. 116-24.
3. Leonard, H. and X. Wen, *The epidemiology of mental retardation: challenges and opportunities in the new millennium.* Ment Retard Dev Disabil Res Rev, 2002. **8**(3): p. 117-34.
4. Roeleveld, N., G.A. Zielhuis, and F. Gabreels, *The prevalence of mental retardation: a critical review of recent literature.* Dev Med Child Neurol, 1997. **39**(2): p. 125-32.
5. Wellesley, D.G., et al., *Prevalence of intellectual handicap in Western Australia: a community study.* Med J Aust, 1992. **156**(2): p. 94-6, 100, 102.
6. Croen, L.A., J.K. Grether, and S. Selvin, *The epidemiology of mental retardation of unknown cause.* Pediatrics, 2001. **107**(6): p. E86.
7. Maulik, P.K., et al., *Prevalence of intellectual disability: a meta-analysis of population-based studies.* Res Dev Disabil, 2011. **32**(2): p. 419-36.
8. Penrose, L., *A clinical study and genetic study of 1280 cases of mental defect.* HSMO, 1938(229).
9. Durkin, M., *The epidemiology of developmental disabilities in low-income countries.* Ment Retard Dev Disabil Res Rev, 2002. **8**(3): p. 206-11.
10. Yeargin-Allsopp, M., et al., *Reported biomedical causes and associated medical conditions for mental retardation among 10-year-old children, metropolitan Atlanta, 1985 to 1987.* Dev Med Child Neurol, 1997. **39**(3): p. 142-9.
11. Bergen, D.C., *Effects of poverty on cognitive function: a hidden neurologic epidemic.* Neurology, 2008. **71**(6): p. 447-51.
12. Cans, C., et al., *Aetiological findings and associated factors in children with severe mental retardation.* Dev Med Child Neurol, 1999. **41**(4): p. 233-9.
13. Drews, C.D., et al., *Variation in the influence of selected sociodemographic risk factors for mental retardation.* Am J Public Health, 1995. **85**(3): p. 329-34.
14. Stromme, P., *Aetiology in severe and mild mental retardation: a population-based study of Norwegian children.* Dev Med Child Neurol, 2000. **42**(2): p. 76-86.
15. Decoufle, P. and C.A. Boyle, *The relationship between maternal education and mental retardation in 10-year-old children.* Ann Epidemiol, 1995. **5**(5): p. 347-53.
16. Breslau, N., et al., *Stability and change in children's intelligence quotient scores: a comparison of two socioeconomically disparate communities.* Am J Epidemiol, 2001. **154**(8): p. 711-7.
17. Rauch, A., et al., *Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation.* Am J Med Genet A, 2006. **140**(19): p. 2063-74.
18. Flint, J. and S. Knight, *The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation.* Curr Opin Genet Dev, 2003. **13**(3): p. 310-6.
19. Stevenson, R.E., et al., *Genetic syndromes among individuals with mental retardation.* Am J Med Genet A, 2003. **123A**(1): p. 29-32.
20. Hoyer, J., et al., *Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays.* J Med Genet, 2007. **44**(10): p. 629-36.
21. Dolk, H., et al., *Trends and geographic inequalities in the prevalence of Down syndrome in Europe, 1980-1999.* Rev Epidemiol Sante Publique, 2005. **53 Spec No 2**: p. 2S87-95.
22. Ropers, H.H., *Genetics of early onset cognitive impairment.* Annu Rev Genomics Hum Genet, 2010. **11**: p. 161-87.
23. Hochstenbach, R., et al., *Array analysis and karyotyping: workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands.* Eur J Med Genet, 2009. **52**(4): p. 161-9.
24. de Vries, B.B., et al., *Diagnostic genome profiling in mental retardation.* Am J Hum Genet, 2005. **77**(4): p. 606-16.
25. Mefford, H.C. and E.E. Eichler, *Duplication hotspots, rare genomic disorders, and common disease.* Curr Opin Genet Dev, 2009. **19**(3): p. 196-204.
26. Jaillard, S., et al., *Identification of gene copy number variations in patients with mental*

retardation using array-CGH: Novel syndromes in a large French series. Eur J Med Genet, 2010. **53**(2): p. 66-75.
27. Koolen, D.A., et al., *Genomic microarrays in mental retardation: a practical workflow for diagnostic applications.* Hum Mutat, 2009. **30**(3): p. 283-92.
28. Vissers, L.E., B.B. de Vries, and J.A. Veltman, *Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis.* J Med Genet, 2010. **47**(5): p. 289-97.
29. Glessner, J.T., et al., *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.* Nature, 2009. **459**(7246): p. 569-73.
30. Zhang, F., et al., *Copy number variation in human health, disease, and evolution.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 451-81.
31. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
32. Najmabadi, H., et al., *Deep sequencing reveals 50 novel genes for recessive cognitive disorders.* Nature, 2011. **478**(7367): p. 57-63.
33. Herbst, D.S. and J.R. Miller, *Nonspecific X-linked mental retardation II: the frequency in British Columbia.* Am J Med Genet, 1980. **7**(4): p. 461-9.
34. Fishburn, J., et al., *The diagnosis and frequency of X-linked conditions in a cohort of moderately retarded males with affected brothers.* Am J Med Genet, 1983. **14**(4): p. 713-24.
35. Ropers, H.H. and B.C. Hamel, *X-linked mental retardation.* Nat Rev Genet, 2005. **6**(1): p. 46-57.
36. de Brouwer, A.P., et al., *Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium.* Hum Mutat, 2007. **28**(2): p. 207-8.
37. Gecz, J., D. Cloosterman, and M. Partington, *ARX: a gene for all seasons.* Curr Opin Genet Dev, 2006. **16**(3): p. 308-16.
38. Mandel, J.L. and J. Chelly, *Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations.* Eur J Hum Genet, 2004. **12**(9): p. 689-93.
39. Vissers, L.E., et al., *A de novo paradigm for mental retardation.* Nat Genet, 2010. **42**(12): p. 1109-12.
40. Molinari, F., et al., *Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation.* Science, 2002. **298**(5599): p. 1779-81.
41. Gschwend, T.P., et al., *Neurotrypsin, a novel multidomain serine protease expressed in the nervous system.* Mol Cell Neurosci, 1997. **9**(3): p. 207-19.
42. Wolfer, D.P., et al., *Multiple roles of neurotrypsin in tissue morphogenesis and nervous system development suggested by the mRNA expression pattern.* Mol Cell Neurosci, 2001. **18**(4): p. 407-33.
43. Reif, R., et al., *Specific cleavage of agrin by neurotrypsin, a synaptic protease linked to mental retardation.* FASEB J, 2007. **21**(13): p. 3468-78.
44. Matsumoto-Miyai, K., et al., *Coincident pre- and postsynaptic activation induces dendritic filopodia via neurotrypsin-dependent agrin cleavage.* Cell, 2009. **136**(6): p. 1161-71.
45. Higgins, J.J., et al., *A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation.* Neurology, 2004. **63**(10): p. 1927-31.
46. Higgins, J.J., et al., *Dysregulation of large-conductance Ca2+-activated K+ channel expression in nonsyndromal mental retardation due to a cereblon p.R419X mutation.* Neurogenetics, 2008. **9**(3): p. 219-23.
47. Jo, S., et al., *Identification and functional characterization of cereblon as a binding protein for large-conductance calcium-activated potassium channel in rat brain.* J Neurochem, 2005. **94**(5): p. 1212-24.
48. Rogaeva, A., K. Galaraga, and P.R. Albert, *The Freud-1/CC2D1A family: transcriptional regulators implicated in mental retardation.* J Neurosci Res, 2007. **85**(13): p. 2833-8.
49. Rogaeva, A., et al., *Differential repression by freud-1/CC2D1A at a polymorphic site in the dopamine-D2 receptor gene.* J Biol Chem, 2007. **282**(29): p. 20897-905.
50. Motazacker, M.M., et al., *A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation.* Am J Hum Genet, 2007. **81**(4): p. 792-8.
51. Contractor, A., G. Swanson, and S.F. Heinemann, *Kainate receptors are involved in short- and long-term plasticity at mossy fiber synapses in the hippocampus.* Neuron, 2001. **29**(1): p. 209-16.
52. Garshasbi, M., et al., *A defect in the TUSC3 gene is associated with autosomal recessive mental retardation.* Am J Hum Genet, 2008. **82**(5): p. 1158-64.

53. Molinari, F., et al., *Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation.* Am J Hum Genet, 2008. **82**(5): p. 1150-7.
54. Mir, A., et al., *Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation.* Am J Hum Genet, 2009. **85**(6): p. 909-15.
55. Philippe, O., et al., *Combination of linkage mapping and microarray-expression analysis identifies NF-kappaB signaling defect as a cause of autosomal-recessive mental retardation.* Am J Hum Genet, 2009. **85**(6): p. 903-8.
56. Mochida, G.H., et al., *A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly.* Am J Hum Genet, 2009. **85**(6): p. 897-902.
57. Hu, W.H., et al., *NIBP, a novel NIK and IKK(beta)-binding protein that enhances NF-(kappa)B activation.* J Biol Chem, 2005. **280**(32): p. 29233-41.
58. Caliskan, M., et al., *Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13.* Hum Mol Genet, 2011.
59. Moon, Y.A. and J.D. Horton, *Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade.* J Biol Chem, 2003. **278**(9): p. 7335-43.
60. Kaufman, L., M. Ayub, and J.B. Vincent, *The genetic basis of non-syndromic intellectual disability: a review.* J Neurodev Disord, 2010. **2**(4): p. 182-209.
61. Ropers, H.H., *Genetics of intellectual disability.* Curr Opin Genet Dev, 2008. **18**(3): p. 241-50.
62. Ropers, F., et al., *Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the WASH complex member SWIP.* Hum Mol Genet, 2011. **20**(13): p. 2585-90.
63. Read, T.S.a.A., *Human Molecular Genetics*. 3rd ed. 2003: Garland Publishing. 647.
64. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees.* Nat Genet, 2002. **30**(1): p. 97-101.
65. Gudbjartsson, D.F., et al., *Allegro, a new computer program for multipoint linkage analysis.* Nat Genet, 2000. **25**(1): p. 12-3.
66. Ruschendorf, F. and P. Nurnberg, *ALOHOMORA: a tool for linkage analysis using 10K SNP array data.* Bioinformatics, 2005. **21**(9): p. 2123-5.
67. Schuelke, M., *An economic method for the fluorescent labeling of PCR fragments.* Nat Biotechnol, 2000. **18**(2): p. 233-4.
68. Li, R., et al., *SOAP2: an improved ultrafast tool for short read alignment.* Bioinformatics, 2009. **25**(15): p. 1966-7.
69. Cubillos-Rojas, M., et al., *Simultaneous electrophoretic analysis of proteins of very high and low molecular mass using Tris-acetate polyacrylamide gels.* Electrophoresis, 2010. **31**(8): p. 1318-21.
70. Derivery, E., et al., *The Arp2/3 activator WASH controls the fission of endosomes through a large multiprotein complex.* Dev Cell, 2009. **17**(5): p. 712-23.
71. Durbin, R.M., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
72. Li, Y., et al., *Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants.* Nat Genet, 2010. **42**(11): p. 969-72.
73. Ferrer-Costa, C., et al., *PMUT: a web-based tool for the annotation of pathological mutations on proteins.* Bioinformatics, 2005. **21**(14): p. 3176-8.
74. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.
75. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
76. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations.* Nat Methods, 2010. **7**(8): p. 575-6.
77. Dawid, I.B., J.J. Breen, and R. Toyama, *LIM domains: multiple roles as adapters and functional modifiers in protein interactions.* Trends Genet, 1998. **14**(4): p. 156-62.
78. Zhao, Y., et al., *Genomic structure, chromosomal localization and expression of the human LIM-homeobox gene LHX5.* Gene, 2000. **260**(1-2): p. 95-101.
79. Hunter, C.S. and S.J. Rhodes, *LIM-homeodomain genes in mammalian development and human disease.* Mol Biol Rep, 2005. **32**(2): p. 67-77.
80. Zhao, Y., et al., *Control of hippocampal morphogenesis and neuronal differentiation by the LIM homeobox gene Lhx5.* Science, 1999. **284**(5417): p. 1155-8.
81. Peng, G. and M. Westerfield, *Lhx5 promotes forebrain development and activates*

transcription of secreted Wnt antagonists. Development, 2006. **133**(16): p. 3191-200.
82. Rottner, K., J. Hanisch, and K.G. Campellone, *WASH, WHAMM and JMY: regulation of Arp2/3 complex and beyond.* Trends Cell Biol, 2010. **20**(11): p. 650-61.
83. Duleh, S.N. and M.D. Welch, *WASH and the Arp2/3 complex regulate endosome shape and trafficking.* Cytoskeleton (Hoboken), 2010. **67**(3): p. 193-206.
84. Gomez, T.S. and D.D. Billadeau, *A FAM21-containing WASH complex regulates retromer-dependent sorting.* Dev Cell, 2009. **17**(5): p. 699-711.
85. Harbour, M.E., et al., *The cargo-selective retromer complex is a recruiting hub for protein complexes that regulate endosomal tubule dynamics.* J Cell Sci, 2010. **123**(Pt 21): p. 3703-17.
86. Jia, D., et al., *WASH and WAVE actin regulators of the Wiskott-Aldrich syndrome protein (WASP) family are controlled by analogous structurally related complexes.* Proc Natl Acad Sci U S A, 2010. **107**(23): p. 10442-7.
87. Scoville, W.B. and B. Milner, *Loss of recent memory after bilateral hippocampal lesions.* J Neurol Neurosurg Psychiatry, 1957. **20**(1): p. 11-21.
88. Murray, E.A. and M. Mishkin, *Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy.* J Neurosci, 1986. **6**(7): p. 1991-2003.
89. Squire, L.R. and P. Alvarez, *Retrograde amnesia and memory consolidation: a neurobiological perspective.* Curr Opin Neurobiol, 1995. **5**(2): p. 169-77.
90. Hobert, O. and H. Westphal, *Functions of LIM-homeobox genes.* Trends Genet, 2000. **16**(2): p. 75-83.
91. Paylor, R., et al., *Learning impairments and motor dysfunctions in adult Lhx5-deficient mice displaying hippocampal disorganization.* Physiol Behav, 2001. **73**(5): p. 781-92.
92. Takenawa, T. and S. Suetsugu, *The WASP-WAVE protein network: connecting the membrane to the cytoskeleton.* Nat Rev Mol Cell Biol, 2007. **8**(1): p. 37-48.
93. Goley, E.D. and M.D. Welch, *The ARP2/3 complex: an actin nucleator comes of age.* Nat Rev Mol Cell Biol, 2006. **7**(10): p. 713-26.
94. Stradal, T.E. and G. Scita, *Protein complexes regulating Arp2/3-mediated actin assembly.* Curr Opin Cell Biol, 2006. **18**(1): p. 4-10.
95. Stradal, T.E., et al., *Regulation of actin dynamics by WASP and WAVE family proteins.* Trends Cell Biol, 2004. **14**(6): p. 303-11.
96. Liu, R., et al., *Wash functions downstream of Rho and links linear and branched actin nucleation factors.* Development, 2009. **136**(16): p. 2849-60.
97. Liu, W., et al., *Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer.* Cancer Res, 2009. **69**(6): p. 2176-9.
98. Coutts, A.S., L. Weston, and N.B. La Thangue, *Actin nucleation by a transcription co-factor that links cytoskeletal events with the p53 response.* Cell Cycle, 2010. **9**(8): p. 1511-5.
99. Campellone, K.G., et al., *WHAMM is an Arp2/3 complex activator that binds microtubules and functions in ER to Golgi transport.* Cell, 2008. **134**(1): p. 148-61.
100. Shikama, N., et al., *A novel cofactor for p300 that regulates the p53 response.* Mol Cell, 1999. **4**(3): p. 365-76.
101. Zuchero, J.B., et al., *p53-cofactor JMY is a multifunctional actin nucleation factor.* Nat Cell Biol, 2009. **11**(4): p. 451-9.
102. Dion, P.A., H. Daoud, and G.A. Rouleau, *Genetics of motor neuron disorders: new insights into pathogenic mechanisms.* Nat Rev Genet, 2009. **10**(11): p. 769-82.
103. Derivery, E. and A. Gautreau, *Generation of branched actin networks: assembly and regulation of the N-WASP and WAVE molecular machines.* Bioessays, 2010. **32**(2): p. 119-31.
104. Valdmanis, P.N., et al., *Mutations in the KIAA0196 gene at the SPG8 locus cause hereditary spastic paraplegia.* Am J Hum Genet, 2007. **80**(1): p. 152-61.
105. Hamdan, F.F., et al., *De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy.* Ann Neurol, 2009. **65**(6): p. 748-53.
106. Tarpey, P.S., et al., *A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation.* Nat Genet, 2009. **41**(5): p. 535-43.
107. D'Adamo, P., et al., *Mutations in GDI1 are responsible for X-linked non-specific mental retardation.* Nat Genet, 1998. **19**(2): p. 134-9.
108. Hackett, A., et al., *CASK mutations are frequent in males and cause X-linked nystagmus and variable XLMR phenotypes.* Eur J Hum Genet, 2010. **18**(5): p. 544-52.
109. Billuart, P., et al., *Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation.* Nature, 1998. **392**(6679): p. 923-6.
110. Manser, E., et al., *Molecular cloning of a new member of the p21-Cdc42/Rac-activated kinase (PAK) family.* J Biol Chem, 1995. **270**(42): p. 25070-8.

111. Daniels, R.H. and G.M. Bokoch, *p21-activated protein kinase: a crucial component of morphological signaling?* Trends Biochem Sci, 1999. **24**(9): p. 350-5.
112. Zheng, Y., et al., *The faciogenital dysplasia gene product FGD1 functions as a Cdc42Hs-specific guanine-nucleotide exchange factor.* J Biol Chem, 1996. **271**(52): p. 33169-72.
113. Allen, K.M., et al., *PAK3 mutation in nonsyndromic X-linked mental retardation.* Nat Genet, 1998. **20**(1): p. 25-30.

Acknowledgements

First of all, I would like to thank Prof Dr. Karl Sperling and Dr. Raymonda Varon-Mateeva. When I was looking for a research project connecting Pediatric Endocrinology to Genetics they generously offered me to study two interesting large consanguineous families with pubertal delay and intellectual disability, respectively. Their experience, help and advice, not only concerning laboratory matters, has been invaluable for my studies and beyond. I am also grateful to their lab technicians and staff, who patiently showed me all techniques, pretended not be annoyed by my questions, and gave me the opportunity to explore and design all facets of my research project.

Many thanks to Dipl.-Ing. Mohsen Karbasiyan, for all his support with sequences, solving computer problems and making jasmine tea, and equally to Dipl.-Ing. Veronique Dutrannoy-Tönsing and Ralf Eckhardt for all technical support and good humour.

Dr. Masoud Garshasbi introduced me to a whole new world of linkage analysis, and Dominik Seelow patiently assisted me with performing all calculations, thereby showing me the fantastic programs he designed, which have greatly facilitated and improved these analyses.

Thanks are also due to PD Ulrich Thomale who scrutinized the MRT images brought from Oman, and was a reliable lunch and coffee partner.

Last but not least I want to thank my direct supervisor at the Department of Pediatrics, Professor Heiko Krude, who works at the interface of Genetics and Pediatric Endocrinology. This may be the reason why he allowed me to choose a genetic subject for my dissertation. Thanks to all of you, I thoroughly enjoyed my time in the laboratory. It was a great experience, and probably not my last in this field.

Resume

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Erklärung

„Ich, Fabienne Ropers, erkläre, dass ich die vorgelegte Dissertation mit dem Thema: *Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: The WASH complex member SWIP* selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.“

27.6.12