Parte 1 - Detailed clinical and results description

Part 2 - Detailed methodology

Part 1 - Detailed clinical and results description

Participants

Patient selection and recruitment

The Life and Health Sciences Research Institute of the School of Health Sciences at University of Minho offers the molecular diagnostic for Rett syndrome in Portugal since 2002 by *MECP2* genetic analysis (Sanger sequencing and dosage analysis by qPCR of all exons). The lab´s work on Rett syndrome includes the setup of a clinical database and detailed clinical analysis of mutation-positive and mutation-negative cases and genoptype-phenotype correlations (and also the behavioral, neuroanatomical and molecular study of a mouse model of the disease, the Mecp2 KO mouse). Due to previous research on this pathology and in intellectual disability as a wider category, a large bank of DNA samples of patients and parents (as well as extensive clinical information) has been obtained. Patients were selected to participate in this project based on clinical suspicion of RTT (typical or atypical) according to the recently revised diagnostic criteria[1]and availability of both parents' DNA. The clinical information was gathered in an anonymous database. The enrollment of the patients and their families was done, after explanation about its purposes, potential pitfalls and eventual benefits by the referring doctor. Written informed consent was obtained for all the participants. This study was approved by the ethics committee of Hospital de Santo António, Centro Hospitalar do Porto.

Clinical description of the patients

Patient 1 – This girl is the middle child of a healthy non-consanguineous couple; her two male sibs are healthy. She was born at full term through vaginal delivery with a weight of 2630g (5-10th centile) and OFC of 33cm (10th centile). The neonatal period had no intercurrences. Head control occurred around 7 months, sitting without support and purposeful grasp at 12 months. Independent walking and language were never acquired. From an early age autistic traits (e.g. gaze avoidance) were noted. Partial complex seizures began when the child was around 3 years old; good control was attained with valproate and clonazepam. At 6 years, purposeful grasp was lost. Currently, at 15 years of age, the patient presents severe intellectual disability, swallowing difficulties, breathing disturbances and laughing spells. On physical exam the patient presents growth retardation and severe scoliosis; head circumference is on the 25th centile. On neurologic exam, hand washing and clapping stereotypies, as well as tremor of the upper limbs and dystonia of the lower limbs were noted. Brain MRI revealed frontal periventricular heterotopies and frontal and temporal atrophy. Because we did not consider these alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient, we still included her in the study. Metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing showed no mutations.

Patient 1 presents a 83,5 kb paternal microdeletion in chromosome 3 (chr3:113,826,308-113,909,889) encompassing the *DRD3* (DOPAMINE RECEPTOR D3) gene. *DRD3* encodes the D3 subtype of the five (D1-D5) dopamine receptors. The activity of the D3 subtype receptor is mediated by G proteins, which inhibit adenylyl cyclase. This receptor is localized to the limbic areas of the brain, which are associated with cognitive, emotional, and endocrine functions. Association studies have linked *DRD3* to risk for

schizophrenia [2] and essential tremor [3]. Deletions encompassing exons of *DRD3* are not observed in DGV, which suggests that preservation of this gene is very relevant for health. *DRD3* has also been pointed out as being one of the most relevant genes in the 3q13.31 deletion syndrome, in which patients invariably present ID and frequently also craniofacial dysmorphic features (high arched palate, short philtrum and protruding lips), skeletal malformations (scoliosis, lordosis, thoracic kyphosis, joint contractures), postnatal overgrowth, agenesis of the corpus callosum and hypoplastic male genitalia. Behavioral problems such as autism and attention deficit disorders have also been described in some affected individuals [4,5]. No variants were found in the the *DRD3* gene using WES. A list of the variants found in the patient is presented in table S1.

Patient 2 – This girl is the only child of a healthy non-consanguineous couple, whose family history is negative for developmental delay/intellectual disability. She was born at full term through vaginal delivery with a weight of 2780g (10^{th} centile) and OFC of 33cm (10^{th} centile). There were no intercurrences in the neonatal period. Head control occurred around 5 months, sitting with support at 6 months. Purposeful grasp was acquired at 3 months and lost around 6 months. Stagnation was noted at 8 months, coinciding with the beginning of complex partial seizures and lack of social interaction; virtually no acquisitions were achieved posteriorly, including gait or sphincter control. Currently, with 17 years of age, the patient presents severe intellectual disability, crying and laughing spells, respiratory dysfunction, frequent psychomotor agitation and sleep disturbances (waking up in the middle of the night). On neurologic exam, dystonia, bradykinesia, continuous stereotypies, tremor of the upper limbs and pyramidal signs of the lower limbs were noted, as well as swallowing difficulties. Brain MRI at the age of 4years and 10 months showed significant striatum atrophy (specially the caudate nuclei) as well as mild atrophy of the cortex and cerebellar vermis. Because we did not

considerthese alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient we still included her in the study. Metabolic screen, *MECP2* and *CDKL5* Sanger sequencing were normal. This patient has two compound heterozygous variants in *HTT* gene. HTT encodes a protein that plays a role in embryogenesis and that, when absent, results in the impairment of specification of ectodermal and mesodermal lineages.[6] Even though RTT is a neurodevelopmental disorder and Huntington Disease is a late onset neurodegenerative disease, they seem to share some features at the molecular and clinical level.[7,8] We believe that it is possible that the combination of two missense variants in *HTT* (though one is a rare polymorphism) in patient 2 may contribute for the RTT-like phenotype. Noteworthy, *HTT* homozygous loss of function variants have not been observed in any individual from the 1000 genomes project cohort suggesting that a *HTT* KO (and likely KD) has significant impact in health in humans, as seen in mice.[9,10] A list of the variants found in the patient is presented in table S2.

Patient 3 – This girl is the youngest daughter of a healthy non-consanguineous couple; her brother and sister are healthy. She was born at full term through vaginal delivery with a weight of 2245g (<5th centile) and OFC of 32cm (<5th centile). The neonatal period had no intercurrences but at 5 months West syndrome was diagnosed. Head control was achieved at 4 months, sitting without support at 11 months, first words at 12 months. Purposeful grasp was noted at 17 months but partially lost at 3 years. At 2 years she started walking (dyspraxic gait) and at 4 years she acquired sphincter control and the capability to say a few sentences (but not always purposeful; jargon and echolalia were frequent). Currently, with 13 years of age, the patient presents severe intellectual disability, poor eye contact, laughing spells, breathing disturbances (hyperpneia) and periods of

psychomotor agitation. In the neurologic exam, dystonia and hand washing stereotypies were observed until the age of 5 years. Metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing were normal. A list of the variants found in the patient is presented in table S3.

Patient 4 – This girl is the daughter of a healthy non-consanguineous couple. She has a healthy brother and the remaining family history is negative for developmental delay/intellectual disability. She was born at full term through vaginal delivery with a weight of 3600g (50-75th centile) and OFC of 36cm (75th centile). Neonatal period had no intercurrences. At 7 months she started having complex partial seizures. Developmental acquisitions occurred very slowly, with sitting without support occurring around 14 months; walking and language were never attained. Purposeful grasp was acquired when she was 5 years old but lost 2 years later. Currently, with 25 years of age, the patient presents severe intellectual disability, autism, laughing spells, eye pointing and sleep disturbances (waking up in the middle of the night). On physical exam kyphosis/scoliosis, peripheral vasomotor disturbances and small cold feet were noted. The neurologic exam revealed stereotypies, dystonia, pyramidal signs. Microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities. This patient carries a *de novo* variant in *SMARCA1* (alias *SNF2L*) gene. This gene encodes a chromatin remodeling ATPase involved in the control of Wnt signaling by regulation of proliferation and cell migration[11] and was also found to function antagonistically with Foxg1 in the regulation of brain size in mice.[12] A list of the variants found in the patient is presented in table S4.

Patient 5 – This girl is the daughter of a healthy non-consanguineous couple, whose remaining family history is negative for developmental delay/intellectual disability; her elder brother is healthy. She was born at full term through vaginal delivery with a weight

of 3260g (25-50th centile) and OFC of 33cm (10-25th centile). Motor delay was present from the beginning, with head control achieved at 4 months, and seems to have been superimposed by regression at 8 months. Sitting with support was achieved at 12 months and tiptoe gait at 26 months. The child never learnt any words. Currently, at 15 years of age, the patient presents severe intellectual disability, screaming spells, bruxism when awake and breathing disturbances (apnea followed by hyperpnoea). The neurologic exam revealed microcephaly, stereotypies, and dystonia. Brain MRI, metabolic screen, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S5.

Patient 6 – This girl is the only child of a healthy non-consanguineous couple; a maternal great aunt had ID and epilepsy reportedly after having had meningitis. She was born at full term through dystocic delivery (requiring forceps) with a weight of 3565g (50-75th centile), length of 49cm (25-50th centile) and OFC of 34,5cm (25-50th centile). Congenital hip dislocation was diagnosed and casts were used until 9 months of age. Head control was achieved around 3 months, sitting without support at 6 months, first words at 12 months and walking without support at 13 months. Regression occurred around 14-20 months of age, with loss of almost all previously acquired motor skills. Currently, at 13 years of age, the patient presents severe intellectual disability. On physical exam growth retardation was clear (weight and height < 5th centile), though the child was normocephalic. Atetosis, pyramidal signs and wallowing difficulties due to oropharyngeal dystonia were noted on the neurologic exam. Brain MRI revealed hyperintensities in the putamen and caudate nuclei. Because we did not considered these alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient we still included her in the

study. Spectroscopy was normal. Metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities.

A *de novo* variant c.A1801G (N601D) in the *LARP4* (LA RIBONUCLEOPROTEIN DOMAIN FAMILY, MEMBER 4) gene was found in this patient, resulting in the substitution of an asparagine for an aspartic acid and predicted to be pathogenic by the SIFT prediction tool. *LARP4* encodes a La-related protein which can bind poly(A) RNA and interact with the poly(A) binding protein (PABP), contributing in this way for mRNA stability and homeostasis [13–15]. So far, no variants were described in *LARP4*. However, due to its participation in such an important biological process, known to be perturbed in several ID syndromes, we cannot ignore its eventual contribution to ID pathogenesis. A list of the variants found in the patient is presented in table S6.

Patient 7 – This girl is the second child of a healthy non-consanguineous couple; her two sibs (one male and one female) are healthy. The couple has lost spontaneously two pregnancies (one in the 2nd and another in the 3rd trimester); gender and presence of other malformations in the fetuses is unknown. She was born full term through vaginal delivery with a weight of 3800g (75th centile) and OFC of 35,0cm (50th centile). Head control occurred around 6 months, sitting without support at 10 months, walking at 3,5 years, girst words with 5 years. Stereotypies were noted when she was 2,5 years and for many years she didn't achieve purposeful grasp. She started having purposeful hand use around 6 years of age, but lost it between 9 and 10 years. When she was 7, she had the first hyperpnea episodes and tonico-clonic generalized seizures as well as nocturnal crying and laughing spells. Currently, as a 31 year old adult, the patient presents severe intellectual disability, bruxism when awake, eye pointing, episodes of psychomotor agitation. On physical exam kyphosis/scoliosis and peripheral vasomotor disturbance

were registered. Brain MRI, metabolic screening, karyotype and *MECP2* Sanger sequencing showed no significant abnormalities.

Patient 8 – This girl is the first child of a healthy non-consanguineous couple. She has a healthy younger sister. A first grade maternal cousin died with spinal muscular atrophy type I. The remaining family history is negative for developmental delay/intellectual disability. She was born at full term through vaginal delivery with a weight of 2870g (10-25th centile) and OFC of 35cm (50th centile). Head control occurred around 6 months, sitting without support at 6 months. Deceleration of head growth was noted at 6 months. Walking without support and first words were achieved at 18 months and building sentences at 24 months. Regression occurred between 2 and 2,5 years with loss of language and purposeful grasp and coincided with the appearance of severe autism (screened with CARS and confirmed with ADI-R). As time went by, a cervical kyphosis became evident. Currently, at the age of 13 , the patient presents severe intellectual disability, epilepsy (well controlled with anti-epileptic drugs), psychomotor agitation, screaming spells and poor eye contact. On neurological exam the patient presents stereotypies, ataxia and dysmetria and peripheral vasomotor disturbances. Brain MRI, karyotype, metabolic screen, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S11.

Patient 9 – This girl is the only child of a healthy non-consanguineous couple; the remaining family history is negative for developmental delay/intellectual disability. She was born at full term through vaginal delivery with a weight of 3140g (25th centile) and OFC of 34,0cm (25-50th centile). Stagnation of psychomotor development was noted early on (7 months) followed by regression, as well as lack if interest in exploring/interacting with the environment. Also at 7 months of life head growth

acceleration was documented with posterior stabilization at 2 years of age around the 98th centile. Head control was attained at around 5 months of age, sitting without support at 18 months and walking at 24 months. Purposeful grasp, language and sphincter control were never acquired. Currently, at 19 years of age, the patient presents severe intellectual disability, episodes of psychomotor agitation with crying spells, bruxism when awake, sleep disturbances (waking up in the middle of the night) and breathing disturbances (hyperventilation). On neurological exam stereotypies, dystonia (lower limbs) and swallowing difficulties were noted. Brain MRI revealed frontoparietal enlargement of CSF spaces. Metabolic screen and *MECP2* Sanger sequencing revealed no significant abnormalities. Patient 9 has a *de novo* variant in *GABBR2* gene. *GABBR2* encodes a Gamma-aminobutyric acid (GABA) type B receptor involved in neuronal activity inhibition and that is co-expressed with FOXG1.[16,17] Because GABA receptors play an important role in maintaining the excitatory-inhibitory balance in brain, the possibility that their deregulation may be associated with epilepsy is not to be dismissed. *GABBR2* expression was found to be reduced in lateral cerebella from subjects with schizophrenia, bipolar disorder and major depression [18] as well as in the cerebella of autistic patients.[19] Common insertion/deletion polymorphisms affecting *GABBR2* were described in patients with autism although without statistically significant enrichment when compared to control populations.[20] Additionally, *de novo* missense variants in *GABRR2* were identified in two different patients with infantile spasms.[21] This information brings new insight into the association of *GABRR2* with neurodevelopmental disorders and highlights the importance of clinical and WES data sharing. A list of the variants found in the patient is presented in table S9.

Patient 10 – This female patient is the second child of a healthy non-consanguineous couple; her male sibling had absence seizures that later evolved to tonic-clonic seizures;

he is otherwise healthy. She was born full term through vaginal delivery with a weight of 2730g (10^{th} centile) and OFC of 30,0cm ($<2^{th}$ centile). In the neonatal period the following dysmorphisms were described: beaked nose, widened space between the nipples, cauliflower anus, low implantation of the 5^{th} toes. Head control occurred around 12 months, sitting without support at 36 months. Walking, talking and fine motor coordination were never attained. Epilepsy (partial seizures) started at 24 months, and so did midline hand stereotypies. When the child was 6 years old, she started having episodes of hyperpena followed by apnea. Currently, with 24 years of age, the patient presents severe intellectual disability, has screaming spells and sleep disturbances (waking up in the middle of the night). On physical exam, growth retardation and microcephaly were apparent, as well as kyphosis/scoliosis and peripheral vasomotor disturbances. On neurological exam dystonia and swallowing difficulties were noted. Brain MRI, metabolic screening, karyotype and *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities.She was classified as a RTT patient (see table 1 in main text) in light of her loss of acquired hand skills, spoken language, gait abnormalities and stereotypical hand movements. A paternally inherited variant c.G2971A (p.V991I) in the *MAGEL2* (MAGE-LIKE 2) gene (OMIM 605283) was found, resulting in the substitution of a valine for an isoleucine, predicted not to have a functional impact by the SIFT, PolyPhen2 and Mutation Assessor prediction tools. However, given the imperfection of these tools it is not possible to exclude this variant as a basis of disease. *MAGEL2* is located at the critical region for Prader-Willi syndrome and paternally expressed in the CNS [22]. Recently, *de novo* truncating variant in the paternal allele of *MAGEL2* were described in patients with Prader-Willi Syndrome features, resulting in the loss of expression of functional protein [23]. The contribution of this alteration to the disease in this patient is not clear since it was not possible to determine the parental origin of the

variant in the father (and/or if it occurred *de novo* in the father, being then inherited). A list of the variants found in the patient is presented in table S8.

Patient 11 – This girl is the second child of a non consanguineous couple. The mother has resting tremor and is suspected of having psychiatric disease, possibly early-onset dementia. The maternal grandmother is bed-ridden and dementiated. The father is suspected of also having psychiatric disease. Formal neurological and psychiatric evaluation of both parents is pending. The older male sibling is healthy and so are the remainder more distant relatives. She was born full term through dystocic delivery (forceps) with a weight of 2880g (10-25th centile), length 46,3cm (5-10th centile) and OFC of 34,3cm (25-50th centile). She gained head control at 3 months and sitting with support at 6 months. Between 6 and 9 months, her development seemed to stagnate and her head growth decelerated. At 9 months generalized epilepsy began, good response to anti-epileptic drugs. Currently, with 6 years of age, the patient presents severe intellectual disability, growth retardation, microcephaly, restless nights with diurnal sleep, bruxism when awake, diminished response to pain and eye pointing. On neurological exam the patient presents stereotypies, resting tremor and peripheral vasomotor disturbances. Brain MRI, metabolic screening, karyotype and *MECP2, CDKL5, UBE3A* and *PCDH19* Sanger sequencing revealed no significant abnormalities.

Patient 11 has a combination of two variants: a *de novo* variant in *RHOBTB2* gene and a homozygous variant in *EIF4G1* gene. *RHOBTB2* belongs to the Rho GTPases family and was found to bind to CUL3 and to be a substrate of the Cul3-based ubiquitin ligase complex, which is necessary for mitotic cell division.[24] Moreover, *de novo* nonsense variants in *CUL3* were identified in two separate next-generation sequencing reports using ASD probands,[25,26] further strengthening the relationship between *RHOBTB2* and neurodevelopmental disorders. RhoBTB2 is also likely to interfere with several

neuron related functions, given its participation in cytoskeleton and membrane trafficking networks.[27] *EIF4G1* encodes a translation initiation factor involved in the recruitment of mRNA to the 40S ribosomal subunit, which is a rate-limiting step during the initiation phase of protein synthesis.[28] Variants in *EIF4G1* have been associated with autosomal dominant forms of Lewy body dementia[29] and Parkinson disease (with and without dementia). However, the true pathogenicity of some of these variants remains unknown as it has been difficult to replicate the findings in populations with different genetic backgrounds.[30–33] In the Human Genome Mutation Database, the mutations that have been clearly associated with Parkinson disease are in heterozigosity, whereas our patient has a variant in homozigosity, which could justify an early and more severe presentation of neurological disease. A list of the variants found in the patient is presented in table S13.

Patient 12 – This girl is the first child of a healthy non-consanguineous couple. She has a healthy younger sister; the remaining family history is also negative for developmental delay/intellectual disability. During pregnancy, intra-uterine growth retardation and single umbilical artery were detected. She was born at full term through vaginal delivery with a weight of 2150g (<5th centile), length of 42cm (<5th centile) and OFC of 30,5cm (2nd centile). During the neonatal period, left ptosis and cleft palate were diagnosed. Head control was achieved at around 6 months, sitting without support at 2 years, dyspraxic gait with support was eventually attained, but never independent walking. Even though she has hand stereotypies, hand use is preserved. At 10 years of age, on physical exam the child presented normal growth but severe microcephaly (weight: 35Kg, height:145cm, OFC:43cm), scoliosis/kyphosis and peripheral vasomotor disturbances. Her significant swallowing difficulties and gastroesophageal reflux prompted the decision of performing a gastrostomy. Currently, at 19 years of age, the patient presents severe intellectual

disability, eye pointing and breathing disturbances (hyperapnea). On neurological exam stereotypies and ataxia were noted. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k) and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S10.

Patient 13 – This girl is the daughter of a healthy but possibly remotely consanguineous couple. She was born at full term through vaginal delivery with a weight of 3610g (50-75th centile) and OFC of 35cm (50th centile).The beginning of epilepsy (well controlled with therapy) at 24 months coincided with regression (namely loss of language) at 24 months. Currently, at the age of 20 years , the patient presents severe intellectual disability, crying and screaming spells, and restless nights with diurnal sleep. The patient is also affected with scoliosis/kyphosis and gastro-esophageal reflux disease. On neurological exam she presents pyramidal signs. Brain MRI, metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities.

Patient 13 presents two compound heterozygous variant in the *TCTN2* (TECTONIC FAMILY, MEMBER 2) (OMIM 613846) gene: a maternal c.C668T (p.T223M), resulting in a substitution of a threonine for a methionine, a very rare variant in the population (rs145374149, MAF T=0.0009/2) and predicted not to be pathogenic by in silico softwares; a paternal c.T1538C (p.I513T), resulting in a substitution of a isoleucine for a threonine predicted not to be pathogenic by in silico softwares. The *TCTN2* gene is a paralogous of *TCTN1* (TECTONIC FAMILY, MEMBER 1) (OMIM 609863) that plays a role in sonic hedgehog signaling pathway modulation [34]. Homozygous and compound heterozygous variants in *TCTN2* gene were described in patients with Meckel-Gruber 8 syndrome and Joubert syndrome [34,35]. A list of the variants found in the patient is presented in table S12.

Patient 14 – This boy is the only child of a healthy non consanguineous couple. The older male sibling is healthy and so are the remainder more distant relatives. He was born at full term through C-section with a weight of 3444g (25-50th centile), length 51,5cm (50-75th centile) and OFC of 35,5cm (25-50th centile). At 1 month West syndrome was diagnosed; control of seizures has been difficult. Development was significantly delayed, with purposeful grasp and sitting being attained at 15 months. At 6 years old he started walking without support but few other psychomotor milestones were conquered. Currently, with 6 years of age, the patient presents severe intellectual disability, psychomotor agitation, restless nights with diurnal sleep and bruxism when awake. On physical exam it was confirmed that growth is within a normal interval. On neurological exam the patient presents poor eye contact, stereotypies, mioclonies, hyperventilation and decreased response to pain. Brain MRI showed moderate enlargement of pericerebral spaces – likely generalized brain atrophy. Because we did not consider these alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient we still included him in the study. Metabolic screening, karyotype and *MECP2, CDKL5, ARX* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S14.

Patient 15 – This boy is the second child of a healthy non consanguineous couple. The older female sibling is healthy and so are the remainder more distant relatives. He was born at full term through vaginal delivery with a weight of 3250g (25-50th centile) and OFC of 33cm (10th centile). Food refusal and recurrent vomiting lead to growth decelaration: around 4 months of life the patient was crossing centiles both in weight (from 50th centile to 10th centile), height (from 25th centile to 10th centile) and OFC (from 10th centile to <3rd centile). Around 10 months the patient had his first apyretic seizure

(difficulty breathing, involuntary movements, gaze deviation). Since then, few psychomotor acquisitions were achieved and autism spectrum features were detected. Currently, with 8 years of age, the patient presents severe intellectual disability, psychomotor agitation, crying spells and disrupted sleep (wakes up in the middle of the night). On physical exam it was verified that the patient had caught up in terms of weight and height but microcephaly was still present. Some minor dysmorphism were observed: sunken eyes, prominent forehead, shortening of the 4th and th fingers, brittle nails. On neurological exam the patient presents stereotypies of the head and hands. Brain MRI (performed at 19 months and repeated at 4 years of age) showed global brain atrophy, thin corpus callosum and bilateral fronto-parietal periventricular heterotopies. Because we did not consider these alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient we still included him in the study. EEG at 7 years showed a disorganized pattern during sleep, with predominantly occipital continuos lentification and paroxystic activity. Metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S15.

Patient 16 – This girl is the conception of a healthy and possibly remotely consanguineous couple. A paternal second grade male cousin is affected with idiopathic intellectual disability. She was born at full term through vaginal delivery with a weight of 3800g (75th centile) and OFC of 34,5cm (25-50th centile). At 6 months, coinciding with inaugural West syndrome there was developmental arrest, lack of interest in interacting/exploring the environment and head growth deceleration (currently being microcephalic). Currently, with 9 years of age, the patient presents severe intellectual disability, psychomotor agitation episodes, screaming and crying spells. On neurological exam the patient presents poor eye contact and lack of interest in social interaction,

repetitive behavior, stereotypies (hand wringing) and ataxia. Brain MRI, metabolic screening, karyotype and *MECP2* and *CDKL5* Sanger sequencing had no significant abnormalities. A list of the variants found in the patient is presented in table S16.

Patient 17 – This girl is the second child of a healthy non-consanguineous couple; her older brother is healthy and so are more distant relatives. She was born at full term through dystocic delivery (vacuum extraction) with a weight of 3700g (50-75th centile). Developmental delay was noticeable all along, with further stagnation at 10 months with the beginning of epilepsy. She sat at 16 months, walked with support at 30 months and without support at 48 months. Her first words were learnt at 36 months but she never constructed sentences. Currently, with 13 years of age, the patient presents severe intellectual disability, autistic behavior, crying spells and bruxism when awake. Disrupted sleep became so problematic that hipnotics were prescribed. On neurological exam the patient presents poor eye contact, stereotypies, respiratory dysfunction (hyperpnoea alternating with apneas) and diminished response to pain. Brain MRI showed global enlargement of the CSF space suggestive of generalized brain atrophy. Because we did not considered these alterations to be a consequence of a peri-postnatal insult, neurometabolic disease or severe infection nor to explain the clinical presentation of the patient we still included her in the study. Metabolic screening, karyotype, FISH for 17p11.2 (Smith-Magenis syndrome), DNA methylation analysis of 15q11.2-q13 (AS/PWS critical region), *MECP2* and *UBE3A* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S17.

Patient 18 – This girl is the first child born to a non-consanguineous couple. Her mother has mild intellectual disability; the remaining family history is negative for developmental delay/intellectual disability. She was born at full term from dystocic delivery (vacuum extraction) with a weight of 2770g (5-10th centile), length 49,0cm (25-

50th centile) and OFC of 33,3cm (10-25th centile). She sat without support at 11 months, said her first words at 12 months (but did not develop language), achieved purposeful grasp at 18 months and walked without support at 30 months. Also around 30 months, ASD was suspected on the basis of hand flapping, temper tantrums, autoagression. Currently, at 9 years old, the patient presents severe intellectual disability, bruxism when awake, crying spells and disrupted sleep (waking up in the middle of the night). On physical exam her growth is normal. On neurological exam the patient presents stereotypies (hand flapping), peripheral vasomotor disturbances, ataxia, diminished response to pain. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* Sanger sequencing revealed no significant abnormalities.

In this patient a maternally inherited variant c.G3226T (A1076S) in *CHD8* (CHROMODOMAIN HELICASE DNA-BINDING PROTEIN 8) (OMIM 610528) was found, resulting in the substitution of an alanine for a serine and predicted by SIFT and Mutation Taster to be damaging. *CHD8* encodes a chromodomain helicase DNA-binding protein that participates in the Wnt-β-catenin signaling pathway trough histone H1 recruitment during development [36]. In rat, duplin (corresponding to *CHD8*), binds to beta-catenin inhibiting the binding of Tcf to beta-catenin and consequently inhibiting the Wnt signaling pathway [37]. Noteworthy, *CHD8* seems to be co-expressed with *MECP2*, the main cause of RTT [38,39]. Rare *de novo* variants in *CHD8* have been described as associated with autism spectrum disorder in recent years [26,40–43]. A list of the variants found in the patient is presented in table S18.

Patient 19 – This boy is the child of a healthy non-consanguineous couple; his younger sister is healthy and so are the remainder distant relatives. He was born at full term through dystocic delivery (vacuum extraction) with a weight of 3300g (25-50th centile), length 49,0cm (25-50th centile) and OFC of 34,0cm (25-50th centile). He sat without

support at 6-9 months, acquired purposeful grasp at 2 years and controlled sphincters at 10 years. Currently, with 14 years of age, the patient presents severe intellectual disability, dyspraxic gate, eye pointing, bruxism when awake and scoliosis/kyphosis. He is also affected with gastroesophageal reflux. On neurologic exam he presents microcephaly, peripheral vasomotor disturbances, diminished response to pain and autistic traits (including lack of interest in social interaction). The patient does not present short stature, hip dislocation or epilepsy. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k) and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S19.

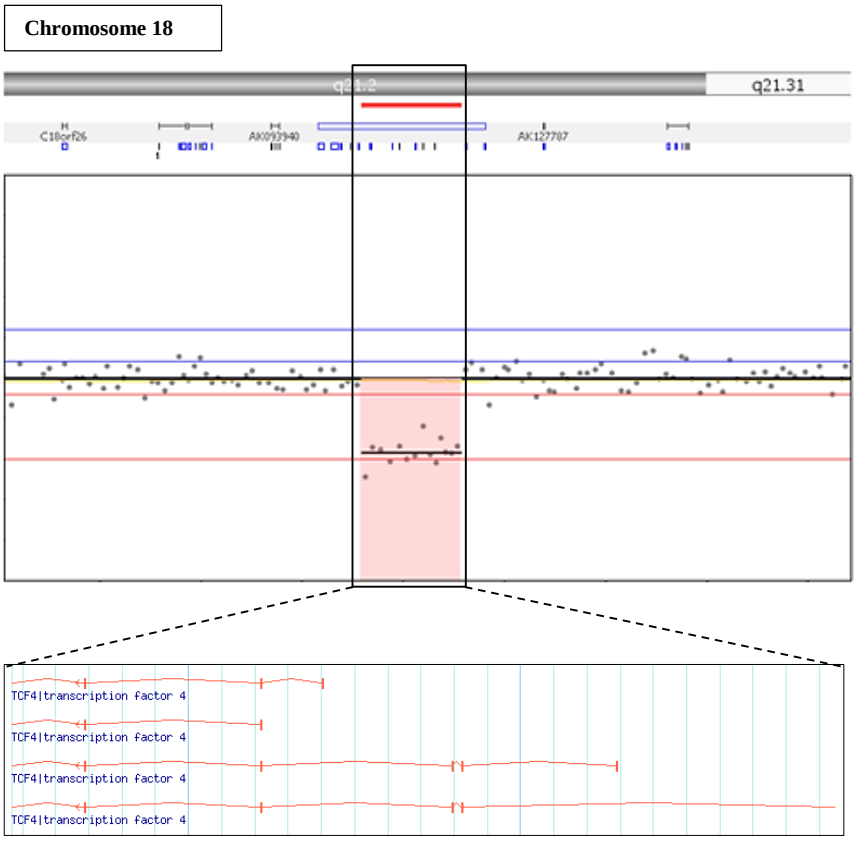


Figure S1.1 – Schematic representation of the 18q21.1 deletion in patient 7. It is possible to observe the decrease in the LRR of the probes contained in that region when compared to neighboring probes (LRR=0). The deletion is 250Kb long and encompasses two genes: *TCF4* and *MIR4529*.

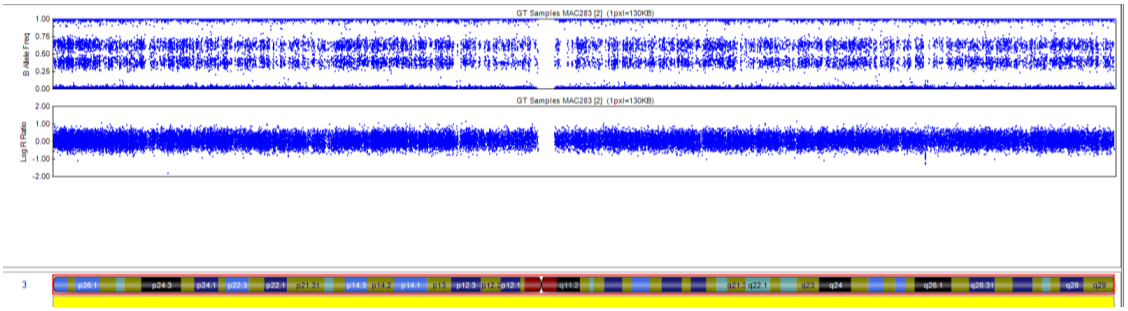


Figure S1.2 – Schematic representation of the UPD mosaic of chromosome 3 in patient 16. It is possible to observe that the entire chromosome 3 has a LRR=0 and a BAF split (0,3 and 0,6), compatible with mosaic UPD.

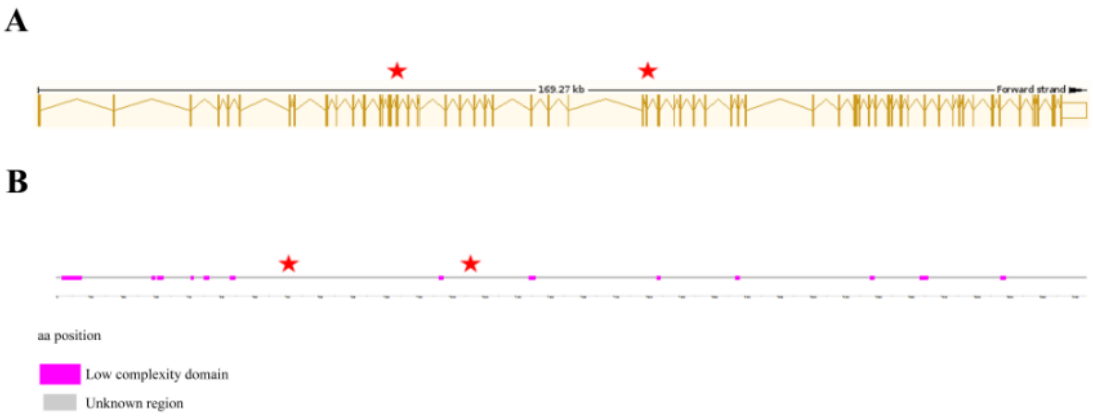


Figure S1.3 – Schematic representation of the *HTT* gene and HTT protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene are indicated by a red star with its respective location at exons and protein levels (aa). Both variants affect and unknown region/domain.

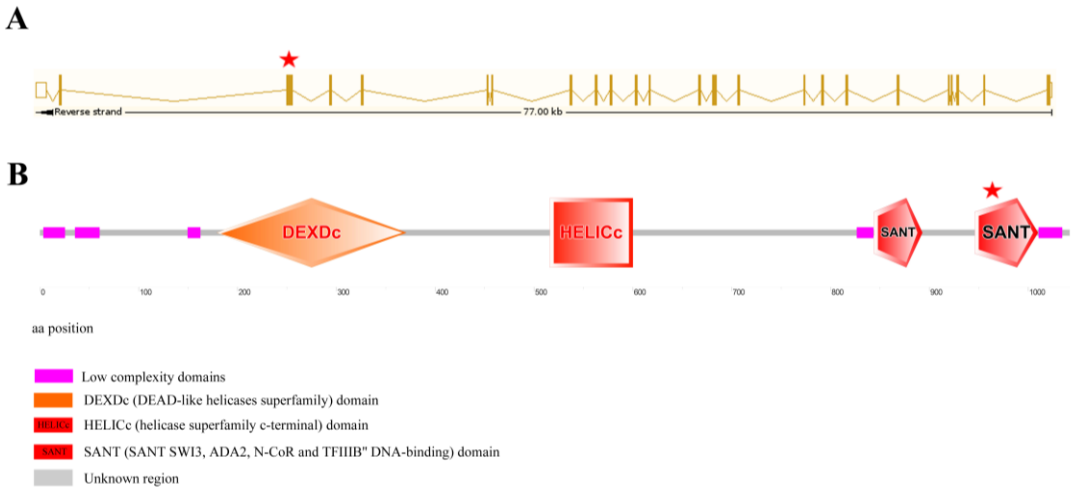


Figure S1.4 – Schematic representation of the *SMARCA1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located id the SANT domain, which is present in the subunits of many chromatin-remodelling complexes [44].

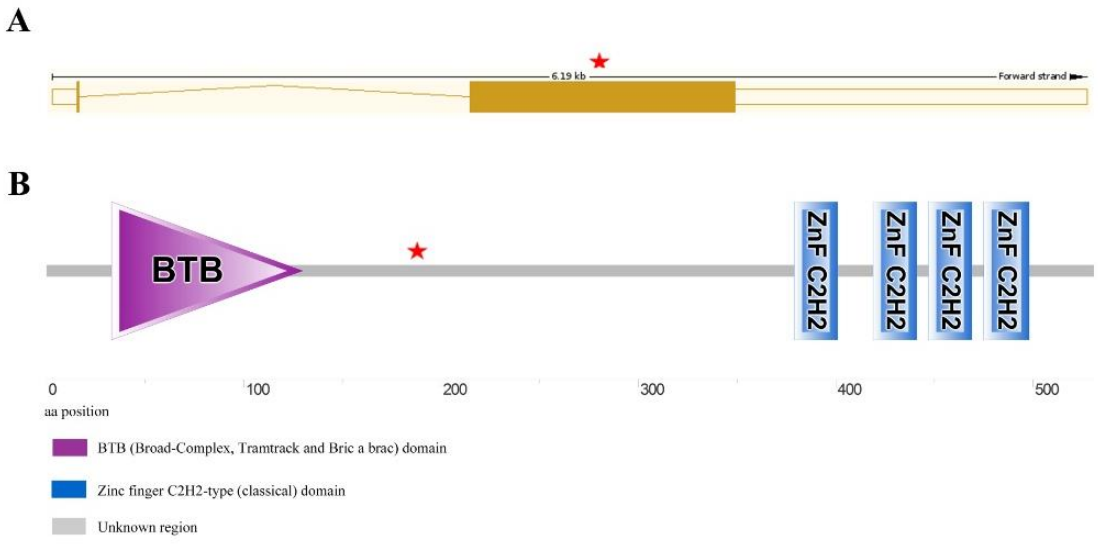


Figure S1.5 – Schematic representation of the *ZNF238* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

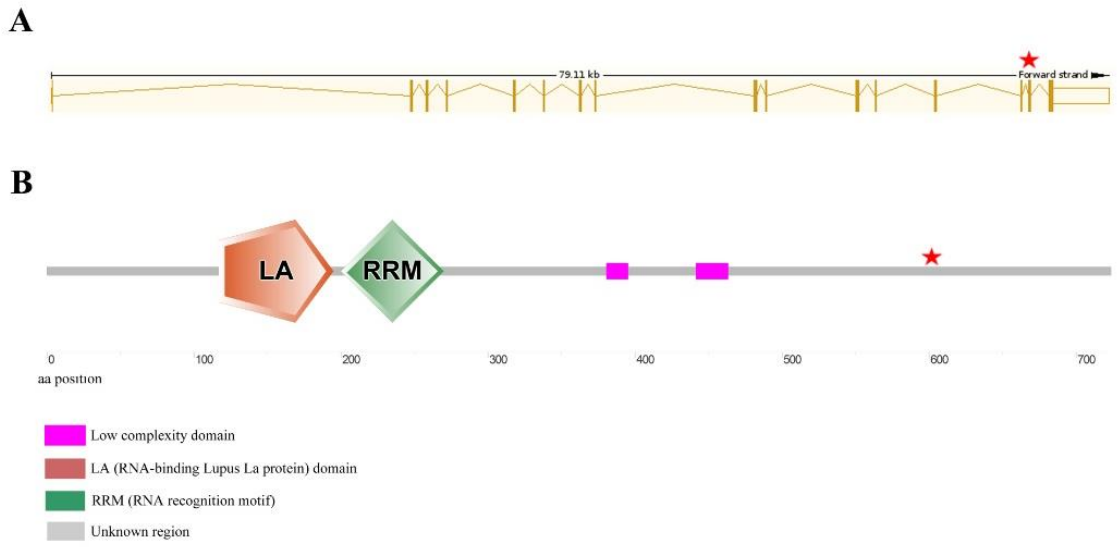


Figure S1.6 – Schematic representation of the *LARP4* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

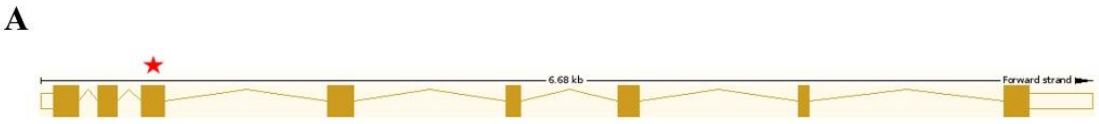


Figure S1.7 – Schematic representation of the *EIF2B2* gene structure. The gene structure was retrieved from Ensembl. The variant identified in this gene is indicated by a red star with its respective location at exon level.

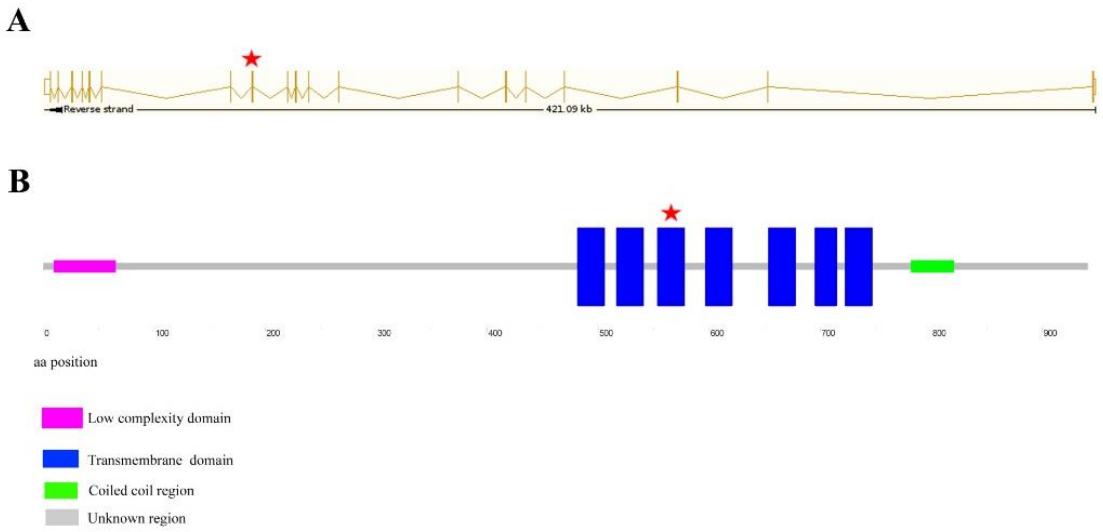


Figure S1.8 – Schematic representation of the *GABBR2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in one of the transmembrane helix regions fo the protein.

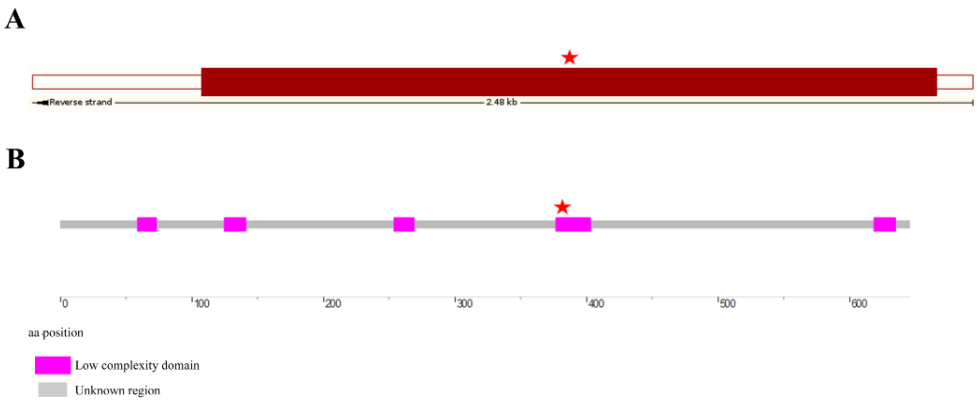


Figure S1.9 – Schematic representation of the *MAGEL2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in one of the low compositional complexity regions fo the protein.

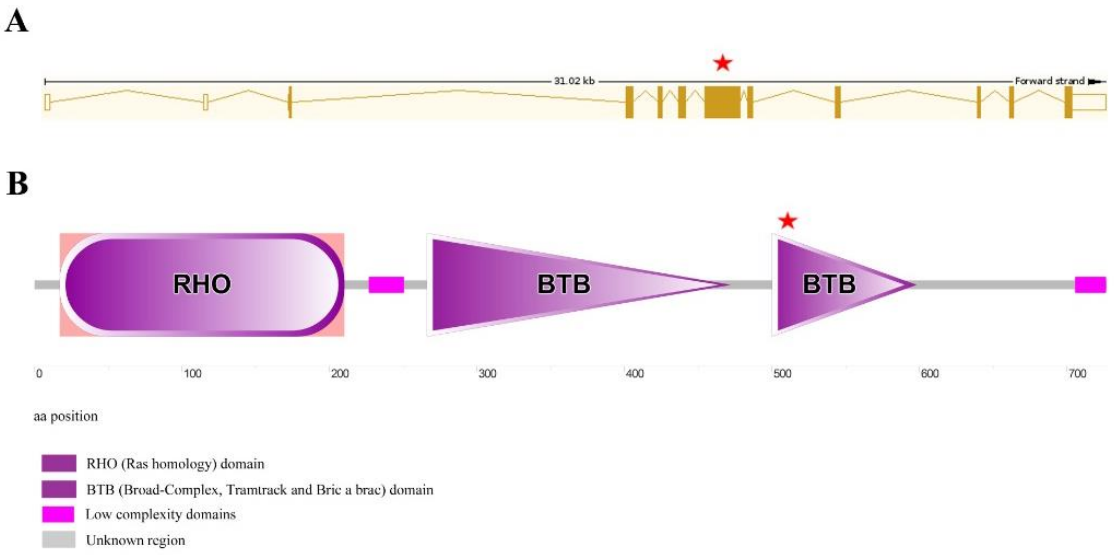


Figure S1.10 – Schematic representation of the *RHOBTB2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in the BTB domain of the protein, which is known to mediate homomeric dimerisation and, in some circunstances, heteromeric dimerization [45].

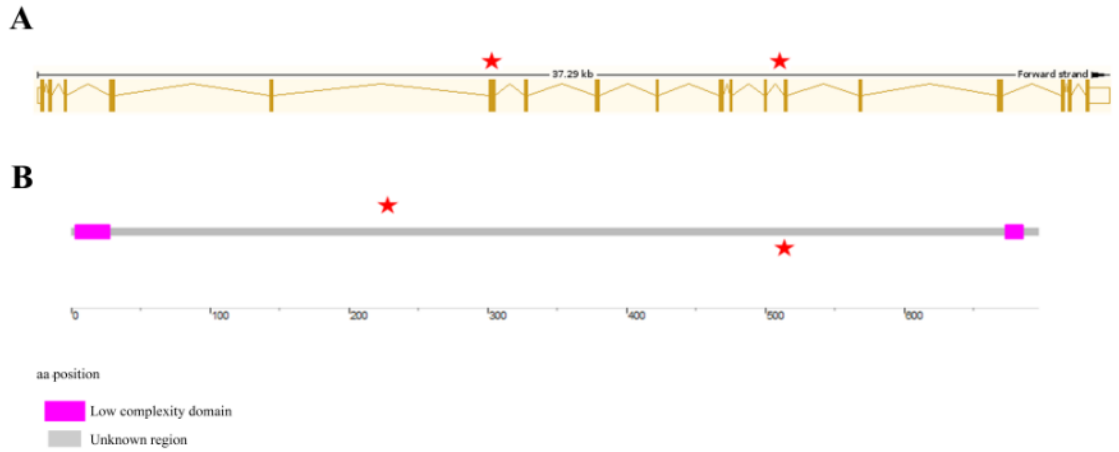


Figure S1.11 – Schematic representation of the *EIF4G1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

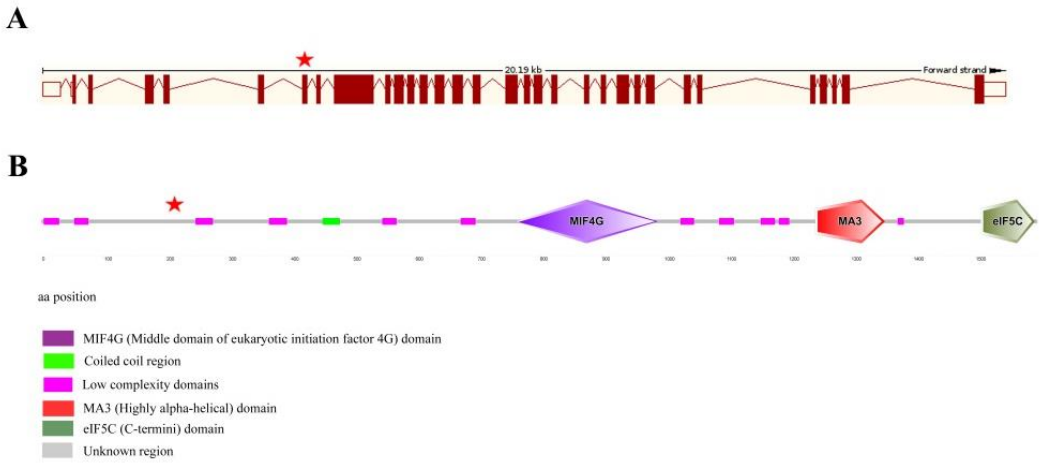


Figure S1.12 – Schematic representation of the *TCTN2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

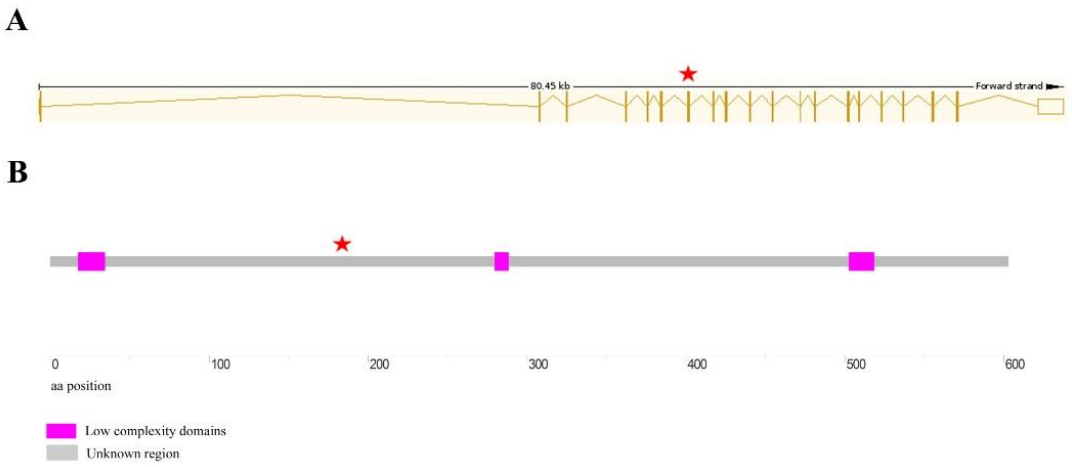


Figure S1.13 – Schematic representation of the *STXBP1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

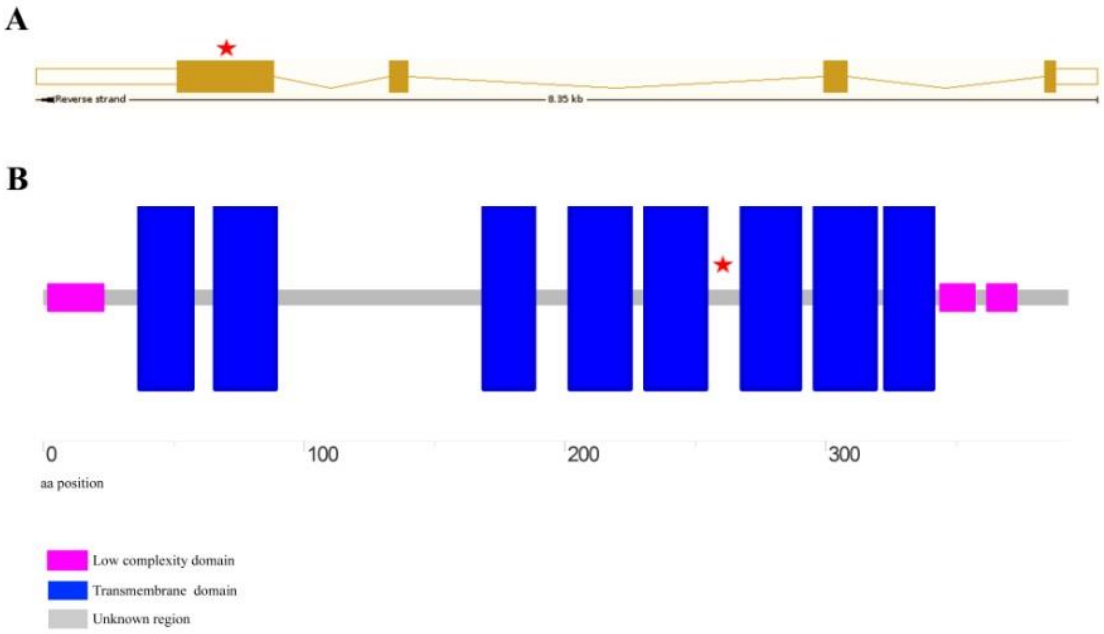


Figure S1.14 – Schematic representation of the *SLC35A2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

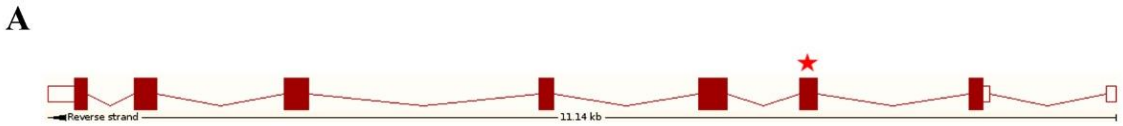


Figure S1.15 – Schematic representation of the *EEF1A2* gene structure. The gene structure was retrieved from Ensembl. The variant identified in this gene is indicated by a red star with its respective location at exon level.

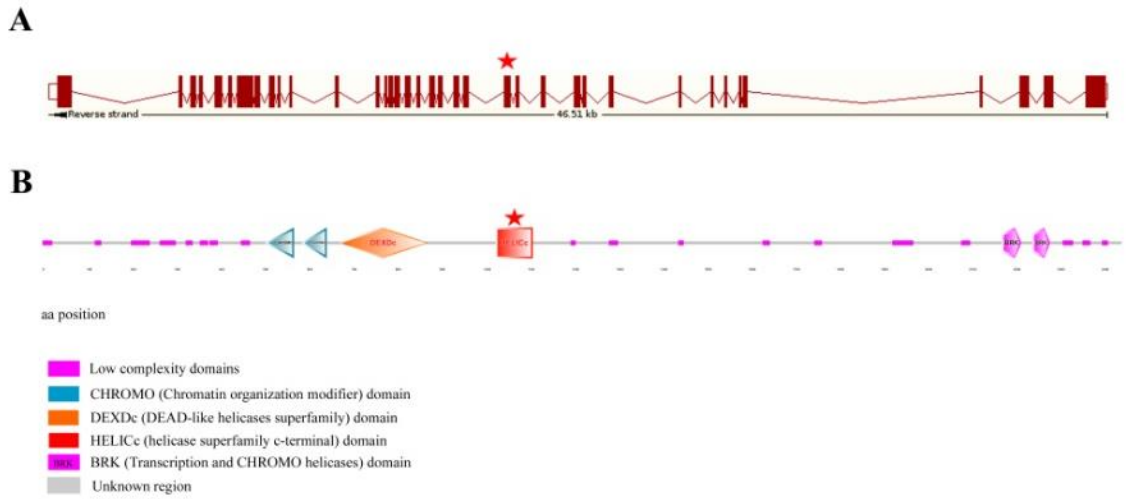


Figure S1.16 – Schematic representation of the C*HD8* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects the HELICc domain, which corresponds to the C-terminal domain found in proteins belonging to the helicase superfamilies 1 and 2 [46].

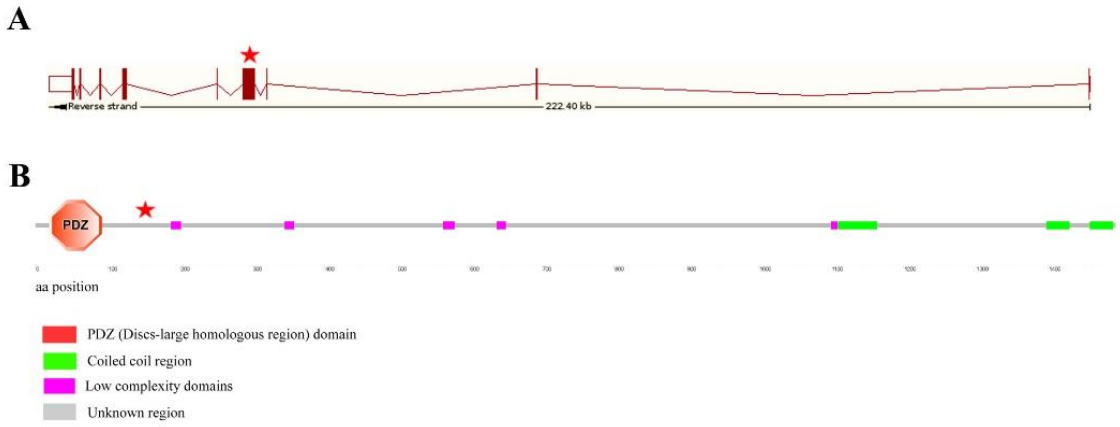


Figure S1.17 – Schematic representation of the *SHROOM4* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

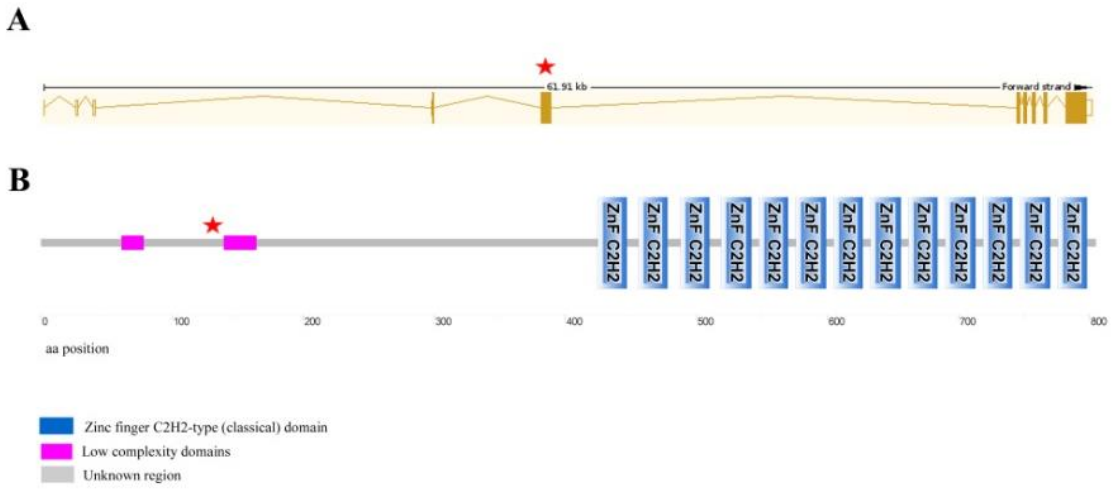


Figure S1.18 – Schematic representation of the *ZFX* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

Supplementary tables

Table S1.1 – Summary of the aCGH findings.

ID	Coordinates (hg19)	Size (bp)	#Probes	Algorithm*	%Rare	Type	Median LRR	Gene symbols	Comments
1	chr2:194920864-194985510	64,646	12	GN,PN,QT	100%	1	-0.46	-	CNV does not encompass any genes
	chr2:42494273-42537995	43,722	16	PN,QT	100%	3	0.42	*EML4*	Plotted on GenomeStudio:waves
	chr3:113826308-113909889	83,581	27	GN,PN,QT	100%	1	-0.37	*DRD3*	inherited from healthy father
2									Not performed
3	-	-	-	-	-	-	-	-	No relevant CNVs to report
4	chr11:38998451-39234089	235,638	23	GN,PN,QT	100%	1	-0.40	-	CNV does not encompass any genes
5									Not performed
6	chr8:22188330-22288995	100,665	31	PN,QT	100%	3	0.2	*PIWIL2; SLC39A14*	*PIWIL2* belongs to the Argonaute family (role in development and maintenance of germline stem cells); *SLC39A14* shows structural characteristics of zinc transporter. DECIPHER patient269297 has a 150kb duplication (*SLC39A14; PPP3CC*).
7	chr18:52996207-53243605	247,399	14	FASST2	100%	1	-0.8	*MIR4529, TCF4*	Loss of functions mutations and exonic or whole-gene deletions of *TCF4* cause Pitt-Hopkins syndrome
8	chr11:37476792-37532517	55,725	18	GN,PN,QT	100%	1	-0.57	-	CNV does not encompass any genes
	chr18:65848177-65897747	49,570	24	GN,PN,QT	100%	1	-0.45	-	CNV does not encompass any genes
9	-	-	-	-	-	-	-	-	No relevant CNVs to report
10	chr6:65669109-65781547	112,438	26	GN,PN,QT	100%	1	-0.41	*EYS*	*EYS* is expressed in the photoreceptor layer of the retina. Mutated in autosomal recessive retinitis pigmentosa. Likely the patietn is a carrier for retinitis pigmentosa.
	chr14:46325257-46435899	110,642	24	GN,PN,QT	100%	1	-0.44	-	CNV does not encompass any genes
11	chr21:44823479-44837555	14,076	12	PN,QT	100%	3	0.22	*SIK1*	*SIK1*: salt-inducible kinase 1. Function not well understood
12	chr4:61339224-61395456	56,232	15	PN,QT	100%	3	0.40	-	CNV does not encompass any genes
	chr13:92874384-92968290	93,906	38	GN,PN,QT	100%	1	-0.49	-	CNV does not encompass any genes
13	chr4:13102351-13221390	119,039	32	GN,PN,QT	100%	1	-0.39	-	CNV does not encompass any genes
14	-	-	-	-	-	-	-	-	No relevant CNVs to report
15	-	-	-	-	-	-	-	-	No relevant CNVs to report
16	chr3:234726-192095111	191,860,385	3032	GN,PN,QT	100%	2,3	0.11	≈1900 genes	Mosaic UPD chr3
17	chr9:5377177-5423103	45,926	22	GN,PN,QT	100%	1	-0.47	-	CNV does not encompass any genes
18	chr9:5373441-5418334	44,893	22	GN,PN,QT	100%	1	-0.51	-	CNV does not encompass any genes
18	chr16:78045800-78092197	46,397	17	GN,PN,QT	100%	1	-0.45	*CLEC3A*	*CLEC3A*: C-type lectin domain family 3, member A. Function not well understood
19	-	-	-	-	-	-	-	-	No relevant CNVs to report

*Algorithms: GN – GNOSIS; PN – PennCNV; QT – QuantiSNP; FASST2 - FASST2 Segmentation

Table S1.2 – Variants found by WES in patient 1.

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	FNDC7			chr1|109265010|G|A	nonsynonymous	chr1	82	P	NP	NP	NP	NP	P	No info	Not performed
Mother	FNDC7			chr1|109265010|G|A	nonsynonymous	chr1	53	P	NP	NP	NP	NP	P	No info	Not performed
Patient	FNDC7			chr1|109273387|T|A	nonsynonymous	chr1	195	P	NP	NP	NP	NP	NP	No info	Not performed
Father	FNDC7			chr1|109273387|T|A	nonsynonymous	chr1	141	P	NP	NP	NP	NP	NP	No info	Not performed

indels_de novo

sample_id	**gene**	**snp137**	**freq**	**indel_id**	**class**	**chr**	**start**	**end**	**Nervous system function**	**type**	**coverage**	**Sanger confirmation**
Patient	NXF1			chr11|62569105|62569105|-|AGCTACAG	splicing	chr11	62569105	62569105	Yes	ins	88	Confirmed. Not *de novo* (maternal)

Table S1.3 - Variants found by WES in patient 2.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RIN2	rs183028833	T=0.0009/2	chr20|19867384|C|T	nonsynonymous	chr20	16	P	NP	NP	NP	NP	NP	Yes/Likely	Confirmed. Not *de novo* (maternal)
Patient	CCRL2			chr3|46449931|A|G	nonsynonymous	chr3	75	P	P	P	P	NP	NP	Unlikely	Not performed

recessive_compound

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	HTT			chr4|3133374|C|T	nonsynonymous	chr4	20	NP	P	P	P	NP	P	Yes	Confirmed. Present
Father	HTT			chr4|3133374|C|T	nonsynonymous	chr4	21	NP	P	P	P	NP	P	Yes	Confirmed. Present
Patient	HTT	rs34315806	T=0.0234/51	chr4|3162034|C|T	nonsynonymous	chr4	114	NP	P	Np	NP	NP	P	Yes	Confirmed. Present
Mother	HTT	rs34315806	T=0.0234/51	chr4|3162034|C|T	nonsynonymous	chr4	94	NP	P	Np	NP	NP	P	Yes	Confirmed. Present
Patient	TG	rs114944116	T=0.0023/5	chr8|133919047|G|T	nonsynonymous	chr8	6	P	P	P	P	P	NP	Unlikely	Not performed
Father	TG	rs114944116	T=0.0023/5	chr8|133919047|G|T	nonsynonymous	chr8	28	P	P	P	P	P	NP	Unlikely	Not performed
Patient	TG	rs150728539	NA	chr8|133920582|C|G	nonsynonymous	chr8	21	P	NP	P	P	NP	NP	Unlikely	Not performed
Mother	TG	rs150728539	NA	chr8|133920582|C|G	nonsynonymous	chr8	9	P	NP	P	P	NP	NP	Unlikely	Not performed

Table S1.4* - Variants found by WES in patient 3.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	CCDC73			chr11|32636452|A|G	nonsynonymous	chr11	23	P	NP	NP	NP	NP	NP	No info	Confirmed. Not *de novo* (maternal)
Patient	PPP1R32			chr11|61254529|A|G	nonsynonymous	chr11	21	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (maternal)
Patient	SYNE1	rs150179494	A=0.0014/3	chr6|152665271|G|A	nonsynonymous	chr6	27	P	P	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (paternal)
Patient	HSDL1			chr16|84163608|C|T	nonsynonymous	chr16	745	NP	NP	NP	NP	NP	NP	Unlikely	Not performed
Patient	ATG10			chr5|81354416|A|G	nonsynonymous	chr5	23	NP	NP	NP	NP	NP	NP	Yes	Not performed

recessive_compound

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	HMCN1	rs140061598	C=0.0014/2	chr1|185956606|T|C	nonsynonymous	chr1	53								Not performed
Mother	HMCN1	rs140061598	C=0.0014/2	chr1|185956606|T|C	nonsynonymous	chr1	52								Not performed
Patient	HMCN1	rs76748242	A=0.0087/19	chr1|186031730|C|A	nonsynonymous	chr1	60								Not performed
Father	HMCN1	rs76748242	A=0.0087/19	chr1|186031730|C|A	nonsynonymous	chr1	60								Not performed
Patient	CKAP5	rs145146116	C=0.0005/1	chr11|46817323|T|C	nonsynonymous	chr11	31								Not performed
Father	CKAP5	rs145146116	C=0.0005/1	chr11|46817323|T|C	nonsynonymous	chr11	29								Not performed
Patient	CKAP5	rs138443179	T=0.0064/14	chr11|46837930|C|T	nonsynonymous	chr11	66								Not performed
Mother	CKAP5	rs138443179	T=0.0064/14	chr11|46837930|C|T	nonsynonymous	chr11	74								Not performed
Patient	AHNAK2	rs116553680	T=0.0087/19	chr14|105406201|C|T	nonsynonymous	chr14	75								Not performed
Mother	AHNAK2	rs116553680	T=0.0087/19	chr14|105406201|C|T	nonsynonymous	chr14	96								Not performed
Patient	AHNAK2	rs144504264	C=0.0114/25	chr14|105419755|G|C	nonsynonymous	chr14	15								Not performed
Father	AHNAK2	rs144504264	C=0.0114/25	chr14|105419755|G|C	nonsynonymous	chr14	15								Not performed
Patient	TRMT61B	rs144501479	NA	chr2|29073096|T|G	nonsynonymous	chr2	26								Not performed
Mother	TRMT61B	rs144501479	NA	chr2|29073096|T|G	nonsynonymous	chr2	28								Not performed
Patient	TRMT61B	rs146318990	A=0.0055/12	chr2|29092933|G|A	nonsynonymous	chr2	46								Not performed
Father	TRMT61B	rs146318990	A=0.0055/12	chr2|29092933|G|A	nonsynonymous	chr2	57								Not performed
Patient	DNAH1			chr3|52366284|C|A	nonsynonymous	chr3	39								Not performed
Father	DNAH1			chr3|52366284|C|A	nonsynonymous	chr3	55								Not performed
Patient	DNAH1	rs147123898	C=0.0046/10	chr3|52400812|A|C	nonsynonymous	chr3	23								Not performed
Mother	DNAH1	rs147123898	C=0.0046/10	chr3|52400812|A|C	nonsynonymous	chr3	24								Not performed
Patient	ARHGAP31	rs751793	T=0.0005/1	chr3|119102053|C|T	nonsynonymous	chr3	138								Not performed
Father	ARHGAP31	rs751793	T=0.0005/1	chr3|119102053|C|T	nonsynonymous	chr3	155								Not performed
Patient	ARHGAP31	rs183837502	C=0.0014/3	chr3|119133106|T|C	nonsynonymous	chr3	36								Not performed
Mother	ARHGAP31	rs183837502	C=0.0014/3	chr3|119133106|T|C	nonsynonymous	chr3	32								Not performed
Patient	SOX30	rs889057	T=0.0005/1	chr5|157053365|C|T	nonsynonymous	chr5	71								Not performed
Mother	SOX30	rs889057	T=0.0005/1	chr5|157053365|C|T	nonsynonymous	chr5	50								Not performed
Patient	SOX30	rs75818287	A=0.0055/12	chr5|157065439|G|A	nonsynonymous	chr5	25								Not performed
Father	SOX30	rs75818287	A=0.0055/12	chr5|157065439|G|A	nonsynonymous	chr5	23								Not performed
Patient	GIF	rs35867471	C=0.0174/38	chr11|59604754|T|C	nonsynonymous	chr11	88								Not performed
Father	GIF	rs35867471	C=0.0174/38	chr11|59604754|T|C	nonsynonymous	chr11	69								Not performed
Patient	GIF	rs11825834	T=0.0041/8	chr11|59611415|C|T	nonsynonymous	chr11	135								Not performed

Mother	GIF	rs11825834	T=0.0041/8	chr11|59611415|C|T	nonsynonymous	chr11	145								Not performed
Patient	TTLL5	rs141614130	A=0.0005/1	chr14|76165549|G|A	nonsynonymous	chr14	17								Not performed
Father	TTLL5	rs141614130	A=0.0005/1	chr14|76165549|G|A	nonsynonymous	chr14	25								Not performed
Patient	TTLL5	rs11848004	A=0.0101/21	chr14|76232470|G|A	nonsynonymous	chr14	22								Not performed
Mother	TTLL5	rs11848004	A=0.0101/21	chr14|76232470|G|A	nonsynonymous	chr14	29								Not performed
Patient	TNFAIP2	rs137885091	NA	chr14|103599106|C|T	nonsynonymous	chr14	45								Not performed
Mother	TNFAIP2	rs137885091	NA	chr14|103599106|C|T	nonsynonymous	chr14	52								Not performed
Patient	TNFAIP2	rs2229727	T=0.0266/57	chr14|103599847|C|T	nonsynonymous	chr14	15								Not performed
Father	TNFAIP2	rs2229727	T=0.0266/57	chr14|103599847|C|T	nonsynonymous	chr14	11								Not performed
Patient	CCDC135	rs116668060	A=0.0087/19	chr16|57738918|G|A	nonsynonymous	chr16	72								Not performed
Father	CCDC135	rs116668060	A=0.0087/19	chr16|57738918|G|A	nonsynonymous	chr16	81								Not performed
Patient	CCDC135	rs147667972	NA	chr16|57761251|C|T	nonsynonymous	chr16	24								Not performed
Mother	CCDC135	rs147667972	NA	chr16|57761251|C|T	nonsynonymous	chr16	26								Not performed
Patient	UIMC1	rs10475633	T=0.0289/62	chr5|176335585|C|T	nonsynonymous	chr5	25								Not performed
Father	UIMC1	rs10475633	T=0.0289/62	chr5|176335585|C|T	nonsynonymous	chr5	34								Not performed
Patient	UIMC1	rs199580368	C=0.0005/1	chr5|176396261|A|C	nonsynonymous	chr5	172								Not performed
Mother	UIMC1	rs199580368	C=0.0005/1	chr5|176396261|A|C	nonsynonymous	chr5	127								Not performed
Patient	ETV7			chr6|36322447|A|T	nonsynonymous	chr6	13								Not performed
Father	ETV7			chr6|36322447|A|T	nonsynonymous	chr6	23								Not performed
Patient	ETV7	rs34306145	T=0.0087/19	chr6|36339176|C|T	nonsynonymous	chr6	21								Not performed
Mother	ETV7	rs34306145	T=0.0087/19	chr6|36339176|C|T	nonsynonymous	chr6	32								Not performed
Patient	VWDE	rs116125922	A=0.0092/20	chr7|12381687|G|A	nonsynonymous	chr7	27								Not performed
Father	VWDE	rs116125922	A=0.0092/20	chr7|12381687|G|A	nonsynonymous	chr7	24								Not performed
Patient	VWDE	rs17165872	A=0.0344/75	chr7|12384089|C|A	nonsynonymous	chr7	29								Not performed
Mother	VWDE	rs17165872	A=0.0344/75	chr7|12384089|C|A	nonsynonymous	chr7	39								Not performed
Patient	CSMD1	rs199914130	NA	chr8|2966208|G|A	nonsynonymous	chr8	72								Not performed
Father	CSMD1	rs199914130	NA	chr8|2966208|G|A	nonsynonymous	chr8	75								Not performed
Patient	CSMD1	rs115021133	G=0.0106/22	chr8|3226864|C|G	nonsynonymous	chr8	31								Not performed
Mother	CSMD1	rs115021133	G=0.0106/22	chr8|3226864|C|G	nonsynonymous	chr8	27								Not performed

*Not a Caucasian patient (both parents are Angolan)

Table S1.5 – Variants found by WES in patient 4.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	SMARCA1			chrX|128599594|C|A	nonsynonymous	chrX	42	P	P	P	P	P	P	Yes	Confirmed. *De novo*
Patient	OR5V1			chr6|29323941|T|A	nonsynonymous	chr6	17	P	P	P	P	P	NP	No Info	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	67	P	P	P	P	NP	NP	Yes	Not performed
Mother	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	75	P	P	P	P	NP	NP	Yes	Not performed
Father	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	39	P	P	P	P	NP	NP	Yes	Not performed
Patient	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	154	P	P	NP	P	NP	NP	No Info	Not performed
Mother	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	166	P	P	NP	P	NP	NP	No Info	Not performed
Father	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	75	P	P	NP	P	NP	NP	No Info	Not performed

Table S1.6 – Variants found by WES in patient 5.

denovo								Functional impact prediction							
sample_id	**Gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	ZNF238			chr1|244217659|C|T	stopgain	chr1	12	P	P	-	-	P	P	Yes	Confirmed. *De novo*
Patient	S100PBP			chr1|33295564|G|A	splicing	chr1	19	-	-	-	-	-	-	No Info	Not performed

Table S1.7 – Variants found by WES in patient 6.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	PLEKHG6			chr12|6436851|C|T	nonsynonymous	chr12	14	NP	NP	NP	NP	NP	NP	Unknown	Confirmed. Not *de novo* (paternal)
Patient	LARP4			chr12|50867915|A|G	nonsynonymous	chr12	76	P	NP	NP	NP	NP	NP	Unknown	Confirmed. *De novo*
Patient	NEK9			chr14|75570699|A|C	nonsynonymous	chr14	34	P	NP	NP	P	P	P	No	Not performed
Patient	AP3B1			chr5|77385269|A|C	nonsynonymous	chr5	23	P	NP	NP	P	P	P	Yes	Not performed

Table S1.8 – Variants found by WES in patient 8.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	MOCOS			chr18|33775244|G|A	nonsynonymous	chr18	18	P	P	P	P	P	P	No	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**		**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	63	NP	NP	NP	NP	NP	P	Yes	Confirmed. In homozygosity
Mother	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	69	NP	NP	NP	NP	NP	P	Yes	Confirmed. In heterozygosity
Father	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	67	NP	NP	NP	NP	NP	P	Yes	Confirmed. In heterozygosity

Table S1.9 – Variants found by WES in patient 9.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RCC2			chr1|17736540|C|T	nonsynonymous	chr1	8	P	P	P	P	P	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	GABBR2			chr9|101133817|C|T	nonsynonymous	chr9	7	P	NP	P	NP	NP	P	Yes	Confirmed. *De novo*
Patient	H2AFY2	rs149708840	T=0.0037/8	chr10|71871347|G|T	nonsynonymous	chr10	25	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	C10orf96			chr10|118084833|G|C	nonsynonymous	chr10	15	P	P	P	NP	P	NP	No Info	Not performed
Patient	NCF4			chr22|37271932|A|C	nonsynonymous	chr22	29	P	NP	-	P	NP	NP	No Info	Not performed
Patient	FYCO1	rs114145679	T=0.0064/14	chr3|46009810|C|T	nonsynonymous	chr3	18	NP	NP	-	-	NP	NP	Yes	Not performed
recessive_compound															
sample_id	**gene**	**snp135**		**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	TTN			chr2|179414062|G|T	nonsynonymous	chr2	108	-	-	-	-	-	NP	No Info	Not performed
Father	TTN			chr2|179414062|G|T	nonsynonymous	chr2	145	-	-	-	-	-	NP	No Info	Not performed
Patient	TTN	rs34819099	T=0.005/11	chr2|179628918|C|T	nonsynonymous	chr2	27	NP	NP	-	-	-	P	No Info	Not performed
Mother	TTN	rs34819099	T=0.005/11	chr2|179628918|C|T	nonsynonymous	chr2	25	NP	NP	-	-	-	P	No Info	Not performed

Table S1.10 – Variants found by WES in patient 10.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	MAGEL2			chr15|23889919|C|T	nonsynonymous	chr15	27	-	-	-	-	-	-	Yes	Performed. Not *de novo* (paternal)
Patient	RPGRIP1L			chr16|53636037|T|C	nonsynonymous	chr16	4	NP	NP	NP	NP	P	NP	Yes	Performed. Not *de novo* (paternal)
Patient	TMEM18			chr2|669832|C|G	nonsynonymous	chr2	18	P	P	P	P	NP	P	Yes	Performed. Not *de novo* (maternal)
Patient	MAST4	rs115056755	T=0.0018/4	chr5|66461954|C|T	nonsynonymous	chr5	6	NP	NP	NP	NP	NP	NP	Yes	Performed. Not *de novo* (maternal)
Patient	TMCC3			chr12|94976229|G|A	nonsynonymous	chr12	29	NP	NP	NP	NP	NP	P	No Info	Not performed

Table S1.11 – Variants found by WES in patient 11.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	LAMB2			chr3|49160344|C|T	nonsynonymous	chr3	7	P	P	NP	P	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	RHOBTB2			chr8|22865220|A|G	nonsynonymous	chr8	70	P	P	NP	P	NP	NP	Yes	Confirmed. *De novo*
Patient	MARCH8			chr10|46028610|A|G	nonsynonymous	chr10	35	NP	NP	NP	NP	NP	NP	Unknown	Not performed
Patient	TTC40			chr10|134736171|G|A	nonsynonymous	chr10	18	P	-	-	-	-	-	No Info	Not performed
Patient	TAS2R30			chr12|11286282|G|A	nonsynonymous	chr12	40	-	-	NP	NP	NP	-	No Info	Not performed
Patient	CIITA			chr16|11001853|T|C	nonsynonymous	chr16	21	P	P	NP	P	P	P	Unknown	Not performed
Patient	ALPPL2			chr2|233273068|T|G	nonsynonymous	chr2	46	NP	NP	NP	NP	NP	NP	Unknown	Not performed
Patient	BCR			chr22|23615846|G|A	nonsynonymous	chr22	14	NP	NP	NP	NP	NP	P	Unlnown	Not performed
Patient	SLC36A2			chr5|150696620|C|T	nonsynonymous	chr5	19	P	P	NP	NP	NP	P	Uknown	Not performed
Patient	TINAG			chr6|54186172|A|G	nonsynonymous	chr6	38	NP	NP	NP	NP	NP	NP	Unknown	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	55	P	P	NP	P	NP	P	Yes	Confirmed. In homozygosity
Mother	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	60	P	P	NP	P	NP	P	Yes	Confirmed. In heterozygosity
Father	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	61	P	P	NP	P	NP	P	Yes	Confirmed. In heterozygosity

Table S1.12 – Variants found by WES in patient 12.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	EEF1A2			chr20|62120361|T|G	nonsynonymous	chr20	21	P	P	P	P	P	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	DNAH1	rs61734653	G=0.005/11	chr3|52381865|A|G	nonsynonymous	chr3	11	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (maternal)
Patient	PRDM5			chr4|121739655|G|A	nonsynonymous	chr4	57	NP	NP	NP	NP	NP	NP	Yes	Not performed

Tables S1.13 – Variants found by WES in patient 13.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RRAS2			chr11|14317388|G|A	nonsynonymous	chr11	20	P	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	BNIP2	rs142043306	NA	chr15|59961478|A|G	nonsynonymous	chr15	11	P	P	-	P	NP	P	Yes	Confirmed. Not *de novo* (paternal)
Patient	MTERFD3			chr12|107371889|C|T	nonsynonymous	chr12	17	P	NP	NP	NP	NP	P	Unknown	Not performed
Patient	ENAM			chr4|71508671|C|A	nonsynonymous	chr4	31	P	P	P	NP	NP	NP	Unknown	Not performed

recessive_homoz

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	45	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in homozygosity
Mother	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	67	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in heterozygosity
Father	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	87	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in heterozygosity
Patient	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	14	NP	NP	NP	NP	P	P	Yes	Not performed
Mother	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	31	NP	NP	NP	NP	P	P	Yes	Not performed
Father	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	42	NP	NP	NP	NP	P	P	Yes	Not performed

recessive_compound

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	TCTN2	rs145374149	T=0.0009/2	chr12|124171486|C|T	nonsynonymous	chr12	56	NP	NP	NP	NP	NP	NP	Yes	Not performed
Mother	TCTN2	rs145374149	T=0.0009/2	chr12|124171486|C|T	nonsynonymous	chr12	99	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	TCTN2			chr12|124184283|T|C	nonsynonymous	chr12	51	NP	NP	NP	NP	P	NP	Yes	Not performed
Father	TCTN2			chr12|124184283|T|C	nonsynonymous	chr12	114	NP	NP	NP	NP	P	NP	Yes	Not performed
Patient	PLCH1	rs150381264	T=0.0018/4	chr3|155199598|C|T	nonsynonymous	chr3	41	NP	NP	NP	NP	NP	NP	Yes	Not performed
Mother	PLCH1	rs150381264	T=0.0018/4	chr3|155199598|C|T	nonsynonymous	chr3	65	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	PLCH1	rs150143990	A=0.0046/10	chr3|155314114|G|A	nonsynonymous	chr3	17	NP	NP	NP	P	P	NP	Yes	Not performed
Father	PLCH1	rs150143990	A=0.0046/10	chr3|155314114|G|A	nonsynonymous	chr3	50	NP	NP	NP	P	P	NP	Yes	Not performed

Table S1.14 – Variants found by WES in patient 14.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	STXBP1			chr9|130425592|T|C	nonsynonymous	chr9	83	P	P	P	P	P	P	Yes	Confirmed. *De novo*
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	HPX	rs150488733	NA	chr11|6452695|G|A	nonsynonymous	chr11	56	P	P	NP	P	P	P	Yes	Not performed
Father	HPX	rs150488733	NA	chr11|6452695|G|A	nonsynonymous	chr11	77	P	P	NP	P	P	P	Yes	Not performed
Patient	HPX			chr11|6459640|C|A	nonsynonymous	chr11	68	P	NP	P	P	NP	P	Yes	Not performed
Mother	HPX			chr11|6459640|C|A	nonsynonymous	chr11	63	P	NP	P	P	NP	P	Yes	Not performed
Patient	HSD17B8			chr6|33172705|G|C	nonsynonymous	chr6	12	P	NP	NP	NP	NP	P	Yes	Not performed
Mother	HSD17B8			chr6|33172705|G|C	nonsynonymous	chr6	5	P	NP	NP	NP	NP	P	Yes	Not performed
Patient	HSD17B8	rs116381506	T=0.0046/10	chr6|33173457|C|T	nonsynonymous	chr6	338	P	P	P	P	P	P	Yes	Not performed
Father	HSD17B8	rs116381506	T=0.0046/10	chr6|33173457|C|T	nonsynonymous	chr6	319	P	P	P	P	P	P	Yes	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	GPR64			chrX|19024128|C|A	nonsynonymous	chrX	8	P	NP	-	-	NP	NP	Yes	Not performed
Mother	GPR64			chrX|19024128|C|A	nonsynonymous	chrX	9	P	NP	-	-	NP	NP	Yes	Not performed
Patient	FAM48B2			chrX|24329737|C|A	nonsynonymous	chrX	5	-	-	-	-	P	-	No info	Not performed
Mother	FAM48B2			chrX|24329737|C|A	nonsynonymous	chrX	16	-	-	-	-	P	-	No info	Not performed

Table S1.15 – Variants found by WES in patient 15.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	CHPF			chr2|220404347|C|G	nonsynonymous	chr2	16	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (paternal)
Patient	ZDBF2			chr2|207172196|A|G	nonsynonymous	chr2	53	P	P	NP	P	P	NP	Unknown	Not performed
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	CAGE1	rs45437691	A=0.0005/1	chr6|7378978|G|A	nonsynonymous	chr6	74	NP	-	-	-	NP	NP	Unknown	Not performed
Father	CAGE1	rs45437691	A=0.0005/1	chr6|7378978|G|A	nonsynonymous	chr6	81	NP	-	-	-	NP	NP	Unknown	Not performed
Patient	CAGE1	rs183206380	A=0.0009/2	chr6|7379149|C|A	nonsynonymous	chr6	71	NP	-	-	NP	NP	NP	Unknown	Not performed
Mother	CAGE1	rs183206380	A=0.0009/2	chr6|7379149|C|A	nonsynonymous	chr6	47	NP	-	-	NP	NP	NP	Unknown	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	SLC35A2			chrX|48762414|C|T	nonsynonymous	chrX	12	P	P	NP	P	NP	P	Yes	Confirmed. Maternal (X-linked)
Mother	SLC35A2			chrX|48762414|C|T	nonsynonymous	chrX	19	P	P	NP	P	NP	P	Yes	Confirmed. In heterozigosity

Table S1.16 – Variants found by WES in patient 17.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	DDX23			chr12|49230558|G|A	nonsynonymous	chr12	120	P	P	P	P	P	P	Unknown	Condirmed. *De novo*
Patient	EEF1A2			chr20|62127259|C|T	nonsynonymous	chr20	37	P	P	NP	P	NP	P	Yes	Condirmed. *De novo*
Patient	LIN9			chr1|226475489|C|T	nonsynonymous	chr1	20	P	NP	NP	NP	NP	NP	Unknwon	Not performed
Patient	L3MBTL3			chr6|130460830|G|C	nonsynonymous	chr6	29	P	P	P	P	P	P	Unknwon	Not performed
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	AHNAK			chr11|62290815|A|G	nonsynonymous	chr11	115	P	P	NP	P	P	NP	Yes	Not performed
Father	AHNAK			chr11|62290815|A|G	nonsynonymous	chr11	117	P	P	NP	P	P	NP	Yes	Not performed
Patient	AHNAK			chr11|62298843|G|A	nonsynonymous	chr11	124	P	NP	NP	NP	NP	NP	Yes	Not performed
Mother	AHNAK			chr11|62298843|G|A	nonsynonymous	chr11	120	P	NP	NP	NP	NP	NP	Yes	Not performed

Table S1.17 – Variants found by WES in patient 18.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	ASPM			chr1|197097663|C|T	nonsynonymous	chr1	13	NP	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	CHD8			chr14|21873449|C|A	nonsynonymous	chr14	35	P	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	LMOD1			chr1|201915453|T|C	nonsynonymous	chr1	85	P	P	NP	NP	NP	NP	Unknown	Not performed
Patient	LAMP1			chr13|113975974|C|T	nonsynonymous	chr13	60	NP	NP	NP	NP	NP	NP	Yes	Not performed

Table S1.18 – Variants found by WES in patient 19.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	SPRED2			chr2|65659108|G|T	nonsynonymous	chr2	17	NP	NP	NP	NP	P	P	Yes	Not performed
Patient	SLC38A9	rs35800744	T=0.005/11	chr5|54945091|C|T	nonsynonymous	chr5	38	NP	NP	NP	NP	NP	NP	No Info	Not performed
Patient	TTC37			chr5|94856512|T|A	nonsynonymous	chr5	15	P	P	P	P	NP	P	Unknown	Not performed
Patient	DTNBP1			chr6|15651554|A|T	nonsynonymous	chr6	15	P	NP	NP	NP	NP	P	Yes	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	ZFX	rs150375972	A=0.0006/1	chrX|24197650|G|A	nonsynonymous	chrX	41	P	NP	NP	NP	-	NP	Yes	Confirmed. Maternal (X-linked)
Mother	ZFX	rs150375972	A=0.0006/1	chrX|24197650|G|A	nonsynonymous	chrX	93	P	NP	NP	NP	-	NP	Yes	Confirmed. In heterozigosity
Patient	SHROOM4	rs189694750	A=0.003/5	chrX|50378637|G|A	nonsynonymous	chrX	12	P	P	NP	P	P	P	Yes	Confirmed. Maternal (X-linked)
Mother	SHROOM4	rs189694750	A=0.003/5	chrX|50378637|G|A	nonsynonymous	chrX	27	P	P	NP	P	P	P	Yes	Confirmed. In heterozigosity
Patient	PHKA2			chrX|18943832|G|A	nonsynonymous	chrX	11	P	P	P	P	P	P	No	Not performed
Mother	PHKA2			chrX|18943832|G|A	nonsynonymous	chrX	39	P	P	P	P	P	P	No	Not performed
Patient	ACOT9			chrX|23722015|G|A	nonsynonymous	chrX	17	P	NP	NP	P	P	-	Unknown	Not performed
Mother	ACOT9			chrX|23722015|G|A	nonsynonymous	chrX	70	P	NP	NP	P	P	-	Unknown	Not performed
Patient	HDX			chrX|83695581|G|C	nonsynonymous	chrX	30	P	NP	NP	P	NP	P	No Info	Not performed
Mother	HDX			chrX|83695581|G|C	nonsynonymous	chrX	47	P	NP	NP	P	NP	P	No Info	Not performed

Supplementary table S1.19 – Genes harboring variants submitted confirmation by to Sanger sequencing

Trio	#Genes with variants selected for Sanger confirmation	Genes with variants selected for Sanger confirmation	#Genes with variants confirmed by Sanger	Gene with variants confirmed by Sanger
1	5	*PSIP1, TRAPPC6A, WASF2, PICALM, NXF1*	5	*PSIP, TRAPPC6A, WASF2, PICALM, NXF1*
2	4	*FXR2, HTT, HEXDC, RIN2*	3	*HTT, HEXDC, RIN2*
3	8	*CCDC73, PPP1R32, GAL, EXOC3L4, CACNA1H, KATNB1, CELSR3, SYNE1*	8	*CCDC73, PPP1R32, GAL, EXOC3L4, CACNA1H, KATNB1, CELSR3, SYNE1*
4	1	*SMARCA1*	1	*SMARCA1*
5	1	*ZNF238*	1	*ZNF238*
6	2	*PLEKHG6, LARP4*	2	*PLEKHG6, LARP4*
8	2	*SLC26A4, EIF2B2*	1	*EIF2B2*
9	3	*GABBR2, RCC2, ITGA7*	2	*GABBR2, RCC2*
10	6	*MAGEL2, DROSHA, TMEM18, DKK3, RPGRIP1L, MAST4*	4	*MAGEL2, TMEM18, RPGRIP1L, MAST4*
11	4	*JMJD1C, RHOBTB2, LAMB2, EIF4G1*	3	*RHOBTB2, LAMB2, EIF4G1*
12	5	*DDX23, EEF1A2, LEPROTL1, DNAH1, TCOF1*	2	*EEF1A2, DNAH1*
13	3	*BNIP2, PSMD1, RRAS2*	3	*BNIP2, PSMD1, RRAS2*
14	1	*STXBP1*	1	*STXBP1*
15	3	*SLC35A2, KCNT2, CHPF*	2	*SLC35A2, CHPF*
17	2	*EEF1A2, DDX23*	2	*EEF1A2, DDX23*
18	3	*CHD8, ASPM, CADPS2*	2	*CHD8, ASPM*
19	3	*SHROOM4, CACNA1G, ZFX*	2	*SHROOM4, ZFX*

Part 2 - Detailed methodology

Array comparative genomic hybridization analysis

Agilent 180K

The aCGH analysis was performed on a human genome CGH Agilent 180K custom array designed by the Low Lands Consortium (LLC, Professor Klass Kok), specifically tailored for the study of children with ID/DD (AMADID:023363; Agilent, Santa Clara, CA). A total of 500 ng of DNA was labeled with Cy3 (test DNA) and Cy5 (reference DNA) using Klenow fragment at 37ºC during 4 hours (http://www.enzolifesciences.com/ENZ-42671/cgh-labeling-kit-for-oligo-arrays/). The removal of uncoupled nucleotides was carried out in a MinElute PCR purification kit (Qiagen) according to the manufacturer´s instructions (http://www.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/MinEluteReactionCleanupKit.aspx?r=1644). As reference DNA we used Kreatech´s MegaPoll Reference DNA (Kreatech Diagnostics). Arrays were then hybridized using the Agilent SurePint G3 Human CGH Microarray Kit. Briefly, samples and controls were hybridized in the presence of Cot-1 DNA and blocking agents for 24 hours, 65ºC and 20rpm. After the hybridization period, the slides were washed and scanned with Agilent Microrray Scanner. Data was extracted with the Agilent Feature Extraction (FE) Software v10.5 using default settings for CGH hybridizations. Image analysis was performed using the across-array methodology described previously [47]. CGH data was analyzed using Nexus Copy Number 6.0 software with FASST2 Segmentation algorithm and a minimum of three probes in a region required to be considered a copy number alteration. The stipulated minimal thresholds for calling a copy number gain were 0.2 (Copy number

Gain) and 0.6 (High Copy Gain) and for calling a copy number loss were -0.2 (Copy number Loss) and -1 (Homozygous Copy Loss).

Illumina HumanOmniExpress

The microarray analysis was performed on a Illumina HumanOmniExpress beadchip array (WG312-1120; Illumina, San Diego, CA). The array contained about 730.525 markers probes with a mean and median resolution of one probe every 4.0kb and 2.1kb, respectively. Samples were processed using the manufacturer's recommended assay: Infinium® HD assay. Using this protocol, around 750 ng each of DNA samples were denatured and neutralized to prepare them for amplification. The denatured DNA was isothermally amplified in an overnight incubation at 37ºC. The amplified product was enzymatically fragmented to 300-600 base pairs, purified by isopropanol precipitation and later resuspended. Hybridization of sheared DNA to the bead chip was carried out in a capillary flow through chamber. Samples were pipetted into the BeadChip and incubated overnight at 48ºC in the Illumina Hybridization Oven. Only the labeled probes are left on the beads. Green fluorescent Streptavidin and red fluorescent Anti-DNP Antibody are used to bind specifically to the labeled probes. The beads arrays are scanned using Illumina HiScan which has a laser that excites the fluorophore of the single-base extension product on the beads. The fluorescence intensity data generated is used to score genotype calls for each SNP. The scanned data are analyzed on Genome Analyzer software provided by Illumina, Inc.

CNVision [48], a pipeline developed for Illumina microarray data, was used to analyze CNVs using three algorithms with two different approaches: PennCNV [49], QuantiSNP [50] and GNOSIS [48]. GNOSIS is a distribution function algorithm that uses sliding windows to detect CNVs whereas PennCNV and QuantiSNP both implement a hidden

Markov model. All algorithms use logRratio (LRR; a measure of total signal intensity of probes) and B allele frequency (BAF; a measure of relative intensity ratio of allelic probes) information from the genotyping arrays to detect CNVs. Priority was given to CNVs with > 40kb (i.e., > 10 probes), rare (frequency<1% in DGV) and *de novo,* called by both QuantiSNP and PennCNV. Accuracy of CNVs detected *in silico* was assessed by visual inspection of LRR and BAF plots in GenomeStudio software, with its built-in algorithm - cnvPartition [51]. CNV interpretation took into consideration the size and number of genes, as well as presence of similar CNVs in databases such as ISCA [52] https://www.iscaconsortium.org/) and DECIPHER [53] and information in OMIM and PubMed.

CNV interpretation

Interpretation of the CNVs found was carried out based on the workflow proposed by [54]. For each patient the total number of CNVs was listed according to the position in the chromosome and classified according to the workflow represented in figure S2.1.

Quantitative PCR analysis

Primers for quantitative PCR analysis (qPCR) were designed using Primer3Plus software [55] (listed in table 22) and taking into account standard recommendations for qPCR primer development. A set of primers were designed for the *TFC4* gene (ENSG00000196628). The reference genes used were *SDC4* (ENSG00000124145) and *ZNF80* (ENSG00000174255) localized in the 20q12-q13 and 3p12 regions, respectively (primers table S2.1). qPCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems, Foster City, CA, USA) using Power SYBR Green® (Applied Biosystems) for a final volume of 20 µl, at the following amplifications conditions: denaturation for 10 minutes at 95ºC, followed by 40 amplification cycles for

15 seconds at 60ºC. The specificity of each of the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage. Quantification was performed as described elsewhere [56]. Ct values obtained for each test were analyzed in DataAssist ™ software from Applied Biosystems.

Exome sequencing and data analysis

Library preparation and SOLiD sequencing

Exome enrichment was performed using 3 μg genomic DNA. DNA samples were sheared by sonication with the Covaris S2 instrument (Covaris, Inc.). Fragment libraries were constructed from the sheared samples using the AB Library Builder System (Life Technologies) and target enrichment was performed according to the manufacturer's protocols (Agilent SureSelect Human All Exon v4 kit). The Agilent SureSelect Human All Exon v4 kit was used for enrichment, containing exonic sequences of 20.965 genes (corresponding to a total of 334.378 exons), covering a total of 51Mb of genomic sequence (as specified by the company). Exome capture was conducted by hybridizing the DNA libraries with biotinylated RNA baits for 24 h followed by extraction using streptavidin coated magnetic beads. Captured DNA was then amplified followed by emulsion PCR using the EZ Bead System (Life Technologies) and sequenced on the SOLiD5500xl system, generating over 40 million reads of length 75bp for each of the samples. Individual libraries were labeled by a post-hybridization barcoding procedure (Agilent, Santa Clara, CA, USA; barcodes compatible with SOLiD sequencing technology).

Mapping and variant calling

Alignment of color space reads to the human reference genome (hg19) was performed using v2.1 of the Lifescope Software (Life Technologies). Around 90% of the reads from each sample were uniquely mapped to the target regions, generating an average coverage of 35X and a median of 30x over the targeted exons across samples. Similar coverage has been described in other studies using a Agilent SureSelect human exome enrichment kit and SOLiD sequencing platform (Life Technologies) [57]. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were subsequently called by the diBayes algorithm available within the Lifescope software. At least 10 reads are required to obtain a 99% probability that at least two reads contain the variant allele (assuming a binomial distribution with probability 0.5 of sequencing the variant allele at a heterozygous position. In our cohort, 80% of the targets were covered at least 10 fold (figure S2.2). Single Nucleotide Variants (SNVs) and indels were filtered as follows: exclusion of synonymous and common variants (MAF>1% dbSNP35, Exome Sequencing Project and 1000 genomes Project), variants present in the CanvasDB in-house exome database[58] in Uppsala (excluding individuals with neurodevelopmental phenotypes), presence of at least 3 unique variant reads (i.e. different start sites), as well as the variant being present in at least 15% of all reads. All called SNVs and indels were imported into a local installation of the CanvasDB database system [58] for annotation and further analysis of the variants.

Analysis pipeline

To avoid false positives (i.e., variant calling in duplicate reads due to PCR artifacts), only variants present in reads with at least 3 different starts were used in downstream analysis. Then, we excluded intronic or synonymous variants. Next we filtered out variants present

in dbSNPv135 or in our in-house variant database. Our database contains variants from 500 in-house performed ‘exomes’ analyzed by the same pipeline. The inhouse database is composed of aproximately 300 healthy individuals and 200 patients with non-neurodevelopmental disorders. Variants with frequency >1% in the 6500 individuals of the Exome Variant Server/National Heart, Lung and Blood Institute [59] or in the 1000 Genomes Project [60] were also excluded This step further reduced the average number of variants (including stopgain, stoploss, missense, splicing, frameshift, in frame) to 26 per patient (which we called “private” variants).

Prioritization of candidate variants

The exome data from both parents was used to provide the parent-of-origin of each candidate variant. Variants that were not identified in either parent with >15% variation reads were considered to be candidate *de novo* variants.

For the male patients the variants located at the X-chromosome and of maternal inheritance were classified as such, allowing an additional filter step for X-linked maternally-inherited variants. For recessive analysis, non-synonymous homozygous, non-synonymous compound heterozygous variants and/or canonical splice site variants in the same gene were selected for evaluation, with each parent being carrier of one of the variants.

Variants were selected for Sanger confirmation if (I) Integrative Genomics Viewer[61] visualization confirmed that the position where the variant occurred was sufficiently covered and matched the pre-defined inheritance model, (II) the variant occurred in a known or candidate ID gene or a gene known to be functionally relevant for the nervous system, (III) the effect of the variant on the gene was predicted to be deleterious. Gene prioritization took into consideration information on biological function available in the

literature (PubMed search) and in the OMIM, Gene Entrez, and GeneCards databases, genetic/protein interactions,[62] brain expression[63,64] and KO mice phenotype.[65]

We also took into consideration the type of variant and gave preference to loss of function variants (i.e., nonsense, splice site, frameshift) followed by missense and in frame variants. Guidance was also obtained from six prediction softwares, which was particularly helpsulf on evauating missense variants. The prediction softwares and cutoffs used were: SIFT (damaging if SIFT score<=0.05) [66], PolyPhen2 (damaging if PolyPhen2 score >0.05) [67], Mutation Assessor (damaging if Mutation assessor score >1.938) [68], Mutation Taster (damaging if Mutation taster classifies as disease causing) [69], PMut (damaging if Pmut score >0.5) [70], Condel (damaging if Condel classifies as deleterious) [71]. Conservation scores were also taken into account, namely: PhyloP (conserved if score >0.95 in a range between 0 and 1) [72], GERP (constrained if score ranges from 0 to 6.18) [73]. Finally, we compared the variants identified in our patients to variants in curated catalogs of pathogenic variants: Human Gene Mutation Database (HGMD) [74] and ClinVar [75]. A workflow for this analysis is represented in figure S2.3.

Prioritization of candidate genes

Gene prioritization took into consideration (I) the biological function according to information available in the literature (PubMed search) [76] and in the Online Mendelian Inheritance in Man (OMIM) [77], Gene Entrez [78], and GeneCards [79] databases; (II) the genetic/protein interactions [80] (preference was given to genes that were on the same pathway or had co-expression, co-localization, physical or genetic interactions with *MECP2*, *CDKL5*, *FOXG1*); (III) brain expression [81,82] (preference was given to genes expressed in the brain); and (IV) KO mice phenotype [83] (preference was given to genes

for which KO mice had a neurological phenotype). A workflow for this analysis is represented in figure S2.4.

Sanger validation of selected candidate genes

Validation of the selected candidate variants was performed using standard Sanger sequencing. Primers were designed to surround the candidate variant using Primer3Plus software [55] and PCR reactions were performed using MasterMix ™ PCR reaction mix (Finnzymmes, ©Thermo Fisher Scientific In) for a final volume of 20 µl, at the following amplifications conditions: denaturation for 5 minutes at 95ºC, followed by 30 amplification cycles (95ªC for 1 minute, amplicom specific annealing temperature for 30 seconds, 72ºC for 1 minute) and final extension for 10 minutes at 72ºC. The products were sequenced in ABI 3730 XL equipment (Applied Biosystems, Life technologies. Carlsbad, CA). Of 56 candidate variants, 44 variants were confirmed by Sanger to be present (see supplementary table S20). Primer sequences are available in supplementary table S2.2.

Network analysis

We performed gene network analysis to: I) verify if our list of candidate genes interacted amongst themselves and with the known RTT genes (*MECP2, CDKL5, FOXG1*), II) study the topology of these interactions, III) predict additional genes that may be involved in RTT if they are shown to interact with a large number of genes in the query set, IV) identify common biological themes by exploring functional enrichment analysis of Gene Ontology (GO) terms.

Network analysis was performed with GeneMANIA (version 3.1.2.7, http://www.genemania.org/).[80,84] Given a set of input genes, GeneMANIA finds related genes using a very large set of functional association data, including protein

interactions (two gene products are linked if they were found to interact in a protein-protein interaction study), genetic interactions (two genes are functionally associated if the effects of perturbing one gene perturb a second gene), pathway (two gene products are linked if they participate in the same reaction within a pathway), co-expression (two genes are linked if their expression levels are similar across conditions in a gene expression study), co-localization (two genes are linked if they are both expressed in the same tissue or if their gene products are both identified in the same cellular location), shared protein domain (two gene products are linked if they have the same protein domain) and predicted functional relationship (a major source of predicted data being known functional relationships from another organism via orthology). GeneMANIA also allows for functional enrichment analysis. Enrichment is calculated using as the foreground set all the genes in either the query list or the related genes discovered by GeneMANIA that have at least one GO annotation and as the background set all genes with GO annotations and at least one interaction in our GeneMANIA's database. For our analysis, we used GeneMANIA's default datasets for *Homo sapiens* (last dataset update June 1st 2014) and equal weighting by network. The genes used as input were the already known RTT genes (*MECP, CDKL5, FOXG1*) as well as the genes selected as likely causing RTT-like phenotype in our cohort. In the results generated by GeneMANIA we allowed for up to 20 related genes and at most 10 related attributes to be displayed.

Supplementary figures

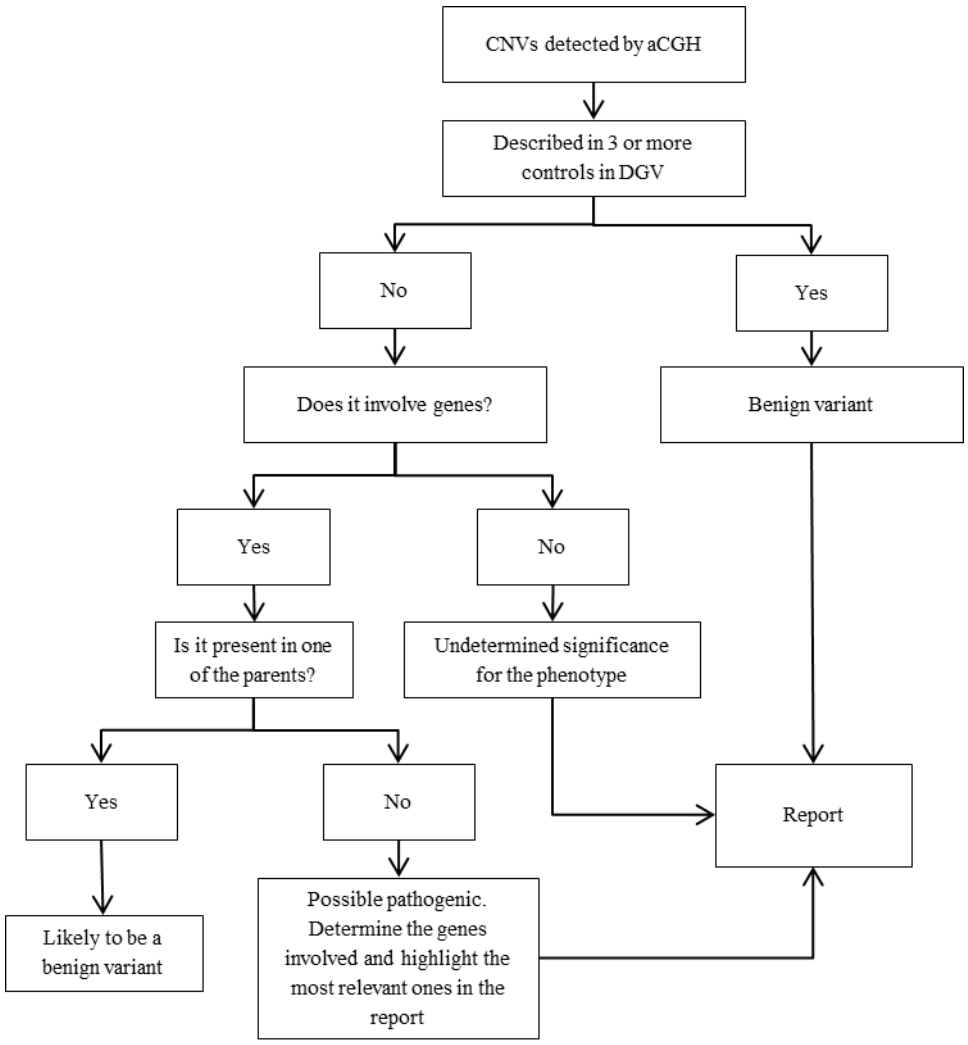


Figure S2.1 – Schematic representation of the workflow followed for the interpretation of the CNVs found in each patient. (DGV: Database of Genomic Variants http://dgv.tcag.ca/dgv/app/home)

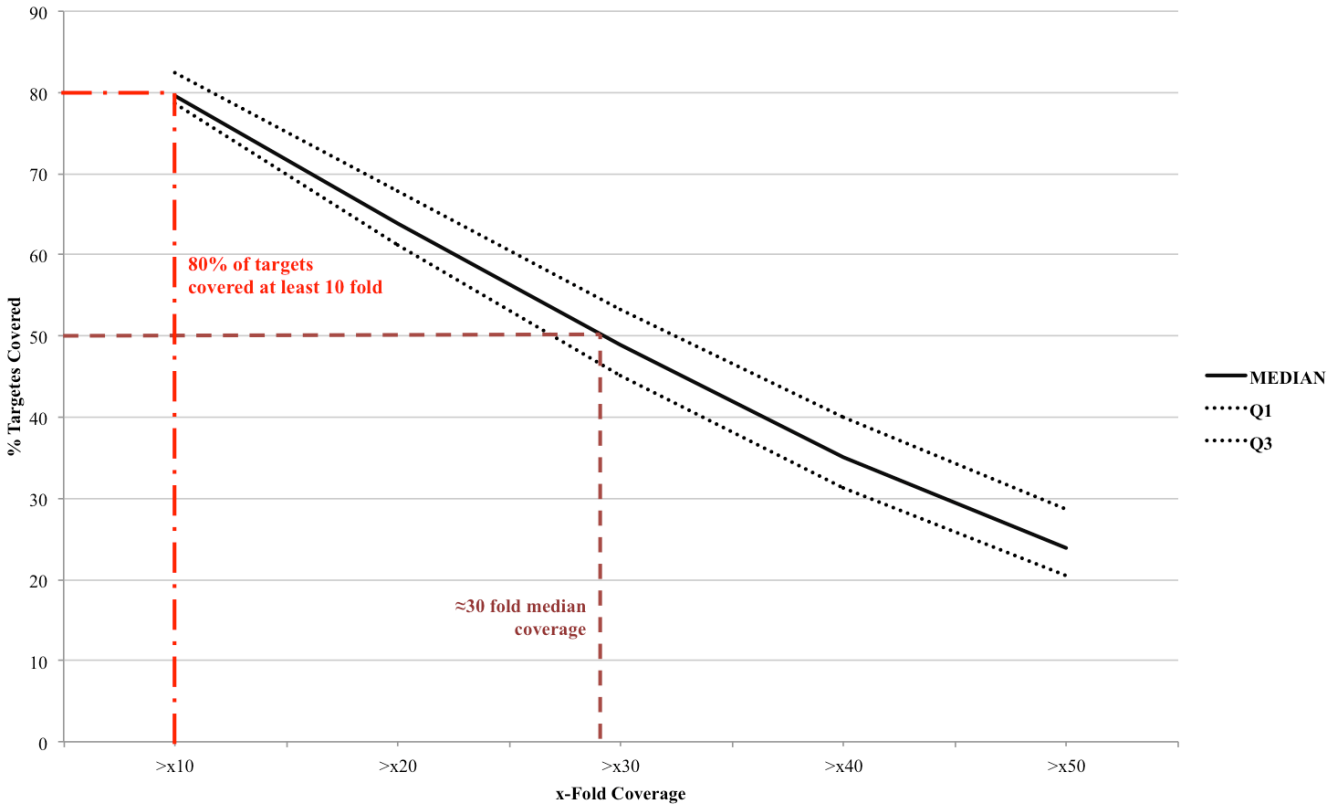
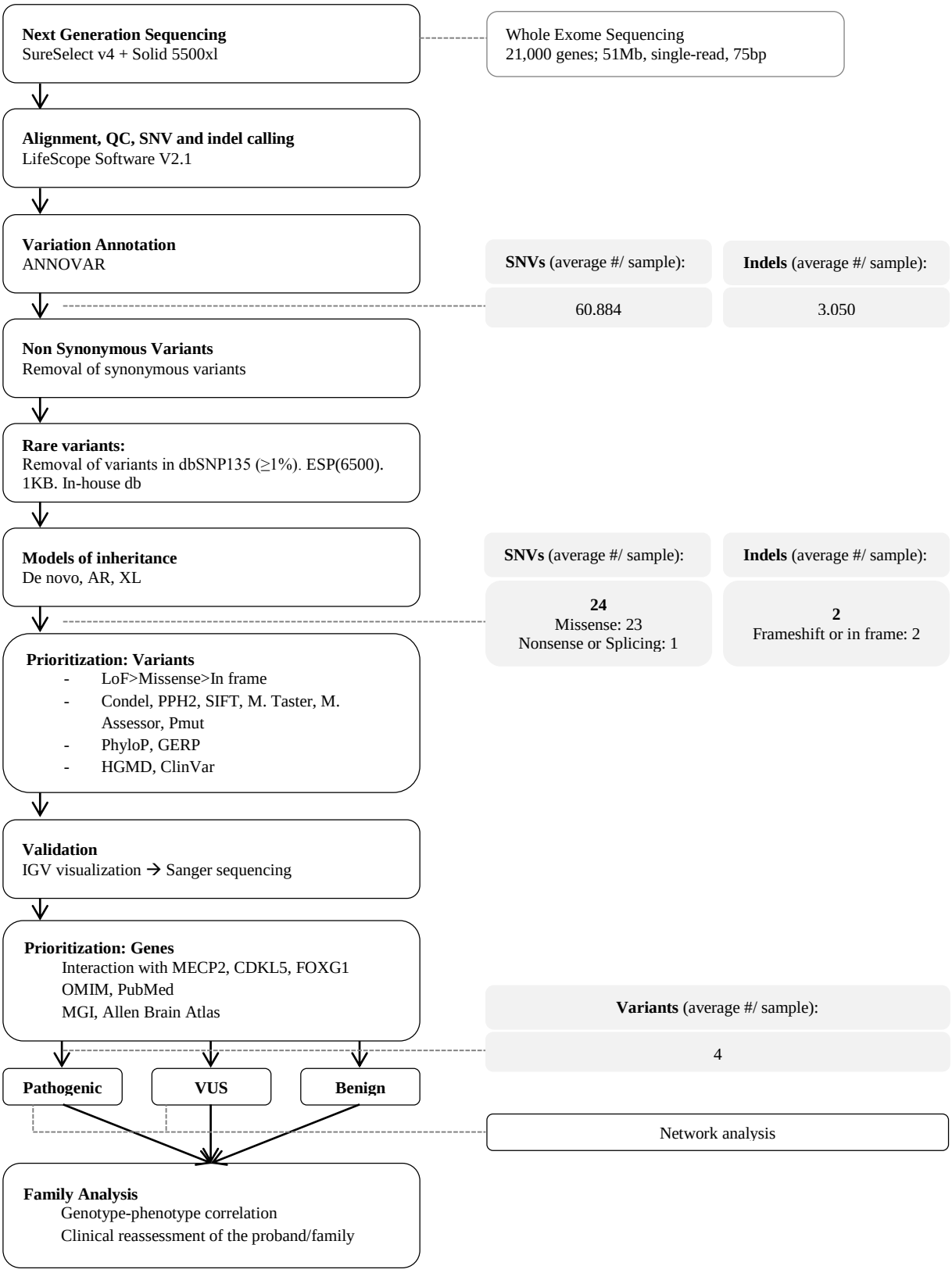


Figure S2.2 – WES coverage distribution. Coverage for all exons targeted by enrichment was evaluated. The median coverage across individuals was approximately 30 fold, with 80% of the targets covered at least 10 fold.


Next Generation Sequencing
SureSelect v4 + Solid 5500xl
Whole Exome Sequencing
21,000 genes; 51Mb, single-read, 75bp
Alignment, QC, SNV and indel calling
LifeScope Software V2.1
Variation Annotation
ANNOVAR
SNVs (average #/ sample):
Indels (average #/ sample):
60.884
3.050
Non Synonymous Variants
Removal of synonymous variants
Rare variants:
Removal of variants in dbSNP135 (≥1%). ESP(6500). 1KB. In-house db
Models of inheritance
De novo, AR, XL
SNVs (average #/ sample):
Indels (average #/ sample):
24
Missense: 23
Nonsense or Splicing: 1
2
Frameshift or in frame: 2
Prioritization: Variants
- LoF>Missense>In frame
- Condel, PPH2, SIFT, M. Taster, M. Assessor, Pmut
- PhyloP, GERP
- HGMD, ClinVar
Validation
IGV visualization → Sanger sequencing
Prioritization: Genes
Interaction with MECP2, CDKL5, FOXG1
OMIM, PubMed
MGI, Allen Brain Atlas
Variants (average #/ sample):
4
Pathogenic
VUS
Benign
Network analysis
Family Analysis
Genotype-phenotype correlation
Clinical reassessment of the proband/family

Exome sequencing and data analysis

Figure S2.3 - Overview of WES bioinformatic analysis pipeline. Abbreviatons: 1KG: 1000 genomes project; AR: autosomal recessive; ESP: Exome Sequencing Project; GERP: Genomic Evolutionary Rate Profiling; HGMD: Human Gene Mutation Database; IGV: Integrative Genomics Viewer; LoF: loss of function; M.ass: Mutation Assessor; M Taster: Mutation Taster; MGI: Mouse Genome Informatics; OMIM: Online Mendelian Inheritance in Man; PPH2: PolyPhen2; PhyloP: phylogenetic p-values; SIFT: Sorting Intolerant From Tolerant; VUS: Variants of Unkown Significance; XL: X-linked

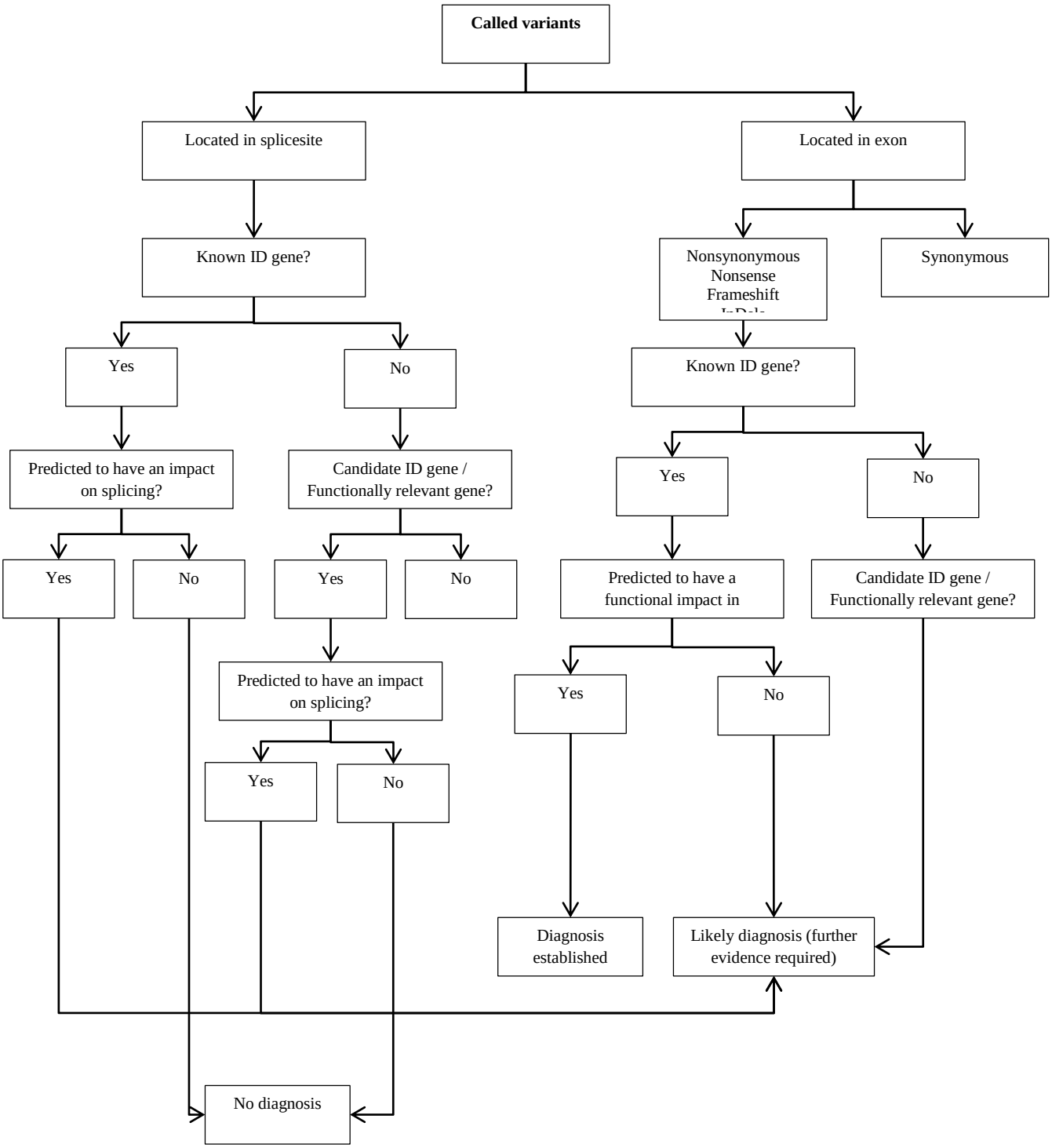


Figure S2.4 – Classification of the variants detected by WES. The workflow used to classify the variants called by WES is represented.

Supplementary tables

Supplementary table S2.1 – Primers used for quantitative PCR confirmation.

Gene	Primer Forward 5'→3'	Size (bp)	%GC	Tm (ºC)	Primer Reverse 5'→3'	Size (bp)	%GC	Tm (ºC)	Amplicon size (bp)
DRD3	CCATCTCACCATGCCTATCC	20	55	60	AAACAGAAGAGGGCAGGACA	20	50	60	168
ZNF80	GCTACCGCCAGATTCACACT	20	55	60	AATCTTCATGTGCCGGGTTA	20	45	60	182
SDC4	ACCGAACCCAAGAAACTAGA	20	45	56	GTGCTGGACATTGACACCT	19	53	57	101

Supplementary table S2.2 – Primers used for Sanger confirmation.

Gene	Primer Forward 5'→3'	Size (bp)	%GC	Tm (ºC)	Primer Reverse 5'→3'	Size (bp)	%GC	Tm (ºC)	Amplicon size (bp)
SLC35A2	AGATCCTCAAAGGCAGCTCA	20	50	60	GACAACCACAGCCACCAGTA	20	55	60	203
KCNT2	TGAACTTCTAGGATCATTTCAAGC	24	37	59	AATGCTGAATTCTCTTCTTTGG	22	36	58	250
CHPF	GACACTGGCCGCTTTGAT	18	55	60	TCCTGTTCAAAGAGTAGCATGG	22	45	59	317
RHOBTB2	CTGTACACGGGGGAGCTAGA	20	60	60	GCTGAGACAGACAGGCAATG	20	55	59	274
LAMB2	ATGCACCCTGTGCTACAAGC	20	55	61	TGGAAGCATTAGCCTTGTCC	20	50	60	261
JMJD1C	GAAATGGTGAGTTTCAGAAGGTT	23	39	59	CGTCGTGATGTAATGCCAAT	20	45	59	300
EIF4G1	TCCTGCTTCCCACTCATCTT	20	50	59	ATCCAGGAAAGCAGGGAAAT	20	45	60	245
PLEKHG6	CCAACCTGGTCTGAGGAAGA	20	55	60	CTCCCGGATCTCTCTGAGC	19	63	60	240
LARP4	TTCCTCTGTTCAGAAGGATGG	21	48	59	CCAACTTGGGCTACTTCCAA	20	50	60	248
DDX23 (exon10)	GCCCAGAGGCTCTACAGAAA	20	55	60	TGTAGCCACACTTATCAATGACC	23	44	59	291
EEF1A2 (exon3)	TTCAAGTATGCCTGGGTGCT	20	50	61	GAGCCAGACTGGGTGAGG	18	67	59	320

ASPM	CCTTTTGGCTTTTTCACGAG	20	45	60	CATTGATGTACCACTTCCCTGA	22	45	60	329
CHD8	AAGCCAATGATGCTGAGAAGA	21	43	60	AAGCATTCCCTCACCATTGA	20	45	60	246
CADPS2	TTTAATTGGGAAAAGAGTATTTTGA	25	24	58	CATCGAGAGGGGTTTAAATTG	21	43	59	300
DDX23 (exon15)	AGTGTTCATAGATGAGGGTTGG	22	45	58	GAAAGGGTGAGACTTGAAAGGA	22	45	60	286
EEF1A2 (exon7)	CAGGTCATCATCCTGAACCA	20	50	59	CGGGTACTGGGAGAAGCTC	19	63	60	231
LEPROTL1	CGTTTTAGGGATTTTTGCTTC	21	38	58	GGTAAGTGAACATTTCTCTGCATT	24	38	59	300
DNAH1	GTCATAGCGCTGGCAGGA	18	61	61	GCAGCACGTACCTTCTCCA	19	58	60	251
TCOF1	AGCTGCTGGAACAGGAAAGA	20	50	60	AGGAATGAGACCAGGTGCTG	20	55	60	473
SLC26A4	AGAATTGATTGTGTGTGTGTGC	22	41	59	TGTTTGTCAACCAAATAATGCTG	23	35	60	287
EIF2B2	TCCTCCCCACCTCTCTCTTT	20	55	60	CCTGTGACCAGTCCCATTTT	20	50	60	230
CCDC73	TTGATACAATAAAGGGAACCCAGT	24	38	60	TCCGTAACATTCGAGGTTTG	20	45	59	325
PPP1R32	CTCCCTTCTGGTGGTTCTGT	20	55	59	CTCCTGCACAAGAGGGGTAG	20	60	60	230
EXOC3L4	AACAAGCAGCAGGAAAGAGC	20	50	60	CTCGGGTTTCAGTTCCTCAG	20	55	60	298
CACNA1H	CATGAACTACCCCACGATCC	20	55	60	GAATGACACACCGGAGACC	19	58	59	249

KATNB1	AGGTGCAGTGCTGAATACCC	20	55	60	CCCCTCTGCAGCCACTTAC	19	63	61	223
SYNE1	AGAGGTAGAAAGCAGTTCTCATAATC	26	38	58	GAATGATTCACAAGAATAACACAAA	25	28	57	239
GAL	CCTGTAGCATGTGTCGTGGT	20	55	60	TGCATAAATTGGCCGAAGAT	20	40	60	235
CELSR3	TCAGGGAGCCTATCTTCGTG	20	55	60	TGCCGAATCAAAAAGTCTGA	20	40	59	298
PSIP1	TTGGTTGGTTTCAGTCTTTTCA	22	36	60	CCCTCAAAACAAGTTTTCAACA	22	36	59	299
TRAPPC6A	GAGGGTTGCATTGCTCCAT	19	53	61	TGGCTTCCCTTAGAGGATCA	20	50	60	373
WASF2	AGTGTGGTCAGCCCAAGC	18	61	60	GGTAGTCAGCTGCTGGTGGT	20	60	60	250
PICALM	TGCACAGATCCTTTCTCTGCT	21	48	60	TGCAGTGCTTTTAAAATAATTAATGG	26	27	60	247
NXF1	GAATGAACTGAAGCCAGAACAA	22	41	59	CTGAGAACTACTGCTGTCAACCA	23	48	60	240
STXBP1	TTTGTGGAAGAAACGGGAAC	20	45	60	AACCTTAAGGCCGGAGGAG	19	58	60	338
SHROOM4	TCTCTTTTGTGGCATGGGTA	20	45	59	TGGTTAGGGTACATGTTCTGGTC	23	48	60	250
CACNA1G	TGGAGCTCTTTGGAGACCTG	20	55	60	CTGTGGAGACTCGGAAGAGG	20	60	60	212
ZFX	GTGCCCAGATATCATGGAAGA	21	48	60	CGTCGGTAGTCAGAGGATCAG	21	57	60	293
TMEM18	TCCACATCAGCTCTCTCCAA	20	50	60	CACATGAAGAAAGACAACTGAAAG	24	38	58	232

DROSHA	AAATCACATTGAGATGAGATAATTTTT	27	22	58	CCACTGAATTGTATGCTTTAAATGAG	26	35	60	248
MAGEL2	CCCTAGCACCTCCAGGAT	18	61	58	CCTCTCATCCAAGGGAGACA	20	55	60	226
DKK3	GACCTCGGCTCCAGTCAA	18	61	60	CGCTCTTCCATGCCTTCC	18	61	62	212
RPGRIP1L	TTTCCTGCTCTTTTCTCCCTA	21	43	58	CCAGCTTCCCATGAATTATAGA	22	41	58	250
MAST4	CACACTGACAGGGCTCCTCTA	21	57	61	CGGGCTTTTGTCTGTCTGTC	20	55	61	249
SMARCA1	TGCACATTCAATTAAAAGCACT	22	32	58	GCAGTCCTAGACTTGATAAACCA	23	43	58	250
FXR2	GCAGCGACAAGGCTGGATA	19	58	62	CCCAGCCAATCAATCACTTT	20	45	60	245
HEXDC	AAAAACGGACCCTGTTAGGC	20	50	60	AGCAGCGCATCCACAGAG	18	61	61	197
HTT	GGAAATGATGGGAGCAGGTA	20	50	60	CCGTGGTGTCAAGAGGAACT	20	55	60	228
RIN2	GGGCTGCCTTCTTCTTCTTC	20	55	60	CTTCTCTTCACCTGCGCTTC	20	55	60	300
BNIP2	TTTTCTTTGTGTGTGTGTGTTTTT	24	29	59	TCAAAATATAGCTCTGTATCCCATAA	26	30	58	231
PSMD1	TGCTACTTTCAGTTGGATGTTTT	23	35	58	GAGCATCAGCCTCTTCCATC	20	55	60	246
TCTN2 (exon 6)	CAGCTCCTGCCTTTATGTTTG	21	48	60	GCGGCTGCTCAGACACTTAC	20	60	61	249
TCTN2 (exon 14)	GTGAAACCCGGACATTCACT	20	50	60	TCACAGCAACCAAGTTACAGG	21	48	59	236

RRAS2	GCCTCCCAGAGCATAGGATT	20	55	61	CGGGCTGCTCTGTCATCTAT	20	55	60	162
RGS4	GCAGAGCGGTCGTCTGAT	18	61	60	GCAATCTTACCTCCTCAAGCA	21	48	60	249
LIG4	CTGCACCTTGCGTTTTCC	18	57	60	TGGCTATCTGTTCCACTCATAA	22	41	57	300
GABBR2	GGGGAGTATTTGTCCCCATC	20	55	60	CCCTGAAACAGAAGGAGAGTG	21	52	59	244
RCC2	AGGAAATGTTGCCGTGATTC	20	45	60	CCCAAACAGACTGCGATGAT	20	50	61	249
ITGA7	GGGCATTGACATTTCCAAAC	20	45	60	CACAGCTCAGCCTCTCCTCT	20	60	60	266
ZNF238	CACATAGCAGGCGATTTGC	19	53	60	CTCACGGAGGTGACAGACCT	20	60	60	245

References

1 Neul JL, Kaufmann WE, Glaze DG, *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* 2010;**68**:944–50. doi:10.1002/ana.22124

2 Spurlock G, Williams J, McGuffin P, *et al.* European Multicentre Association Study of Schizophrenia: a study of the DRD2 Ser311Cys and DRD3 Ser9Gly polymorphisms. *Am J Med Genet* 1998;**81**:24–8.

3 Lucotte G, Lagarde JP, Funalot B, *et al.* Linkage with the Ser9Gly DRD3 polymorphism in essential tremor families. *Clin Genet* 2006;**69**:437–40. doi:10.1111/j.1399-0004.2006.00600.x

4 Vuillaume M-L, Delrue M-A, Naudion S, *et al.* Expanding the clinical phenotype at the 3q13.31 locus with a new case of microdeletion and first characterization of the reciprocal duplication. *Mol Genet Metab* 2013;**110**:90–7. doi:10.1016/j.ymgme.2013.07.013

5 Wiśniowiecka-Kowalnik B, Kastory-Bronowska M, Bartnik M, *et al.* Application of custom-designed oligonucleotide array CGH in 145 patients with autistic spectrum disorders. *Eur J Hum Genet EJHG* 2013;**21**:620–5. doi:10.1038/ejhg.2012.219

6 Nguyen GD, Molero AE, Gokhan S, *et al.* Functions of Huntingtin in Germ Layer Specification and Organogenesis. *PLoS ONE* 2013;**8**:e72698. doi:10.1371/journal.pone.0072698

7 McFarland KN, Huizenga MN, Darnell SB, *et al.* MeCP2: a novel Huntingtin interactor. *Hum Mol Genet* 2014;**23**:1036–44. doi:10.1093/hmg/ddt499

8 Reiss AL, Faruque F, Naidu S, *et al.* Neuroanatomy of Rett syndrome: A volumetric imaging study. *Ann Neurol* 1993;**34**:227–34. doi:10.1002/ana.410340220

9 MacArthur DG, Balasubramanian S, Frankish A, *et al.* A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science* 2012;**335**:823–8. doi:10.1126/science.1215040

10 Duyao MP, Auerbach AB, Ryan A, *et al.* Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* 1995;**269**:407–10.

11 Eckey M, Kuphal S, Straub T, *et al.* Nucleosome Remodeler SNF2L Suppresses Cell Proliferation and Migration and Attenuates Wnt Signaling. *Mol Cell Biol* 2012;**32**:2359–71. doi:10.1128/MCB.06619-11

12 Yip DJ, Corcoran CP, Alvarez-Saavedra M, *et al.* Snf2l Regulates Foxg1-Dependent Progenitor Cell Expansion in the Developing Brain. *Dev Cell* 2012;**22**:871–8. doi:10.1016/j.devcel.2012.01.020

13 Hussain RH, Zawawi M, Bayfield MA. Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. *Nucleic Acids Res* 2013;**41**:8715–25. doi:10.1093/nar/gkt649

14 Merret R, Martino L, Bousquet-Antonelli C, *et al.* The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins. *RNA N Y N* 2013;**19**:36–50. doi:10.1261/rna.035469.112

15 Yang R, Gaidamakov SA, Xie J, *et al.* La-related protein 4 binds poly(A), interacts with the poly(A)-binding protein MLLE domain via a variant PAM2w motif, and can promote mRNA stability. *Mol Cell Biol* 2011;**31**:542–56. doi:10.1128/MCB.01162-10

16 Blein S, Hawrot E, Barlow P. The metabotropic GABA receptor: molecular insights and their functional consequences. *Cell Mol Life Sci CMLS* 2000;**57**:635–50.

17 Wang Q. Integrative Genomics Identifies Distinct Molecular Classes of Neuroblastoma and Shows That Multiple Genes Are Targeted by Regional Alterations in DNA Copy Number. *Cancer Res* 2006;**66**:6050–62. doi:10.1158/0008-5472.CAN-05-4618

18 Fatemi SH, Folsom TD, Thuras PD. Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. *Schizophr Res* 2011;**128**:37–43. doi:10.1016/j.schres.2010.12.025

19 Fatemi SH, Folsom TD, Reutiman TJ, *et al.* Expression of GABA(B) receptors is altered in brains of subjects with autism. *Cerebellum Lond Engl* 2009;**8**:64–9. doi:10.1007/s12311-008-0075-3

20 Hedges DJ, Hamilton-Nelson KL, Sacharow SJ, *et al.* Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci. *Mol Autism* 2012;**3**:2. doi:10.1186/2040-2392-3-2

21 EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet* 2014;**95**:360–70. doi:10.1016/j.ajhg.2014.08.013

22 Lee S, Kozlov S, Hernandez L, *et al.* Expression and imprinting of MAGEL2 suggest a role in Prader-willi syndrome and the homologous murine imprinting phenotype. *Hum Mol Genet* 2000;**9**:1813–9.

23 Schaaf CP, Gonzalez-Garay ML, Xia F, *et al.* Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. *Nat Genet* 2013;**45**:1405–8. doi:10.1038/ng.2776

24 Wilkins A, Ping Q, Carpenter CL. RhoBTB2 is a substrate of the mammalian Cul3 ubiquitin ligase complex. *Genes Dev* 2004;**18**:856–61. doi:10.1101/gad.1177904

25 Kong A, Frigge ML, Masson G, *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;**488**:471–5. doi:10.1038/nature11396

26 O'Roak BJ, Vives L, Girirajan S, *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012;**485**:246–50. doi:10.1038/nature10989

27 Siripurapu V, Meth J, Kobayashi N, *et al.* DBC2 significantly influences cell-cycle, apoptosis, cytoskeleton and membrane-trafficking pathways. *J Mol Biol* 2005;**346**:83–9. doi:10.1016/j.jmb.2004.11.043

28 Villa N, Do A, Hershey JWB, *et al.* Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J Biol Chem* 2013;**288**:32932–40. doi:10.1074/jbc.M113.517011

29 Fujioka S, Sundal C, Strongosky AJ, *et al.* Sequence variants in eukaryotic translation initiation factor 4-gamma (eIF4G1) are associated with Lewy body dementia. *Acta Neuropathol (Berl)* 2013;**125**:425–38. doi:10.1007/s00401-012-1059-4

30 Blanckenberg J, Ntsapi C, Carr JA, *et al.* EIF4G1 R1205H and VPS35 D620N mutations are rare in Parkinson's disease from South Africa. *Neurobiol Aging* 2014;**35**:445.e1–3. doi:10.1016/j.neurobiolaging.2013.08.023

31 Li K, Tang B, Guo J, *et al.* Analysis of EIF4G1 in ethnic Chinese. *BMC Neurol* 2013;**13**:38. doi:10.1186/1471-2377-13-38

32 Puschmann A. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. *Parkinsonism Relat Disord* 2013;**19**:407–15. doi:10.1016/j.parkreldis.2013.01.020

33 Sudhaman S, Behari M, Govindappa ST, *et al.* VPS35 and EIF4G1 mutations are rare in Parkinson's disease among Indians. *Neurobiol Aging* 2013;**34**:2442.e1–3. doi:10.1016/j.neurobiolaging.2013.04.025

34 Shaheen R, Faqeih E, Seidahmed MZ, *et al.* A TCTN2 mutation defines a novel Meckel Gruber syndrome locus. *Hum Mutat* 2011;**32**:573–8. doi:10.1002/humu.21507

35 Sang L, Miller JJ, Corbit KC, *et al.* Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* 2011;**145**:513–28. doi:10.1016/j.cell.2011.04.019

36 Nishiyama M, Skoultchi AI, Nakayama KI. Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-β-catenin signaling pathway. *Mol Cell Biol* 2012;**32**:501–12. doi:10.1128/MCB.06409-11

37 Kobayashi M, Kishida S, Fukui A, *et al.* Nuclear localization of Duplin, a beta-catenin-binding protein, is essential for its inhibitory activity on the Wnt signaling pathway. *J Biol Chem* 2002;**277**:5816–22. doi:10.1074/jbc.M108433200

38 Kang H, Chen I-M, Wilson CS, *et al.* Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood* 2010;**115**:1394–405. doi:10.1182/blood-2009-05-218560

39 Smirnov DA, Morley M, Shin E, *et al.* Genetic analysis of radiation-induced changes in human gene expression. *Nature* 2009;**459**:587–91. doi:10.1038/nature07940

40 Krumm N, O'Roak BJ, Shendure J, *et al.* A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014;**37**:95–105. doi:10.1016/j.tins.2013.11.005

41 Neale BM, Kou Y, Liu L, *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;**485**:242–5. doi:10.1038/nature11011

42 O'Roak BJ, Vives L, Fu W, *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012;**338**:1619–22. doi:10.1126/science.1227764

43 Talkowski ME, Rosenfeld JA, Blumenthal I, *et al.* Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012;**149**:525–37. doi:10.1016/j.cell.2012.03.028

44 Aasland R, Stewart AF, Gibson T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem Sci* 1996;**21**:87–8.

45 Bardwell VJ, Treisman R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* 1994;**8**:1664–77.

46 Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* 1989;**17**:4713–30.

47 Buffart TE, Israeli D, Tijssen M, *et al.* Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. *Genes Chromosomes Cancer* 2008;**47**:994–1004. doi:10.1002/gcc.20605

48 Sanders SJ, Ercan-Sencicek AG, Hus V, *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011;**70**:863–85. doi:10.1016/j.neuron.2011.05.002

49 Wang K, Li M, Hadley D, *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;**17**:1665–74. doi:10.1101/gr.6861907

50 Colella S, Yau C, Taylor JM, *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007;**35**:2013–25. doi:10.1093/nar/gkm076

51 Illumina. DNA Copy Number and Loss of Heterozygosity Analysis Algorithms. 2013.http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDEQFjAA&url=http%3A%2F%2Fres.illumina.com%2Fdocuments%2Fproducts%2Ftechnotes%2Ftechnote_cnv_algorithms.pdf&ei=XYFWU62wIuKk0QX-yIDgAw&usg=AFQjCNGPmHgIQLQ0YFdD7hya3mv_5lp8TQ&bvm=bv.65177938,d.bGQ&cad=rja (accessed 22 Apr2014).

52 Welcome to the ISCA Consortium web site. https://www.iscaconsortium.org/ (accessed 24 Jul2014).

53 Firth HV, Richards SM, Bevan AP, *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009;**84**:524–33. doi:10.1016/j.ajhg.2009.03.010

54 Edelmann L, Hirschhorn K. Clinical utility of array CGH for the detection of chromosomal imbalances associated with mental retardation and multiple congenital anomalies. *Ann N Y Acad Sci* 2009;**1151**:157–66. doi:10.1111/j.1749-6632.2008.03610.x

55 Primer3Plus. http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi (accessed 5 Aug2014).

56 Hoebeeck J, van der Luijt R, Poppe B, *et al.* Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab Investig J Tech Methods Pathol* 2005;**85**:24–33. doi:10.1038/labinvest.3700209

57 Hoischen A, van Bon BWM, Gilissen C, *et al.* De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;**42**:483–5. doi:10.1038/ng.581

58 Ameur A, Bunikis I, Enroth S, *et al.* CanvasDB: a local database infrastructure for analysis of targeted- and whole genome re-sequencing projects. *Database J Biol Databases Curation* 2014;**2014**. doi:10.1093/database/bau098

59 Exome Variant Server. http://evs.gs.washington.edu/EVS/ (accessed 24 Jul2014).

60 1000 Genomes Project Consortium, Abecasis GR, Auton A, *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65. doi:10.1038/nature11632

61 Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. doi:10.1038/nbt.1754

62 Warde-Farley D, Donaldson SL, Comes O, *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;**38**:W214–220. doi:10.1093/nar/gkq537

63 Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012;**489**:391–9. doi:10.1038/nature11405

64 Microarray Data :: Allen Brain Atlas: Human Brain. http://human.brain-map.org/ (accessed 16 Jun2014).

65 Blake JA, Bult CJ, Eppig JT, *et al.* The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res* 2014;**42**:D810–817. doi:10.1093/nar/gkt1225

66 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;**11**:863–74. doi:10.1101/gr.176601

67 Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9. doi:10.1038/nmeth0410-248

68 Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011;**39**:e118. doi:10.1093/nar/gkr407

69 Schwarz JM, Rödelsperger C, Schuelke M, *et al.* MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;**7**:575–6. doi:10.1038/nmeth0810-575

70 Ferrer-Costa C, Gelpí JL, Zamakola L, *et al.* PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinforma Oxf Engl* 2005;**21**:3176–8. doi:10.1093/bioinformatics/bti486

71 González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011;**88**:440–9. doi:10.1016/j.ajhg.2011.03.004

72 Pollard KS, Hubisz MJ, Rosenbloom KR, *et al.* Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;**20**:110–21. doi:10.1101/gr.097857.109

73 Davydov EV, Goode DL, Sirota M, *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010;**6**:e1001025. doi:10.1371/journal.pcbi.1001025

74 Stenson PD, Ball EV, Mort M, *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;**21**:577–81. doi:10.1002/humu.10212

75 Landrum MJ, Lee JM, Riley GR, *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;**42**:D980–985. doi:10.1093/nar/gkt1113

76 Home - PubMed - NCBI. http://www.ncbi.nlm.nih.gov/pubmed (accessed 5 Aug2014).

77 Home - OMIM - NCBI. http://www.ncbi.nlm.nih.gov/omim (accessed 5 Aug2014).

78 Home - Gene - NCBI. http://www.ncbi.nlm.nih.gov/gene (accessed 5 Aug2014).

79 GeneCards - Human Genes | Gene Database | Gene Search. http://www.genecards.org/ (accessed 5 Aug2014).

80 Warde-Farley D, Donaldson SL, Comes O, *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;**38**:W214–220. doi:10.1093/nar/gkq537

81 Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012;**489**:391–9. doi:10.1038/nature11405

82 Microarray Data :: Allen Brain Atlas: Human Brain. http://human.brain-map.org/ (accessed 5 Aug2014).

83 Blake JA, Bult CJ, Eppig JT, *et al.* The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res* 2014;**42**:D810–817. doi:10.1093/nar/gkt1225

84 Mostafavi S, Ray D, Warde-Farley D, *et al.* GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008;**9 Suppl 1**:S4. doi:10.1186/gb-2008-9-s1-s4