De novo variants disrupting the HX repeat motif of ATN1 cause a non-progressive neurocognitive disorder with recognisable facial features and congenital malformations

Elizabeth E. Palmer[#1,2,3,4], Seungbeom Hong[#5], Fatema AlZahrani[6], Mais O Hashem[6], Fajr A. Aleisa[5], Heba M Jalal Ahmed[5], Tejaswi Kandula[1,2], Rebecca Macintosh[1], Andre E. Minoche[3], Clare Puttick[3], Velimir Gayevskiy[3], Alexander P. Drew[3], Mark J. Cowley[3,7], Marcel Dinger[3,7], Jill A. Rosenfeld[8], Rui Xiao[8,9], Megan T. Cho[10], Suliat F. Yakubu[5], Lindsay B. Henderson[10], Maria J. Guillen Sacoto[10], Amber Begtrup[10], Muddathir Hamad[11], Marwan Shinawi[12], Marisa V. Andrews[12], Marilyn C. Jones[13], Kristin Lindstrom[14], Ruth E. Bristol[14] Saima Kayani[15], Molly Snyder[16], María Mercedes Villanueva[17] , Angeles Schteinschnaider[17], Laurence Faivre[18,19], Christel Thauvin[18], Antonio Vitobello[18], Tony Roscioli[1,20,21], Edwin P Kirk[1,2,20], Ann Bye[1,2], Jasmeen Merzaban[5], Lukasz Jaremko[5], Mariusz Jaremko[5], Rani K Sachdev[1,2], Fowzan S Alkuraya[$6,17,22], Stefan T. Arold[$5]

1. Sydney Children's Hospital, Randwick, NSW 2031, Australia

2. School of Women's and Children's Health, University of New South Wales, Randwick, NSW 2031, Australia

3. The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia

4 Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW 2298, Australia

5 King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Division of Biological and Environmental Sciences and Engineering (BESE), Thuwal, 23955-6900, Saudi Arabia

6 Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia

7. St Vincent's Clinical School, University of New South Wales, Darlinghurst, NSW 2010, Australia

8. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA

9. Baylor Genetics, Houston, Texas, 77021, USA

10. GeneDx, Gaithersburg, Maryland 20877, USA

11. King Khalid University Hospital, King Saud University, Riyadh 11472, Saudi Arabia

12. Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, USA

13. Division of Genetics, Department of Pediatrics, University of California, San Diego and Rady Children's Hospital, San Diego, California 92123, USA

14. Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, Arizona, 85016, USA

15. University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA

16. Department of Neurology, Children's Health, Dallas, Texas, 75235, USA

17. Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia, Montañeses, Buenos Aires, 2325, Argentina.

18 INSERM U1231, LNC UMR 1231 GAD, Burgundy University, Dijon, 21079, France

19 Reference Center for Developmental Anomalies, Department of Medical Genetics,

Dijon University Hospital, Dijon, 21079, France

20 New South Wales Health Pathology Genomic Laboratory, Prince of Wales Hospital, Randwick NSW 2031, Australia

21 Neuroscience Research Australia, University of New South Wales, NSW 2031, Australia

22 Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia

Correspondence: falkuraya@kfshrc.edu.sa, stefan.arold@kaust.edu.sa

ADDITIONAL FOOTNOTES $ # These authors contributed equally to this work.

ABSTRACT

Polyglutamine expansions in the transcriptional co-repressor Atrophin-1, encoded by *ATN1,* cause the neurodegenerative condition dentatorubral-pallidoluysian atrophy (DRPLA) *via* a proposed novel toxic gain of function. We present detailed phenotypic information on eight unrelated individuals with *de novo* missense and insertion variants within a conserved 16 amino acid 'HX repeat' motif of ATN1. Each of the subjects had severe cognitive impairment and hypotonia, a recognisable facial gestalt and variable congenital anomalies. However, they lack the progressive symptoms typical of DRPLA neurodegeneration. To distinguish this subset of affected individuals from the DRPLA diagnosis, we suggest using the term CHEDDA (congenital hypotonia, epilepsy, developmental delay, digit abnormalities) to classify the condition. CHEDDA-related variants alter the particular structural features of the HX repeat motif, suggesting that CHEDDA results from perturbation of the structural and functional integrity of the HX repeat. We found several non-homologous human genes containing similar motifs of eight to ten HX repeat sequences, including *RERE* where disruptive variants in this motif have also been linked to a separate neurocognitive-congenital anomalies condition. These findings suggest that perturbation of the HX motif may explain other Mendelian human conditions.

REPORT

The combination of unbiased chromosomal analysis (chromosomal microarray) and more recently next generation sequencing approaches (exome and whole genome sequencing: ES/WGS) combined with the use of databases that promote sharing of information on genotype and phenotype is enabling the identification and validation of genetic conditions and improved diagnostic rates for complex congenital conditions[1]. Such unbiased genetic approaches can also unveil the complexity of how different types of genetic variation in a particular gene can result in varied and sometimes distinct phenotypic presentations[2-6].

ATN1 (MIM: 607462), located at chromosome 12p13.31, comprises ten exons and has two transcript variants (NM_001007026.1 and NM_001940.3) that differ only in their untranslated exons. *ATN1* encodes the Atrophin-1 protein (ATN1), a member of a class of evolutionarily conserved transcriptional corepressors involved in nuclear signalling[7]. The normal roles of ATN1 are incompletely understood; however, converging evidence supports its role as a nuclear transcriptional regulator important in the control of brain and other organ system development[8-10]. Although *Atn1*$^{-/-}$ mice are neurologically normal[8], Zhang et al.,[9] demonstrated that knockdown of the atrophin-1-like gene in rat neuronal progenitor cells (NPC) led to significant abnormalities in brain development, which could be largely rescued by co-transfection with a human *ATN1* construct. That study also demonstrated that ATN1 is a direct target of the lysine-specific histone demethylase 1A (LSD1), a protein known to have key developmental roles including controlling embryonic stem cell differentiation, cortical neuronal migration and adult NPC proliferation[9]. *ATN1* transcripts are widely expressed, including in brain, heart, lung, kidney and skeletal muscle: expression is higher in fetal tissues, especially in the brain[11]. In the human adult brain, *ATN1* is broadly expressed in multiple regions, including the amygdala, corpus callosum, hippocampus, hypothalamus,

caudate nucleus, substantia nigra, subthalamic nucleus, and thalamus, consistent with ATN1 playing a role in central nervous system development and function[11].

The only human condition definitively associated with *ATN1* to date is the autosomal dominant neurodegenerative condition dentatorubral-pallidoluysian atrophy (DRPLA: MIM #125370)[12; 13] caused by a polyglutamine expansion in exon 5. DRPLA is characterised by the progressive neurological features of choreoathetosis, myoclonus, epilepsy, ataxia and dementia. Age of onset ranges between infancy to late adulthood, dependent on size of the expansion[14; 15]. Congenital anomalies are not a feature. The underlying pathogenic mechanism whereby polyglutamine expansion of ATN1 causes DRPLA is incompletely understood: a toxic gain of function effect of the expanded polyglutamine tract causing neurotoxicity rather than simple loss of function is postulated. These toxic effects may include formation of peri and intranuclear inclusions, abnormal protein cleavage or abnormal phosphorylation of ATN1, and downstream suppression of cAMP response element-binding protein (CREB)-dependent transcriptional activation, which is required for neuronal plasticity and survival[16-19].

We report here a cohort of eight individuals affected with overlapping severe primarily neurocognitive phenotypes, all of whom harboured *de novo* variants in a specific and highly evolutionarily conserved and invariant 16 amino acid motif, consisting of a histidine-rich 16 amino acid motif encoded by exon 7 of *ATN1* (Figure 2). This motif is distal from the Gln-rich region involved in DRPLA, and, the affected individuals lacked the progressive neurodegenerative features characteristic of DRPLA[14].

For all affected individuals concerns arose within the first three months of life regarding significant hypotonia, feeding difficulties, seizures, congenital malformations and distinctive facial features and all have severe to profound global developmental delay/ intellectual disability, truncal hypotonia, global motor disability and very limited verbal communication

(see **Table 1** for an overview of the clinical data, and **Table S2** for further clinical details). Five have a seizure disorder: for four of these, the seizure disorder could be described as a neonatal or infantile onset developmental encephalopathy. Hearing and visual impairments and functional gastrointestinal disorders were common and frequently severe with four individuals requiring orogastric feeding or total parenteral nutrition. Growth parameters were within the normal range other than suboptimal weight gain in those with more significant feeding difficulties. Individual 8 was born prematurely at 33 weeks and died at 2 months of age due to respiratory distress in the setting of severe multiple congenital anomalies. Congenital structural anomalies were common but variable between individuals: four individuals had cardiac malformations including atrial and ventricular septal defects, and abnormalities of the aorta and superior vena cava, two individuals had palatal clefts, three individuals had congenital renal anomalies and two had an anteriorly placed anus. Common neuroanatomical abnormalities were evident on examination of available MRI in one centre (individuals 2, 5 7 and 8) (**Supplementary Figure 1**). Several individuals have cranio-skeletal abnormalities: in particular two individuals had stenosis of the craniocervical junction which prompted screening for this complication in all individuals. When assessed as a group, a similarity in facial features was apparent (**Figure 1**): a particularly striking feature being sparsity of the lateral forehead hair and low-set posteriorly rotated ears. Characteristic hand and feet features were overlapping toes, camptodactyly, persistent fetal fingertip pads and abnormalities of the palmar creases (**Figure 1**).

This clinical cohort was collated by identifying individuals with *de novo ATN1* variants listed in the ClinVar database, as well as *via* contact with individual diagnostic laboratories and networking at Human Genetics conferences. All individuals with a *de novo ATN1* variant were undiagnosed prior to ES or WGS despite clinical genetic assessment and prior genetic screening which in all individuals included chromosomal microarray (further details of prior

genetic studies provided in Supplementary **Table S1**). Individual 5 was previously included in a large exome sequencing study[20] and Individual 8 was previously described clinically, without a molecular diagnosis[21]. All *ATN1* variants in the clinical cohort were rare, in that they were absent from the 125,748 exomes and 15,708 genomes listed in the gnomAD 2.1 database, a database depleted of individuals with severe pediatric disease[22], or from the BRAVO database of 62,785 healthy individuals. All missense variants were predicted to be pathogenic by the majority of *in silico* pathogenicity scoring tools (see **Table 2**). No affected individual had another plausible cause for their neurocognitive condition or congenital anomalies after ES/WGS variant filtering and prioritisation (see **Table S2**). Segregation analysis was consistent with the variant being *de novo* for all individuals with no evidence of mosaicism. Details of the methods for sequencing, variant filtering and prioritization are provided in the **Materials and Methods section** of the supplement and **Supplementary Table S2.** Genetic studies were approved by local ethics committees and written informed consent was obtained from the participants' legal guardians for molecular genetic analysis and for the publication of clinical and radiological data and photographs, which were obtained as part of standard diagnostic procedures.

The histidine-rich motif perturbed in all individuals of our cohort is located in the C-terminal part of ATN1 (residues 1049-1065). It consists of eight HX repeats, where H is a histidine and X is any amino acid (Figure **Figure 2A,B,C)**. The DNA region encoding for this HX repeat motif was covered in the sequencing of all individuals to a depth of at least 20 reads (**Table S2**). A search for the $(HX)_8$ pattern in the human proteome (using PatternSearch[23]) yielded 71 sequences (**Supplementary Data**), 20 of which appeared distinct from ATN1's HX repeat in that the 'X' position only contained histidines or prolines (see PHLDA1, **Figure 2B**). The remaining 51 sequences were isoforms of ATN1 and its paralogous arginine-glutamic acid dipeptide repeats protein RERE (which is also a transcriptional repressor), or

isoforms of the paralogous autism susceptibility gene 2 protein (AUTS2, a component of the PRC1-like complex involved in maintaining the transcriptional repressive state of many genes during development), fibrosin (FBRS) and fibrosin-like proteins (FBRSL1), or belonged to the ZIP family of zinc transporters (**Figure 2**). Within these 51 sequences, the 'X' position showed limited variability, mostly consisting of Gln and Thr (**Figure 2D**).

To probe the impact of a pathological variant on the molecular behavior of the ATN1 protein we studied two ATN1-derived polypeptides containing the HX repeat motif in solution by NMR. The peptides 1046-NVTPHHHQHSHIHSHLHLHQQD-1067 ($ATN1_{1046\text{-}1067}$) and another bearing the variant His1060Tyr 1046-NVTPHHHQHSHIHS**<u>Y</u>**LHLHQQD-1067 ($ATN1_{1046\text{-}1067}^{His1060Tyr}$) were commercially synthesized, and dissolved in 500 μL of 100% D_2O at a concentration of 2.2 mg/mL. Nuclear magnetic resonance (NMR) experiments were performed on a 700 MHz Bruker spectrometer at 25 ˚C and a pD (the pH for D_2O) between 5.51 and 6.4. The NMR data were processed by NMRPipe[24] and analyzed with SPARKY. ^{1}H and ^{13}C resonances for the two peptides were analyzed using standard procedures[25] based on 2D homonuclear 2D ^{1}H-^{1}H TOCSY (with mixing times 10 and 80 ms) and 2D ^{1}H-^{1}H ROESY as well as 2D ^{1}H-^{1}H NOESY (with mixing times 300 and 500 ms) supported by the natural abundance 2D ^{1}H-^{13}C HSQC (separately tuned for the aliphatic and aromatic regions) experiments. For details, see **Supplemental Data**.

For the wild-type $ATN1_{1046\text{-}1067}$ peptide, we found the cross-peaks corresponding to the imidazole Hδ2/Cδ2 and Hε1/Cε1 atoms of all histidines clustered around one position (**Figure 2E**). Moreover, 2D homonuclear ^{1}H -^{1}H ROESY and 2D ^{1}H -^{1}H spectra showed that the trivial (|j-i| = 0) and short range (|j-i| < 2) NOE cross-peaks were absent or of only very weak intensity (**Figure 2F**). Both types of experiments indicated that the regularly-spaced occurrence of histidines introduces a spatial and dynamical synchronization of the histidines.

This synchronization was lost in $ATN1_{1046\text{-}1067}^{His1060Tyr}$, which displayed dispersed cross-peaks of the side-chain aromatic region and stronger intensity of the NOE cross-peaks.

Side-chain imidazole rings of histidines are known to coordinate metal ions[26; 27]. Zn^{2+} binding was assessed by the stepwise addition of $ZnCl_2$ stock solution (500 mM in 100% D_2O) to the solution of peptides reaching the peptide:Zn^{2+} molar ratios of 1:0.5, 1:1, 1:4, 1:8, 1:16, 1:32 and 1:48 respectively. For each peptide:Zn^{2+} ratio the pD was checked and corrected if required, and the same 1D and 2D NMR spectra as for the free peptides were recorded (see **Supplemental Data)**. We found that only the histidines of the His1060Tyr mutant but not of the wild type peptide bound Zn^{2+} at a pD of 5.5, as demonstrated by histidine $^{1}H\beta2,3/^{13}C\beta$ chemical shift and signal intensity changes upon addition of Zn^{2+} (**Figure 2G,H)**. This particular feature was lost upon deprotonation of the histidines at higher pD (**Figure S2A, B)**. Thus, under the conditions used, the HX repeat motif created specific and unusual pH-dependent zinc-binding properties for the histidine-rich sequence, which were abolished by the variant.

Herein, we describe a recognisable constellation of severe neurocognitive impairment, distinctive facial features and pleiotropic but overlapping congenital anomalies, in eight affected individuals with *de novo* variants in the 'HX repeat motif' encoded within exon 7 of *ATN1*. This static (non-progressive) syndromic phenotype is distinct from the neurodegenerative condition of DRPLA, caused by a triplet repeat expansion in exon 5 of *ATN1* which is thought to result in a toxic gain of function. We propose the name CHEDDA (congenital hypotonia, epilepsy, developmental delay, digit abnormalities) to distinguish this previously unreported condition from *ATN1*-related DRPLA. Our clinical observations are consistent with ATN1 having a role as a key nuclear transcriptional regulator involved in the regulation of organ development, including the brain and heart[9; 10; 28], and hint at a critical role of the ATN1 HX repeat motif in the control of human embryonic development. That the

variants cause a simple haploinsufficiency of *ATN1* is unlikely, given the presence of an (albeit very small) number of healthy individuals with heterozygous stop gain, frameshift and canonical splice variants in gnomAD and BRAVO, and the clustering of the variants in CHEDDA within a specific restricted protein motif.

The HX repeat motif was only briefly described in 1991 in a purely bioinformatics-based 'hypothesis' publication[29]. The authors reported the presence of $(HX)_n$ repeat motifs (where n is the number of times the motif is repeated) in certain *Drosophila* transcription factors, and suggested that this motif might be used to coordinate zinc binding. Interestingly, the wild-type ATN1 HX repeat sequence showed a strong pH dependency for zinc binding *in vitro*. This feature appears to be linked to the regular spacing of the histidines which introduces a specific synchronization of the histidine side chains. The introduction of the His1060Tyr variant present in an affected individual in the cohort disrupted this synchronization, and allowed zinc binding, as expected for poly-histidine motifs, thus endowing ATN1 with a novel property albeit with unclear molecular consequences. On a molecular level, the ATN1 HX repeat motif appears therefore to give rise to specific features that distinguish it from other poly-histidine motifs. The *ATN1* HX repeat may serve as a specific pH-dependent interaction motif for ions and/or proteins or other biomolecules.

That disruption in the spacing of the histidines in HX domains will affect critical functioning is supported by our observation that nine of the 19 individuals published with Neurodevelopmental Disorder with or without anomalies of the Brain, Eye and Heart (NEDBEH MIM:#616975:) have *de novo* variants disrupting the HX motif of RERE and that those individuals with variants in the HX motif, as opposed to the rest of the protein, are more likely to have congenital anomalies including septal cardiac, eye and brain anomalies[30]. We also note that three rare *de novo* variants in the HX domain of AUTS2 are listed in

ClinVar as likely pathogenic (SCV000571291.3; SCV000837721.1 and SCV000493076.1) and occurred in individuals with an intellectual disability and congenital anomaly phenotype.

A possible link between dysregulation of *ATN1* expression and another congenital syndromic neurocognitive condition, Pallister Killian syndrome (PKS, mosaic tetrasomy 12p MIM: # 601803), has also been previously postulated by Kaur et al.,[31]. *ATN1* lies within the PKS critical region on 12p13.31, and dysregulation of *ATN1* expression, amongst other genes, was demonstrated in fibroblasts from affected individuals with PKS. Kaur *et al.,* postulated that *ATN1* overexpression may be a key driver of the phenotype of PKS *via* dysregulation of the key developmental *HOX* genes through the action of the master transcriptional regulator *CREBBP*, although direct evidence was lacking. This speculation is intriguing, given certain similarities in phenotype between individuals in our cohort and individuals affected by PKS, including severe cognitive impairment, hypotonia, distinctive facial features including high forehead and sparse fronto-temporal hair at birth (the latter represents a relatively rare clinical finding), and variable congenital anomalies including high arched or cleft palate, polymicrogyria, limb and genitourinary anomalies and congenital heart defects[32]. Indeed, PKS was considered as a differential diagnosis in individual 7 of our cohort.

The *de novo* variants in the HX motif of ATN1 reported here account for 1/6100 individuals with a neurological phenotype who had exome sequencing through Baylor Genetics and 5/13,640 individuals with neurodevelopmental delay who had exome sequencing through GeneDx, including another affected individual with the variant NM_001007026.1(ATN1):c.3178C>T (p.His1060Tyr) (ClinVar accession number SCV000620232.1) whose family did not give consent for inclusion of clinical data. This data suggests a frequency of CHEDDA between 1.6 - 3.7 $x10^{-4}$ individuals with neurocognitive/neurological disorders. Despite this apparent rarity of CHEDDA, we postulate more affected individuals may have already had a variant detected in this region through

diagnostic ES/WGS, but the lack of the progressive neurological phenotype characteristic of DRPLA may have led to the variants being unreported or classified as variants of uncertain clinical significance. This is a similar situation to other clinically distinct conditions our groups have recently described[2-4], and highlights genotype-phenotype complexity and the importance of rigorous evaluation of the possible pathogenicity of unreported variants through international clinical and basic science collaborations[33; 34]. To assist with the dissemination of accessible information about CHEDDA to clinicians and families of affected individuals we have adopted *ATN1* on the Human Disease Gene Webseries.

Further work is required to clarify the biological role of the HX repeat motif in the ATN1 and AUTS2 protein families, and to understand how this function may be altered in CHEDDA and potentially linked conditions such as PKS. Such work may lead to the development of targeted therapies: for example, Zhang et al. demonstrated that the clinical LSD1 inhibitor, tranylcypromine, could suppress *ATN1* expression, and suggested that this agent may have potential therapeutic implications for conditions resultant from aberrant *ATN1* expression[33; 35]. Further studies into the primary gene regulatory functions of ATN1, and how this may be altered in CHEDDA and potentially linked conditions such as PKS, are warranted.

DECLARATION OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGEMENTS

The authors thank the affected individuals and their families for participation in this study and the clinical neurology, general paediatric, and clinical genetics teams involved with their

clinical care and support. We thank Toby Baldwin, Toni Saville and Michael Buckley at SEALS genetics laboratory for their technical assistance.

The research was supported by the National Health and Medical Research Council (GNT11149630 to E.E.P.; GNT0512123 to T.R.), Office for Health and Medical Research, NSW Health (to E.E.P.) King Abdulaziz City for Science and Technology (13-BIO1113-20 and 15-BIO3688-20 to F.S.A.), King Salman Center for Disability Research (FSA), Saudi Human Genome Program (F.S.A.), and the King Abdullah University of Science and Technology through baseline funds (S.T.A., S.H., J.M., H.A., F.A., L.J., M.J.) and the Award No. FCC1/1976-25 form the Office of Sponsored Research (S.T.A.).

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing done at Baylor Genetics Laboratory.

WEB RESOURCES

The URLs for data presented herein are as follows:

BRAVO, https://bravo.sph.umich.edu/freeze5/hg38/

CADD, http://cadd.gs.washington.edu

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/

dbSNP, https://www.ncbi.nlm.nih.gov/SNP/

DECIPHER, https://decipher.sanger.ac.uk/

Ensembl, https://www.ensembl.org/index.html

ExAC database, http://exac.broadinstitute.org

gnomAD database, http://gnomad.broadinstitute.org

Human Disease Gene Webseries, http://humandiseasegenes.nl/

Mendelian Inheritance in Man, http://www.omim.org

PROVEAN, http://provean.jcvi.org

SPARKY, https://www.cgl.ucsf.edu/home/sparky

UCSC Genome Browser, http://genome.ucsc.edu

ACCESSION NUMBERS

ClinVar accession numbers:

NM_001007026.1(ATN1):c.3178C>T (p.His1060Tyr): SCV000221666.1;

NM_001007026.1(ATN1):c.3177_3178insGACCTG (p.Ser1059_His1060insAspLeu): SCV000619648.1;

NM_001007026.1(ATN1):c.3177_3178insAACCTG (p.Ser1059_His1060insAsnLeu): SCV000571353.3

NM_001007026.1(ATN1):c.3184C>G (p.His1062Asp):SCV000528073.3

NM_001007026.1(ATN1):c.3160C>A (p.His1054Asn):SCV000678263.1

NM_001007026.1(ATN1):c.3172C>T (p.His1058Tyr):SCV000678264.1

NM_001007026.1(ATN1):c.3188T>G (p.Leu1063Arg): SCV000678265.1

NM_001007026.1(ATN1):c.3185A>G (p.His1062Arg): SCV000853264

REFERENCES

1. Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., den Dunnen, J.T., et al. (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet 100, 695-705.
2. Palmer, E.E., Kumar, R., Gordon, C.T., Shaw, M., Hubert, L., Carroll, R., Rio, M., Murray, L., Leffler, M., Dudding-Byth, T., et al. (2017). A Recurrent De Novo Nonsense Variant in ZSWIM6 Results in Severe Intellectual Disability without Frontonasal or Limb Malformations. Am J Hum Genet 101, 995-1005.
3. Patel, N., Faqeih, E., Anazi, S., Alfawareh, M., Wakil, S.M., Colak, D., and Alkuraya, F.S. (2015). A novel APC mutation defines a second locus for Cenani–Lenz syndrome. Journal of medical genetics, jmedgenet-2014-102850.
4. Aldahmesh, M.A., Mohamed, J.Y., Alkuraya, H.S., Verma, I.C., Puri, R.D., Alaiya, A.A., Rizzo, W.B., and Alkuraya, F.S. (2011). Recessive mutations in ELOVL4 cause ichthyosis, intellectual disability, and spastic quadriplegia. The American Journal of Human Genetics 89, 745-750.
5. Alkuraya, F.S. (2016). Discovery of mutations for Mendelian disorders. Human genetics 135, 615-623.
6. Monies, D., Maddirevula, S., Kurdi, W., Alanazy, M.H., Alkhalidi, H., Al-Owain, M., Sulaiman, R.A., Faqeih, E., Goljan, E., Ibrahim, N., et al. (2017). Autozygosity reveals recessive mutations and novel mechanisms in dominant genes: implications in variant interpretation. Genet Med 19, 1144-1150.
7. Wang, L., and Tsai, C.C. (2008). Atrophin proteins: an overview of a new class of nuclear receptor corepressors. Nucl Recept Signal 6, e009.
8. Shen, Y., Lee, G., Choe, Y., Zoltewicz, J.S., and Peterson, A.S. (2007). Functional architecture of atrophins. J Biol Chem 282, 5037-5044.
9. Zhang, F., Xu, D., Yuan, L., Sun, Y., and Xu, Z. (2014). Epigenetic regulation of Atrophin1 by lysine-specific demethylase 1 is required for cortical progenitor maintenance. Nat Commun 5, 5815.
10. Wood, J.D., Nucifora, F.C., Jr., Duan, K., Zhang, C., Wang, J., Kim, Y., Schilling, G., Sacchi, N., Liu, J.M., and Ross, C.A. (2000). Atrophin-1, the dentato-rubral and pallido-luysian atrophy gene product, interacts with ETO/MTG8 in the nuclear matrix and represses transcription. J Cell Biol 150, 939-948.
11. Onodera, O., Oyake, M., Takano, H., Ikeuchi, T., Igarashi, S., and Tsuji, S. (1995). Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expanded alleles in the CNS. Am J Hum Genet 57, 1050-1060.
12. Nagafuchi, S., Yanagisawa, H., Ohsaki, E., Shirayama, T., Tadokoro, K., Inoue, T., and Yamada, M. (1994). Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). Nat Genet 8, 177-182.
13. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6, 9-13.
14. Hasegawa, A., Ikeuchi, T., Koike, R., Matsubara, N., Tsuchiya, M., Nozaki, H., Homma, A., Idezuka, J., Nishizawa, M., and Onodera, O. (2010). Long-term disability and prognosis in dentatorubral-pallidoluysian atrophy: a correlation with CAG repeat length. Mov Disord 25, 1694-1700.
15. Veneziano, L., and Frontali, M. (1993). Drpla. In GeneReviews((R)), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, andA. Amemiya, eds. (Seattle (WA).
16. Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nat Med 10 Suppl, S10-17.
17. Nucifora, F.C., Jr., Ellerby, L.M., Wellington, C.L., Wood, J.D., Herring, W.J., Sawa, A., Hayden, M.R., Dawson, V.L., Dawson, T.M., and Ross, C.A. (2003). Nuclear localization of a non-

caspase truncation product of atrophin-1, with an expanded polyglutamine repeat, increases cellular toxicity. J Biol Chem 278, 13047-13055.
18. Okamura-Oho, Y., Miyashita, T., Nagao, K., Shima, S., Ogata, Y., Katada, T., Nishina, H., and Yamada, M. (2003). Dentatorubral-pallidoluysian atrophy protein is phosphorylated by c-Jun NH2-terminal kinase. Hum Mol Genet 12, 1535-1542.
19. Fischbeck, K. (1997). Kennedy disease. Journal of inherited metabolic disease 20, 152-158.
20. Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H., Patel, N., Hashem, M., Ibrahim, N., and Abdulwahab, F. (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Molecular psychiatry 22, 615.
21. Mosca, A.L., Laurent, N., Guibaud, L., Callier, P., Thauvin-Robinet, C., Mugneret, F., Huet, F., Grimaldi, M., Gouyon, J.B., Sandre, D., et al. (2007). Polymicrogyria, cerebellar vermis hypoplasia, severe facial dysmorphism and cleft palate: a new syndrome? Eur J Med Genet 50, 48-53.
22. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
23. Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol 430, 2237-2243.
24. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277-293.
25. Wüthrich, K. (1986). NMR with proteins and nucleic acids. Europhysics News 17, 11-13.
26. Witkowska, D., Rowinska-Zyrek, M., Valensin, G., and Kozlowski, H. (2012). Specific poly-histidyl and poly-cysteil protein sites involved in Ni2+ homeostasis in Helicobacter pylori. Impact of Bi3+ ions on Ni2+ binding to proteins. Structural and thermodynamic aspects. Coordination Chemistry Reviews 256, 133-148.
27. Remelli, M., Brasili, D., Guerrini, R., Pontecchiani, F., Potocki, S., Rowinska-Zyrek, M., Watly, J., and Kozlowski, H. (2017). Zn(II) and Ni(II) complexes with poly-histidyl peptides derived from a snake venom. Inorganica Chimica Acta.
28. Zhang, C.L., Zou, Y., Yu, R.T., Gage, F.H., and Evans, R.M. (2006). Nuclear receptor TLX prevents retinal dystrophy and recruits the corepressor atrophin1. Genes Dev 20, 1308-1320.
29. Janknecht, R., Sander, C., and Pongs, O. (1991). (HX)n repeats: a pH-controlled protein-protein interaction motif of eukaryotic transcription factors? FEBS Lett 295, 1-2.
30. Fregeau, B., Kim, B.J., Hernandez-Garcia, A., Jordan, V.K., Cho, M.T., Schnur, R.E., Monaghan, K.G., Juusola, J., Rosenfeld, J.A., Bhoj, E., et al. (2016). De Novo Mutations of RERE Cause a Genetic Syndrome with Features that Overlap Those Associated with Proximal 1p36 Deletions. Am J Hum Genet 98, 963-970.
31. Kaur, M., Izumi, K., Wilkens, A.B., Chatfield, K.C., Spinner, N.B., Conlin, L.K., Zhang, Z., and Krantz, I.D. (2014). Genome-wide expression analysis in fibroblast cell lines from probands with Pallister Killian syndrome. PLoS One 9, e108853.
32. Izumi, K., and Krantz, I.D. (2014). Pallister-Killian syndrome. Am J Med Genet C Semin Med Genet 166C, 406-413.
33. Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Rost, H.L., Teruel, M.N., and Barna, M. (2017). Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. Mol Cell 67, 71-83 e77.
34. Galan, A., Lozano, G., Pineiro, D., and Martinez-Salas, E. (2017). G3BP1 interacts directly with the FMDV IRES and negatively regulates translation. FEBS J 284, 3202-3217.
35. Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 145, 383-397.

36. Thomsen, M.C., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res 40, W281-287.
37. Ng, P.C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research 31, 3812-3814.
38. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One 7, e46688.
39. Quang, D., Chen, Y., and Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 31, 761-763.
40. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2018). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res.
41. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17, 405-424.

FIGURES

Figure 1. Clinical images of affected individuals with CHEDDA

A: Facial images of affected individuals 1-7. Common facial features include tall foreheads with bitemporal narrowing, deep set eyes, sparsity of the lateral forehead hair, low-set posteriorly rotated ears, a bulbous, slightly overhanging nasal tip, longer philtrum, prominent columella and thin upper lip.

B: Hand and foot images of affected individuals 1, 2, 3, 5, 6 and 7. Common features include abnormalities of the palmar creases, bulbous endings to the fingers and toes and overlapping toes.

Figure 2. *De novo* variants in *ATN1* affect a highly invariant motif and alter the subcellular localization

A: Schematic overview of ATN1, with amino acid positions of the Gln-repeat region expanded in DRPLA and the HX repeat motif illustrated.

B: Human proteins with their (HX)n repeat motifs, with n ≥ 8. Start and end residue numbers are given and histidines are highlighted. For ATN1, residues found mutated in this study are underlined. The variant substitutions present in affected individuals are indicated above the sequence alignment (green). The arrow head marks the ~~region~~ position of the two-amino acid insertions Asn-Leu (insNL) and Asp-Leu (insDL), observed in two individuals of our cohort. Proteins and database accession numbers are: ATN1: atrophin-1, NP_001007027.1; RERE: arginine-glutamic acid dipeptide repeats protein isoform a, NP_001036146.1; AUTS2: autism susceptibility gene 2 protein, NP_056385.1; FBRS: probable fibrosin-1, NP_001098549.2; FBRSL1: fibrosin-1-like protein, NP_001136113.1; ZIP10: zinc transporter ZIP10 precursor, NP_001120729.1; PHLADA1: pleckstrin homology-like domain family A member 1, NP_031376.3.

C: 3D structural visualization of the ATN1 HX repeat motif. The 3D structure is not derived experimentally, but only chosen to illustrate the localization of the histidines.

D: Amino acid enrichment within the HX regions of the 51 human sequences most similar to the ATN1 HX repeat motif (excluding His-only or His-Pro motifs). Figure was produced with Seq2Logo 2.0[36]

E-F: 2D heteronuclear ^{1}H - ^{13}C HSQC correlation spectra. Peptides were recorded using natural abundance of ^{1}H and ^{13}C in synthesized peptides. ATN1: $ATN1_{1046\text{-}1067}$; ATN1mut: $ATN1_{1046\text{-}1067}^{His1060Tyr}$. C-D: "1:16" indicates a peptide: Zn^{2+} ion ratio of 1:16. pD corresponds to the pH in D_2O.

Table 1. Comparison of clinical features of affected individuals with missense variants in the poly HX domain of *ATN1*.

Abbreviations: AA: aortic arch; abn: abnormal; AED: antiepileptic drug; Ant anus: anteriorly placed anus; AmA: amino acid; ASD: atrial septal defect; bg: background; bl: bilateral; CC: corpus callosum; C+OSA: central and obstructive sleep apnoea; CVI: cortical visual impairment; CoA: coarctation of the aorta; DDH: developmental disorder of the hips; EE: epileptic encephalopathy; F: female; GDD: global developmental delay; GERD: gastroesophageal reflux disease; G tube: gastrostomy tube; HC: head circumference; inf spasms: infantile spasms; inv: inverted; L: length; LV: left ventricle; MEA: multifocal epileptiform activity; MT: monotherapy (i.e. controlled on monotherapy); M: male; mo: month; mod: moderate; movts: movements; N: normal; NA: not available; ND: not done; NK: not known; NEC: acute necrotising enterocolitis; PC: palmar crease; PROM: premature rupture of membranes; OSA: obstructive sleep apnoea; OME: otitis media with effusions PVL: periventricular leukomalacia; sn: sensorineural; SVC: superior vena cava; TPN: total parenteral nutrition; rUTI: recurrent urinary tract infections; r: recurrent; UL: upper limb; ul: unilateral; USS: ultrasound; VUR: vesicoureteric reflux; VSD: ventriculoseptal defect; W: weight; yr: year.

[1] based on transcript NM_001007026.1

[2]Protein ID NP_001007027.1

Case ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8** ***(clinical report Mosca et al., 2007)***
Current age	3 yr	1 year	5 yr	7 yr	9 yr	4 yr	5 yr	2 mo
Gender	M	M	F	F	F	F	F	F
Ethnicity	Argentinian	Hispanic	Hispanic	Hispanic	Saudi	Mexican	Australian	French
Variant cDNA[1]	c.3160C>A	c.3172C>T	c.3177_3178insAACCTG	c.3177_3178insGACCTG	c.3178C>T	c.3184C>G	c.3188T>G	c.3185A>G
AA change[2]	p.His1054Asn	p.His1058Tyr	p.Ser1059_His1060insAsnLeu	p.Ser1059_His1060insAspLeu	p.His1060Tyr	p.His1062Asp	p.Leu1063Arg	p.His1062Arg
Antenatal findings	Increased nuchal translucency (karyotype N)	Oligohydramnios and partial urinary obstruction	Normal antenatal USS	Normal antenatal USS	No	Ambiguous genitalia and cardiac malformation	Normal antenatal USS.	Normal antenatal USS. Breech. PROM.
Gestation	Term	Term	Term	31^{+6} weeks	Term	Term	Term	33 weeks
Growth: birth centiles	L:NA W:2.65kg HC: NA	L:>90% W:50% HC: NA	L:NA W:3% HC:15-50%	L:50-75% W:25-50% HC:50%	L:NA W:NA HC:NA	L:50% W:15% HC:50%	L:15% W:3% HC:15-50%	L:50% W:50% HC:75%
Growth: Current centiles	L:3% W: 3% HC: 50%	L:85% W:40% HC: 45%	L:25% W:85% HC: 60%	L: 5% W: 3% HC: 85%	L:25% W:2% HC:75%	L:30% W:85% HC:20%	L:3-10% W: 3% HC:25%	NK
Infantile hypotonia	+	+	+	+	+	+	+	+
Current neurology	Global hypotonia	Central hypotonia, appendicular spasticity	Global hypotonia	Global hypotonia. Hyperkinetic UL movts	Central hypotonia and appendicular spasticity	Global hypotonia	Global hypotonia	Axial hypotonia and appendicular hypertonia
Overt seizure disorder	+ Lennox Gestaut onset 1yr	- no clinical seizures	-	+ Inf spasms controlled 2 AED	+ controlled MT	+ intractable neonatal onset EE (mixed types)	+ EE onset 7 mo controlled MT	-
EEG	MEA with slow spike and wave and periods of voltage attenuation	Bitemporal epileptiform discharges	Focal theta slowing	Hyps arrhythmia + bg slowing	Diffuse slowing	Diffuse slowing	MEA	Not done
Level of DD/ID	Severe-profound GDD	GDD	Severe GDD	Profound	Severe	Severe-profound	Severe	Severe

Visual impairment	+ Does not fix or follow Corneal leukoma	+ CVI	No	+ CVI	No	+ CVI	+ CVI	+ microphthalmia
Hearing impairment	+	+ bl mod (hearing aides)	+ OME (grommets)	- OME (grommets)	- OME (grommets)	+ bl sn	+ bl mod (hearing aides and grommets)	NK
Verbal ability	Non verbal	Coos	Single words	Non-verbal	Non-verbal	Non-verbal	Babble	None
Gross motor ability	No head control, cannot roll.	Rolls to side	Walks few steps unsupported	Sits with support	Immobile	Sits with support	Sits with support	None
Fine motor ability			Holds small objects	Grasps objects			Grasps objects	
MRI brain	Parenchymal atrophy, Unilateral PVL Left cerebellar hyperintensity	Peri-sylvian polymicrogyria, parenchymal atrophy, thin CC, absent falx cerebri	Normal	Normal	Peri-sylvian polymicrogryri, thin CC, partial absence falx cerebri, parenchymal atrophy	Vermian hypoplasia	Peri-sylvian polymicrogyria, thin CC, absent falx cerebri	Polymicrogyria of the rt Sylvian fissure, vermian hypoplasia, thin CC
MRI cervical spine	ND	Craniocervical stenosis	Craniocervical stenosis	ND	ND	Normal	Normal	ND
Respiratory system	No OSA	+ severe OSA	-	-	+ Asthma	+ ul choanal stenosis. O+CSA (tracheostomy)	+ OSA (CPAP)	++ respiratory distress
Orofacial clefting	- High narrow palate	+ Small hard palate cleft	-	-	-	-	- High, narrow palate	+ Cleft palate and gingiva
GI abnormalities	+ Pyloric hypertrophy Dysphagia GERD	+ GERD NEC	+ Dysphagia constipation	-	+ GERD	+ Dysphagia GERD. Ant anus	+ Dysphagia GERD, constipation	+ Ant anus

Nutrition	Oral feeding	TPN	Self feeds	Oral feeding	G tube	G tube	Oral feeding	Orogastric feeding
Congenital heart disease	-	+ ASD+	-	-	+ VSD+ASD	+ CoA + hypoplasia LV and AA	-	+ large foramen ovale, persistence left SVC
Genitourinary disease	Right renal agenesis	Cryptorchidism, VUR, ul hydrouteronephrosis rUTI	-	- N USS	+ r UTI N USS	- N USS	- N USS	+ Left non-dysplastic renal hypoplasia
Skeletal system	Joint hypermobility Scoliosis	Hip dysplasia	Short trunk DDH Joint laxity	Mild scoliosis	DDH	Sagittal craniosynostosis.	N	Phalangeal hypoplasia
Facial Gestalt	+	+	+	+	+	+	+	+
Hands and feet	Overlapping digits	Overlapping digits	Overlapping digits	Overlapping toes	Abnormal PC	Overlapping toes, single PC, Fetal pads	Overlapping toes, fetal pads,	Overlapping toes, proximally implanted thumbs, single PC
Others						Inv nipples	Inv nipples	

Table 2: *De novo ATN1* variants reported in this clinical cohort.

Case ID	Genomic Location (GRCh37)	Variant cDNA change (NM_001007026.1)	Amino acid change (NP_001007027.1)	Present in gnomAD database?	Segregation	SIFT[37] (score)	PROVEAN[38] (score)	DANN[39] score	CADD[40] score	Classification as per ACMG guidelines[41]
1	NC_000012.11:g.7048286C>A	c.3160C>A	p.His1054Asn	no	*de novo*	Damaging (0)	Damaging (-6.1)	0.994	29.6	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PP3: Pathogenic Supporting (8/8 pathogenic predictions)
2	NC_000012.11:g.7048298C>T	c.3172C>T	p.His1058Tyr	no	*de novo*	Damaging (0)	Damaging (-5.26)	0.992	23	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PP3: Pathogenic Supporting (6/8 pathogenic predictions)
3	NC_000012.11:g.7048303-7048304insAACCTG	c.3177_3178insAACCTG	p.Ser1059_His1060insAsnLeu	no	*de novo*	NA	NA	NA	NA	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PM4: Pathogenic Moderate (in frame variant in *ATN1*, and is not in a repeat region.) PP3: Pathogenic Supporting: 1 pathogenic prediction from GERP (vs no benign predictions).
4	NC_000012.11:g.7048303-	c.3177_3178insGACCTG	p.Ser1059_His1060insAspLeu	no	*de novo*	NA	NA	NA	NA	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage)

	7048304insGACCTG									PM4: Pathogenic Moderate (in frame variant in *ATN1*, and is not in a repeat region.) PP3: Pathogenic Supporting: 1 pathogenic prediction from GERP (vs no benign predictions).
5	NC_000012.11:g.7048304C>T	c.3178C>T	p.His1060Tyr	no	*de novo*	Damaging (0)	Damaging (-5.21)	0.998	28.7	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PP3: Pathogenic Supporting (8/8 pathogenic predictions)
6	NC_000012.11:g.7048310C>G	c.3184C>G	p.His1062Asp	no	*de novo*	Damaging (0)	Damaging (-7.86)	0.992	26.2	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PP3: Pathogenic Supporting (8/9 pathogenic predictions)
7	NC_000012.11:g.7048314T>G	c.3188T>G	p.Leu1063Arg	No	*de novo*	Damaging (0)	Damaging (-5.33)	0.997	29.6	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage)
8	NC_000012.11:g.7048311A>G	c.3185A>G	p.His1062Arg	no	*de novo*	Damaging (0)	Damaging (-6.89)	0.9938	24.1	Variant of uncertain significance. PM2:Pathogenic Moderate (absent GnomAD despite good coverage) PP3: Pathogenic Supporting (8/8 pathogenic predictions)

Supplemental Data

Supplemental Material & Methods

Next generation sequencing and variant filtering and analysis

Pattern Search for $(HX)_8$ motifs in the human proteome

Detailed NMR analysis

Supplemental Figures

Figure S1: Magnetic resonance images of affected individuals in CHEDDA cohort, where available

Figure S2: Supplemental NMR analysis of the HX repeat motif

Figure S3: Evolutionary conservation of HX repeat motif

Supplemental Tables (provided as separate excel files)

Table S1: Detailed clinical and variant data

Table S2: Data on exome/ genome sequencing and variant analysis