## 1 Expanding the Phenotypic spectrum of *NAA10*-related neurodevelopmental syndrome

## 2 and NAA15-related neurodevelopmental syndrome

- 3
- 4 Gholson J. Lyon<sup>1,2,3\*</sup>, Marall Vedaie<sup>1,#</sup>, Travis Besheim<sup>1,#</sup>, Agnes Park<sup>1,#</sup>, Elaine Marchi<sup>1</sup>, Leah
- 5 Gottlieb<sup>4,5</sup>, Tzung-Chien Hsieh<sup>6</sup>, Hannah Klinkhammer<sup>6,7</sup>, Katherine Sandomirsky<sup>1</sup>, Hanyin
- 6 Cheng<sup>8</sup>, Lois J. Starr<sup>9</sup>, Isabelle Preddy<sup>1</sup>, Marcellus Tseng<sup>1</sup>, Quan Li<sup>10</sup>, Yu Hu<sup>11</sup>, Kai Wang<sup>11</sup>, Ana
- 7 Carvalho<sup>12</sup>, Francisco Martinez<sup>13</sup>, Alfonso Caro-Llopis<sup>14</sup>, Maureen Gavin<sup>2</sup>, Karen Amble<sup>2</sup>, Peter
- 8 Krawitz<sup>6</sup>, Ronen Marmorstein<sup>4,5,15</sup>, Ellen Herr-Israel<sup>2</sup>
- 9
- <sup>1</sup>Department of Human Genetics, New York State Institute for Basic Research in Developmental
- 11 Disabilities, Staten Island, New York, United States of America
- 12 <sup>2</sup>George A. Jervis Clinic, New York State Institute for Basic Research in Developmental
- 13 Disabilities, Staten Island, New York, United States of America
- <sup>14</sup> <sup>3</sup>Biology PhD Program, The Graduate Center, The City University of New York, New York,
- 15 United States of America
- <sup>4</sup>Department of Chemistry, University of Pennsylvania, Philadelphia, United States of America
- <sup>17</sup> <sup>5</sup>Abramson Family Cancer Research Institute, Perelman School of Medicine, University of
- 18 Pennsylvania, Philadelphia, United States of America
- <sup>19</sup> <sup>6</sup>Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Rheinische
- 20 Friedrich-Wilhelms-Universität Bonn, Bonn, Germany
- 21 <sup>7</sup>Institute for Medical Biometry, Informatics and Epidemiology,
- 22 University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität
- 23 Bonn, Bonn, Germany
- 24 <sup>8</sup>HematoLogics Inc. Seattle, Washington, United States of America
- <sup>9</sup>Department of Pediatrics, University of Nebraska Medical Center; Omaha, Nebraska
- <sup>10</sup>Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto,
- 27 Ontario, M5G2C1, Canada

- <sup>11</sup>Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of
- 29 Philadelphia, Philadelphia, PA 19104, United States of America
- 30 <sup>12</sup> Department of Medical Genetics, Pediatric Hospital, Coimbra Hospital and University Centre,
- 31 Coimbra, Portugal
- <sup>13</sup> Unidad de Genetica, Hospital Universitario y Politecnico La Fe, Valencia 46026, Spain
- <sup>14</sup> Grupo de Investigacion Traslacional en Genetica, Instituto de Investigacion Sanitaria La Fe,
- 34 Valencia 46026, Spain
- <sup>15</sup>Department of Biochemistry and Biophysics, Perelman School of Medicine, University of
- 36 Pennsylvania, Philadelphia, United States of America
- 37
- 38 #co-equal second authors
- 39 \* Corresponding author
- 40 E-mail: gholsonjlyon@gmail.com / gholson.j.lyon@opwdd.ny.gov (GJL)
- 41

## 42 Abstract

43 Amino-terminal (Nt-) acetylation (NTA) is a common protein modification, affecting 80% 44 of cytosolic proteins in humans. The human essential gene. NAA10, encodes for the enzyme 45 NAA10, which is the catalytic subunit in the N-terminal acetyltransferase A (NatA) complex, also 46 including the accessory protein, NAA15. The full spectrum of human genetic variation in this 47 pathway is currently unknown. Here we reveal the genetic landscape of variation in NAA10 and 48 NAA15 in humans. Through a genotype-first approach, one clinician interviewed the parents of 49 56 individuals with NAA10 variants and 19 individuals with NAA15 variants, which were added 50 to all known cases (N=106 for NAA10 and N=66 for NAA15). Although there is clinical overlap 51 between the two syndromes, functional assessment demonstrates that the overall level of 52 functioning for the probands with NAA10 variants is significantly lower than the probands with 53 NAA15 variants. The phenotypic spectrum includes variable levels of intellectual disability. 54 delayed milestones, autism spectrum disorder, craniofacial dysmorphology, cardiac anomalies,

seizures, and visual abnormalities (including cortical visual impairment and microphthalmia).
One female with the p.Arg83Cys variant and one female with an *NAA15* frameshift variant both
have microphthalmia. The frameshift variants located toward the C-terminal end of *NAA10* have
much less impact on overall functioning, whereas the females with the p.Arg83Cys missense in
NAA10 have substantial impairment. The overall data are consistent with a phenotypic spectrum
for these alleles, involving multiple organ systems, thus revealing the widespread effect of
alterations of the NTA pathway in humans.

62

# 63 Introduction

64 Targeting 40% of the human proteome, NatA, the major N-terminal acetyltransferase 65 (NAT) complex, acetylates Ser-, Ala-, Gly-, Thr-, Val-, and Cys- N-termini following cleavage of 66 the initiator methionine <sup>1,2</sup>. The canonical human NatA consists of two main subunits, the 67 catalytic subunit N- $\alpha$ -acetyltransferase 10 (NAA10) (Ard1) and the auxiliary subunit NAA15 (Nat1) and engages with a regulatory subunit, HYPK <sup>3-5</sup>, NAA15 function has been linked to cell 68 69 survival, tumor progression, and retinal development <sup>6,7</sup>. N-terminal (Nt-) acetylation (NTA) is 70 one of the most common protein modifications, occurring co- and post-translationally<sup>8</sup>. 71 Approximately 80% of cytosolic proteins are N-terminally acetylated in humans and ~50% in veast<sup>1</sup>, while NTA is less common in prokarvotes and archaea<sup>8</sup>. 72

73 NTA is catalyzed by a set of enzymes, the NATs, which transfer an acetyl group from 74 acetyl coenzyme A (Ac-CoA) to the free  $\alpha$ -amino group of a protein's N-terminus. To date, eight 75 distinct NATs (NatA – NatH) have been identified in metazoan (NatA-F and NatH) and plant 76 (NatG) species that are classified based on different conserved subunit compositions and 77 substrate specificities. NTA has been implicated in steering protein folding, stability or 78 degradation, subcellular targeting, and complex formation <sup>9</sup>. Particularly, Naa10-catalyzed N-79 terminal acetylation has been reported to be essential for development in many species and 80 although NatA is not essential in S. cerevisiae, depletion of Naa10 or Naa15 has strong effects,

81 including slow growth and decreased survival when exposed to various stresses. In addition, it 82 has been recently shown that mice have a compensating enzyme Naa12, which prevents embryonic lethality in the *Naa10* knockouts <sup>10</sup>, but a similar gene has not been found in humans. 83 84 Furthermore, NAA10 was also identified in screens for essential genes in human cell lines <sup>11,12</sup>. 85 Ogden syndrome (OS) was first reported in 2011, and it was given this name due to the location of the first affected family residing in Ogden. Utah. USA <sup>13,14</sup>. In that first family, five 86 87 males were born to four different mothers and all the males died in early infancy, with a range of 88 cardiac and other defects, including lethal cardiac arrhythmias. The underlying genetic defect 89 was characterized: this variant involved a single missense change coding for Ser37Pro in the X-90 linked gene, NAA10, which was found in a second independent family in California, USA with 91 three males that had also died during infancy. The identical variant was recently reported in a 92 third family<sup>15</sup>. There is a S. cerevisiae model for the Naa10 Ser37Pro mutant, in which that 93 variant impairs NatA complex formation and leads to a reduction in both NatA catalytic activity 94 and functionality <sup>16,17</sup>. Furthermore, OS patient-derived cells have impaired *in vivo* NTA of a few 95 NatA substrates <sup>18</sup>.

96 Since the initial discovery of OS in 2011, multiple groups have reported additional variants either in *NAA10* in both males and females<sup>19–27</sup> or in the heterodimeric protein partner 97 98 encoded by NAA15<sup>20,28,29</sup> (although reference limitations at the journal do not allow us to cite all 99 relevant papers, so we are only citing a few here, but see **Supplementary Discussion**). Given 100 that NAA10 and NAA15 are part of the NatA complex, it is very likely that there might be shared 101 phenotypes in individuals with pathogenic variants in either of them. As such, the goal of the 102 present study was for one clinician-scientist (G.J.L.) to meet many families with NAA10 and 103 NAA15 variants directly and prospectively to uncover patterns that might have been missed by 104 the prior retrospective reviews of medical records and/or summaries provided by the referring 105 clinicians. The prospective collection of phenotyping information presented here, through a set 106 of direct interviews with the families, has produced a better overall understanding of the range of

107 organ systems affected and the overall natural history of these conditions. In addition, a subset

108 of these families underwent functional assessment of the probands, with one psychologist (E.I.)

administering the Vineland-3 to all patients coming to the Jervis clinic for developmental

- 110 assessments.
- 111

### 112 Materials and Methods

### 113 Clinical features methodology

114 New families were either self-referred, referred by their clinicians, or referred by social 115 media Facebook group parent managers to the corresponding author (GJL). Consent forms were 116 provided to the families for a research study collecting medical information and, for those families 117 who provided written consent, an in-person assessment or videoconference assessment was 118 performed with each family by G.J.L. The in-person assessments included nine visits by NAA10 119 families from September 2019 to March 2020, but these visits stopped with the onset of the 120 COVID19 pandemic in March 2020. During these nine visits, the research participants were 121 examined by a medical geneticist, a neurologist, a psychiatrist, and an optometrist. Phenotypic 122 information was also obtained from clinical records with varying data available, ranging from a list 123 of the key clinical features to detailed history and examination findings. A separate written 124 informed consent was obtained for publication of photographs.

125 A summary template was formulated and filled in for each research participant, where 126 this template was first completed by a research coordinator (E.M.) and then reviewed by a 127 second person for accuracy. These summaries are available from the corresponding author, as 128 MedRxiv does allow posting of such information, and these summaries will be submitted as 129 Supplementary Information for eventual publication in a peer-reviewed journal. The phenotype 130 information from all published and unpublished cases was uploaded to the Human Disease Gene website series <sup>30</sup> with secondary data quality control completed by a second person as 131 132 well, to ensure accuracy and completeness. Human Phenotype Ontology (HPO) terms were

133 selected in the Human Disease Gene website and the percentages represent the collation of 134 groups of multiple terms falling under one feature archetype. For example, "Motor delay" 135 included all probands listed as having "Gait disturbance", "Fine Motor Delay", "Gross Motor 136 Delay", "Difficulty Standing" or "Inability to stand" entered into their summary. While the intent of 137 this method was to create a more simplified and comprehensive gestalt of the OS patient, the 138 natural limitation to this method is the lack of nuance in distinguishing minute differences in 139 presentation from one patient to another. Another limitation may be the overlap of symptoms 140 falling between different feature archetypes, such as "Poor eye contact" being heavily 141 associated with (but not exclusive to) "autistic behavior", the latter intended to reflect more 142 stereotyped and repetitive behavior. Future research may wish to parse out the fine distinctions 143 between these behaviors for greater clarity. Both pertinent positive and negative features were 144 noted, with phenotypes classified as "pertinent negatives" only if the provided clinical 145 information explicitly stated that a phenotype was denied by the parents, was noted to be 146 absent during physical examinations, or not reported during diagnostic procedures. Clinical 147 features were only reported as pertinent negatives if this was explicitly mentioned in the records. 148 otherwise, these were classified as "unknown" or "not available". The relative prevalence of 149 each phenotype was calculated by dividing the number of individuals positive for the phenotype 150 by the sample size. Percentiles for head circumference, weight and height were calculated 151 using CDC growth charts. On the website, for NAA10, the protein variants were numbered with 152 three significant digits, thus allowing sorting in order from 1 to 235 amino acids; for example, 153 p.Arg83Cys. For NAA15, the cDNA and protein variants were numbered with four digits, also to 154 allow sorting.

155

### 156 Variant identification and Bioinformatics Methodology:

Variants were identified primarily using exome sequencing through clinical diagnostic
 testing. The sequencing kits and technology varied based on the different companies involved,

and the variants of interest were highlighted and verified in the collected clinical diagnostic testreports.

161 For the annotation of ClinVar variants, the combined annotation dependent depletion 162 (CADD) scores, MPC scores, Polyphen-2, SIFT scores, M-CAP score and GERP++ RS were 163 retrieved from the dbNSFP4.3a database. Consensus scoring for "pathogenic" was assigned 164 when CADD-scores were above 23, MPC scores were>=2, GERP++ RS score were >=2, and 165 SIFT, Polyphen and M-CAP all provided supportive evidence for pathogenicity for these 166 variants. The variant was assigned by consensus as benign when more than 3 above 167 approaches provided evidence of benign or tolerance. We also classified variants according to the ACMG Guidelines and the Association for Molecular Pathology (ACMG/AMP) framework <sup>31</sup>. 168 169 For gnomAD frequency, the population maximum frequency is set at 4.066e-06 in Annovar, so a 170 period in that column (Supplementary Table 4) indicates that the allele is less frequent than 171 this. 172

### 173 Molecular Analysis of NAA10 and NAA15 Mutations

Molecular effects of the mutations in NAA10 and NAA15 subunits were discussed based upon the human NatA crystal structure (PDB: 6C9M) and the human NatA/HYPK crystal structure (PDB: 6C95), with the heterodimeric NatA crystal structure being used as a representative model for the prepared figures. The structures were visually inspected using PyMol, which was also used in conjunction with Adobe Illustrator CC to prepare **Figure 4** and **Figure 5**. The discussion was also informed by numerous biochemical and biophysical studies of NAA10 and NAA15 mutant proteins <sup>13,16,18–21,23,32,33</sup>.

# 181 Patient clustering analysis description

medRxiv preprint doi: https://doi.org/10.1101/2022.08.22.22279061; this version posted February 14, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Phenotypes (manually curated HPO terms and identifiers) of patients were extracted using Perl and taken as input of clustering analysis. Resnik similarity score <sup>34</sup> was calculated between any two HPO terms using the Python module ssmpy <sup>35</sup>. Given each pair of patients, the similarity score is formalized as:

186 
$$Sim(P_1, P_2) := avg\left(\sum_{t \in P_1} \left(max_{t' \in P_2} S im(t, t')\right)\right)$$

where  $P_1$  and  $P_2$  represent patient 1 and 2, and t and t' represent HPO terms in patient 1 and 2. 187 188 Similarity score matrix between patients were taken as input of software R. Patient-patient 189 difference were calculated based on 1/Resnik similarity score. Hierarchical clustering analysis 190 was performed using hclust function in R based on patient-patient difference. Hierarchical 191 clustering results was visualized as dendrogram in R. Patients from NAA10 and NAA15 were 192 highlighted using different colors. In addition to tree structure, we visualized the clustering results 193 using multidimensional scaling (MDS) based on patient-patient difference matrix. Specifically, we 194 assume  $x_i = (x_{i1}, ..., x_{in})$  in n-dimensional space for each patient *i* and estimate coordinate matrix 195  $X = (x_1, ..., x_N)$  for all patients by minimizing strain function:

196 
$$Strain(x_1, ..., x_N) = \left(\frac{\sum_{i,j} (b_{ij} - x_i^T x_j)^2}{\sum_{i,j} b_{ij}^2}\right)^{1/2}$$

197 where  $b_{ij}$  are element of the double centralized patient-patient difference matrix B =

198  $-\frac{1}{2}(I - J_n)D_n(I - J_n)$ , where *I* is identity matrix,  $J_n$  is all-one matrix and  $D_n$  is patient-patient 199 distance square matrix. Then, to minimize the strain function above, singular value 200 decomposition is applied to matrix *B*. Patient-coordinate matrix  $X = E_m \Lambda_m^{1/2}$  where  $E_m$  is the 201 matrix of *m* eigenvectors of *B* and  $\Lambda_m$  is the diagonal of *m* eigenvalue of *B*. To visualize the 202 patient clustering results, we reduce dimension of *X* to two by selecting top two largest 203 eigenvalue of  $E_m$ . All process is implemented in R.

## 205 Facial Gestalt Methodology

206 To validate whether two cohorts share a similar facial phenotype, we conducted a statistical analysis in the clinical face phenotype space of GestaltMatcher<sup>36,37</sup>. In this analysis, we analyze 207 208 78 NAA10 patients and 33 NAA15 patients, including the patients reported in this work and the 209 previously published patients in the GestaltMatcher Database (GMDB; 210 https://db.gestaltmatcher.org). The information on all patients performed in this analysis is shown 211 in Supplementary Table 6. The mean pairwise cosine distance between patients from the two 212 cohorts is examined to measure the similarity. We sampled control distributions of mean pairwise 213 distances between two cohorts stemming a) from the same syndrome and b) from two different 214 syndromes (shown in blue (a) and red (b) in **Supplementary Figure 8**. Cohorts were sampled 215 from patients in the GMDB that were not included in the training of GestaltMatcher. Then, a ROC 216 analysis was conducted to derive a threshold to decide whether two cohorts stem from the same 217 or different syndromes. The control distributions and resulting threshold were tuned via a five-fold 218 cross-validation resulting in a final threshold of c = 0.9176. To test the similarity of the two cohorts 219  $C_1$  and  $C_2$ , their mean pairwise cosine distance  $d(C_1, C_2)$  is calculated and compared to c. 220 Additionally, we sample two subcohorts from  $C_1$  and  $C_2$  respectively and calculate their mean 221 pairwise cosine distance for 100 times. If at least 50% of those 100 subsampled comparisons fall 222 above the threshold c, it is considered as evidence for the two cohorts stemming from different 223 syndromes. The approach was tested on the validation fold not used in the derivation of c 224 comprising 66 syndromes. The method correctly detected two cohorts stemming from the same 225 syndrome in 80.30% of the comparisons and correctly identified two cohorts stemming from 226 different syndromes in 91.98% of the comparisons.

227

#### 228 Rating scales

The Vineland-3 Comprehensive Interview Form was administered to the parents by one psychologist (E.I.), and provides norm-referenced scores at three levels: subdomains, domains,

231 and the overall Adaptive Behavior Composite (ABC). Adaptive behavior subdomains make up 232 the most fine-grained score level. The primary norm-referenced scores for the subdomains are 233 v-scale scores, which have a mean of 15 and standard deviation (SD) of 3. The v-scale score 234 for each subdomain is included in the narrative interpretation. Standard scores have a mean of 235 100 and SD of 15. Confidence intervals reflect the effects of measurement error and provide, for 236 each standard score, a range within which the proband's true standard score falls with a certain 237 probability or confidence. The confidence level chosen for the report is the 90% confidence 238 interval. A percentile rank is the percentage of individuals in the age group who scored the 239 same or lower than the proband. For example, a percentile rank of 41 indicates that the 240 examinee scored higher than (or the same as) 41% of the age-matched norm sample.

241

### 242 **Results**

243 An in-person assessment was performed on nine families with NAA10 variants from 244 October 2019 – March 2020 and an additional 26 families with NAA10 variants were interviewed 245 by videoconferencing from April 2020 until September 2020 due to the COVID19 pandemic. All 246 families seen by September 2020 were scheduled for cognitive evaluations by one gualified 247 psychologist (E.I.). For families with NAA15 variants, none of the families were seen in-person 248 prior to pandemic onset and ten families were seen via videoconferencing from April 2020 until 249 September 2020, also with referral for cognitive evaluation. After this time, from October 2020 250 until July 2021, an additional 20 families with NAA10 variants and 9 families with NAA15 251 variants were seen via videoconferencing, and cognitive evaluations are being conducted on an 252 ongoing basis for these families. One other family with a recently deceased child with a NAA10 253 variant provided extensive medical records and consent to publish but was unable to meet for 254 videoconferencing. All summary data and percentages for phenotype data for these and other 255 published (and some unpublished) NAA10 and NAA15 cases was curated at the Human 256 Disease Gene website series <sup>30</sup>, under the gene names NAA10 and NAA15, at the following

links using the "Graph and Chart" tab:

- 258 <u>https://humandiseasegenes.nl/naa10/</u>
- 259 and
- 260 <u>https://humandiseasegenes.nl/naa15/</u>
- 261

The data were downloaded from the website on May 2, 2022; these results are shown in Supplementary Table 1 (an excel file, with tabs for *NAA10* and *NAA15*). The website series is meant to be an open-source venue for such data <sup>30</sup>. World maps showing the location of the

265 cases for *NAA10* and *NAA15* are included in the public-facing portion of the websites.

266

267 NAA10 and NAA15 variants

268 The numbering of individuals and the variants are shown in Figure 1, Supplementary Table 2

269 for *NAA10* and in **Supplementary Table 3** for *NAA15*. The overwhelming majority of individuals

seen were female, where one missense variant (c.247C>T, p.Arg83Cys) in NAA10 occurs more

271 frequently and is most recurrent in females, totaling to 25 females seen in this study. However,

272 consistent with this X-linked disease being more severe in males, one male (Individual 9) with

the p.Arg83Cys missense in NAA10 died at 11 months of age. Other variants were recurrent,

but at a much lower frequency, including p.Tyr43Ser, p.Ile72Tyr, p.Ala87Ser, p.Phe128Leu, and

p.Met147Thr. In this cohort of individuals, novel variants include: p.His16Pro, p.Tyr31Cys,

276 p.Ala104Asp, p.Arg116Trp, p.His120Pro, p.Leu121Val, p.Ser123Pro, p.Phe128Ser



277

Figure 1. Pathogenic variants in *NAA10* and *NAA15* reported herein. Abbreviations: UTR:
untranslated region.

280

(same position as p.Phe128Leu), p.Thr152Argfs\*6, and p.Glu181Alafs\*67. New information was
gathered on every individual reported herein, while a shorter summary was published previously
describing some of these individuals with *NAA10* variants (specifically individuals 8, 11, 13-21,
35-38, 41, 47, and 51)<sup>20</sup>.

285 One child (Individual 1) was found to have two *de novo* variants on the same sequencing 286 reads: [c.22C>T;30C>G] p.[Pro8Ser;Asp10Glu]. The c.022C>T variant affects the first base of 287 exon 2 with possible interference on the splicing of messenger RNA. Clinical studies of the 288 messenger RNA expression in the patient confirmed that mRNA from this allele was not present

289 (**Supplementary Figure 1**). The clinical presentation for this child is very similar to many other

290 individuals with other NAA10 variants, further supporting a haploinsufficiency model in which

291 loss of function of one allele seems to mirror the effects of missense variants in other292 individuals.

293 Most of the *NAA10* variants are *de novo*, with the exception of p.Tyr43Ser in a 294 previously reported family <sup>38</sup>, p.Ile72Thr in two families (one of which, Individual 8, was 295 previously reported <sup>20,23</sup> and Individual 54 with frameshift variant p.Thr152Argfs\*6, with this 296 same p.Thr152Argfs\*6 variant being reported in two other cases <sup>20,39</sup>.

For *NAA15*, further information about the variants is supplied in **Supplementary Table** 3. Two of these individuals (4 and 19) were published previously and referred to as Individuals 14 and 21 in that paper <sup>28</sup>. Almost all variants are *de novo*, with the exception of a splice site variant with an unknown inheritance pattern.

301 As of April 2022, there are 58 putative missense or in-frame deletion, substitution or 302 insertion variants in NAA10 and 64 putative missense or in-frame deletion, substitution or 303 insertion variants in NAA15 submitted to ClinVar, with many of these listed as variants of 304 uncertain significance (Supplementary Table 4). The bioinformatic analyses described in the 305 Methods section reports consensus classification for each variant, with an upgrade to 306 pathogenic for several of these. However, many of these entries come from clinical laboratories, 307 with very little uploaded phenotype information, and future work will involve an attempt to 308 contact and interview these other families. Of course, all of the information in ClinVar is de-309 identified, so this requires contacting the clinical testing laboratories and/or other sources of the 310 data, and then getting referred to the clinicians who are involved in these cases. For now, it is 311 worth at least cataloging what is currently in ClinVar, along with the new bioinformatic 312 annotations.

313

### 314 Clinical Features

315 Although some variants will be discussed in detail below, most individuals with variants 316 involving either *NAA10* or *NAA15* have variable degrees of neurodevelopmental disabilities,

317 including impaired motor abilities (HP:0001270), ID (HP:0001249), impaired verbal abilities 318 (HP:0000750) and autism spectrum disorder (ASD) (HP:0000729) (Table 1). These data include 319 previously published cases with variants in NAA10, so the overall numbers are greater than just 320 the families that were videoconferenced. Due to having only one X-chromosome, the males with 321 variants in NAA10 are usually much more severely affected with cardiac issues to the point where 322 some of these individuals died in infancy. However, it is notable that one male. Individual 7 with 323 p.lle72Thr has no apparent cardiac issues and his level of functioning (as assessed by Vineland-324 3) was better than some of the females. This is in contrast to a different child (Individual 8) with 325 the exact same variant who experienced a sudden cardiac death around 5 years of age and had prior development of a medulloblastoma <sup>20</sup>. In the cohort, the only other individual developing any 326 327 type of cancer was Individual 2 with a p.His16Pro missense in NAA10, with leukemia cured by 328 chemotherapy, although a case report of a child with the same exact variant did not report any 329 leukemia or other cancers <sup>19</sup>. As such, this very low level of cancer with just one case of 330 medulloblastoma and one case of leukemia in this current cohort could be explained simply by 331 coincidence or other unknown genetic causes not yet identified, but, given the literature regarding 332 the involvement of NAA10 perhaps in cancer development <sup>40</sup>, this likely warrants further 333 investigation. Overall, individuals with NAA10 variants more often have much more impaired 334 motor function, including fine motor difficulties, abnormality of movement, motor delay, and 335 hypotonia, in comparison to individuals with NAA15 variants.

- 336
- 337

System	Feature	Prevalence (% positive/(negative + unknown)
Neurologic	<ul> <li>Intellectual disability</li> <li>Delayed speech</li> <li>Motor delay</li> <li>Developmental delay</li> <li>Hypotonia</li> <li>Seizures</li> <li>Corpus callosum hypoplasia</li> <li>Microcephaly</li> <li>Periventricular leukomalacia</li> </ul>	92/95 (96.8%) 91/95 (95.8%) 89/95 (93.7%) 77/95 (81.1%) 57/95 (60.0%) 34/95 (35.8%) 27/95 (28.4%) 25/85 (26.3%) 17/95 (17.9%)
Psychiatric	<ul> <li>Harming behaviors (self or others)</li> <li>Attention deficits</li> <li>Eye contact deficits</li> <li>Autistic behavior</li> <li>Compulsive behavior</li> <li>Impulsive behavior</li> </ul>	37/95 (38.9%) 31/85 (32.6%) 28/95 (29.5%) 26/85 (27.4%) 16/95 (16.8%) 11/95 (11.6%)
Cardiovascular	<ul> <li>Prolonged QT interval</li> <li>Atrial septal defects</li> <li>Ventricular septal defects</li> <li>Hypertrophic cardiomyopathy</li> <li>Arrythmias</li> <li>Murmurs</li> <li>Cardiac arrest</li> </ul>	17/95 (17.9%) 17/95 (17.9%) 13/95 (13.7%) 9/95 (9.5%) 9/95 (9.5%) 8/95 (8.4%) 6/95 (6.3%)
Respiratory	<ul> <li>Neonatal respiratory distress</li> <li>Pneumonia &amp; other recurrent upper respiratory infections</li> <li>Apnea</li> <li>Pulmonary hypertension</li> <li>Bronchiolitis</li> </ul>	14/95 (14.7%) 14/95 (14.7%) 5/95 (5.3%) 5/95 (5.3%) 4/95 (4.2%)
Gastrointestinal	<ul> <li>Feeding difficulties in infancy</li> <li>Gastrostomy tube feeding</li> <li>Nasogastric tube feeding</li> <li>Feeding difficulties after infancy</li> <li>Dysphagia</li> <li>Gastroesophageal reflux</li> <li>Vomiting</li> <li>Bowel incontinence</li> <li>Constipation</li> <li>Diarrhea</li> </ul>	57/95 (60.0%) 12/95 (12.6%) 10/95 (10.5%) 40/95 (42.1%) 27/95 (28.4%) 21/95 (22.1%) 19/95 (20.0%) 16/95 (16.8%) 9/95 (9.5%) 7/95 (7.4%)
Eye abnormalities	•Abnormal eyelashes •Thick eyebrows •Abnormal palpebral fissure •Strabismus •Hypertelorism	42/95 (44.2%) 35/95 (36.8%) 22/95 (23.1%) 18/95 (18.9%) 17/95 (17.9%)

# 338 Table 1. Phenotype feature frequency in Individual with NAA10 variants

Visual abnormalities	•Astigmatism •Myopia •Cortical visual impairment	22/95 (23.1%) 17/95 (17.9%) 13/95 (13.7%)
Facial features	<ul> <li>Ear abnormalities <ul> <li>Low-set</li> <li>Other outer ear abnormalities</li> <li>Middle and internal ear abnormalities</li> </ul> </li> <li>Nose abnormalities <ul> <li>Abnormal nasal bridge</li> <li>Abnormal nasal tip</li> <li>Ala or nares abnormalities</li> </ul> </li> <li>Mouth abnormalities <ul> <li>Abnormal philtrum, vermillion</li> <li>Abnormal palate</li> <li>Teeth and gingival abnormalities</li> </ul> </li> <li>Chin abnormalities</li> </ul>	50/95 (52.6%) 32/95 (33.7%) 35/95 (36.8%) 6/95 (6.3%) 64/95 (67.4%) 39/95 (41.1%) 48/95 (50.5%) 12/95 (12.6%) 69/95 (72.6%) 59/95 (62.1%) 29/95 (30.5%) 29/95 (30.5%)
Metabolic & Endocrine	<ul> <li>Sleep disturbance <ul> <li>Sleep apnea</li> </ul> </li> <li>Abnormality of temperature regulation</li> <li>Slender build</li> <li>Hirsutism</li> <li>Precocious puberty</li> </ul>	28/95 (29.5%) 4/95 (4.2%) 28/95 (29.5%) 25/95 (26.3%) 14/95 (14.7%) 6/95 (6.3%)

339

Various levels of ID are reported in almost all study subjects with available data, including mild, moderate or severe ID, and learning difficulties with or without behavioral issues. The severity of the intellectual disability, when known, for *NAA10* and *NAA15* is shown in **Figure 2** and shows that *NAA10* cases are usually much more severe than *NAA15* cases, which is once again consistent with X-linked inheritance for *NAA10*, along with the fact that NAA10 is the acetyltransferase enzyme, whereas NAA15 is the dimeric binding partner that helps to localize the NatA complex to the ribosome.



348

Figure 2. Severity of Intellectual Disability. A) Individuals with *NAA10* variants (n=95); B)
Individuals with *NAA15* variants (n=57).

351

352 The two male individuals with frameshift variants in NAA10 (Individuals 55 and 56) are 353 generally much higher functioning than the other individuals with missense variants. For example, 354 Individual 56 was reported to be cognitively better than individuals his age (for example, he can 355 count from 1 to 10 and knows almost all the letters of the alphabet and is only 29 months of age). 356 Although it was noted that he is a couple of months behind in gross motor skills (he still cannot 357 run or jump even though he tries. He does attend a typical nursery and the staff report that he is 358 intelligent, and "he always wants to learn something else". His teacher says that while she "has 359 to explain something to individuals in his age 10 times to get them to do it, he already does it the 360 second time". These frameshift variants occur toward the C-terminus of the protein, with Individual 361 56 being the most C-terminal with Glu181Alafs\*67 in NAA10. Therefore, it is likely that some 362 functional N-terminal acetyltransferase enzymatic activity remains.

Most individuals with *NAA10* variants have verbal issues including complete absence of speech, delayed language development, required use of sign language, or other speech difficulties related to articulation, which some parents report as possibly being connected to low muscle tone in the pharyngeal muscles. Some research participants also present with either ASD and/or other behavioral challenges. The birth weight was low ( $\leq$  5th percentile) in a few individuals

and it was noted that some individuals failed to grow, causing their weights to fall below the 5th percentile over time. The percentages for some clinical features are listed in **Table 1** and can be found on the Human Disease Gene Website series.

371 The most notable new information relates primarily to the development and maturation of 372 these individuals, as it has become apparent that there are several hurdles to overcome during 373 development, and that failure to overcome each hurdle significantly impacts the trajectory of 374 development. First and foremost, these individuals must overcome the neonatal lethality of 375 possible cardiac arrythmias and cardiovascular or respiratory complications. This seems to be 376 more common in males, but one female (Individual 3) with a p.Tyr31Cys missense in NAA10 died 377 at 6.5 months with such issues after spending some time in the neonatal intensive care unit 378 (NICU) (see **Supplementary Case** information), which also occurred with male infants in the past. 379 including the male Individual 9 reported herein. The second hurdle involves feeding difficulties in 380 infancy including swallowing food and liquids, as some individuals have severe hypotonia, and 381 ultimately require feeding tube insertions <sup>41</sup>. The third hurdle involves remaining seizure-free, as 382 it seems that the cognitive development of the individuals is markedly delayed if seizures begin 383 and are not well-controlled. Lastly, the fourth hurdle involves learning to walk, as some of the 384 individuals remain wheelchair-bound, due to a lack of coordination and/or an inability to stay 385 balanced when walking, which seems to also be related to low muscle tone. As such, the clinical 386 presentation can be quite variable, as someone who requires tube-feeding, is wheelchair bound, 387 and is suffering from seizures will likely have a far different cognitive outcome than someone who 388 does not experience these setbacks.

389

### **390** Facial features

Although it was previously noted that "a recognizable, regular pattern of dysmorphic
 facial features was not appreciated amongst the cases" <sup>36,37</sup>, this is debatable now that one
 clinician has met and examined these individuals via videoconferencing, as there are commonly

394 seen and sometimes recognizable traits in individuals with NAA10 variants, including thicker 395 eyebrows, long eyelashes, upturned nose, and broad nasal bridge (Table 1, Supplementary 396 **Figure 2-4**). While the facial features in the males and females can also be quite variable. 397 some, including an upturned nose, along with bushy eyebrows and long eyelashes, are guite 398 common. Eye abnormalities were found in almost all individuals, which sometimes included 399 large, prominent eyes and large down-slanting palpebral fissures. Other possible features 400 include delayed closing of anterior and posterior fontanels, presence of a broad and/or 401 prominent forehead, and sparse anterior scalp hair. The most common ear findings were low-402 set ears and large ears, with a broad range of hearing impairment and abnormalities of the outer 403 ear. Nose features included broad nasal bridge, prominent or broad nasal tip, and anteverted or 404 flared nares. Lastly, mouth abnormalities were present in a majority of the individuals with a high 405 palate being the main feature, but there was a range of abnormalities in the vermillion and 406 philtrum, as well as teeth abnormalities with particular issues with eruption and disorganized 407 growth. There were several cases with short chins, protruding upper lip, and microretrognathia. 408 Facial feature abnormalities also vary in many individuals with NAA15 variants 409 (Supplementary Figure 5). The most common are abnormalities of the evebrows (horizontal, 410 thick, or long), long eyelashes, hypertelorism, prominent epicanthal folds, abnormalities of the 411 palpebral fissure, amblyopia, astigmatism, and/or strabismus. Mouth features were also 412 common with abnormalities of the philtrum (long, short, broad, or smooth), teeth, upper lip 413 vermillion, chin, and/or presence of a high palate. Abnormalities of the nose included prominent 414 nasal tip or abnormal nasal bridge. Ear features were least common but presented mostly as 415 outer ear abnormalities or low-set ears.

To investigate further the similarity between the two cohorts based on facial gestalt, we conducted a facial analysis using the GestaltMatcher approach<sup>41</sup>, using all available facial photographs for *NAA10* and *NAA15* patients, including previously published cases (**Supplementary Table 6**). We first utilized tSNE <sup>42</sup> to visualize the distribution of *NAA10* and

420 *NAA15* patients in the two-dimensional space. In **Supplementary Figure 7**, we can see that 421 NAA10 and NAA15 patients do not separate clearly. Moreover, comparing the cohorts NAA10 422 and NAA15 in the clinical face phenotype space resulted in a mean pairwise cosine distance of 423 d(NAA10, NAA15) = 0.9146 between two patients. The mean pairwise distance falls below the 424 threshold c = 0.9176 defined to decide whether two cohorts stem from the same syndrome 425 (mean pairwise cosine distance smaller than c) or from two different syndromes (mean pairwise 426 cosine distance greater than c). A subsampling approach vielded that 67% of sampled 427 subcohorts showed similar phenotypes. Overall, our analysis indicates an overlap in the facial 428 gestalt of NAA10 patients and NAA15 patients (Supplementary Figure 8). 429 430 **Cardiovascular features** 

The majority of females and even some males have no obvious cardiac issues.

However, several of the males had structural anomalies of their hearts, including ventricular

433 septal defect, atrial septal defect, and pulmonary artery stenosis. For those who died,

434 arrhythmias at the time of death included torsade de pointes, premature ventricular contraction

435 (PVC), premature atrial contraction (PAC), supraventricular tachycardia (SVtach), and

436 ventricular tachycardia (Vtach). Death can result from cardiogenic shock following arrhythmia,

437 which was noted in some affected individuals.

Baseline echocardiography is recommended as septal defects have been identified in

439 several cases. Medical (pharmaceutical) management has been employed to slow the

440 progression of congestive heart failure (CHF) when manifesting in some individuals.

441 Electrocardiography (EKG) is also obtained at the time of diagnosis, as fatal dysrhythmias can

442 develop with time. In several cases, the initial EKG has been normal, with evolution of

443 arrhythmias during later periods of life. It may be necessary to evaluate on an annual basis and

444 be vigilant for signs of and symptoms of dysrhythmia/CHF. When present, the arrhythmias have

445 proven difficult to manage, though the full spectrum of anti-arrhythmics has not been tested.

Individuals have responded to episodic cardioversion/defibrillation, but long-term treatment has
not yet been determined. Pacemakers and implantable defibrillators have been used in a few
cases, including in one published family <sup>38</sup>.

In addition to gross anatomical defects that subsequently resulted in arrythmias, EKG analysis of OS patients has demonstrated a large proportion of patients demonstrating an elongated QT interval congruent with previously published literature <sup>38</sup>. Our data shows an equal proportion of patients with elongated QT intervals as patients with atrial septal defects in patients with variants in *NAA10* (**Table 1**). These seem to be well-controlled with pharmacotherapy in these patients and caught routinely with Holter monitor assessment.

455

## 456 **Respiratory features**

457 The vast majority of probands do not have any major respiratory issues, but there are a 458 few isolated cases of interstitial lung disease and/or respiratory distress (**Table 1**). For example, 459 Individual 3 with p.Tyr31Cys missense in NAA10 died in the first year of life with local interstitial 460 lung disease, neonatal respiratory distress, respiratory failure requiring assisted ventilation, 461 lymphangiectasia, and pulmonary fibrosis. This included repeated apnea episodes requiring 462 CPR after one such episode. Radiological CT imaging of the thorax was performed at 6 weeks 463 of age and reported "around glass changes throughout both lungs". The etiology remained 464 uncertain and a lung biopsy reported mild thickening of the interstitium without significant 465 interstitial inflammatory infiltrate, but expansion of the inter-lobular septae with lymphangiectasia 466 and fibrosis. Focal interstitial glycogenosis was confirmed with PAS staining. The Individual 9, a 467 male with p.Arg83Cys missense in NAA10, died around one year of age, also with respiratory 468 arrest, neonatal respiratory distress, respiratory failure requiring assisted ventilation, tachypnea, 469 and interstitial lung disease. Of course, the original OS males also had respiratory distress and frequent pneumonias <sup>13</sup>, seen also in a third family <sup>15</sup>, but these other individuals without primary 470

471 cardiac arrhythmias seem to demonstrate that this respiratory distress can be primarily caused 472 by interstitial lung disease, rather than being entirely secondary to cardiovascular complications. 473 As mentioned above, most individuals did not have any notable respiratory issues, but a 474 recognizable proportion of our cohort did demonstrate a pattern of recurrent respiratory 475 infections similar to the original OS males <sup>13</sup>, with varying severity and frequency. A larger 476 proportion of patients with NAA10 variants had respiratory infections as opposed to individuals 477 with NAA15 variants, and this pattern is replicated with the other respiratory features. A few 478 patients also had neonatal respiratory distress, pulmonary hypertension and/or apnea (such as 479 Individuals 12, 38, 45, 47 and 56). Given the relative lack of respiratory distress in most cases, 480 many clinical tests, such as CT scanning and/or lung biopsy were not clinically indicated, so 481 these data are simply not available. As such, it is not known if these individuals never had any 482 interstitial lung disease or if they perhaps had a mild case which subsequently resolved over 483 time. In this regard, it was written that one female with p.His16Pro in NAA10 needed postnatally 484 to "have an oxygen mask applied at night because of oxygen desaturations", and "because of 485 symptoms resembling interstitial lung disease (chronic tachy-dyspnea, recurrent pneumonia and 486 bronchitis), a lung biopsy was performed at the age of 3 years without revealing any specific findings" <sup>40</sup>. As such, it is possible that some of these individuals might have interstitial lung 487 488 disease neonatally, but that in many cases, this might resolve over time.

489

### 490 **Gastrointestinal features**

The gastrointestinal pathology associated with OS, in order from most to least prevalent, includes feeding difficulties in infancy, dysphagia, GERD/silent reflux, vomiting, constipation, diarrhea, bowel incontinence, and presence of eosinophils on esophageal endoscopy. The percentages for some of these are shown in **Table 1** and/or are available on the Human Disease Gene website series. Additionally, the gastrointestinal symptom profile for individuals with this syndrome has been expanded to include eosinophilic esophagitis, cyclic vomiting

497 syndrome, Mallory Weiss tears, abdominal migraine, esophageal dilation, and subglottic 498 stenosis<sup>41</sup>. A recent analysis including nine G-tube or GJ-tube fed probands demonstrated that 499 G/GJ-tubes are overall efficacious with respect to improvements in weight gain and caregiving 500 <sup>41</sup>. A recommendation has been made that OS individuals not tracking above the failure to thrive 501 (FTT) range past 1 year of age should promptly undergo G-tube placement to avoid prolonged 502 growth failure. If G-tubes are not immediately helping to increase weight gain after insertion. 503 recommendations include altering formula, increasing caloric input, or exchanging a G-tube for 504 a GJ-tube by means of a minimally invasive procedure. 505 506 Growth, including height, weight and head circumference 507 Many of the OS individuals and a few of the individuals with NAA15 variants have low 508 height and weight, as detailed in a recent study <sup>41</sup>. As demonstrated in that study, poor growth

509 could not entirely be explained by inadequate caloric intake, inability to properly chew or 510 swallow, growth hormone deficiency, or low appetite; past the age of 6-12 months, calorie 511 tracking, caloric supplementation, and using feeding therapy to successfully teach individuals 512 how to chew, swallow, and no longer choke on food did not induce adequate weight gain. Low 513 appetite was also not an observable cause of poor growth, as multiple proband parents claimed 514 that their individuals had great appetites, were constantly hungry, and that "desire to eat has 515 never been a problem." Growth hormone deficiency was treated with growth hormone (GH) 516 administration in two probands but these efforts kept these probands in the failure to thrive 517 (FTT) range, despite some slight weight improvements. Effect of GH administration on weight 518 gain among GH-deficient OS individuals with growth failure should be further investigated in 519 subsequent studies.

520

### 521 Endocrine and Metabolic features

522 Some of these features are shown in **Table 1**. Extensive laboratory panels assessing 523 digestive function, screening for metabolic decompensation, and hormonal dysregulation have 524 not been significantly useful in improving the care or outcome for these individuals. Although 525 some parents report that their children are guite lean with thermoregulation abnormalities, there 526 is no metabolic or quantitative data yet to prove this claim. Although evidence from mouse models<sup>43</sup> indicates dysregulation of glucose and insulin levels, the one human proband 527 528 (Individual 46) tested thus far with extensive glucose monitoring over 14 days did not show any 529 major abnormalities. Mouse models also show decreased brown adjpose tissue, which might offer an explanation for the thermoregulation issues <sup>43</sup>, but brown adipose tissue amounts have 530 531 not yet been assessed in affected individuals, particularly as there are logistical challenges with 532 getting children and adults with severe intellectual disability to cooperate with imaging exams 533 and it is very difficult to justify the use of sedation in a research setting. It is therefore up to each 534 individual treating clinician to decide whether to pursue clinical testing to explore whether these 535 individuals are indeed lean and/or have decreased brown adipose tissue.

536 There are occasional findings in particular individuals that may or may not be related to 537 the variants in *NAA10* or *NAA15*. For example, Individual 8 with an *NAA15* variant has been 538 found to have corneal calcifications and persistent serum hypercalcemia, although there is no 539 other laboratory-based evidence for hyperparathyroidism.

540

### 541

### Ophthalmologic and visual features

The majority of *NAA10* individuals present with visual abnormalities, which often include astigmatism, myopia, strabismus, and/or amblyopia (also see **Table 1**). Further, there are a subset of individuals who have exotropia/esotropia, microphthalmia, are blind, and/or have abnormalities to the optic disc or nerve. In *NAA15* individuals, visual abnormalities were present in a small subset as strabismus, amblyopia, astigmatism, and other visual impairment. Many individuals with this syndrome wear eyeglasses and due to this high prevalence of visual

548 impairment, eye examination by ophthalmology is warranted. Further, thirteen individuals with 549 NAA10 variants and one male with p.Asn864Ser in NAA15 presented with cortical or cerebral 550 visual impairment (CVI). CVI can be described as a "verifiable visual dysfunction which cannot 551 be attributed to disorders of the anterior visual pathways or any potentially co-occurring ocular 552 impairment"<sup>44</sup>. CVI affects the processing of visual information, which has a profound impact on 553 their ability to learn and must be thoroughly assessed to cater to the unique needs of each child 554 and their impairment. Generally, most individuals have learning disabilities, and particularly 555 seem to have difficulty with depth perception and discerning contrast, which may contribute to 556 their difficulty using stairs or walking off of curbs. CVI was not thoroughly investigated for this 557 paper and should be further explored with more examinations of these individuals to understand 558 how their CVI impacts their learning and/or visuomotor abilities.

559 Individual 23 is a female with the common p.Arg83Cys missense in NAA10, yet she is the 560 only female published to date who was born with microphthalmia, with a very small right eyeball 561 along with a smaller, misshapen pupil, leading to the need for a prosthetic eyeball (see 562 Supplementary Figure 2). She also has stigmatism in her left eye, along with myopia and 563 possible cortical visual impairment. Her exome sequencing did not reveal any other possible 564 cause for the microphthalmia. Individual 54 (a male) with p.Thr152Argfs\*6 in NAA10 also has microphthalmia (see Supplementary Figure 3), similarly to a previously published case <sup>39</sup>. 565 566 whereas the male in Japan did not have microphthalmia <sup>44</sup>. Individual 55 has p.Glu181Alafs\*67 567 and does not have microphthalmia, indicating that even frameshift variants toward the C-terminus 568 of NAA10 can have quite variable outcomes.

After the posting of an earlier version of this manuscript as a preprint <sup>45</sup>, one clinician (A.C.) reported to us the discovery of one female (19 years old, Individual 20 in Supplementary Information) with unilateral microphthalmia with a *NAA15* variant (NM\_057175.4:c.852G>A p.(Trp284\*), heterozygous). This is further proof that the expression of many phenotypes is likely titrated by the overall amounts of monomeric NAA10 and/or NatA activity, whereby

574 microphthalmia occurs in both syndromes when some threshold is crossed during development 575 in which likely some substrate is not acetylated enough.

576

### 577 Neurologic features

578 The prototypical subject displays intellectual disability, global developmental delays, 579 gross and/or fine motor delay, hypotonia, and speech delay. In many cases speech delay was 580 severe enough for speech to be considered absent. Additionally, the presence of tonic-clonic 581 seizures was not uncommon. Virtually all individuals have neonatal hypotonia, and several had 582 neurogenic scoliosis. For Individual 45, intervention for neurogenic scoliosis was attempted with 583 placement of rods. The percentages for some of the neurologic features are shown in **Table 1** 584 and/or are listed on the Human Disease Gene website series.

585 Neuroimaging has been obtained in a number of these individuals, primarily to screen for 586 brain anomalies associated with their hypotonia and growth failure, and there is a paper in 587 preparation regarding detailed analysis of brain MRI findings. The majority possessed at least 588 one brain abnormality with corpus callosum hypoplasia and microcephaly being by far the most 589 common findings followed by periventricular leukomalacia (see Supplementary Table 1 for 590 exact figures). In some cases, there was a finding of cerebral atrophy. However, there was no 591 radiographic evidence to explain their hypotonia or growth failure. It seems warranted to obtain 592 brain MRI with diffusion tensor imaging, to check for white matter delays, in the first few years of 593 life, and then a follow-up MRI around puberty. Given that some of the individuals develop 594 seizures, it seems that baseline EEG should be obtained to check for any seizure activity and, 595 ideally, an annual EEG, or at least one every 2-3 years, should be performed up until their early 596 20's. Prophylactic treatment with anti-epileptic medication should be considered if EEG shows 597 any possible seizures. Lastly, given that some of the individuals have developed autism, it is 598 recommended to perform standard autism screening assessments during childhood, which can

include Autism Diagnostic Inventory (ADI) or Autism Diagnostic Observation Schedule (ADOS-screening.

- 601
- 602 **Cognitive and psychiatric features**

The psychiatric profile of subjects with OS typically includes harmful behavior which is either outwardly directed (agitation or violent behavior toward others) or inwardly directed / selfinjurious (hair pulling, self-biting, head-banging, etc) in addition to impulsive or compulsive behavior. The presence of any combination of short attention span, poor eye contact, and autistic, stereotyped or repetitive behaviors (tics, hand-wringing, hand-flapping, etc) were common. The percentages for some of these features are shown in **Table 1** and/or are available on the Human Disease Gene website series.

610 Vineland-3 functional assessment revealed substantial variability among the NAA10 and 611 NAA15 probands, although the overall level of functioning for the NAA15 probands was 612 significantly higher than the NAA10 probands (Figure 3A, Supplementary Table 5). This is 613 consistent with the results presented previously in **Figure 2** demonstrating that more NAA10 614 probands were given a "severe" intellectual disability diagnosis. When the results for NAA10 615 were separated by sex, inheritance pattern, and variant type, it became clear that the frameshift 616 variants located toward the C-terminal end of NAA10 have much less impact on overall 617 functioning. One of the males (Individual 7) with a maternally inherited missense variant 618 (c.215T>C, p.Ile72Thr) in NAA10 had an almost normal level of functioning, with an ABC 619 standard score of 94, which is less than one standard deviation away from the "normal" of 100. 620 However, Individual 8 (now deceased) had the exact same variant, but his ABC standard score 621 was 59, demonstrating again that there is substantial variability even among the exact same 622 variants. For the females with the p.Arg83Cys missense in NAA10, there is also substantial 623 phenotypic variability with functioning as well, which appears to be equivalent in scope to the 624 many other variants in females (Figure 3B). As such, there are clearly other factors affecting

- 625 the overall trajectory of these individuals, in addition to just the one variant they have in NAA10
- 626 or NAA15, which as mentioned above could include, among other things, the cardiac
- 627 arrhythmias, feeding difficulties, delayed speech, delayed motor development, seizures, and
- 628 cortical visual impairment. Comprehensive analyses of the many different Vineland sub-scale,
- 629 age-equivalent scores (AES), and growth-scale values (GSVs) <sup>46</sup> falls outside the scope of the
- 630 current paper and will be reported in a future paper.
- 631



Figure 3. Vineland results. Adaptive Behavior Composite (ABC) Standard scores have a mean of 100 in the "normal" population and SD of 15. A) All scores are plotted for *NAA10* and *NAA15*, without regard for sex, inheritance pattern, or type of variant. B) These same scores are now separated out by sex, inheritance pattern and type of variant. In the case of *NAA15*, which is an autosomal gene, the sexes are still grouped together.

- 638
- 639
- 640

## 641 Human Phenotype Ontology-based clustering analyses

642 Hierarchical clustering analyses were performed using the HPO terms downloaded from 643 the Human Disease Gene website series in October 2022 (see **Supplementary Table 1**). The 644 two syndromes for NAA10 and NAA15 do cluster separately, albeit with some overlap (see 645 Supplementary Figure 6). Most importantly, and consistent with the clinical impression of 646 these two syndromes, there was no obvious sub-clustering among the different variants within 647 each syndrome. This provides quantitative and supportive data regarding the phenotypic 648 spectrum for the presentation for variants in NAA10 and NAA15, and this is also consistent with 649 the presentation of NAA15-related neurodevelopmental syndrome being somewhat distinct (and 650 usually much more mild) than NAA10-related neurodevelopmental syndrome. 651 652 **Molecular Analysis of the Variants** 653 Most NAA10 variants identified to date in NAA10-related neurodevelopmental diseases reduce NAA10- and NatA-type activity<sup>20</sup>. Although novel variants clinically described here have 654 655 not been characterized biochemically, it is possible to anticipate biochemical implications 656 resulting from variants in the conserved region of NAA10 (residues 1-160) based on previous studies<sup>13,16,18–21,23,32,33</sup>. However, variants and truncations occurring in the C-terminus of NAA10 657 658 require additional studies to understand the role of the NAA10 C-terminal domain. The 659 observation that frameshifts occurring at the C-terminal region of NAA10 are associated with a 660 lower impact on cognitive function is consistent with the conserved NAA10 domain being the 661 minimal region necessary for in vitro activity. 662 To contextualize the phenotypes associated with the NAA10 and NAA15 mutant 663 proteins, we inspected two published human NatA crystal structures: the heterodimeric complex 664 composed of the NAA10 and NAA15 subunits (PDB: 6C9M) and the heterotrimeric complex 665 composed of NAA10, NAA15, and the regulatory subunit, HYPK (PDB: 6C95). Based on these

666 structures, the NAA10 variants p.Pro8Ser/Asp10Glu, p.Tyr31Cys, p.His120Pro, p.Ser123Pro,

667 p.Phe128Ser, and p.Arg149Trp would likely compromise the folding and/or thermal stability of 668 NAA10 (Figure 4) <sup>5</sup>. Specifically, the mutations to proline (p.His120Pro, p.Ser123Pro) occur in 669 helices. likely destabilizing the helical structure, and the p.Tvr31Cvs and p.Phe128Ser 670 mutations likely disrupt hydrophobic core interactions that are mediated by Tyr31 and Phe128. 671 The double variant, [c.22C>T;30C>G] p.[Pro8Ser;Asp10Glu] in the exon/intron boundary, leads 672 to a reduction in NAA10 formation by altering the splicing of the mRNA from this allele 673 (Supplementary Figure 1). 674 The biochemistry of the p.Arg83Cys variant is unique amongst the NAA10-type variants. 675 The monomeric NAA10 p.Arg83Cys protein is less active than wild type NAA10<sup>33</sup> and the 676 heterodimeric p.Arg83Cys NatA complex was found have enhanced activity, while HYPK 677 binding returns p.Arg83Cys NatA activity back to wild type levels <sup>20</sup>. In the context of the 678 heterodimeric NatA complex, the binding of acetyl CoA to p.Arg83Cys NAA10 is compromised, 679 as indicated by the loss of acetyl CoA-mediated stabilization <sup>20</sup>. For the wild type NAA10, the 680 positively charged p.Arg83Cys side chain interacts with the negatively charged acetyl CoA 681 pantothenic arm, which would be absent with the Cys variant. In addition, it is possible that this 682 complex NatA phenotype arises from the formation of a disulfide bond between NAA10 683 p.Arg83Cys with the nearby NAA15 p.Cys322 residue. This disulfide bond has been proposed 684 to anchor the complex so that the complex is better positioned to catalyze the N-terminal 685 acetylation reaction. This disulfide bond would not form in the monomeric NAA10 state. 686 rendering mutant NAA10 p.Arg83Cys less active than the wild type monomeric NAA10. 687

689



Figure 4. Crystal structure of human NaA complex labeled with relevant NAA10 residues.
NAA10 (dark blue) and NAA15 (light cyan) are shown in cartoon with wild-type residues in stick
(orange). A) Top view of complex annotated with relevant NAA10 residues outlined by dashed
boxes: B) P8 and D10; C) Y31; D) H120, S123, and F128; E) R149; and F) R83 and NAA15
C322 (light cyan, stick), where the acetyl CoA binding pocket is indicated by a dash-outlined
circle.

In contrast to the reported *NAA10* variants, *NAA15* variants tend to confer a range of
biochemical and biophysical effects on NatA function <sup>20</sup>. Similar to *NAA10* variants, several
novel variants have been included that have not yet been biochemically characterized
(p.Arg27Gly, p.Ala585Thr, p.Ala719Thr, and p.Asn864Ser). Variants occurring in NAA15 could
affect NAA15 folding and/or thermostability, as well as intermolecular interactions with NAA10
(directly or through a bridging inositol hexaphosphate, IP<sub>6</sub>, molecule) or HYPK – and, thereby,
NatA catalytic activity, or NatA localization to the ribosome.

704 Based on structural considerations of the human NatA and NatA/HYPK crystal structures (Figure 5) and previous *in vitro* studies, <sup>46</sup> we anticipate that the NAA15 p.Lys450Glu mutant 705 706 would disrupt the NAA10-IP<sub>6</sub> interaction and, thus, destabilize the NAA10/NAA15 complex, 707 while the p.Lys338Asn mutant could disrupt NAA15 thermostability by disrupting a hydrogen 708 bond between two adjacent helices in NAA15. p.Ala585Thr is disordered in both crystal 709 structures and mutants p.Arg27Gly and p.Asp112Asn are solvent-exposed on NAA15, so it is 710 unclear how these mutations impact NatA function. However, it is possible that they disrupt the 711 NatA-ribosome interaction. In particular, p.Arg27 and p.Ala585 are both located in regions 712 previously implicated in ribosome-binding, thereby altering NatA co-translational activity <sup>47</sup>. 713 P.Ala719Thr is buried within a bundle of helices in the metazoan-conserved NAA15 C-terminus, 714 which is critical for NatA heterodimer stabilization and HYPK regulatory subunit binding.<sup>47</sup> 715 Finally, the molecular basis for the p.Asn864Ser mutant is unclear because published NatA 716 structures are not resolved past residue 841. However, there are two published phosphorylation 717 sites p.Ser855 and p.Ser856 in this region that have not been evaluated but may be impacted 718 by this variant.



720 Figure 5. Crystal structure of human NatA complex labeled with relevant NAA15 residues.

721 NAA10 (dark blue) and NAA15 (light cyan) are shown in cartoon with wild-type residues in stick

- 722 (orange). A) Side view of complex annotated with relevant NAA15 residues outlined by dashed
- boxes: **B)** K450 with bound inositol hexaphosphate (IP<sub>6</sub>, stick); **C)** K338N; **D)** R27; **E)** D112; and
- 724 **F)** A719.
- 725

719

# 727 Discussion

728 Mendelian diseases are anything but simple, as the presentations can be quite complex 729 and variable due to modification by genetic background and environmental influences <sup>48</sup>. This is 730 quite apparent in the case of NAA10-related neurodevelopmental syndrome, as the variability is 731 enormous, with different trajectories that are impacted by whether the child overcomes certain 732 hurdles, such as learning to feed themselves and to walk. Males with NAA10-related 733 neurodevelopmental syndrome may have symptoms at different ages: those with arrhythmias or 734 other serious cardiac conditions may be embryonic lethal; others are apparent at birth 735 experiencing cardiac concerns, hypotonia and dysmorphic features; and others come to 736 attention later with relatively nonspecific developmental/intellectual and growth impairments. 737 Likewise, some females can present with similar issues as males, including rarely 738 microphthalmia and/or prolonged QT interval, and the majority of the females with NAA10 739 variants have severe to profound intellectual disability, whereas the individuals with NAA15 740 variants are much less severely impaired (see Figure 3). 741 An area for future investigation concerns the amount of physical, speech, occupational, 742 and other therapies that should be recommended. Based on these extensive interviews with the 743 families, the anecdotal information seems to suggest that individuals with NAA10-related 744 neurodevelopmental syndrome benefit from the following therapies on at least a weekly basis in 745 at least the first 7 years of life, so as to ensure that they can eat, walk, and ideally verbalize: 746 Mobility therapy; speech and language therapy; sensory integration therapy; hydrotherapy; 747 occupational therapy with a focus on fine motor skills; and cortical visual impairment therapy (if 748 warranted). The trajectory of these individuals seems to vastly improve with intensive therapy, 749 whereas some of these individuals do not make adequate progress without adequate therapy. 750 As mentioned, future studies can investigate this guestion much more guantitatively with a 751 natural history study, documenting the exact modalities and frequencies of therapies.

752 One limitation of our study is that only the first nine research participants were 753 interviewed in-person and physically examined by medical doctors, including by a medical 754 geneticist, a neurologist, a psychiatrist, and an optometrist. After this time, all interviews were 755 converted to videoconferencing starting in March 2020, with the onset of the COVID19 756 pandemic, and the families met with the psychiatrist only (GJL). However, we did not observe 757 any major differences in the information obtained from the in-person visits as compared to the 758 videoconference visits, and all visits included review of available medical records. Please see 759 the **Supplementary Discussion** for additional details.

760 To date, most NAA10 and NAA15 variants have been studied in vitro using either immunopurified overexpressed full-length NAA10 (235 residues) protein from veast <sup>16</sup> or 761 762 mammalian culture <sup>19,21,23</sup>, S. frugiperda (Sf9)-expressed recombinant NatA complex containing 763 a C-terminally truncated NAA10 (1-160)<sup>20</sup>, or an affinity-tagged full-length NAA10 expressed in 764 *E. coli*<sup>48</sup>. Each of these methodologies have their limitations. Protein isolated from mammalian 765 cell culture offers a unique advantage, where, depending on the immunoprecipitation conditions, NatA likely co-purifies with relevant binding partners, such as the NatA regulatory protein, 766 767 HYPK, and the NatE catalytic subunit, NAA50. However, this does not allow for selective 768 isolation of the NAA10 catalytic subunit or the NatA complex without either of its binding 769 partners. By contrast, Sf9 overexpression of the human NatA complex can be performed with 770 and without the co-transfection of the HYPK virus to allow for selective production of the NatA 771 and NatA/HYPK complexes. In addition, it is also possible to prepare purified NatE complex with 772 or without HYPK<sup>49</sup>. However, this approach has its limitations because the Sf9-expressed NatA 773 complex construct contains a C-terminally truncated NAA10 (1-160 of 235 total residues). This C-terminal region can be phosphorylated at 6 sites <sup>50</sup>, and could potentially influence the activity 774 775 and the stability of NAA10, although the interplay of these phosphorylation events with NAA10 776 and NAA15 variants has not yet been evaluated. While recombinant NAA10 overexpression in 777 E. coli allows for the studying of the full-length monomeric NAA10, proteins produced in E. coli

are not post-translationally modified. Overall, these *in vitro* assays are limited in scope as there
 could very well be cellular and organismal phenotypes that result from decreased (or otherwise
 altered) amino-terminal acetylation of a wide range of substrates.

781 NatA plays a complex role in the cell with a range of functional roles and varying 782 substrate types, depending on the enzyme's oligomeric state. This is particularly apparent with 783 the NAA10 p.Arg83Cys variant. The heterodimeric p.Arg83Cys NatA complex has enhanced 784 activity, HYPK binding returns p.Arg83Cys NatA activity back to wild type levels <sup>20,25,49</sup> while the 785 monomeric mutant NAA10 features a diminished level of activity<sup>20,25,50</sup>. Despite this, the 786 phenotype of the individuals with this particular mutation is not appreciably different from the 787 other individuals reported herein. There are multiple ophthalmologic abnormalities in these 788 females, and individual 23 is a female with the p.Arg83Cys missense in NAA10, with 789 microphthalmia, just like Individual 54 (a male) with p.Thr152Argfs\*6 in NAA10 also with 790 microphthalmia, and similar in phenotype to previously published cases <sup>20,25,51</sup>. The cognitive 791 functioning of the females with p.Arg83Cys missense in NAA10 is also similarly impaired as the 792 other females with different missense changes (Figure 3), and there is a phenotypic spectrum 793 including multiple organ systems in all of these females with missense changes in NAA10. 794 These pathologies highlight the need to build off the groundwork laid by focused biochemical 795 studies into model systems that can account for NatA's multiple oligomeric states and cellular 796 functions. One such multidimensional approach has been to study the human mutations in the 797 human genes expressed in a S. cerevisiae model knocked out for endogenous yeast Naa10 798 and/or Naa15<sup>50</sup>. However, this approach has its own caveats. For example, S. cerevisiae is a 799 much simpler unicellular organism that does not represent the complexity of a human cell or a 800 human being, including the absence of expression of the NatA regulatory subunit, HYPK. 801 Instead, these mutations are perhaps best studied using patient-derived or cell lines with the mutations engineered into them, as has been done recently with NAA15<sup>51</sup>, and/or with animal 802 803 models with the human variants engineered into NAA10 or NAA15.

804 In conclusion, we have presented phenotypic information on many new variants in 805 NAA10 and NAA15, along with additional phenotypic information on some previously published 806 NAA10 variants. We document an extensive phenotypic spectrum caused by likely decreased 807 expression and/or function of NAA10 and/or the NatA complex, with the cases involving NAA10 808 variants usually presenting as more severe than cases involving NAA15. Although some papers 809 have also suggested that there might be different allelic presentations or mechanisms of action 810 for different variants involving NAA10, such as with microphthalmia present in males with splice-811 site <sup>51</sup> or frameshift <sup>53</sup> variants, the present study demonstrates that this is much more likely to 812 be a phenotypic spectrum of one unitary disease. Although some do refer to this entire disease 813 entity as Ogden syndrome, another name could be NAA10-related neurodevelopmental 814 syndrome. The latter name is longer and thus more cumbersome, whereas the name Ogden 815 syndrome is more memorable and easier to introduce when describing the disease to non-816 experts, but the longer name does follow the recently suggested dyadic nomenclature  $5^4$ . 817 Further information about this can be found in **Supplementary Discussion**. We further suggest 818 that the disease entity involving NAA15 variants should be labeled as NAA15-related 819 neurodevelopmental syndrome. 820 821 Acknowledgements 822 The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups 823 that provided exome and genome variant data to this resource. A full list of contributing groups 824 can be found at http://gnomad.broadinstitute.org/about. Early in the project, Drs. Milen Velinov, 825 Ricardo Madrid, and Roseanne Ricciardi graciously performed medical genetic, neurology, and

826 optometry examinations, respectively, on the nine families with NAA10 variants seen in-person

827 at the Jervis Clinic. GJL would also like to thank the Clinical and Administrative staff of the

828 Jervis Clinic for their assistance, and the many clinicians who performed tests and/or wrote

829	medical records that were obtained and summarized. This study has made use of data
830	generated by Human Disease Genes Website series, www.humandiseasegenes.nl.
831	
832	Author Contributions
833	GJL was responsible for all videoconferencing and primary data collection, with secondary
834	summaries performed by EM, MV, TB, AP, KS, IB, and MT. Data analysis was performed by
835	GJL MV, TB, and AP. Bioinformatics analyses were performed by QL and KW. In-person
836	examinations at the Jervis Clinic were aided by MG and KA. Biomedical interpretation of the
837	human NatA structure were provided by LG and RM. Vineland-3 psychological testing was
838	performed by EHI. The first draft of the manuscript was written by GJL, MV, TB, and AP, with
839	input thereafter from all other authors.
840	
841	Funding
842	This work is supported by New York State Office for People with Developmental Disabilities
843	(OPWDD) and NIH NIGMS R35-GM-133408. NIH grants R01 GM060293 and R35 GM118090
844	awarded to R.M. and T32 GM071339 grant awarded to L.G. also supported this work.
845	
846	Ethical Approval
847	Both oral and written patient consent were obtained for research and publication, with approval
848	of protocol #7659 for the Jervis Clinic by the New York State Psychiatric Institute - Columbia
849	University Department of Psychiatry Institutional Review Board. Written family consent was
850	given for publication of any photography of the children.
851	
852	Database Deposition and Access— All data produced in the present study are available upon
853	reasonable request to the authors. The sequencing data were generated as part of clinical
854	testing, so the underlying raw data are not consented for deposition to a public database.

## 855

### 856 **Competing Interests**

- The authors do not declare any competing interests.
- 858

## 859 **References**

- 860 1. Arnesen, T. *et al.* Proteomics analyses reveal the evolutionary conservation and divergence
- 861 of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. U. S. A.*
- 862 **106**, 8157–8162 (2009).
- 863 2. Starheim, K. K., Gevaert, K. & Arnesen, T. Protein N-terminal acetyltransferases: when the
  864 start matters. *Trends Biochem. Sci.* **37**, 152–161 (2012).
- 3. Arnesen, T. *et al.* Identification and characterization of the human ARD1-NATH protein
  acetyltransferase complex. *Biochem. J.* **386**, 433–443 (2005).
- 4. Arnesen, T. *et al.* The chaperone-like protein HYPK acts together with NatA in
- 868 cotranslational N-terminal acetylation and prevention of Huntingtin aggregation. *Mol. Cell.*
- 869 *Biol.* **30**, 1898–1909 (2010).
- 5. Gottlieb, L. & Marmorstein, R. Structure of Human NatA and Its Regulation by the
- 871 Huntingtin Interacting Protein HYPK. *Structure* **26**, 925-935.e8 (2018).
- 872 6. Arnesen, T. et al. Characterization of hARD2, a processed hARD1 gene duplicate,
- encoding a human protein N-alpha-acetyltransferase. *BMC Biochem.* **7**, 13 (2006).
- 874 7. Gendron, R. L. *et al.* Loss of tubedown expression as a contributing factor in the
- 875 development of age-related retinopathy. *Invest. Ophthalmol. Vis. Sci.* **51**, 5267–5277
- 876 (2010).
- 877 8. Dörfel, M. J. & Lyon, G. J. The biological functions of Naa10 from amino-terminal 878 acetylation to human disease. *Gene* **567**, 103–131 (2015).
- 9. Aksnes, H., Ree, R. & Arnesen, T. Co-translational, Post-translational, and Non-catalytic
- 880 Roles of N-Terminal Acetyltransferases. *Mol. Cell* **73**, 1097–1114 (2019).

- 10. Kweon, H. Y. *et al.* Naa12 compensates for Naa10 in mice in the amino-terminal acetylation
  pathway. *Elife* **10**, (2021).
- 11. Blomen, V. A. et al. Gene essentiality and synthetic lethality in haploid human cells.
- 884 Science **350**, 1092–1096 (2015).
- 12. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- 13. Rope, A. F. *et al.* Using VAAST to identify an X-linked disorder resulting in lethality in male
  infants due to N-terminal acetyltransferase deficiency. *Am. J. Hum. Genet.* **89**, 28–43
- 889 (2011).
- 890 14. Lyon, G. J. Personal account of the discovery of a new disease using next-generation

891 sequencing. Interview by Natalie Harrison. *Pharmacogenomics* **12**, 1519–1523 (2011).

- 15. Gogoll, L. et al. Confirmation of Ogden syndrome as an X-linked recessive fatal disorder
- due to a recurrent NAA10 variant and review of the literature. *Am. J. Med. Genet. A* 185,
  2546–2560 (2021).
- 895 16. Van Damme, P., Støve, S. I., Glomnes, N., Gevaert, K. & Arnesen, T. A Saccharomyces

896 cerevisiae model reveals in vivo functional impairment of the Ogden syndrome N-terminal

- 897 acetyltransferase NAA10 Ser37Pro mutant. *Mol. Cell. Proteomics* **13**, 2031–2041 (2014).
- 898 17. Dörfel, M. J. *et al.* Proteomic and genomic characterization of a yeast model for Ogden
  899 syndrome. *Yeast* 34, 19–37 (2017).

Myklebust, L. M. *et al.* Biochemical and cellular analysis of Ogden syndrome reveals
downstream Nt-acetylation defects. *Hum. Mol. Genet.* 24, 1956–1976 (2015).

- 902 19. Bader, I. *et al.* Severe syndromic ID and skewed X-inactivation in a girl with NAA10
- 903 dysfunction and a novel heterozygous de novo NAA10 p.(His16Pro) variant a case report.
- 904 BMC Med. Genet. **21**, 153 (2020).

- 905 20. Cheng, H. *et al.* Phenotypic and biochemical analysis of an international cohort of
- 906 individuals with variants in NAA10 and NAA15. *Hum. Mol. Genet.* (2019)
- 907 doi:10.1093/hmg/ddz111.
- 908 21. McTiernan, N. et al. NAA10 p.(N101K) disrupts N-terminal acetyltransferase complex NatA
- 909 and is associated with developmental delay and hemihypertrophy. *Eur. J. Hum. Genet.* **29**,
- 910 280–288 (2021).
- 911 22. Ree, R. et al. A novel NAA10 p.(R83H) variant with impaired acetyltransferase activity
- 912 identified in two boys with ID and microcephaly. *BMC Med. Genet.* **20**, 101 (2019).
- 913 23. Støve, S. I. et al. A novel NAA10 variant with impaired acetyltransferase activity causes
- 914 developmental delay, intellectual disability, and hypertrophic cardiomyopathy. *Eur. J. Hum.*
- 915 Genet. **26**, 1294–1305 (2018).
- 916 24. Afrin, A. *et al.* NAA10 variant in 38-week-gestation male patient: a case study. *Cold Spring*
- 917 Harb Mol Case Stud 6, (2020).
- 918 25. Johnston, J. J. *et al.* NAA10 polyadenylation signal variants cause syndromic
- 919 microphthalmia. J. Med. Genet. jmedgenet-2018 (2019).
- 920 26. Gupta, A. S., Saif, H. A., Lent, J. M. & Couser, N. L. Ocular Manifestations of the NAA10-
- 921 Related Syndrome. *Case Rep. Genet.* **2019**, 8492965 (2019).
- 922 27. Maini, I. et al. Clinical Manifestations in a Girl with NAA10-Related Syndrome and
- 923 Genotype–Phenotype Correlation in Females. *Genes* **12**, 900 (2021).
- 924 28. Cheng, H. et al. Truncating Variants in NAA15 Are Associated with Variable Levels of
- 925 Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies. *Am. J. Hum.*
- 926 Genet. **102**, 985–994 (2018).
- 927 29. Ward, T. *et al.* Mechanisms of congenital heart disease caused by NAA15
- 928 haploinsufficiency. *Circ. Res.* **128**, 1156–1169 (2021).

- 929 30. Dingemans, A. J. M. *et al.* Human disease genes website series: An international, open and
- 930 dynamic library for up-to-date clinical information. *Am. J. Med. Genet. A* 185, 1039–1046
  931 (2021).
- 932 31. Richards, S. *et al.* Standards and guidelines for the interpretation of seguence variants: a
- 933 joint consensus recommendation of the American College of Medical Genetics and
- 934 Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- 935 32. McTiernan, N. et al. Biochemical analysis of novel NAA10 variants suggests distinct
- 936 pathogenic mechanisms involving impaired protein N-terminal acetylation. *Hum. Genet.*
- 937 (2022) doi:10.1007/s00439-021-02427-4.
- 938 33. Saunier, C. *et al.* Expanding the Phenotype Associated with NAA10-Related N-Terminal
- 939 Acetylation Deficiency. *Hum. Mutat.* **37**, 755–764 (2016).
- 34. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. *arXiv [cmp-lg]* (1995).
- 942 35. Couto, F. M. & Lamurias, A. Semantic Similarity Definition. in *Encyclopedia of*
- 943 Bioinformatics and Computational Biology 870–876 (Elsevier, 2019).
- 36. Hsieh, T.-C. *et al.* GestaltMatcher facilitates rare disease matching using facial phenotype
  descriptors. *Nat. Genet.* (2022) doi:10.1038/s41588-021-01010-x.
- 946 37. Hustinx, A. *et al.* Improving Deep Facial Phenotyping for Ultra-rare Disorder Verification
- 947 Using Model Ensembles. in *Proceedings of the IEEE/CVF Winter Conference on*
- 948 Applications of Computer Vision 5018–5028 (2023).
- 949 38. Casey, J. P. et al. NAA10 mutation causing a novel intellectual disability syndrome with
- 950 Long QT due to N-terminal acetyltransferase impairment. *Sci. Rep.* **5**, 16022 (2015).
- 951 39. Shishido, A. *et al.* A Japanese boy with NAA10-related syndrome and hypertrophic
- 952 cardiomyopathy. *Hum. Genome Var.* **7**, 23 (2020).
- 953 40. Chaudhary, P., Ha, E., Vo, T. T. L. & Seo, J. H. Diverse roles of arrest defective 1 in cancer
- 954 development. Arch. Pharm. Res. **42**, 1040–1051 (2019).

- 955 41. Sandomirsky, K., Marchi, E., Gavin, M., Amble, K. & Lyon, G. J. Phenotypic variability and
- 956 Gastrointestinal Manifestations/Interventions for growth in Ogden syndrome (also known as
- 957 NAA10-related Syndrome). *medRxiv* (2022) doi:10.1101/2022.03.16.22272517.
- 958 42. Gmail, L. @. & Hinton, G. Visualizing Data using t-SNE.
- 959 https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl (2008).
- 960 43. Lee, C.-C. et al. Naa10p inhibits beige adipocyte-mediated thermogenesis through N-α-
- 961 acetylation of Pgc1α. *Mol. Cell* **76**, 500-515.e8 (2019).
- 962 44. Chokron, S., Kovarski, K. & Dutton, G. N. Cortical Visual Impairments and Learning
- 963 Disabilities. *Front. Hum. Neurosci.* **15**, 713316 (2021).
- 964 45. Lyon, G. J. *et al.* Expanding the Phenotypic spectrum of Ogden syndrome (NAA10-related
- 965 neurodevelopmental syndrome) and NAA15-related neurodevelopmental syndrome.
- 966 *medRxiv* (2022) doi:10.1101/2022.08.22.22279061.
- 967 46. Farmer, C. A. *et al.* Person Ability Scores as an Alternative to Norm-Referenced Scores as
- 968 Outcome Measures in Studies of Neurodevelopmental Disorders. *Am. J. Intellect. Dev.*
- 969 Disabil. **125**, 475–480 (2020).
- 970 47. Magin, R. S., Deng, S., Zhang, H., Cooperman, B. & Marmorstein, R. Probing the
- 971 interaction between NatA and the ribosome for co-translational protein acetylation. *PLoS*
- 972 One **12**, e0186278 (2017).
- 973 48. Lyon, G. J. & O'Rawe, J. Human genetics and clinical aspects of neurodevelopmental
- 974 disorders. in *The Genetics of Neurodevelopmental Disorders*. (ed. Mitchell, K.) (Wiley),
- 975 2015).
- 976 49. Deng, S., McTiernan, N., Wei, X., Arnesen, T. & Marmorstein, R. Molecular basis for N977 terminal acetylation by human NatE and its modulation by HYPK. *Nat. Commun.* **11**, 818
  978 (2020).
- 979 50. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation
  980 site occupancy during mitosis. *Sci. Signal.* **3**, ra3 (2010).

- 981 51. Esmailpour, T. et al. A splice donor mutation in NAA10 results in the dysregulation of the
- 982 retinoic acid signalling pathway and causes Lenz microphthalmia syndrome. J. Med. Genet.
- 983 **51**, 185–196 (2014).
- 984 52. Wu, Y. & Lyon, G. J. NAA10-related syndrome. *Exp. Mol. Med.* 50, 85 (2018).
- 985 53. Sidhu, M., Brady, L., Tarnopolsky, M. & Ronen, G. M. Clinical Manifestations Associated
- 986 With the N-Terminal-Acetyltransferase NAA10 Gene Mutation in a Girl: Ogden Syndrome.
- 987 *Pediatr. Neurol.* **76**, 82–85 (2017).
- 988 54. Biesecker, L. G. et al. A dyadic approach to the delineation of diagnostic entities in clinical
- 989 genomics. Am. J. Hum. Genet. **108**, 8–15 (2021).
- 990