ARTICLE

doi:10.1038/nature10423

Deep sequencing reveals 50 novel genes for recessive cognitive disorders

Hossein Najmabadi[1,2], Hao Hu[3]*, Masoud Garshasbi[1,3]*, Tomasz Zemojtel[4], Seyedeh Sedigheh Abedini[1], Wei Chen[3,5], Masoumeh Hosseini[1], Farkhondeh Behjati[1], Stefan Haas[4], Payman Jamali[6], Agnes Zecha[3], Marzieh Mohseni[1], Lucia Püttmann[3], Leyla Nouri Vahid[1], Corinna Jensen[3], Lia Abbasi Moheb[1,3], Melanie Bienek[3], Farzaneh Larti[1], Ines Mueller[3], Robert Weissmann[3], Hossein Darvish[1], Klaus Wrogemann[3,7], Valeh Hadavi[2], Bettina Lipkowitz[3], Sahar Esmaeeli-Nieh[3], Dagmar Wieczorek[8], Roxana Kariminejad[2], Saghar Ghasemi Firouzabadi[1], Monika Cohen[9], Zohreh Fattahi[1], Imma Rost[10], Faezeh Mojahedi[11], Christoph Hertzberg[12], Atefeh Dehghan[13], Anna Rajab[14], Mohammad Javad Soltani Banavandi[1], Julia Hoffer[3], Masoumeh Falah[1], Luciana Musante[3], Vera Kalscheuer[3], Reinhard Ullmann[3], Andreas Walter Kuss[3]†, Andreas Tzschach[3], Kimia Kahrizi[1] & H. Hilger Ropers[3]

Common diseases are often complex because they are genetically heterogeneous, with many different genetic defects giving rise to clinically indistinguishable phenotypes. This has been amply documented for early-onset cognitive impairment, or intellectual disability, one of the most complex disorders known and a very important health care problem worldwide. More than 90 different gene defects have been identified for X-chromosome-linked intellectual disability alone, but research into the more frequent autosomal forms of intellectual disability is still in its infancy. To expedite the molecular elucidation of autosomal-recessive intellectual disability, we have now performed homozygosity mapping, exon enrichment and next-generation sequencing in 136 consanguineous families with autosomal-recessive intellectual disability from Iran and elsewhere. This study, the largest published so far, has revealed additional mutations in 23 genes previously implicated in intellectual disability or related neurological disorders, as well as single, probably disease-causing variants in 50 novel candidate genes. Proteins encoded by several of these genes interact directly with products of known intellectual disability genes, and many are involved in fundamental cellular processes such as transcription and translation, cell-cycle control, energy metabolism and fatty-acid synthesis, which seem to be pivotal for normal brain development and function.

Early-onset cognitive impairment, or intellectual disability, is an unresolved health care problem and an enormous socio-economic burden. Most severe forms of intellectual disability are due to chromosomal abnormalities or defects in specific genes. For many years, research into the genetic causes of intellectual disability and related disorders has focused on X-chromosome-linked intellectual disability (XLID). It has become clear, however, that X-linked forms account for only 10% of intellectual disability cases, which means that the vast majority of the underlying genetic defects must be autosomal[1]. For severe forms of intellectual disability, autosomal-dominant inheritance is rare because most affected individuals do not reproduce, but recent observations suggest that in outbred Caucasian populations, a significant portion of the sporadic cases may be due to dominant *de novo* mutations[2–4]. So far, relatively little is known about the role of autosomal recessive intellectual disability (ARID), because in Western societies, where most of the research takes place, its investigation has been hampered by infrequent parental consanguinity and small family sizes.

In most Northern African countries, and also in the Near and Middle East, parental consanguinity and large families are common; for example, in Iran, 40% of the families are consanguineous and about two-thirds of the population is 30 years of age or younger. Since 2004, we have performed systematic array-based consanguinity mapping in 272 consanguineous Iranian families. In several dozen families, we have defined single linkage intervals and mapped the underlying gene defects[5,6], and by subsequent mutation screening of candidate genes from these intervals, we and others identified several novel ARID genes (for review see refs 1, 7).

Recently, exome enrichment and next-generation sequencing have been introduced as a cost-effective and fast strategy for comprehensive mutation screening and disease-gene identification in the coding portion of the human genome[8–10]. To unravel the molecular basis of ARID in a systematic fashion, we have now used a related, but more targeted, approach. Instead of sequencing entire exomes in consanguineous families, we have focused on the exons from homozygous linkage intervals known to carry the genetic defect. Before sequencing, these exons were enriched by hybrid capture using custom-made oligonucleotide arrays as baits. All patients had cognitive impairment (mostly moderate or severe, see Supplementary Table 1), and in a subset of the families there were signs of autism spectrum disorder. More information about the families and their clinical features, quality controls performed to validate the sequence variants observed and to assess their pathogenicity, as well as other methodological details are provided in Supplementary Information.

[1]Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, 19857 Tehran, Iran. [2]Kariminejad-Najmabadi Pathology & Genetics Centre, 14667 Tehran, Iran. [3]Department Human Molecular Genetics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. [4]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. [5]Max-Delbrueck-Centrum für Molekulare Medizin, 13092 Berlin, Germany. [6]Shahroud Welfare Organization, 36156 Semnan, Iran. [7]Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba R3E0J9, Canada. [8]Institut fuer Humangenetik, Universitaetsklinikum, 45122 Essen, Germany. [9]Kinderzentrum Muenchen, 81377 Muenchen, Germany. [10]Zentrum fuer Humangenetik und Laboratoriumsmedizin Dr Klein und Dr Rost, 82152 Martinsried, Germany. [11]Mashhad Medical Genetic Counseling Center, 91767 Mashhad, Iran. [12]Kinderneurologie und Sozialpaediatrie, Vivantes-Klinikum Neukölln, 12351 Berlin, Germany. [13]Yazd Welfare Organization, 89178 Yazd, Iran. [14]Genetics Unit, Ministry of Health, Directorate General of Health Affairs, Royal Hospital, Muscat 113, Oman. †Present address: Institute for Human Genetics, University Medicine Greifswald & Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University, 17489 Greifswald, Germany.
*These authors contributed equally to this work.

Mutations in known and novel intellectual disability genes

In 115 out of 136 families studied, plausible causal defects were observed, and in 78 of these, a single, apparently disease-causing mutation could be identified (see Supplementary Fig. 1, Tables 1 and 2 and Supplementary Table 2). Twenty-eight protein-truncating changes were found, including frameshift, splice-site and nonsense mutations, as well as whole-exon deletions, plus several smaller in-frame deletions of varying size. In 26 families listed in Table 1, we identified known, mostly syndromic forms of ARID, including rare metabolic defects and storage disorders, such as an atypical form of Tay–Sachs' disease and Sanfilippo's syndrome (mucopolysaccharidosis IIIb), as well as intellectual disability with congenital abnormalities, such as a Joubert-like syndrome resulting from *AHI1* mutations, observed in two unrelated families. Two families were also found with allelic *PRKCG* mutations, implicated previously in spinocerebellar ataxia, and two families carried different allelic mutations in the *SRD5A3* gene, associated with Kahrizi's syndrome, a recently elucidated congenital glycosylation disorder[11,12].

Two mutations involving the adaptor protein complex 4 were observed, namely in the *AP4M1* and *AP4E1* genes, which encode different AP-4 subunits. AP-4 is involved in the recognition and sorting of cargo protein transported from the trans-Golgi network to the endosomal–lysosomal system. Another possibly pathogenic change was found in the *AP4B1* gene, but its effect may be obscured by a *PEX6* mutation in the same family, which causes a severe peroxisome biosynthesis disorder[13] and probably accounts for most of the clinical features. In highly inbred families, coexistence of two different recessive defects is not unexpected and is the most plausible explanation for the complex phenotypes in at least two families with novel forms of ARID (M154 and M189, see Table 2).

Mutations in the *SLC2A1* gene, which encodes a glucose transporter, the *PRKRA* gene with a role in dysautonomia, and the *MED13L* gene, previously associated with intellectual disability and cardiac symptoms, were the only plausible causes of intellectual disability in three families with non-syndromic intellectual disability. None of the respective families showed signs of dysautonomia or cardiac abnormalities. In all other families, the phenotype was characteristic for the molecular defect, including family M198 with folate receptor deficiency, a rare syndromic form of ARID that can often be treated by oral administration of folinic acid[14]. Further details are provided in Table 1.

Apparently pathogenic changes were also found in 50 genes that had not been previously implicated in ARID (see Table 2). Thirty of the relevant families had non-syndromic forms of intellectual disability, whereas 22 exhibited syndromic forms. Only two of the novel ARID genes were mutated in more than a single family. Two different missense mutations with high pathogenicity scores were detected in *ZNF526*, which encodes a krüppel-type zinc-finger protein. One of these changes was observed in DNA samples collected from two distinct families with non-syndromic intellectual disability, but closer inspection revealed that these families, which live in the same city in the northwestern part of Iran, share a common haplotype and thus must be distantly related. In these families, no other potentially disease-causing and co-segregating change could be identified. Zinc-finger proteins are transcriptional regulators, and other krüppel-type zinc-finger genes have been implicated in intellectual disability before[15]. Recent protein interaction studies have indicated a role for ZNF526 in promoting messenger RNA translation and cell growth (N. Hubner *et al.*, personal communication). Another gene within which disease-causing mutations were found in two families was *ELP2*. It encodes a subunit of the RNA polymerase II elongator complex, which is a histone acetyltransferase component of RNA polymerase II. This gene is involved in the acetylation of histones H3 and probably H4, and it may have a role in chromatin remodelling.

Mutations affecting housekeeping genes

In the *LARP7* gene, we found a frameshift mutation in a family with intellectual disability and microcephaly. *LARP7* is a negative transcriptional regulator of polymerase II genes, acting by means of the 7SK RNP system. Within the 7SK RNP complex, the positive transcription elongation factor b (P-TEFb) is sequestered in an inactive form, preventing RNA polymerase II phosphorylation and subsequent transcriptional elongation. Hitherto, no disease association has been reported for *LARP7*.

Presumably causative homozygous mutations were also found in *KDM5A* and *KDM6B*. These genes encode histone demethylases that specifically demethylate histone H3 at lysine 4 and lysine 27, respectively, and they both have a central role in the histone code. We have

Table 1 | Mutations identified in known genes for intellectual disability or related disorders

Family	Gene	Mutation	LOD score	Length (Mb)	OMIM no.	Diagnosis, clinical features
8500306	*AHI1*	R329X	2.65	10.35	608629	Joubert's syndrome 3
M332	*AHI1*	R495H	3.2	11.1	608629	Joubert's syndrome 3
M254	*AP4E1*	V454fs	2.5	13.57	607244	Microcephaly, paraplegia
M004	*AP4M1*	E193K	1.9	16.75	602296	Microcephaly, paraplegia
M324	*BBS7*	533del2aa	3.24	8.2	209900	Bardet–Biedl's syndrome
M107	*CA8*	R237Q	2.4	4.02	613227	Ataxia, cerebellar hypoplasia
M175	*COL18A1*	L1587fs	2.1	9.8	267750	Knobloch's syndrome (eye and brain development)
G026	*FAM126A*	Splice site*	2.4	15.46	610532	Hypomyelination-cataract
M198	*FOLR1*	Splice site*	2.1	16.95	136430	Folate receptor deficiency
M165	*HEXA*	C58Y	2.7	15.91	272800	Psychomotor delay, mild Tay–Sachs' disease
8600276†	*L2HGDH*	R335X	5.1	13.39	609584	Hydroxyglutaric aciduria
M142	*MED13L*	R1416H	1.9	9.17	608808	Non-syndromic ID, no cardiac involvement
8600486	*NAGLU*	R565Q	2.8	13.25	252920	Sanfilippo's syndrome, MPS IIIB
8500234	*PDHX*	R15H	3.13	35.17	245349	Pyruvate dehydrogenase defect
M331	*PEX6*	L534P	3.8	10.83	601498	Peroxisome biogenesis disorder
8307998	*PMM2*	Y106F	2.67	6.71	212065	Glycosylation disorder CDG Ia
8600273	*PRKCG*	V177fs	2.53	0.72	605361	Spinocerebellar ataxia 14
M146	*PRKCG*	D480Y	2.1	7.45	605361	Spinocerebellar ataxia 14
8600162	*PRKRA*	S235T	2.1	40.02	612067	Non-syndromic ID
8600042	*SLC2A1*	V237M	3.73	16.7	606777	Non-syndromic ID
8700017	*SRD5A3*	Y169C	4.8	10.5	612713	Kahrizi's syndrome, CDG
M069†	*SRD5A3*	A68fs	3.01	10.44	612713	Kahrizi's syndrome, CDG
G008	*SURF1*	W227R	1.8	4.59	185620	Leigh's syndrome, very mild form
8600041	*TH*	R202H	2.1	7.23	605407	infantile parkinsonism, Segawa's syndrome
M017N	*VRK1*	R133C	3.4	3	607596	Pontocerebellar hypoplasia
M196	*WDR62*	G705G	2.1	18.33	600176	Microcephaly, cerebellar atrophy

CDG, congenital disorder of glycosylation; fs, frameshift; ID, intellectual disability; LOD, logarithm of the odds; MPS, mucopolysaccharidosis; OMIM, Online Mendelian Inheritance in Man.
* See Supplementary Information for further details.
† Remotely related, degree of consanguinity is not clear, analysis performed under conservative assumption of second degree consanguinity.

previously shown that mutations in another lysine-specific histone demethylase, *KDM5C* (also called *JARID1C*), are a relatively frequent cause of X-linked intellectual disability[16]. In two other families, we observed apparently pathogenic mutations that involved histones directly: a frameshift mutation in the *HIST1H4B* gene which belongs to the histone 4 family, and a *HIST3H3* missense mutation with high pathogenicity scores that was the only plausible change in a family with non-syndromic intellectual disability. Together, at least ten of the novel candidate genes for ARID involve histone structure, histone modification, chromatin remodelling or the regulation of transcription, and many of these genes are functionally linked to known and novel intellectual disability genes, as shown in Fig. 1a.

Several other mutated genes are directly or indirectly involved in the regulation of translation. A homozygous frameshift mutation inactivating the *TRMT1* gene was detected in a family with non-syndromic intellectual disability. *TRMT1* is an RNA methyltransferase that dimethylates a single guanine residue at position 26 of most tRNAs. Previously we and others have shown that inactivation of the X-linked gene *FTSJ1*, another RNA methyltransferase, also gives rise to non-syndromic intellectual disability[17,18], and we have recently identified several ARID families with truncating mutations in a third RNA methyltransferase (L.A.M. *et al.*, manuscript in preparation). A large deletion in the *EEF1B2* gene was the only detectable defect in another family with non-syndromic intellectual disability. *EEF1B2* encodes the elongation factor 1β, which is involved in the transport of aminoacyl-tRNAs to the ribosomes. In yet another family with non-syndromic intellectual disability, a missense change was found in *ADRA2B*. This gene encodes a brain-expressed G-protein-coupled receptor that associates with EIF2B, a guanine exchange factor regulating translation[19]; notably, ADRA2B also interacts with the 14-3-3 protein, which in turn associates with RGS7, another novel ARID gene product that regulates G-protein signalling. Finally, in a family with a syndromic form of intellectual disability, a missense change was found in the *POLR3B* gene, involving a nucleotide with a very high conservation score and predicted to be pathogenic by Mutation Taster[20]. *POLR3B* encodes the second-largest core component of RNA polymerase III, which synthesizes small RNAs such as tRNAs and 5S rRNAs[21] and also interacts with ENTPD1, the product of a novel candidate gene for intellectual disability (see GeneCards, http://www.genecards.org/cgi-bin/cardsearch.pl?search=POLR3B and Table 2). Together, these observations indicate that gene defects interfering with transcription and translation are particularly important causes of intellectual disability.

However, we also found pathogenic mutations affecting other fundamental cellular functions and pathways such as cell-cycle control, as illustrated by a mutation inactivating *CCNA2*, and another one truncating *SCAPER*, a specific regulator of the CCNA2–CDK2 complex (see Fig. 1b). The *C11orf46* gene encodes TTI2, a subunit of the Triple T complex, which is required for the establishment of cell-cycle checkpoints and for DNA-damage signalling[22]. Other mutations involved fatty-acid synthesis and turnover (*ACBD6*, *FASN* and *PECR*; see Table 2), protein degradation (*UBR7*), splicing (*ZCCHC8*) and cell migration (*LAMA1*).

Intellectual disability genes with brain-specific functions

Not surprisingly, several mutations involved genes with neuron- or brain-specific functions. For example, we found a frameshift mutation abolishing the function of *CACNA1G*, a T-type calcium channel with a critical role in the generation of $GABA_B$ receptor-mediated spike and wave discharges in the thalamocortical pathway[23,24]. A nonsense mutation inactivated *ZBTB40*, which has a role in glia cell differentiation[25], and other observed changes are expected to interfere with the regulation of neurotransmission, exocytosis or neurotransmitter release. Our study also adds several novel intellectual-disability-associated genes to the Ras and Rho pathway (see Fig. 1c); for example, a convincing missense mutation in the *RALGDS* gene was the only variant detected in one family with non-syndromic intellectual disability. This gene encodes an effector of the Ras-related GTPase Ral, which stimulates the dissociation of GDP from the Ras-related RalA and RalB GTPases, thereby allowing GTP binding and activation of the GTPases[26]. Regulators of small GTPases were among the first genes to be implicated in non-syndromic intellectual disability[27,28]. We also found a homozygous frameshift mutation in *CNKSR1*, which is physically associated with *RALGDS*. Homozygous carriers of this mutation have a severe syndromic phenotype with quadrupedal gait. CNKSR1 binds to rhophilin (Online Mendelian Inheritance in Man (OMIM) 601031), a Rho effector, suggesting that it acts as a scaffold protein and mediates crosstalk between the Ras and Rho GTPase signalling pathways[29]. Neither *RALGDS* nor *CNKSR1* had been implicated in intellectual disability so far; thus, both are novel ARID genes.

Genes without obvious link to intellectual disability

For several of the sequence variants, there is no obvious functional link between the molecular defect and intellectual disability. This applies to *LINS1* and *NDST1*, and it is not easy to understand why

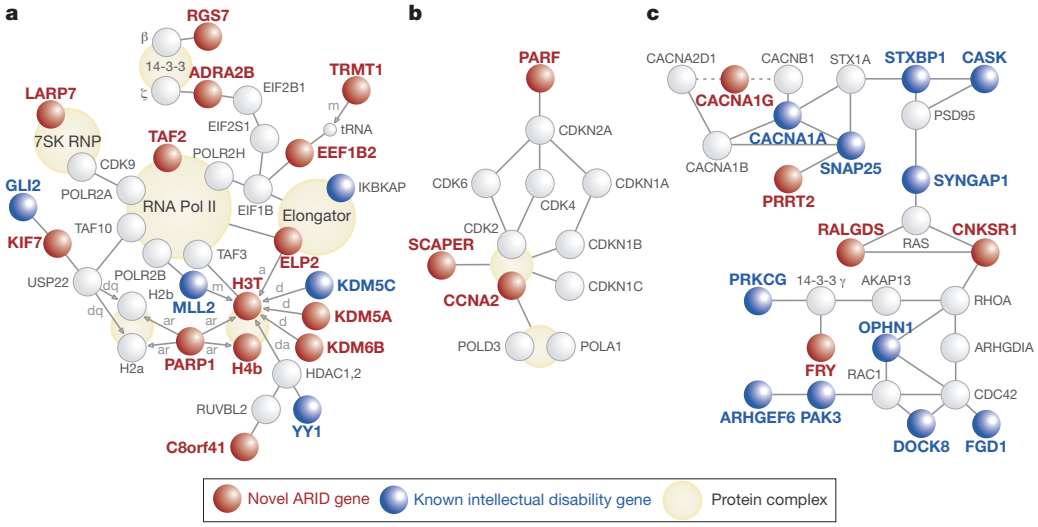

Figure 1 | Known and novel intellectual disability genes form protein and regulatory networks. **a**, Transcriptional/translational network. **b**, Cell-cycle-related network. **c**, Ras/Rho/PSD95 network. Connecting edges in the figure stand for protein–protein interactions. Arrows define direction of post-translational protein modifications: a, acetylation; ar, ADP-ribosylation; d, demethylation; da, deacetylation; dq, deubiquitination; m, methylation. Dotted lines indicate modulation of gene function. Data were obtained in part by using the INGENUITY software package (http://www.ingenuity.com) and by literature mining. More details about these proteins and their interactions are provided in Table 2 and in Supplementary Information.

Table 2 | Apparently causative variants in novel (candidate) genes for intellectual disability

Family	Phenotype	Gene	Mutation	LOD score	Length (Mb)	Supporting evidence
M008†	S	*ACBD6*	G22fs	2.65	6.46	P; binds long-chain acyl-CoA molecules, role in fatty acid synthesis or turnover[44].
M173	NS, ASD	*ADK*	H324R	5.1	9.68	S, P; only change in family. Adenosine kinase, regulates adenosine levels in the brain. Overexpression leads to learning impairment in mice[45]; knockout mice develop lethal neonatal liver steatosis[30]. In human, a different gene defect has been found in this condition.
M266-2	NS	*ADRA2B*	R440G	2.53	24.97	S, P; GPCR regulating adrenergic neurons in the CNS. Associates with EIF2B, a GEF regulating translation[19]. Also associates with 14-3-3, which interacts with RGS7, mutated in family 8700136.
M226	NS	*ASCC3*	S1564P	3.2	62.80	S, P, E; helicase that is part of the activating signal co-integrator complex, enhances NF-κB and AP1. Interacts with RARS2, implicated in pontocerebellar hypoplasia 6[46].
M007L‡	NS	*ASCL1*	A41S	2.4	18.13	Encodes the bHLH factor MASH1, critical role in neuronal commitment and differentiation[47,48].
M182	NS	*C11orf46*	R236H	2.1	12.39	P, E; encodes subunit of the Triple T complex, role in regulation of DNA damage response[22].
G001	NS	*C12orf57*	M1V	2.5	11.19	S; function hitherto unknown. May overlap neighbouring *ANT1* (DRPLA) gene (see UCSC Genome Browser, hg18; OMIM 125370).
M100	NS	*C8orf41*	P367L	3.3	6.44	S, P, E; C8orf41 associates with RUVBL2[49], which is involved in regulation of transcription and interacts with HDACs[50].
G015	NS	*C9orf86*	A562P	3.3	2.17	P; encodes Rab-like GTP-binding protein PARF, which interacts with ARF (or CDKN2A). Other Rab has been implicated in ID[49].
8500031	S	*CACNA1G*	S1346fs	2.7	18.76	P, E; encodes a low-voltage-activated calcium channel which may also modulate the firing patterns of neurons[23,24].
8600057	S	*CAPN10*	138ins5aa	2.1	2.09	E; calcium-regulated non-lysosomal endopeptidase with a role in cytoskeletal remodelling and signal transduction, involved in long-term potentiation[51].
8600495	NS	*CASP2*	Q392X	2.5	29.62	P; caspase 2, role in apoptosis, abnormal in CASP2-deficient mice, particularly for motor and sympathetic neurons[52]. Motor abnormalities not observed in family.
M346	NS	*CCNA2*	Splice site*	3.3	52.17	S, P; cyclin A2 is essential for cell cycle control[53]. In mice, targeted deletion of this gene is lethal[54]. Regulated by *SCAPER*, mutated in family 8600277.
8500235‡	S	*CNKSR1*	T282fs	2.53	15.83	P; regulates Raf in the MAPK pathway, acts as scaffold protein linking Ras and Rho signal transduction pathways[29]. Interacts with RALGDS, which is mutated in family 8500155.
M144	NS	*COQ5*	G118S	1.8	15.10	P, E; methyltransferase with pivotal role in coenzyme Q biosynthesis. Interacts with NAB2 which controls length of poly(A) tail (see http://thebiogrid.org/35094/summary/saccharomyces-cerevisiae/coq5.html). The human orthologue of NAB2 is implicated in ARID[32].
M178	NS	*EEF1B2*	Splice site*	2.6	13.84	S, P, E; controls translation by transferring aminoacyl-tRNAs to the ribosome. Interacts with UNC51-like kinase 2 which is involved in axonal elongation translation[55].
G017	NS	*ELP2*	T555P	2.4	14.33	P, E; encodes subunit of the RNA polymerase II elongator complex[56]. ELP3 subunit implicated in motor neuron degeneration. Allelic *ELP2* mutation found in family M8500061.
8500061	NS	*ELP2*	R462L	2.7	16.98	P, E; involved in transcriptional elongation, see also family G017 with allelic *ELP2* mutation.
M263	NS	*ENTPD1*	Y65C	2.65	12.12	P, E; ectonucleoside triphosphate diphosphohydrolase, expressed in CNS; knockout mice display abnormal synaptic transmitter release[57].
M050†	S	*ERLIN2*	R36K	3.73	12.72	S, P,E; involved in the ER-associated degradation of inositol 1,4,5-triphosphate receptors[58].
8500058	NS	*FASN*	R1819W	3.3	4.50	P; gene product synthesizes long-chain fatty acids from acetyl-CoA and malonyl-CoA. Expressed in post-synaptic density. In mice, FASN deficiency leads to embryonic lethality[59].
M269	S	*FRY*	R1197X	2.8	12.68	P; regulates actin cytoskeleton, limits dendritic branching. In HeLa cells, FRY binds to microtubules and localizes on the spindle and is crucial for the alignment of mitotic chromosomes[60].
M251	S	*GON4L*	Splice site*	3.01	40.19	P, E; cloned from brain. Encodes a transcription factor thought to function in cell cycle control[61].
M189‡	S	*HIST1H4B*	K9fs	2.1	48.87	P, E; encodes a member of the histone H4 family; analogy to histone H3 mutation in family G002. Ehlers–Danlos-related symptoms are probably due to *TNXB* mutation.
G002	NS	*HIST3H3*	R130C	2.53	26.74	P; role in spindle assembly and chromosome bi-orientation[62–64]. See also family M189 with *HIST1H4B* mutation.
8500064	NS	*INPP4A*	D915fs	2.4	46.16	P, E; encodes inositol polyphosphate-4-phosphatase, only plausible change in family. Regulates localization of synaptic NMDA receptors, protects neurons from excitotoxic cell death[65]. Knockout mice develop locomotor instability; not observed in this family.
M061	S	*KDM5A*	R719G	2.3	6.06	P, E; encodes histone demethylase specific for Lys 4 of histone H3, role in transcriptional regulation[66]. Other histone demethylase has been implicated in X-linked ID[16]. See also family M8303971 with *KDM6B* mutation.

in humans, adenosine kinase deficiency should lead to intellectual disability, whereas in the mouse, overexpression of *Adk* causes neurological symptoms, and *Adk* deficiency gives rise to early lethal liver steatosis[30]. Nothing is known yet about the function of the *C12orf57* gene, apart from its apparent overlap with *ATN1* (see UCSC Genome Browser, NCBI36/hg18). CAG trinucleotide expansion in the *ATN1* gene is the cause of dentatorubral pallidoluysian atrophy (DRPLA), another syndromic form of intellectual disability. A comprehensive list of families with single, probably disease-causing mutations is shown in Table 2.

Despite exhaustive validation of our data and stringent filtering against all known neutral and pathogenic sequence variants (see Supplementary Information and Supplementary Tables 3–6), it is still possible that not all of these changes will turn out to be causative. Particularly for the numerous missense mutations observed, functional studies will be required to rule out rare polymorphisms that are unrelated to intellectual disability. In a previous study, 1% of the protein-truncating mutations on the X chromosome were found to be unrelated to disease[31], and in our study, 12 observed inactivating mutations did not co-segregate with intellectual disability (see Supplementary Table 4). However, we believe that the vast majority of the changes presented here as probably pathogenic will be confirmed, even if they have been observed only once, because most of the proteins encoded by these novel candidate genes interact with the

Table 2 | Continued

Family	Phenotype	Gene	Mutation	LOD score	Length (Mb)	Supporting evidence
8303971	S	*KDM6B*	P888S	3.1	5.08	S, P; demethylase 6B specifically targeting Lys 27 of histone H3, has a central role in regulation of posterior development by regulating HOX gene expression[67]. Mutation of *KDM5A* gives rise to ID (see family M061).
M154	S	*KIF7*	E758K	2.1	7.46	P, E; knockout mouse model with complex picture involving brain and other neurological abnormalities[68]. Stickler-like clinical features in this family can be explained by co-existing *COL9A1* mutation.
M183	S	*LAMA1*	G1572fs	2.1	5.82	S, P; codes for subunit of laminin, role in attachment, migration and organization of cells during embryonic development. Required for normal retinal development in mice[69].
G030	S	*LARP7*	K276fs	1.93	8.94	S,P; encodes negative transcriptional regulator of polymerase II genes[70].
7903104	S	*LINS1*	H329fs	2.65	7.87	S, P; similar to *lin*, a *Drosophila* gene having important roles in the development of the epidermis and the hindgut. Link with ID unclear.
8600060†	NS	*MAN1B1*	R334C	3.13	2.49	P, E; encodes mannosidase that targets misfolded glycoproteins for degradation. *MAN1B1* frameshift mutation observed in another ARID family by Canadian group (J. Vincent, personal communication).
8600277	NS	*NDST1*	R709Q	2.1	10.18	S, P; only change in family. Encodes heparan *N*-deacetylase/*N*-sulphotransferase, deficiency is lethal in mice due to respiratory distress[71]. No obvious link with ID.
M158	S	*PARP1*	L293F	1.8	16.76	P; poly(ADP-ribose) polymerase involved in histone 1 modification; role in memory stabilization in mice[72].
M194	NS, ASD	*PECR*	L57V	2.5	11.27	P; brain-expressed peroxisomal *trans*-2-enoyl-CoA reductase involved in the biosynthesis of unsaturated fatty acids[73].
8401214	S	*POLR3B*	T199K	1.93	24.89	E; second-largest core component of RNA polymerase III, which synthesizes small RNAs such as tRNAs and 5S rRNAs[21].
8500302	NS	*PRMT10*	G189R	2.65	9.75	P, E; protein arginine methyltransferase 10. Protein arginine methylation affects chromatin remodelling leading to transcriptional regulation, RNA processing, DNA repair and cell signalling[74].
M010	NS	*PRRT2*	A214fs	5.2	25.59	P; interacts with SNAP25 which in turn assembles with syntaxin-1 and synaptobrevin to form exocytotic fusion complex in neurons[55].
8500155	NS	*RALGDS*	A706V	4.0	5.56	S, E; effector of Ras-related RalA and RalB GTPases, role in synaptic plasticity[26]. Interacts with CNKSR1, inactivated in family 8500235.
8700136	NS, ASD	*RGS7*	N304fs	2.53	24.34	P; regulator of G protein signalling. Interacts with 14-3-3 protein, tau and snapin, a component of the SNARE complex required for synaptic vesicle docking and fusion[75]. Indirectly linked with *ADRA2B*, mutated in family M266_2.
8600086	NS	*SCAPER*	Y118fs	3.9	17.45	S, E; interacts with CCNA2/CDK2 complex, transiently maintains CCNA2 in cytoplasm[76]. *CCNA2* is mutated in family M346.
8600012	S	*SLC31A1*	R90G	2.1	13.85	P, E; encodes one of two genes involved in copper import. Deficiency of the SLC31A1 orthologue in mice is early lethal, heterozygotes have progressive neurological disorder[77], similar to patients in this family.
M177	S	*TAF2*	W649R	2.1	19.16	P, E; TATA-box-associated gene is very important regulator of transcription (see OMIM 604912). Other TAF genes have been implicated in X-linked ID (V.K. *et al.*, manuscript in preparation). MAL2 is another, less likely, candidate in this family.
M160	S	*TMEM135*	C228S	2.4	16.89	S, P, E; transmembrane protein involved in fat metabolism and energy expenditure[78].
M300	NS	*TRMT1*	I230fs	3.4	10.34	P, E; encodes dimethylguanosine tRNA methyltransferase[79]. At least two other RNA methyltransferases have been implicated in ID (ref. 17 and L.A.M., manuscript in preparation).
M168	NS, ASD	*UBR7*	N124S	2.5	8.78	P, E; encodes n-regognin 7, a component of E3 ubiquitin ligase[80]. Involved in protein degradation, which has been implicated in ID.
8500320	S	*WDR45L*	R109Q	1.93	2.55	P, E; WD repeat domain, phosphoinositide-interacting protein 3, ILF1-like[81], specific function unknown.
M169	S	*ZBTB40*	Q525X	3.5	14.56	S, P, E; krüppel-type zinc finger, highly expressed in brain. Regulator of glia differentiation[25].
M156	NS	*ZCCHC8*	L90X	2.3	7.64	P; zinc-finger protein, identified in the spliceosome C complex. Interacts with BRCA1 and RBM7[82,83]. RBM10 has been implicated in X-linked ID (V.K. *et al.*, manuscript in preparation).
M025	NS	*ZNF526*	R459Q	4.5	6.13	P; zinc-finger protein, only remaining change in family. Functional relevance supported by 3D modelling. Probable activator of mRNA translation. Allelic *ZNF526* mutation observed in family 8500156.
8500156	NS	*ZNF526*	Q539H	4.04	11.33	P; see family M025 with allelic *ZNF526* mutation.

References 44–83 are listed in Supplementary Information. E, high evolutionary conservation score; P, high pathogenicity score, includes truncating mutations; S, only change found in family. ASD, autism spectrum disorder; GPCR, G-protein-coupled receptor; ID, intellectual disability; NS, non-syndromic; S, syndromic.
* See Supplementary Information for further details.
† Parents are distantly related. LOD scores provided are minimum estimates, calculated on the assumption that they are second cousins.
‡ In ethnically matching healthy controls a single heterozygous carrier was found (for details, see Supplementary Table 3).

products of known or novel genes associated with intellectual disability, as shown in Fig. 1.

Most ARID genes are not synapse specific

We have previously shown that ARID is an extremely heterogeneous disorder[6]. In contrast to non-syndromic hearing impairment or X-linked intellectual disability, common forms of ARID do not seem to exist, although there is evidence for regional clustering of the underlying gene defects[5]. Extrapolating from the number of known X-chromosomal intellectual disability genes argues for the involvement of several hundred genes in non-syndromic ARID, and the total number of ARID genes may well run into the thousands[1]. Identification of most or all of these genes is a prerequisite for early diagnosis, prevention and, eventually, therapy of intellectual disability, but at the present pace, many years would be required to accomplish this task. Here, we have combined homozygosity mapping, targeted exon enrichment and next-generation sequencing to speed up the molecular elucidation of ARID. In 78 out of 136 consanguineous families investigated, we have found apparently pathogenic mutations in single genes. Fifty of these genes had not been implicated in ARID before, and only two of these novel intellectual disability genes were found to be mutated in two independent families. None of the ~10 previously known genes for non-syndromic ARID, including those that were identified in Iranian families[32–36], was observed in our present

cohort, thereby corroborating previous evidence that ARID is extremely heterogeneous.

Much of the research into the molecular causes of intellectual disability has focused on the synapse and synapse-specific genes (for example, see refs 2, 37). In the present study, relatively few of the novel defects identified involve synapse- or neuron-specific genes, and they are vastly outnumbered by ubiquitously expressed genes with indispensable cellular functions, such as DNA transcription and translation, protein degradation, mRNA splicing, energy metabolism as well as fatty-acid synthesis and turnover. Many of these defects were found to be associated with non-syndromic ARID. It is not immediately clear why the clinical consequences of defects involving such a wide spectrum of basic cellular processes should be confined to the brain, but this conceivably reflects the complexity of the central nervous system which may render it particularly vulnerable to damage.

We expect that these findings will have direct implications for the diagnosis and prevention of intellectual disability, and perhaps also for autism, schizophrenia and epilepsy, which often co-exist in intellectual disability patients and are frequently associated with mutations in the same genes (for example, see ref. 38; reviewed in ref. 1). Further investigation of the novel genes and networks presented here should significantly deepen our insight into the pathogenesis of intellectual disability and related disorders. Moreover, this study illustrates the power of large-scale next-generation sequencing in families as a general strategy to shed light on the aetiology of complex disorders and on the function of the underlying genes.

Note added in proof: While this work was in the press, two unrelated groups reported on inactivating ERLIN2 mutations in patients with recessive intellectual disability and progressive motor dysfunction[39,40]. Moreover, syndromic forms of intellectual disability have been described in patients with AP4B1 and AP4E1 (ref. 41) and MAN1B1 (ref. 42) mutations, respectively. Finally, mutations inactivating the *KIF7* gene were identified as the cause of the recessive fetal hydrolethalus and acrocallosal syndromes that include brain malformations[43].

METHODS SUMMARY

Most families studied were from Iran, and less than 10% had a Turkish or Arabic background. Wechsler Intelligence Scales for Children (WISC) and WAIS were used to assess the IQ in children and parents. Many of the pedigrees, as well as the methods used for autozygosity mapping, have been described previously.

Exons from homozygous intervals were enriched with custom-made Agilent SureSelect DNA capture arrays and sequenced on an Illumina Genome Analyser II yielding 76-bp single reads. >98% of the targeted exons were covered by at least four non-redundant sequence reads, each with a PHRED-like quality score of 20 or above (mean, 0.984; median, 0.993; for details, see Supplementary Table 5).

To assess the reliability of this procedure for calling homozygous mutations, we looked up SNP markers from homozygous intervals of five selected families that had been analysed with high-resolution SNP arrays. For 773 out of 776 markers, next-generation sequencing and array-based SNP typing yielded identical results.

To detect single nucleotide variants, high-quality reads were aligned to the human reference genome (hg18) by SOAP2.20 with default settings, typically gap-free. Homozygous exon-spanning deletions were assumed if the sequence coverage of the relevant exon(s) was reduced to <5% of the mean. Details about the detection of smaller deletions and insertions are provided in Methods. All variants were validated by high-resolution array CGH, Sanger sequencing, or both.

Homozygous variants were filtered against dbSNP130/131, whole genomes from 185 healthy individuals studied by the 1000 Genomes Project and exomes from 200 Danish individuals, and found to be absent in at least 100 chromosomes from Iranian controls (see Supplementary Tables 1 and 3). To select and prioritize apparently disease-causing variants, various criteria were used (for more details, see Methods). All putative mutations co-segregated with intellectual disability in the respective families.

Received 9 March; accepted 5 August 2011.
Published online 21 September 2011.

1. Ropers, H. H. Genetics of early onset cognitive impairment. *Annu. Rev. Genomics Hum. Genet.* **11,** 161–187 (2010).
2. Hamdan, F. F. *et al.* Excess of *de novo* deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88,** 306–316 (2011).
3. Hamdan, F. F. *et al.* Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *N. Engl. J. Med.* **360,** 599–605 (2009).
4. Vissers, L. E. *et al.* A *de novo* paradigm for mental retardation. *Nature Genet.* **42,** 1109–1112 (2010).
5. Najmabadi, H. *et al.* Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum. Genet.* **121,** 43–48 (2007).
6. Kuss, A. W. *et al.* Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Hum. Genet.* **129,** 141–148 (2011).
7. Kaufman, L., Ayub, M. & Vincent, J. B. The genetic basis of non-syndromic intellectual disability: a review. *J. Neurodev. Disord.* **2,** 182–209 (2010).
8. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genet.* **42,** 790–793 (2010).
9. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42,** 30–35 (2010).
10. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362,** 1181–1191 (2010).
11. Cantagrel, V. *et al.* SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell* **142,** 203–217 (2010).
12. Kahrizi, K. *et al.* Next generation sequencing in a family with autosomal recessive Kahrizi syndrome (OMIM 612713) reveals a homozygous frameshift mutation in *SRD5A3*. *Eur. J. Hum. Genet.* **19,** 115–117 (2011).
13. Raas-Rothschild, A. *et al.* A PEX6-defective peroxisomal biogenesis disorder with severe phenotype in an infant, versus mild phenotype resembling Usher syndrome in the affected parents. *Am. J. Hum. Genet.* **70,** 1062–1068 (2002).
14. Steinfeld, R. *et al.* Folate receptor alpha defect causes cerebral folate transport deficiency: a treatable neurodegenerative disorder associated with disturbed myelin metabolism. *Am. J. Hum. Genet.* **85,** 354–363 (2009).
15. Shoichet, S. A. *et al.* Mutations in the *ZNF41* gene are associated with cognitive deficits: identification of a new candidate for X-linked mental retardation. *Am. J. Hum. Genet.* **73,** 1341–1354 (2003).
16. Jensen, L. R. *et al.* Mutations in the *JARID1C* gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76,** 227–236 (2005).
17. Freude, K. *et al.* Mutations in the *FTSJ1* gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* **75,** 305–309 (2004).
18. Ramser, J. *et al.* A splice site mutation in the methyltransferase gene *FTSJ1* in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). *J. Med. Genet.* **41,** 679–683 (2004).
19. Klein, U., Ramirez, M. T., Kobilka, B. K. & von Zastrow, M. A novel interaction between adrenergic receptors and the α-subunit of eukaryotic initiation factor 2B. *J. Biol. Chem.* **272,** 19099–19102 (1997).
20. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* **7,** 575–576 (2010).
21. Geiduschek, E. P. & Kassavetis, G. A. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310,** 1–26 (2001).
22. Hurov, K. E., Cotta-Ramusino, C. & Elledge, S. J. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev.* **24,** 1939–1950 (2010).
23. Zamponi, G. W., Lory, P. & Perez-Reyes, E. Role of voltage-gated calcium channels in epilepsy. *Pflugers Arch.* **460,** 395–403 (2010).
24. Singh, B. *et al.* Mutational analysis of *CACNA1G* in idiopathic generalized epilepsy. *Hum. Mutat.* **28,** 524–525 (2007).
25. Södersten, E., Lilja, T. & Hermanson, O. The novel BTB/POZ and zinc finger factor Zbtb45 is essential for proper glial differentiation of neural and oligodendrocyte progenitor cells. *Cell Cycle* **9,** 4866–4875 (2010).
26. Zhu, X. *et al.* Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc. Natl Acad. Sci. USA* **99,** 10293–10298 (2002).
27. Billuart, P. *et al.* Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392,** 923–926 (1998).
28. D'Adamo, P. *et al.* Mutations in GDI1 are responsible for X-linked non-specific mental retardation. *Nature Genet.* **19,** 134–139 (1998).
29. Jaffe, A. B., Aspenstrom, P. & Hall, A. Human CNK1 acts as a scaffold protein, linking Rho and Ras signal transduction pathways. *Mol. Cell. Biol.* **24,** 1736–1746 (2004).
30. Boison, D. *et al.* Neonatal hepatic steatosis by disruption of the adenosine kinase gene. *Proc. Natl Acad. Sci. USA* **99,** 6985–6990 (2002).
31. Tarpey, P. S. *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature Genet.* **41,** 535–543 (2009).
32. Pak, C. *et al.* Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in *Drosophila* and humans. *Proc. Natl Acad. Sci. USA* **108,** 12390–12395 (2011).
33. Garshasbi, M. *et al.* A defect in the *TUSC3* gene is associated with autosomal recessive mental retardation. *Am. J. Hum. Genet.* **82,** 1158–1164 (2008).
34. Mir, A. *et al.* Identification of mutations in *TRAPPC9*, which encodes the NIK- and IKK-β-binding protein, in nonsyndromic autosomal-recessive mental retardation. *Am. J. Hum. Genet.* **85,** 909–915 (2009).
35. Moheb, L. A. *et al.* Identification of a nonsense mutation in the very low-density lipoprotein receptor gene (VLDLR) in an Iranian family with dysequilibrium syndrome. *Eur. J. Hum. Genet.* **16,** 270–273 (2008).

36. 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.* **81,** 792–798 (2007).
37. Laumonnier, F., Cuthbert, P. C. & Grant, S. G. The role of neuronal complexes in human X-linked brain diseases. *Am. J. Hum. Genet.* **80,** 205–220 (2007).
38. Ullmann, R. *et al.* Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum. Mutat.* **28,** 674–682 (2007).
39. Yildirim, Y. *et al.* A frameshift mutation of ERLIN2 in recessive intellectual disability, motor dysfunction and multiple joint contractures. *Hum. Mol. Genet.* **20,** 1886–1892 (2011).
40. Alazami, A. M., Adly, N., Al Dhalaan, H. & Alkuraya, F. S. A nullimorphic *ERLIN2* mutation defines a complicated hereditary spastic paraplegia locus (SPG18). *Neurogenetics* doi:10.1007/s10048-011-0291-8 (2011).
41. Abou Jamra, R. *et al.* Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am. J. Hum. Genet.* **88,** 788–795 (2011).
42. Rafiq, M. A. *et al.* Mutations in the alpha 1,2-mannosidase gene, *MAN1B1*, cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* **89,** 176–182 (2011).
43. Putoux, A. *et al.* KIF7 mutations cause fetal hydrolethalus and acrocallosal syndromes. *Nature Genet.* **43,** 601–606 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We express our gratitude to the patients and their families for their participation in the study. We thank S. Nakhee and K. Javan for their support; S. Arzhangi, S. Banihashemi, M. Kasiri, H. Khodaee, M. Schlicht and M. Gerloff for contributing to this project in various ways; G. Eder for her assistance with the preparation of the manuscript; and S. Shoichet for critically reading the manuscript. This work was supported by the Max Planck Innovation Funds, the German Federal Ministry of Education and Research through the MRNET (grant 01GS08161, to H.H.R.), the Iranian National Science Foundation and the EU-FP7 project GENCODYS.

Author Contributions H.H.R. and H.N. initiated and directed this study. H.H., M.G., W.C., S.H., K.W., V.K., R.U., K.K. and A.W.K. contributed to its design and coordination. H.N., K.K., A.T., P.J., V.H., D.W., M.C., I.R., F.M., C.H., A.D., A.R., M.J.S.B., M.F. and H.D. recruited patients and families, and K.K. and A.T. were responsible for the clinical investigations. F.B., S.G.F. and R.K. did the karyotyping. M.G. performed the linkage analyses and together with R.W. and H.H., he established data management tools. R.U. and I.M. performed exon enrichments, and C.J. and M.B. did the deep sequencing experiments. H.H. analysed the sequencing data and provided bioinformatics support. M.G., S.S.A., M.H., A.Z., M.M., L.P., L.N.V., L.A.M., F.L., B.L., S.E.-N., Z.F., J.H., L.M. and A.W.K. participated in the validation of the results and the segregation analyses. T.Z. performed the pathway analyses. H.H., M.G., A.Z., L.P., R.W., T.Z., L.M., A.W.K., A.T., K.K., H.N. and H.H.R. evaluated and interpreted the results, and H.H.R., H.H. and M.G. wrote the manuscript.

Author Information Raw sequencing data can be retrieved from the Sequence Read Archive (SRA), accession number SRA036250. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.H.R. (ropers@molgen.mpg.de) or K.K. (kkahrizi@uswr.ac.ir).