Exome Sequencing Fails to Identify the Genetic Cause of Aicardi Syndrome

Caroline Lund^a Pasquale Striano^c Hanne Sørmo Sorte^b Pasquale Parisi^e
Michele Iacomino^d Ying Sheng^b Magnus D. Vigeland^b Anne-Marte Øye^b
Rikke Steensbjerre Møller^{f,g} Kaja K. Selmer^{a,b} Federico Zara^d

^aNational Centre for Rare Epilepsy-Related Disorders, and ^bDepartment of Medical Genetics, Oslo University Hospital, Oslo, Norway; ^cPediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, and ^dLaboratory of Neurogenetics, Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, University of Genoa, 'G. Gaslini' Institute, Genoa, and ^eNESMOS Department, c/o Sant'Andrea Hospital, Faculty of Medicine and Psychology, University of Rome Sapienza, Rome, Italy; ^fDanish Epilepsy Centre, Dianalund, and ^gInstitute for Regional Health Services, University of Southern Denmark, Odense, Denmark

Key Words

Agenesis of corpus callosum · Aicardi syndrome · Whole-exome sequencing · X chromosome

Abstract

Aicardi syndrome (AS) is a well-characterized neurodevelopmental disorder with an unknown etiology. In this study, we performed whole-exome sequencing in 11 female patients with the diagnosis of AS, in order to identify the disease-causing gene. In particular, we focused on detecting variants in the X chromosome, including the analysis of variants with a low number of sequencing reads, in case of somatic mosaicism. For 2 of the patients, we also sequenced the exome of the parents to search for de novo mutations. We did not identify any genetic variants likely to be damaging. Only one single missense variant was identified by the de novo analyses of the 2 trios, and this was considered benign. The failure to identify a disease gene in this study may be due to technical limitations of our study design, includ-

ing the possibility that the genetic aberration leading to AS is situated in a non-exonic region or that the mutation is somatic and not detectable by our approach. Alternatively, it is possible that AS is genetically heterogeneous and that 11 patients are not sufficient to reveal the causative genes. Future studies of AS should consider designs where also non-exonic regions are explored and apply a sequencing depth so that also low-grade somatic mosaicism can be detected.

© 2016 S. Karger AG, Basel

Aicardi syndrome (AS; OMIM 304050) is a rare neurodevelopmental disorder originally characterized by the triad of infantile spasms, agenesis of the corpus callosum, and chorioretinal lacunae [Aicardi et al., 1965]. Despite a well-recognized picture for 50 years, the etiology of AS is still unknown. Considering AS as a monogenic X-linked dominant disease caused by a de novo mutation in a gene in the X chromosome has been a leading hypothesis.

During the past decades, technological advances have revolutionized the field of genetics. Massively parallel sequencing has been used in several studies to identify mutations underlying rare Mendelian disorders [Ng et al., 2010; Guerreiro et al., 2014]. In particular, analysis of trios has been successful in finding de novo mutations in sporadic neurodevelopmental disorders [Epi4K Consortium et al., 2013; Veeramah et al., 2013].

In this study, we applied whole-exome sequencing in 11 Aicardi females in search for a shared disease gene. Furthermore, 2 of the patients in the cohort were analyzed together with their parents, making a detection of de novo mutations possible.

Patients and Methods

Patients

Eleven patients with the diagnosis of AS were identified through a collaborative network involving different European neuropediatric epilepsy centers with the following inclusion criteria: (1) infantile seizures, (2) chorioretinal lacunae, and (3) agenesis of the corpus callosum (total or partial). Additional symptoms supportive of AS were: occurrence of malformations of cortical development, gross hemispheric asymmetry, vertebral and costal abnormalities, microphthalmia as well as other eye structural abnormalities. All patients were unrelated and of European ancestry. Clinical features of the patients are summarized in table 1.

Exome Sequencing

Whole-exome sequencing of 11 patients with AS and 2 pairs of parents (15 individuals) was performed on genomic DNA from whole blood (fig. 1). The exome was enriched by using the Agilent SureSelect Human All Exon Capture kit (50 Mb v.3 and v.5). Sequencing was performed with 100-bp paired-end reads on an Illumina HiSeq2000, yielding an average coverage above 100× (10 samples), and ~40× (5 samples).

Bioinformatic Pipeline

For all 15 samples, the reads were aligned to the reference human genome (hg19) by using Novoalign (Version 2.07.17, <http://www.novocraft.com>). The alignments were refined by the Genome Analysis Toolkit (GATK, Version 2.8-1-g932cd3a) [McKenna et al., 2010; DePristo et al., 2011] and PCR duplicates were removed by Picard (Version 1.74, <http://broadinstitute.github.io/picard>). Eighty-six percent of the reads mapped uniquely to the reference sequence. Joint variant calling was performed on trios and probands by using GATK. Finally, variant annotation was done using ANNOVAR (2012May25) [Wang et al., 2010].

Bioinformatic Analysis of Alleles with Low Frequencies

In order to detect possible somatic mutations, the data from the X chromosome was also scrutinized for variants with a low allele ratio, following the approach described by Tapper et al. [2014]. In this approach, alignment was performed by Novoalign (Version 3.02.00), while PCR duplicates removal and variant calling was done using SAMtools (Version 1.19) [Li, 2011] with a flat distribu-

tion of initial allele frequency spectrum. The variants were filtered on alternate base count (minimum 3), strand bias, base quality bias, mapping quality bias, and end distance bias. The alternate base should be aligned on both strands [Tapper et al., 2014].

Filtering of Variants, de novo Analysis, and Analysis Approach

An overview of analysis approaches is given in figure 1. The variants were considered to be rare if they were absent from or had a frequency <0.01 in the publicly available databases (1000 Genomes Project, ESP, and ExAC) and in our in-house database with variants from 430 exomes. To identify potentially damaging variants, we focused our analysis on nonsynonymous variants, splice site variants, and deletions or insertions causing a frameshift. We excluded variants of low quality and variants in genes known to harbor false positives [Fuentes Fajardo et al., 2012]. For variant filtering and de novo analysis, we used the program FILTUS [Vigeland et al., 2016]. Potential pathogenic variants were prioritized by scores obtained from PolyPhen2, MutationTaster, and GERP.

Results

After analysis of the X chromosome, *SLC7A3* was the only remaining candidate gene containing rare variants with pathogenic predictions in 2 individuals. Validation by Sanger sequencing revealed that one variant was inherited from the father, whereas the other could not be validated by Sanger. To evaluate the quality and reliability of this variant, the exome data in the region surrounding the variant was inspected in the Integrative Genomics Viewer, where the variant was observed in 3/20 reads of good quality (online suppl. fig. 1; see www.karger.com/doi/10.1159/000448367 for all online suppl. material).

In the autosomes, a gene where >5 patients shared rare potentially damaging variants was not identified. In the de novo analysis of the 2 trios, only 1 variant was identified in 1 trio, a heterozygous mutation in neuronal PAS domain protein 2 (*NPAS2*) (NM_002518:c.2465C>T, p.P822L). We could not detect any de novo variants in the other trio, or any de novo variants in the X chromosome. None of the remaining 11 patients had rare potentially pathogenic variants in the *NPAS2* gene. De novo variants of uncertain clinical significance in *TEAD1* and *OCEL1* have recently been reported in Aicardi patients [Schrauwen et al., 2015], so particular analysis of these genes were performed. However, no rare variants were observed in any of our patients.

A complete list of all missense/nonsense/splice variants shared by at least 3 patients and with frequency <1% in 1000GP and ESP6500 is provided in online supplementary table 1.

Table 1. Clinical features of the patients

ID	Age, ys	Onset IS, m	Other seizure types	Brain MRI findings	Ocular findings	Additional findings	Intellectual disability	Current therapy	Mobility outcome	Previous therapy
1	3	2	focal, tonic-clonic	CCA, NPH	CL	–	profound	LTG, VPA	unable to walk	ACTH, VGB, CZP
2	6	4	focal, tonic-clonic	CCA, MCD	CL	–	profound	LTG, ACZ, PB, LEV	unable to walk	ACTH, VGB, CZP, NTZ
3	12	1	focal, tonic-clonic	CCA	CL, OC	–	profound	VPA, LEV	unable to walk	ACTH, VGB, LCM, TPM
4	14	3	focal, tonic-clonic	CCA	CL	–	profound	VPA, CLB, TPM	unable to walk	ACTH, CBZ, VGB, PB, LTG
5	17	1	focal, tonic-clonic	CCA	CL, OC	scoliosis, microphthalmia	profound	VGB, CBZ, PB	unable to walk	ACTH, LCM, TPM
6	7	1.5	focal, tonic	CCA, interhemispheric cysts	CL, OC	scoliosis, vertebral arthrodesis, hemivertebrae	moderate	VPA, CLB	hemiplegic gait	VGB, Pred, TPM
7	13	3	tonic, atypical absences, tonic-clonic	CCA, interhemispheric cysts, brain asymmetry	CL	microphthalmia, microcephaly, scoliosis	severe	VPA, NTZ, VNS, KD	unable to walk	PB, VGB, LTG, TPM, CZP, CLB, STM
8	20	2.5	focal, tonic, tonic-clonic, myoclonia	CCA, enlarged cisterna magna	CL, OC	microphthalmia, microcephaly, scoliosis	severe	VPA, RUF, VNS	unable to walk	CZP, NTZ, ACTH, Pred, VGB, CBZ, KD, LTG, ACZ, OXC, TPM, LEV, TGB, ZNS
9	21	2.5	tonic spasms, myoclonia, atypical absences, tonic-clonic	CCA	CL, OC	scoliosis, microcephaly	severe	VPA, LTG	unable to walk	CBZ, ACTH, NTZ, CLB, CZP, VGB, KD, VNS
10	7	3.25	tonic seizures, tonic spasms	CCA	CL	–	profound	LCM, VNS	unable to walk	TPM, VPA, CZP, VGB, ZNS, Cort
11	2	0.25	focal, tonic seizures, tonic spasms	CCA, NPH, MCD	CL, OC	microcephaly	profound	RUF, VGB, ZNS	able to walk	TPM, PER, LCM

ACTH = Adrenocorticotrophic hormone; ACZ = acetazolamide; CBZ = carbamazepine; CCA = corpus callosum agenesis (total or partial); CL = chorioretinal lacunae; CLB = clonazepam; Cort = hydrocortisone; CZP = clonazepam; ID = patient identity number; IS = infantile spasms; KD = ketogenic diet; LCM = lacosamide; LEV = levetiracetam; LTG = lamotrigine; m = months; MCD = malformation of cortical development; NPH = nodular periventricular heterotopia; NTZ = nitrazepam; OC = optic disc coloboma; OXC = oxcarbazepine; PB = phenobarbital; PER = perampanel; Pred = prednisolone; RUF = rufinamide; STM = sulthiame; TGB = tiagabine; TPM = topiramate; VGB = vigabatrin; VNS = vagus nerve stimulation; VPA = valproate; ys = years; ZNS = zonisamide.

Fig. 1. Overview of patients and the analysis approach.

Discussion

Whole-exome sequencing of 11 affected individuals, including de novo analysis of 2 trios, did not reveal the underlying genetic etiology of AS. Our approach was based on the assumptions that AS is a relatively homogeneous genetic disorder and that the disease-causing variant is a germline point mutation residing in an exon.

The only remaining candidate gene after filtering and analysis was the *SLC7A3* gene. This gene encodes a sodium-independent cationic amino acid transporter, CAT-3, which is a member of the solute carrier family 7 [Verrey et al., 2004]. Studies of CAT-3 expression in brain development are conflicting; however, a recent study reported expression in the mouse brain during embryonic development [Nava et al., 2015]. Variation in the *SLC7A3* gene is sparse, and loss-of-function variants are completely absent in the ExAC database (<http://exac.broadinstitute.org/gene/ENSG00000165349>), demonstrating the importance of a functioning CAT-3 protein. Nonetheless, several patients, including one 46,XY male with a nonlethal phenotype, have been reported to have deletions encompassing the *SLC7A3* gene in the DECIPHER database of copy number variations (<https://decipher.sanger.ac.uk/>). In our study, we found 2 patients with missense variants which are absent in all public databases and which have never previously been described. The substituted amino acids are highly conserved, and in silico predictions suggest severe effects on the function of the protein. However, one of the variants was detected by our low allele frequency approach and could not be validated by Sanger sequencing. As the variant was found in 3 out of only 20 reads, we were unable to conclude whether this was a true somatic variant or a false-positive finding. Thus, whether somatic mutations in this gene are involved in AS remains to be explored, and will probably require a study design with more patients and a sequencing depth that permits detection of somatic mutations with a lower allele frequency.

Trio analysis and detection of de novo mutations has previously been performed in 10 patients with AS [Schrauwen et al., 2015]. In 2 patients, they identified a de novo variant of interest, and suggested *TEAD1* and *OCEL1* as candidate genes for involvement in AS. We investigated our patients for all variants in these genes; however, none of the variants had a frequency below 0.01 in the public databases. The trio analysis in our study retrieved only one de novo variant in the *NPAS2* gene. This variant has a population frequency of 0.003 in the ExAC database and has even been observed in a homozygous state 9 times, making it unlikely that it is monogenically involved in AS.

This study shows that the etiology of AS will most probably not be found by a regular exome sequencing approach of patients. There are several potential explanations for this. One possibility is that AS is genetically heterogeneous and maybe even not X-linked after all. A non-X-linked cause of AS has been suggested previously, partly due to a lack of nonrandom X chromosome inactivation, which would be expected in a disorder with an X-linked genetic defect [Wang et al., 2009]. However, this is not always present in X-linked disorders, such as e.g., Rett syndrome [Ropers et al., 1982; Yilmaz et al., 2007; Morleo and Franco, 2008]. Furthermore, a causal mutation in an autosomal gene with a sex-limited effect that causes AS in females has also been suggested [Prontera et al., 2013]. In the case of an autosomal dominant inheritance combined with strong genetic heterogeneity, our study design would probably fail to detect the genetic origin. Also, additional technical limitations are to be considered. Whole-exome sequencing has limitations regarding coverage [Gilissen et al., 2014] and causative variants outside of the targeted exome; indels, copy number variations, inversions, or translocations can remain undetected. Furthermore, despite our efforts to also discover somatic mutations in our data, our analysis was limited by both the read depth in our study, where only variants with an allele frequency of ~0.10 or higher would be detected, and also by the availability of only DNA from whole blood, in which the mutation may not even be present. Results from a study of somatic mutations in tuberous sclerosis complex showed that a majority of the patients with somatic mosaicism had an allele frequency <0.10 in blood [Tyburczy et al., 2015].

Conclusions

In this study, exome sequencing failed to identify the genetic cause of AS. Future studies should aim at designs where also non-exonic regions are explored and apply a sequencing depth where low-grade somatic mosaicism can be detected.

Acknowledgments

We thank the patients and their families, whose help and participation made this study possible. The financial support of Telethon, Italy (grant GEP14118 to F.Z.) and the South-Eastern Norway Regional Health Authority is gratefully acknowledged.

Statement of Ethics

Parents or guardians of the patients have given informed written consent. The study protocol in each country has been approved by the local IRB.

Disclosure Statement

The authors have no conflicts of interest to disclose.

References

- Aicardi J, Lefebvre J, Lerique-Koechlin A: A new syndrome: spasms in flexion, callosal agenesis, ocular abnormalities. *Electroencephalogr Clin Neurophysiol* 19:609–610 (1965).
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498 (2011).
- Epi4K Consortium; Epilepsy Phenome/Genome Project; Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, et al: De novo mutations in epileptic encephalopathies. *Nature* 501:217–221 (2013).
- Fuentes Fajardo KV, Adams D; NISC Comparative Sequencing Program; Mason CE, Sincan M, et al: Detecting false-positive signals in exome sequencing. *Hum Mutat* 33:609–613 (2012).
- Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, et al: Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511:344–347 (2014).
- Guerreiro R, Brás J, Hardy J, Singleton A: Next generation sequencing techniques in neurological diseases: redefining clinical and molecular associations. *Hum Mol Genet* 23:R47–R53 (2014).
- Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993 (2011).
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303 (2010).
- Morleo M, Franco B: Dosage compensation of the mammalian X chromosome influences the phenotypic variability of X-linked dominant male-lethal disorders. *J Med Genet* 45:401–408 (2008).
- Nava C, Rupp J, Boissel JP, Mignot C, Rastetter A, et al: Hypomorphic variants of cationic amino acid transporter 3 in males with autism spectrum disorders. *Amino Acids* 47:2647–2658 (2015).
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al: Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 42:30–35 (2010).
- Prontera P, Bartocci A, Ottaviani V, Isidori I, Rogaia D, et al: Aicardi syndrome associated with autosomal genomic imbalance: coincidence or evidence for autosomal inheritance with sex-limited expression? *Mol Syndromol* 4:197–202 (2013).
- Ropers HH, Zuffardi O, Bianchi E, Tiepolo L: Agensis of corpus callosum, ocular, and skeletal anomalies (X-linked dominant Aicardi's syndrome) in a girl with balanced X/3 translocation. *Hum Genet* 61:364–368 (1982).
- Schrauwen I, Szelinger S, Siniard AL, Corneveaux JJ, Kurdoglu A, et al: A de novo mutation in *TEAD1* causes non-X-linked Aicardi syndrome. *Invest Ophthalmol Vis Sci* 56:3896–3904 (2015).
- Tapper WJ, Foulds N, Cross NC, Aranaz P, Score J, et al: Megalencephaly syndromes: exome pipeline strategies for detecting low-level mosaic mutations. *PLoS One* 9:e86940 (2014).
- Tyburczy ME, Dies KA, Glass J, Camposano S, Chekaluk Y, et al: Mosaic and intronic mutations in *TSC1/TSC2* explain the majority of TSC patients with no mutation identified by conventional testing. *PLoS Genet* 11:e1005637 (2015).
- Veeramah KR, Johnstone L, Karafet TM, Wolf D, Sprissler R, et al: Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* 54:1270–1281 (2013).
- Verrey F, Closs EI, Wagner CA, Palacin M, Endou H, Kanai Y: CATs and HATs: the SLC7 family of amino acid transporters. *Pflugers Arch* 447:532–542 (2004).
- Vigeland MD, Gjøtterud KS, Selmer KK: FILTUS: a desktop GUI for fast and efficient detection of disease-causing variants, including a novel autozygosity detector. *Bioinformatics* 32:1592–1594 (2016).
- Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164 (2010).
- Wang X, Sutton VR, Eble TN, Lewis RA, Gunaaratne P, et al: A genome-wide screen for copy number alterations in Aicardi syndrome. *Am J Med Genet A* 149A:2113–2121 (2009).
- Yilmaz S, Fontaine H, Brochet K, Grégoire MJ, Devignes MD, et al: Screening of subtle copy number changes in Aicardi syndrome patients with a high resolution X chromosome array-CGH. *Eur J Med Genet* 50:386–391 (2007).