

1
2
3 **NAA10 polyadenylation signal variants cause syndromic**
4
5
6 **microphthalmia**
7

8 Jennifer J. Johnston^{1*}, Kathleen A. Williamson², Christopher M. Chou^{3^}, Julie C. Sapp¹, Morad
9
10 Ansari^{2,4}, Heather M. Chapman⁵, David N. Cooper⁶, Tabib Dabir⁷, Jeffrey N. Dudley¹, Richard J.
11
12 Holt⁸, Nicola K. Ragge^{8,9}, Alejandro A. Schäffer^{10,11}, Shurjo K. Sen¹², Anne M. Slavotinek¹³, David
13
14 R. FitzPatrick², Tom M. Glaser⁵, Fiona Stewart⁷, Graeme C.M. Black^{14,15}, Leslie G. Biesecker^{1**}
15
16
17
18
19
20

- 21 1. National Human Genome Research Institute, National Institutes of Health, Bethesda,
22 MD, USA
- 23
- 24 2. MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU UK
- 25
- 26 3. Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI,
27 USA
- 28
- 29 4. DNA Diagnostic Laboratory, South East Scotland Regional Genetics Services, Western
30 General Hospital, Edinburgh EH4 2XU UK
- 31
- 32 5. Department of Cell Biology and Human Anatomy, U.C. Davis, Davis, CA, USA
- 33
- 34 6. Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK
- 35
- 36 7. Northern Ireland Regional Genetics Service (NIRGS), Belfast City Hospital, Belfast,
37 Northern Ireland, UK
- 38
- 39 8. Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK
- 40
- 41 9. West Midlands Regional Clinical Genetics Service and Birmingham Health Partners,
42 Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1
2
3 10. Computational Biology Branch, National Center for Biotechnology Information, NIH,
4
5 Bethesda, Maryland, USA
6
7
8 11. Cancer Data Science Laboratory, National Cancer Institute, NIH, Bethesda, MD, USA
9
10
11 12. Leidos Biomedical Research, Inc, Basic Science Program, Cancer & Inflammation
12
13 Program, Frederick National Laboratory for Cancer Research, Bethesda, Maryland
14
15 20892, USA
16
17
18 13. Department of Pediatrics and Institute for Human Genetics, University of California San
19
20 Francisco, San Francisco, CA, USA
21
22
23 14. Manchester Centre for Genomic Medicine, Manchester Academic Health Sciences
24
25 Centre, Manchester University NHS Foundation Trust, St Mary's Hospital, Manchester,
26
27 UK
28
29
30 15. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of
31
32 Biology, Medicine and Health, University of Manchester, Manchester, UK
33
34
35
36
37

38 Corresponding Authors

39
40 *Jennifer J. Johnston, PhD

41
42 50 South Drive Room 5139, Bethesda, MD 20892

43
44
45 Tel: 301-594-3981 Fax: 301-402-2170

46
47 Email: jjohnsto@mail.nih.gov
48

49
50 **Leslie G. Biesecker, MD

51
52 50 South Drive Room 5140, Bethesda, MD 20892

53
54
55 Tel: 301-402-2041 Fax: 301-402-2170
56
57
58
59
60

1
2
3 Email: lesb@mail.nih.gov
4

5
6 ^Current affiliation, Kaiser-Permanente Health System, Sacramento, CA, USA
7

8 Word Count: 4,000
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 CONFLICT OF INTEREST PAGE
4

5 LGB receives royalties from Genentech Corp, is an advisor to the Illumina Corp, received
6
7 honoraria from Wiley-Blackwell and receives honoraria from Cold Spring Harbor Press. LGB, JJJ,
8
9 JND, and JCS received support from the Intramural Research Program of the National Human
10
11 Genome Research Institute grants HG200328 11 and HG200388 03. RJH and NKR were
12
13 supported by grants from Baillie Gifford, Visually Impaired Children Taking Action (VICTA)
14
15 (<http://www.victa.org.uk/>), and Microphthalmia, Anophthalmia, Coloboma Support (MACS)
16
17 (www.macs.org.uk). AAS received support from the Intramural Research Program of the
18
19 National Library of Medicine. DNC received funding from Qiagen Inc. through a License
20
21 Agreement with Cardiff University. AMS received support from the extramural research
22
23 program of the National Human Genome Research Institute grant 5U01HG009599 02, and
24
25 received honoraria from Wiley-Blackwell, Inc., Oxford University Press and UptoDate, Inc. TMG
26
27 and HMC received support from the National Institutes of Health grant EY19497 and HMC
28
29 received support from the Hartwell Foundation. DRF and KAW were funded via the MRC
30
31 University Unit award to the University of Edinburgh for the MRC Human Genetics Unit. MA
32
33 was partially funded by the MRC IGMM Translational Science Initiative.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **Abstract**
4

5 **Background:** A single variant in *NAA10* (c.471+2T>A), the gene encoding N-acetyltransferase 10,
6
7
8 has been associated with Lenz microphthalmia syndrome. In this study, we aimed to identify
9
10 causative variants in families with syndromic X-linked microphthalmia.
11

12 **Methods:** Three families, including 15 affected individuals with syndromic X-linked
13
14
15 microphthalmia, underwent analyses including linkage analysis, exome sequencing and
16
17
18 targeted gene sequencing. The consequences of two identified variants in *NAA10* were
19
20
21 evaluated using quantitative PCR and RNAseq.
22

23 **Results:** Genetic linkage analysis in family 1 supported a candidate region on Xq27-28, which
24
25
26 included *NAA10*. Exome sequencing identified a hemizygous *NAA10* polyadenylation signal
27
28 (PAS) variant, chrX:153,195,397T>C, c.*43A>G, which segregated with the disease. Targeted
29
30
31 sequencing of affected males from families 2 and 3 identified distinct *NAA10* PAS variants,
32
33
34 chrX:g.153,195,401T>C, c.*39A>G and chrX:g.153,195,400T>C, c.*40A>G. All three variants
35
36
37 were absent from gnomAD. Quantitative PCR and RNAseq showed reduced *NAA10* mRNA levels
38
39
40 and abnormal 3' UTRs in affected individuals. Targetted sequencing of *NAA10* in 376 additional
41
42
43 affected individuals failed to identify variants in the PAS.
44

45 **Conclusion:** These data show that PAS variants are the most common variant type in *NAA10*-
46
47
48 associated syndromic microphthalmia, suggesting reduced RNA is the molecular mechanism by
49
50
51 which these alterations cause microphthalmia/anophthalmia. We reviewed recognized variants
52
53
54 in PAS associated with Mendelian disorders and identified only 23 others, indicating that *NAA10*
55
56
57 harbors more than 10% of all known PAS variants. We hypothesize that PAS in other genes
58
59
60 harbor unrecognized pathogenic variants associated with Mendelian disorders. The systematic

1
2
3 interrogation of PAS could improve genetic testing yields.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Congenital anophthalmia and microphthalmia result from failure of early eye development. One-third of affected individuals are syndromic[1],[2] and it is genetically heterogeneous.[3] X-linked syndromic anophthalmia and microphthalmia have been shown to result from pathogenic variants in four genes: *BCOR*[4] (MIM:300485, MCOPS2), *HCCS*[5] (MIM:300056, MCOPS7), *HMGB3*[6] (MIM:300193, MCOPS13), and *NAA10*[7] (MIM:309800, MCOPS1). A fifth locus, MCOPS4 (MIM:301590, also designated as ANOP1), with linkage to Xq27-q28, was specified without identification of an associated gene. *HMGB3* and *NAA10* reside on Xq27-Xq28. A single *HMGB3* variant has been reported in one family to cause syndromic colobomatous microphthalmia.[6] Forrester et al. described a family segregating what they termed Lenz microphthalmia syndrome (LMS) and linkage to Xq27-q28.[8] Affected individuals had severe microphthalmia, renal anomalies, high-arched palate, cutaneous syndactyly of the hands, and severe intellectual disability (ID). A c.471+2T>A splice variant in *NAA10* was identified as causative, the only study to date associating *NAA10* variants with syndromic microphthalmia.[7]

The MCOPS4 locus was originally defined based on a family reported by Graham et al.[9] Their manifestations included microphthalmia/anophthalmia and ID in multiple affected males. Here, we re-analyze this family using genetic linkage analysis and exome sequencing, verified our findings in two additional families, and functionally characterized the variants.

Materials and Methods

This study was approved by Institutional Review Boards at the National Institutes of Health, the UK Multicentre Research Ethics Committee (MREC), the University of Michigan, the University

1
2
3 of California, Davis, and the Huntingdon Ethics Committee, UK. All DNA analyses were
4
5 performed using standard techniques.
6

7 **Linkage and haplotype analyses**

8
9
10 Resampling and repeat genetic linkage analyses were performed in family 1 (Figure 1A) using 30
11
12 microsatellite markers on the X chromosome. Initial analyses of 22 markers spanning the X
13
14 chromosome were followed by fine mapping with an additional eight markers from the region
15
16 near Xqtel. LOD scores were computed using FASTLINK using equal marker allele
17
18 frequencies.[10], [11], [12] For multi-marker analyses, inter-marker recombination fractions
19
20 were derived from the Rutgers map.[13] The disease locus was modeled as X-linked recessive.
21
22 Analyses were performed under two additional possible scenarios, first that the individual I-2
23
24 was gonadal mosaic and second that penetrance was less than 100%. Haplotype analysis was
25
26 performed assuming a model of X-linked recessive inheritance.
27
28
29
30
31

32 **Sequence analyses**

33
34 The coding region and consensus splice sites of two candidate genes, *NAA10* and *HMGB3*, were
35
36 analyzed in family 1 using Sanger sequencing. Exome sequencing was performed on individual
37
38 IV-3 and his parents at the NIH Intramural Sequencing Center (NISC), as described.[14]
39
40 Identified variants in the linkage region were filtered for quality, absence in ExAC males and
41
42 absence in 473 males from an in-house dataset.[15],[16] Sanger sequence analyses were
43
44 performed in family 1 to confirm segregation. Targeted Sanger sequencing of *NAA10* was
45
46 performed in the proband of family 2 (Figure 1B). Independently, sequence analysis of a panel
47
48 of 41 genes causally associated with severe eye malformations was carried out on the proband
49
50 in family 3 (Figure 1C) using a custom-designed Agilent SureSelect panel (Agilent Technologies,
51
52
53
54
55
56
57
58
59
60

1
2
3 Inc., Santa Clara, CA) and paired-end DNA sequencing using an Illumina MiSeq platform
4
5 (Illumina, San Diego, CA), as per the manufacturer's instructions. Targeted sequencing of the
6
7 *NAA10* polyadenylation signal (PAS) was performed in 250 unrelated
8
9
10 microphthalmia/anophthalmia/coloboma patients selected without sex bias from the MRC
11
12 Human Genetics Unit Eye Malformation cohort. A further 126 males with developmental eye
13
14 disorders underwent targeted sequencing of the *NAA10* PAS. Forty-four of these individuals
15
16 with bilateral developmental eye disorders were analysed for variants in the coding region and
17
18 consensus splice sites of *NAA10*.
19
20
21

22 **X-inactivation analyses**

23
24 X-inactivation studies were performed on genomic DNA from 11 carrier females (family 1
25
26 individuals I-2, II-2, II-6, II-17, III-15; family 2 individuals III-2, III-3, IV-1, IV-5, V-8; family 3
27
28 individual I-2) using the human androgen receptor (HUMARA) assay as described in Allen et
29
30 al[17] with minor modifications. Briefly, genomic DNA was digested with *Rsa1* either alone or
31
32 with the addition of *HpaII* (New England Biolabs, Ipswich, MA). Digested DNA was PCR amplified
33
34 using primers HUMAR_F: /56-FAM/TCCAGAATCTGTTCCAGAGCGTGC and Humara_R:
35
36 GCTGTGAAGGTTGCTGTTCCCTCAT (Integrated DNA Technologies, Inc. Coralville, Iowa). Products
37
38 were run on an 3130xl genetic analyzer (ThermoFisher Scientific, Waltham, MA) and analyzed
39
40 using GeneMapper v4.0 (ThermoFisher Scientific, Waltham, MA).
41
42
43
44
45
46

47 **RNA Analyses**

48
49 Total RNA was isolated from whole blood from affected and carrier individuals in family 1 and
50
51 from unrelated male and female controls using PAXgene Blood RNA tubes (BD Biosciences, San
52
53 Jose, CA). Total RNA was isolated from lymphoblasts from one affected and one unaffected
54
55
56
57
58
59
60

1
2
3 individual in family 2 using the RNeasy Mini kit (Qiagen, Germantown, MD). RNA samples were
4
5 not available from family 3. RNA-Seq libraries were constructed from 0.5-1 µg total RNA after
6
7 rRNA depletion using Ribo-Zero Globin (Illumina, San Diego, CA). The Illumina TruSeq Stranded
8
9 Total RNA Kit was used according to manufacturer's instructions. The cDNA inserts were ~200
10
11 bp after chemical shearing. PCR amplification was performed using 10 cycles. Unique barcode
12
13 adapters were applied to each library. Libraries were pooled in equimolar ratio for sequencing.
14
15 The pooled libraries were sequenced on multiple lanes of a HiSeq 2500 (Illumina, San Diego, CA)
16
17 using version 4 chemistry to achieve a minimum of 69 million 125 base read pairs. The data
18
19 were processed using RTA version 1.18.64 and CASAVA 1.8.2. RNA-Seq reads were aligned with
20
21 the STAR aligner,[18] using the hg19 genome assembly and the Ensembl transcript database
22
23 (release 74). The STAR BAM output file was converted to WIG format using the bam2wig.py
24
25 script from the RSeqC suite,[19] and these WIG files were further converted to bigwig format
26
27 using the wigToBigWig tool from the UCSC Genome Browser's binary utilities directory.
28
29 For quantitative PCR (qPCR), 400 ng total RNA was converted to cDNA using the High Capacity
30
31 RNA-to-cDNA kit (Applied Biosystems, Beverly, MA). TaqMan assays were performed using 10
32
33 ng cDNA, gene specific TaqMan probes (*NAA10*: HS01125831_g1 and HA00185854_m1; beta-
34
35 Actin: Hs99999903_m1 (Thermo Fisher Scientific, Waltham, MA)) and TaqMan Gene Expression
36
37 Master Mix (Applied Biosystems, Beverly, MA) on the OneStepPlus Real-Time PCR System
38
39 (Thermo Fisher Scientific, Waltham, MA). Samples were run in triplicate and relative expression
40
41 levels in affected and carrier individuals as compared to control individuals were calculated
42
43 using the delta-delta C_t method.
44
45
46
47
48
49
50
51
52
53

54 **Results**

1
2
3 **Clinical evaluations**
4

5 **Family 1**
6

7
8 Family 1 (Figure 1A) was originally described by Graham et al. in support of a fifth
9
10 microphthalmia/anophthalmia locus MCOPS4. The family was reported to include seven
11
12 affected males with microphthalmia/anophthalmia and ID.[9] One male, II-18, was reported to
13
14 have normal eyes with cleft soft palate and a single ear tag and was considered unaffected with
15
16 respect to syndromic anophthalmia/microphthalmia. An additional affected male, IV-3, with a
17
18 milder phenotype as compared to his affected uncles, was born after publication of the Graham
19
20 et al paper.[9] Briefly, this male infant was born following a pregnancy in which scans were
21
22 assessed to be normal. However, at birth he was noted to have a large open neural tube defect
23
24 (NTD), right sided anophthalmia, and bilateral 2-3 cutaneous syndactyly of his toes. When his
25
26 teeth erupted they were irregular. He was sociable and did not show signs of the severe
27
28 developmental delay or behavioral problems, most notably self-mutilation, seen in other
29
30 affected males in the family.
31
32
33
34
35
36

37 **Family 2**
38

39
40 Family 2 (Figure 1B) was originally described by Slavotinek et al.[20] as a four-generation family
41
42 with X-linked anophthalmia consistent with linkage to Xq27. The intelligence of two living males
43
44 in family 2 was apparently normal. Carrier females III-2 and IV-1 had unremarkable eye
45
46 examinations, individual IV-1 was diagnosed with dyslexia but her intelligence was apparently
47
48 normal.
49
50

51 **Family 3**
52
53
54
55
56
57
58
59
60

1
2
3 The male proband (individual II-2, Figure 1C) was the second child of a non-consanguineous
4 couple, referred for diagnostic screening of genes associated with severe eye malformations.
5
6 Family history was remarkable for a maternal uncle with anencephaly. The proband presented
7
8 with spina bifida at 20 weeks gestation. He was a term delivery with birth weight 3.2 Kg. At
9
10 birth, he was noted to have widely spaced eyes, unilateral phthisis bulbi, downturned corners
11
12 of the mouth, small penis, and small feet with upturned nails. At eight months his length,
13
14 weight, and head circumference were recorded at 9th centile. He had a repaired
15
16 myelomeningocele, hydrocephalus with a ventriculoperitoneal shunt, and left grade 2 vesico-
17
18 ureteric reflux with a normal renal ultrasound. MRI studies showed ventriculomegaly with a
19
20 Chiari type 2 malformation. Developmental milestones were reported to be normal at 8
21
22 months. Array CGH was reported as normal.
23
24
25
26
27
28
29

30 **Molecular Data**

31 **Family 1**

32
33 LOD score analyses reported in the original paper showed a peak logarithm of odds (LOD) score
34
35 of 1.9, between the disease and markers on Xq27-q28, below the 2.0 threshold recommended
36
37 to declare X-chromosome linkage.[21] The evidence of linkage was weak unless one assumed
38
39 non-penetrance, pleiotropy, variable expressivity, or gonadal mosaicism as the mutant
40
41 haplotype was shared by II-18 and his affected relatives.
42
43
44

45 Repeat multi-marker analysis using four informative markers DXS8091, DXS1193, DXS8086, and
46
47 DXS1073 achieved a peak LOD score of 1.43 with the disease locus placed on top of DXS1073
48
49 and assuming full penetrance. The peak multi-marker score was 2.19 with 75% penetrance,
50
51 2.46 with 50% penetrance, and 3.31 assuming gonadal mosaicism. Haplotype analysis was
52
53
54
55
56
57
58
59
60

1
2
3 performed assuming a model of X-linked recessive inheritance and showed a recombination
4
5 event in affected individual III-2 between markers DXS8043 and DXS8086 restricting the
6
7 possible linkage region telomeric to position ChrX(GRCh37):g.144,028,513.
8
9

10 Both *HMGB3* and *NAA10* lie within this linkage region. The coding regions and
11
12 consensus splice sites of *HMGB3* and *NAA10* were interrogated without identification of a
13
14 pathogenic variant. Trio exome sequencing was performed on individual IV-3 and his parents
15
16 and variants in the linked region were filtered for absence in male controls. A single variant was
17
18 identified in the 3'UTR of *NAA10*, chrX:153,195,397T>C, c.*43A>G (reference cDNA
19
20 NM_003491.3), altering the consensus PAS from AATAAA to AATAGA. The *NAA10* variant was
21
22 verified by Sanger sequencing and shown to be present in all affected individuals available for
23
24 analysis, the matriarch (I-2), as well as in individual II-18, the male previously thought to be
25
26 unaffected. Querying the gnomAD database for variants in the consensus PAS for *NAA10* did
27
28 not identify any variants in over 170,000 alleles.
29
30
31
32
33

34 35 **Family 2**

36
37 A reexamination of the linkage data showed that family 1 and family 2 had overlapping regions
38
39 near Xqtel consistent with linkage. Based on the findings in family 1, targeted Sanger
40
41 sequencing of *NAA10* was performed in the proband of family 2. A second distinct variant in the
42
43 the 3'UTR of *NAA10*, chrX:g.153,195,401T>C, c.*39A>G, was identified, which altered the
44
45 consensus PAS from AATAAA to GATAAA. The variant was verified by Sanger sequencing in
46
47 carrier females and was not present in two unaffected males.
48
49
50

51 52 **Family 3**

1
2
3 Targeted sequencing of 41 genes causally associated with severe eye malformations in the
4
5 proband from family 3 identified a third variant in *NAA10*, chrX:g.153,195,400T>C , c.*40A>G,
6
7 that altered the consensus PAS from AATAAA to AGTAAA. The *NAA10* variant in family 3 was
8
9 confirmed by Sanger sequencing. DNA from the uncle with anencephaly was not available for
10
11 testing.
12
13

14 15 ***NAA10* Sequence Analysis in Microphthalmia/Anophthalmia/Coloboma Cohorts**

16
17 Targeted sequencing of the *NAA10* PAS in 250 unrelated
18
19 microphthalmia/anophthalmia/coloboma patients selected without sex bias from the MRC
20
21 Human Genetics Unit Eye Malformation cohort did not identify additional variants. No variants
22
23 affecting the *NAA10* PAS were identified in a further 126 males with developmental eye
24
25 disorders. In addition, no pathogenic variants in the coding region and consensus splice sites of
26
27 *NAA10* were identified in 44 of these individuals who presented with bilateral developmental
28
29 eye disorders.
30
31
32
33

34 35 **X-Inactivation**

36
37 Four of 11 carrier females showed greater than 90% skewing of X-inactivation, individual II-2
38
39 from family 1 and individuals III-2, IV-1 and IV-5 from family 2. However, females did not show
40
41 consistent skewing of X-inactivation.
42
43

44 45 **RNA analyses**

46
47 The consequences of the PAS variants for *NAA10* mRNA expression level and structure were
48
49 investigated using qPCR and RNAseq in families 1 and 2. qPCR showed carrier females from
50
51 family 1 to have similar levels of *NAA10* RNA as compared to control individuals. In contrast,
52
53
54
55
56
57
58
59
60

1
2
3 affected individuals from families 1 and 2 demonstrated a decrease in the quantity of *NAA10*
4
5 RNA of approximately 50% when compared to controls (Figure 2).
6

7
8 Transcript structure was investigated by mapping RNAseq reads to the UCSC Genome
9
10 Browser. RNAseq data from unaffected (family 2, individual V-9) and control individuals
11
12 indicated the normal 3'UTR of *NAA10* ended at the approximate position of the predicted
13
14 polyadenylation site (Figure 3G-I). In contrast, for affected individuals with the PAS variants
15
16 (family 1, individuals III-2, III-5, IV-3 and family 2, individual IV-6), the read depth did not
17
18 decrease as expected at the polyadenylation site in the 3'UTR, but instead declined
19
20 approximately 600 bp further 3' at a second polyadenylation site predicted by bioinformatic
21
22 analysis of genomic sequence (Figure 3A-C, F).[22] A similar result was observed in carrier
23
24 females (family 1, individuals II-2 and III-15) (Figure 3D-E).
25
26
27
28
29

30 **Discussion**

31
32 Six protein complexes, NatA through NatF, carry out N^α acetylation of proteins in the
33
34 cell.[23],[24] *NAA10* is the primary subunit of the NatA complex and an auxiliary subunit of the
35
36 NatE complex.[24],[25] The acetylation targets of the six complexes are believed to be mostly
37
38 distinct and dependent on the amino acids that follow the methionine in position 1.[24]
39
40 Additionally, data from a knock-out mouse model suggests that *NAA10* plays a role in the
41
42 regulation of methylation through direct DNA binding.[26] Prior to the present publication, the
43
44 mutational spectrum in *NAA10* comprised nine missense alterations and the single splice site
45
46 variant identified previously in a single family with what those authors called LMS. The first
47
48 reported variant in *NAA10*, c.109T>C; p.(Ser37Pro), was identified in two families segregating a
49
50 sex-linked recessive male lethal syndrome with an aged appearance and cardiac
51
52
53
54
55
56
57
58
59
60

1
2
3 arrhythmias.[27] Eight additional missense variants have since been reported in heterozygous
4
5 females or hemizygous males with developmental delay and/or ID with or without cardiac
6
7 involvement and without anophthalmia.[28],[29],[30],[31],[32] The majority of these variants
8
9 were identified as *de novo* alterations.[28],[30],[31],[32] Functional studies of the missense
10
11 alterations have demonstrated a variable effect on the ability of NAA10 to acetylate substrates
12
13 and the clinical phenotype may correlate with the level of acetylation dysfunction and affected
14
15 substrates. Alternatively, clinically relevant predicted missense variants in *NAA10* have been
16
17 shown to disrupt DNA binding[26] and it is likely that different missense alterations will variably
18
19 contribute to this phenomenon.
20
21
22
23

24
25 The polyA_DB database contains human mRNA polyadenylation sites based on EST/cDNA
26
27 evidence[33] and predicts four sites in the 3' region of *NAA10*. Based on the *NAA10* Poly(A)
28
29 sequence data in the UCSC Genome Browser, the PAS at c.*39 – c.*44 is the most commonly
30
31 used signal with polyadenylation occurring between 12-19 nucleotides 3' of that hexamer.
32
33
34 Three distinct nucleotides within this six-nucleotide signal are mutated in the families described
35
36 here. Previous in vitro studies have demonstrated that these changes in the PAS hexamer
37
38 disrupt both cleavage and polyadenylation (Figure 4).[34] Furthermore, these variants are
39
40 analogous to variants associated with abnormal phenotypes in other genes. The *NAA10* 3'UTR
41
42 variant identified in family 1, c.*43A>G, alters the AATAAA consensus to AATAGA, similar to the
43
44 c.*112A>G (reference cDNA NM_000518.4) pathogenic variant in the beta globin (*HBB*) gene
45
46 (HGMD CR900266) (See Table 1).[35] The 3'UTR variant of *NAA10* identified in family 2,
47
48 c.*39A>G, alters the AATAAA consensus to GATAAA, similar to another reported pathogenic
49
50 variant in the *HBB* gene, (c.*108A>G, HGMD CR127145).[36] The *NAA10* 3'UTR variant
51
52
53
54
55
56
57
58
59
60

1
2
3 identified in family 3 , c.*40A>G, alters the AATAAA consensus to AGTAAA, similar to the variant
4
5 in the *ARSA* gene (HGMD CR890137) reported by Gieselmann et al[37] and a variant in the
6
7 *SLC6A4* gene (HGMD CR102248) reported by Gyawali et al[38] although it is important to note
8
9 that the wild-type PAS consensus sequence in both *ARSA* and *SLC6A4* genes is AATAAC, (See
10
11 Table 1). That the *NAA10* variant disrupts a consensus PAS rather than a non-canonical PAS
12
13 could be consistent with the observation that the *NAA10* variants are associated with a
14
15 Mendelian disorder whereas the *ARSA* and *SLC6A4* variants are associated with lower
16
17 penetrance traits.
18
19
20
21

22 All three families presented here manifested microphthalmia/anophthalmia with
23
24 variable additional features. In family 1, individuals had variable expressivity ranging from
25
26 syndromic microphthalmia/anophthalmia with severe ID to isolated cleft palate and an ear tag.
27
28 A single individual, IV-3, had an open NTD. Affected individuals in family 2 had anophthalmia
29
30 and additional findings including minor skeletal anomalies and attention deficit disorder, but
31
32 notably all affected males were reported to have normal intelligence. The proband in family 3
33
34 had syndromic microphthalmia/anophthalmia with complex features including spina bifida. By
35
36 report, a maternal uncle (deceased) to the proband had anencephaly, however, his genotype
37
38 was unknown. Although the numbers are small, two out of 15 individuals in this study had NTDs
39
40 (three if the uncle in family 2 is included). Additional families will need to be studied to
41
42 determine if NTDs are a feature of *NAA10* PAS variants. The mouse *NAA10* knock-out model
43
44 had variable expressivity ranging from embryonic lethality to normal body size and Lee et al.
45
46 suggested this variable expressivity may result from the role of *NAA10* in global
47
48 methylation.[26] The role of *NAA10* in global methylation may contribute to the observed
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 variable expressivity in these families as well. Specifically, the observed phenotype may result
4
5 from the regions of the genome experiencing defective methylation in any one individual. As
6
7 methylation is important in the etiology of NTDs it is interesting to speculate that the role of
8
9 NAA10 in global methylation may account for the NTDs seen in these families. Future studies
10
11 will be needed to determine if methylation or other factors play a role in the variable
12
13 expressivity seen with these variants.
14
15
16

17
18 Cleavage and polyadenylation signals reside in the 3' UTRs of mRNAs and include a
19
20 number of key *cis* elements including the AAUAAA PAS which typically resides 10-30
21
22 nucleotides upstream of the cleavage site (reviewed in [39],[40]). Ninety percent of identified
23
24 PAS conform to the hexanucleotide sequence AAUAAA and most of the remaining 10% differ by
25
26 only a single substitution.[39] The AAUAAA hexamer is critical both for cleavage and poly(A)
27
28 addition,[41] and the effects of point mutations in this hexamer have been determined.[34]
29
30 Disease-associated variants in PAS are rare and consequently we were surprised to identify
31
32 three such variants in this rare disorder. To identify all known consensus PAS variants, we
33
34 performed a directed search of the Human Gene Mutation Database (HGMD).[42] We
35
36 identified 31 reported PAS variants in 19 genes (Table 1), 23 associated with Mendelian
37
38 disorders and eight associated with a functional polymorphism or susceptibility. Overall, PAS
39
40 variants account for 23 out of the ~195,000 DM variants in HGMD, or 0.012% of the total. As
41
42 well as being individually uncommon, such variants have a highly non-random distribution
43
44 across genes. Of the 23 variants associated with Mendelian disorders, nine are in *HBB*, four are
45
46 in *HBA2*, and two are in *FOXP3*, the remainder being singletons. Some of this distribution can be
47
48 explained by the depth and detail to which the various genes have been studied, *HBB* being one
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 of the best studied of all human genes. To allow for this, the proportion of polyadenylation
4
5 variants to all known pathogenic variants in a given gene can be tabulated. When expressed in
6
7 this way, a very different pattern emerges. *NAA10* has the highest proportion of PAS variants,
8
9 at 25%. The next highest cluster of genes includes *IGF1*, *NAT1*, *BMP1*, and *SLC6A4*, each of
10
11 which has a single PAS variant, but a relatively small overall number of variants, yielding a high
12
13 proportion. *HBB* and *HBA2* lie in the middle of the range at about 1% of variants, reflecting the
14
15 large number of variants in these well-studied genes. Most of the clinical phenotypes
16
17 associated with PAS variants are typical of those seen for loss of function variants in the same
18
19 gene although they may be found with milder forms of the phenotypes – for example, in
20
21 *HBB*, [36], [43] the phenotypes of PAS variants are described as either typical or mild
22
23 thalassemia. The contribution to milder phenotypes has two implications, first in disorders
24
25 where loss of gene function would be lethal, PAS variants may allow sufficient function for
26
27 viability. Second, for disorders where the phenotype is due to complete loss of function, PAS
28
29 variants may be missed if they lead to a milder condition. [36] We conclude from these data
30
31 that the phenotypic consequence of pathogenic PAS variants is likely due to loss or partial loss
32
33 of function, consistent with the experimental observations of these variants in other genes and
34
35 the qPCR data we show here. Specifically, for *NAA10*, this suggests that partial loss of protein
36
37 function due to reduced mRNA levels leads to a distinct phenotype, as compared to missense
38
39 variants where the functional deficiency may only affect specific acetylation or DNA binding
40
41 functions. The previously reported *NAA10* splice site variant [7] also showed reduced
42
43 mRNA/protein supporting reduced function as causal for the microphthalmia phenotype.
44
45
46
47
48
49
50
51
52
53
54 Alternatively, sequences within the elongated 3' end, or within the retained intron 7 for the
55
56
57
58
59
60

1
2
3 splice variant, may have a novel functional effect. Possible novel effects include: altered RNA or
4
5 protein localization[44], altered 3'UTR-directed protein-protein interactions[45], or acquisition
6
7 of novel regulatory functions. However, since the retained sequences are distinct between the
8
9 splice site variant and the PAS variants, a novel property of the mRNA is unlikely to be the
10
11 mechanism accounting for the overlapping feature of microphthalmia. Additionally, the
12
13 absence of a phenotype in carrier females without skewed X-inactivation suggests that any
14
15 effect would need to fit a recessive model of inheritance. The general dearth of recognized PAS
16
17 pathogenic variants may be in part due to the inherent difficulty in identifying PAS. In addition,
18
19 many such variants may not be included in next generation sequence data due to the
20
21 limitations of exome sequencing kits and the predominant use of exome over genome
22
23 sequencing.
24
25
26
27
28
29

30 The assessment of pathogenicity for non-coding variants in Mendelian disorders is
31
32 challenging. The ACMG/AMP [46] framework includes few criteria that are relevant to 3' UTR
33
34 variants, reflecting a focus on high penetrance coding variants. For the variants we identified
35
36 here, we invoked criteria PS3 (well-established in vitro or in vivo functional studies), PM1
37
38 (located in a mutational hot spot and/or critical and well-established functional domain), PM2
39
40 (absent from controls), and PP1 (co-segregation with disease in multiple affected family
41
42 members). As qPCR is not “a well-established functional assay” because qPCR for *NAA10* is not
43
44 established in the diagnostic realm, the PS3 criteria can be downgraded to a moderate level of
45
46 evidence (PS3_Mod). Additionally, our segregation data are arguably less than robust
47
48 considering that we have invoked variable expressivity to explain the occurrence of the variant
49
50 in II-18 in family 1 and thus PP1 might arguably be dropped. That would leave PS3_Mod, PM1,
51
52
53
54
55
56
57
58
59
60

1
2
3 and PM2, which combine to yield a likely pathogenic assessment. Based on these pieces of
4
5 evidence, we conclude that these three PAS variants are likely pathogenic.
6
7

8 These data have important implications for the overall process of identifying pathogenic
9 variation. We hypothesize that PAS variants are candidate pathogenic variants for many
10 Mendelian disorders and may be a component of the mutational spectrum of patients who are
11 currently 'mutation-negative'. A full assessment of this hypothesis will first require the
12 continued development of genome (as opposed to exome) population sequence reference
13 databases, allowing a greater understanding of the evolutionary constraints operating at 3' UTR
14 loci. Second, robust methods to identify PAS for all genes will need to be developed. While
15 trivial for many genes, this is challenging for others as some use multiple polyadenylation sites
16 and often the PAS sequence does not conform to the consensus, as demonstrated for a number
17 of the genes that we reviewed. Third, to the extent that exome analysis continues to outpace
18 genome analysis, it will be necessary to target these sequences in exome capture reagents.
19 Fourth, testing laboratories will need to direct their attention to detecting and interpreting such
20 variants. Finally, criteria for pathogenicity of non-coding variants will need to be updated. We
21 believe that the accomplishment of these improvements in variant detection, interrogation,
22 and interpretation will allow the findings based on the three variants reported here as a cause
23 of a rare disorder to yield benefits for the diagnosis of patients with many other genetic
24 disorders and contribute to the overall improvement in clinical genomic sequencing diagnostic
25 yield.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 52 53 54 **Contributions** 55 56 57 58 59 60

1
2
3 All authors fulfil the criteria for authorship. AAS performed and interpreted linkage analysis of family 1.
4 JJJ performed haplotype analysis, interpreted NGS data and performed Sanger confirmation in family 1.
5 JCS, FS and LGB provided clinical management for family 1. HMC performed sequence analysis of *NAA10*
6 in family 2. CMC and AMS provided clinical management for family 2. JND performed and interpreted
7 qPCR experiments in families 1 and 2. SKS analyzed RNA-seq data for families 1 and 2. KAW and MA
8 interpreted NGS data in family 3 and analyzed sequence data in a cohort of individuals with
9 anophthalmia/microphthalmia for variants in *NAA10*. TD provided clinical management for family 3. RJH
10 and NKR performed and analyzed sequencing of *NAA10* in a cohort of individuals with
11 anophthalmia/microphthalmia/coloboma. DNC provided an analysis of polyadenylation signal variants
12 within the Human Gene Mutation Database. JJJ and LGB drafted the manuscript and all authors
13 contributed to the manuscript. All authors read and approved the final manuscript. LGB (family 1),
14 GCMB (family 1), TMG (family 2) and DRF (family 3) planned the study. JJJ submitted the study.
15
16

17 **Funding**

18
19
20 This work was supported by funding from the Intramural Research Program of the National
21 Human Genome Research Institute (LGB, JJJ, JCS, grants HG200328 11 and HG200388 03). This
22 work was supported in part by funding from the Intramural Research Program of the National
23 Library of Medicine (AAS, grant LM00097), the extramural research program of the National
24 Human Genome Research Institute (AMS, grant 5U01HG009599 02), the National Institutes of
25 Health (TMG, HMC, grant EY19497), the Hartwell Foundation (HMC), the MRC University Unit
26 award to the University of Edinburgh for the MRC Human Genetics Unit (DRF, KAW), and the
27 MRC IGMM Translational Science Initiative (MA). This work was supported in part by grants
28 from Baillie Gifford, Visually Impaired Children Taking Action (VICTA)
29 (<http://www.victa.org.uk/>), and Microphthalmia, Anophthalmia, Coloboma Support (MACS)
30 (www.macs.org.uk).
31
32
33

34 **Competing Interests**

35
36
37 LGB receives royalties from Genentech Corp, is an advisor to the Illumina Corp, received
38 honoraria from Wiley-Blackwell, and receives honoraria from Cold Spring Harbor Press. DNC is
39 in receipt of funding from Qiagen Inc through a License Agreement with Cardiff University. AMS
40 receives honoraria from Wiley-Blackwell, Inc., Oxford University Press and UptoDate, Inc.
41
42

43 **Acknowledgments**

44
45
46 We thank the families for their participation in this study and the NIH Intramural Sequencing
47 Center for performing the exome sequencing on family 1 and the RNAseq data generation. We
48 thank MaryPat Jones and Ursula Harper for generating STRP genotypes for this project. We are
49 grateful to Dorine Bax for assisting with the coordination of the project. We thank Dr. Michael
50 Sheets and Dr. Marvin Wickens for thoughtful review of the manuscript and permission to use
51 Figure 4 reproduced from Sheets et al. 1990. The image was used by permission of Oxford
52 University Press. We thank Catherine Driscoll for help in editing the manuscript.
53
54
55
56
57
58
59
60

1
2
3 **WEB RESOURCES**
4

5 Online Mendelian Inheritance in Man (OMIM): <http://www.omim.org/>
6

7
8 Human Gene Mutation Database (HGMD): <http://www.hgmd.cf.ac.uk>
9

10 University of California Santa Cruz genome browser (UCSC): <https://genome.ucsc.edu/>
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Tucker S, Jones B, Collin R. Systemic anomalies in 77 patients with congenital anophthalmos or microphthalmos. *Eye (Lond)* 1996;**10 (Pt 3)**:310-4 doi: 10.1038/eye.1996.65published Online First: 1996/01/01].
2. Verma AS, Fitzpatrick DR. Anophthalmia and microphthalmia. *Orphanet J Rare Dis* 2007;**2**:47 doi: 10.1186/1750-1172-2-47published Online First: 2007/11/28].
3. Gerth-Kahlert C, Williamson K, Ansari M, Rainger JK, Hingst V, Zimmermann T, Tech S, Guthoff RF, van Heyningen V, Fitzpatrick DR. Clinical and mutation analysis of 51 probands with anophthalmia and/or severe microphthalmia from a single center. *Mol Genet Genomic Med* 2013;**1**:15-31 doi: 10.1002/mgg3.2published Online First: 2014/02/06].
4. Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tiff C, Zhang L, Wilkie AO, van der Smagt JJ, Gorlin RJ, Burgess SM, Bardwell VJ, Black GC, Biesecker LG. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in *BCOR*. *Nat Genet* 2004;**36**:411-6 doi: 10.1038/ng1321published Online First: 2004/03/09].
5. Wimplinger I, Morleo M, Rosenberger G, Iaconis D, Orth U, Meinecke P, Lerer I, Ballabio A, Gal A, Franco B, Kutsche K. Mutations of the mitochondrial holocytochrome c-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. *Am J Hum Genet* 2006;**79**:878-89 doi: 10.1086/508474published Online First: 2006/10/13].
6. Scott AF, Mohr DW, Kasch LM, Barton JA, Pittiglio R, Ingersoll R, Craig B, Marosy BA, Doheny KF, Bromley WC, Roderick TH, Chassaing N, Calvas P, Prabhu SS, Jabs EW. Identification of an *HMGB3* frameshift mutation in a family with an X-linked colobomatous microphthalmia syndrome using whole-genome and X-exome sequencing. *JAMA Ophthalmol* 2014;**132**:1215-20 doi: 10.1001/jamaophthalmol.2014.1731published Online First: 2014/07/06].
7. Esmailpour T, Riazifar H, Liu L, Donkervoort S, Huang VH, Madaan S, Shoucri BM, Busch A, Wu J, Towbin A, Chadwick RB, Sequeira A, Vawter MP, Sun G, Johnston JJ, Biesecker LG, Kawaguchi R, Sun H, Kimonis V, Huang T. A splice donor mutation in *NAA10* results in the dysregulation of the retinoic acid signalling pathway and causes Lenz microphthalmia syndrome. *J Med Genet* 2014;**51**:185-96 doi: 10.1136/jmedgenet-2013-101660published Online First: 2014/01/17].
8. Forrester S, Kovach MJ, Reynolds NM, Urban R, Kimonis V. Manifestations in four males with and an obligate carrier of the Lenz microphthalmia syndrome. *Am J Med Genet* 2001;**98**:92-100 Online First: 2001/06/28].
9. Graham CA, Redmond RM, Nevin NC. X-linked clinical anophthalmos. Localization of the gene to Xq27-Xq28. *Ophthalmic Paediatr Genet* 1991;**12**:43-8 Online First: 1991/03/01].
10. Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 1984;**81**:3443-6 Online First: 1984/06/01].
11. Cottingham RW, Jr., Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1993;**53**:252-63 Online First: 1993/07/01].

12. Schaffer AA, Gupta SK, Shriram K, Cottingham RW, Jr. Avoiding recomputation in linkage analysis. *Hum Hered* 1994;**44**:225-37 Online First: 1994/07/01].
13. Kong X, Murphy K, Raj T, He C, White PS, Matisse TC. A combined linkage-physical map of the human genome. *Am J Hum Genet* 2004;**75**:1143-8 doi: 10.1086/426405published Online First: 2004/10/16].
14. Johnston JJ, Sanchez-Contreras MY, Keppler-Noreuil KM, Sapp J, Crenshaw M, Finch NA, Cormier-Daire V, Rademakers R, Sybert VP, Biesecker LG. A point mutation in *PDGFRB* causes autosomal-dominant Penttinen syndrome. *Am J Hum Genet* 2015;**97**:465-74 doi: 10.1016/j.ajhg.2015.07.009published Online First: 2015/08/19].
15. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, DeFlaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285-91 doi: 10.1038/nature19057published Online First: 2016/08/19].
16. Patton J, Brewer C, Chien W, Johnston JJ, Griffith AJ, Biesecker LG. A genotypic ascertainment approach to refute the association of *MYO1A* variants with non-syndromic deafness. *Eur J Hum Genet* 2016;**25**:147-49 doi: 10.1038/ejhg.2016.140published Online First: 2016/10/21].
17. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of *HpaII* and *HhaI* sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 1992;**51**:1229-39 Online First: 1992/12/01].
18. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15-21 doi: 10.1093/bioinformatics/bts635published Online First: 2012/10/30].
19. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 2012;**28**:2184-5 doi: 10.1093/bioinformatics/bts356published Online First: 2012/06/30].
20. Slavotinek A, Lee SS, Hamilton SP. A family with X-linked anophthalmia: exclusion of *SOX3* as a candidate gene. *Am J Med Genet A* 2005;**138A**:89-94 doi: 10.1002/ajmg.a.30872published Online First: 2005/08/23].
21. Ott J. *A Short Guide to Linkage Analysis*, 1986.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
22. Cheng Y, Miura RM, Tian B. Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics* 2006;**22**:2320-5 doi: 10.1093/bioinformatics/btl394published Online First: 2006/07/28].
 23. Kalvik TV, Arnesen T. Protein N-terminal acetyltransferases in cancer. *Oncogene* 2013;**32**:269-76 doi: 10.1038/onc.2012.82published Online First: 2012/03/07].
 24. Polevoda B, Arnesen T, Sherman F. A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. *BMC Proc* 2009;**3 Suppl 6**:S2 doi: 10.1186/1753-6561-3-S6-S2published Online First: 2009/08/08].
 25. Myklebust LM, Van Damme P, Stove SI, Dorfel MJ, Abboud A, Kalvik TV, Grauffel C, Jonckheere V, Wu Y, Swensen J, Kaasa H, Liszczak G, Marmorstein R, Reuter N, Lyon GJ, Gevaert K, Arnesen T. Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum Mol Genet* 2015;**24**:1956-76 doi: 10.1093/hmg/ddu611published Online First: 2014/12/10].
 26. Lee CC, Peng SH, Shen L, Lee CF, Du TH, Kang ML, Xu GL, Upadhyay AK, Cheng X, Yan YT, Zhang Y, Juan LJ. The Role of N-alpha-acetyltransferase 10 Protein in DNA Methylation and Genomic Imprinting. *Mol Cell* 2017;**68**:89-103 e7 doi: 10.1016/j.molcel.2017.08.025published Online First: 2017/09/26].
 27. Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, Swensen JJ, Johnson WE, Moore B, Huff CD, Bird LM, Carey JC, Opitz JM, Stevens CA, Jiang T, Schank C, Fain HD, Robison R, Dalley B, Chin S, South ST, Pysher TJ, Jorde LB, Hakonarson H, Lillehaug JR, Biesecker LG, Yandell M, Arnesen T, Lyon GJ. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet* 2011;**89**:28-43 doi: 10.1016/j.ajhg.2011.05.017published Online First: 2011/06/28].
 28. Popp B, Stove SI, Ende S, Myklebust LM, Hoyer J, Sticht H, Azzarello-Burri S, Rauch A, Arnesen T, Reis A. De novo missense mutations in the NAA10 gene cause severe non-syndromic developmental delay in males and females. *Eur J Hum Genet* 2015;**23**:602-9 doi: 10.1038/ejhg.2014.150published Online First: 2014/08/08].
 29. Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, Stray-Pedersen A, Kury S, Mercier S, Lessel D, Denecke J, Wiszniewski W, Penney S, Liu P, Bi W, Lalani SR, Schaaf CP, Wangler MF, Bacino CA, Lewis RA, Potocki L, Graham BH, Belmont JW, Scaglia F, Orange JS, Jhangiani SN, Chiang T, Doddapaneni H, Hu J, Muzny DM, Xia F, Beaudet AL, Boerwinkle E, Eng CM, Plon SE, Sutton VR, Gibbs RA, Posey JE, Yang Y, Lupski JR. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med* 2017;**9**:26 doi: 10.1186/s13073-017-0412-6published Online First: 2017/03/23].
 30. Saunier C, Stove SI, Popp B, Gerard B, Blenski M, AhMew N, de Bie C, Goldenberg P, Isidor B, Keren B, Leheup B, Lampert L, Mignot C, Tezcan K, Mancini GM, Nava C, Wasserstein M, Bruel AL, Thevenon J, Masurel A, Duffourd Y, Kuentz P, Huet F, Riviere JB, van Slegtenhorst M, Faivre L, Piton A, Reis A, Arnesen T, Thauvin-Robinet C, Zweier C. Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency. *Hum Mutat* 2016;**37**:755-64 doi: 10.1002/humu.23001published Online First: 2016/04/21].

- 1
2
3 31. Cherot E, Keren B, Dubourg C, Carre W, Fradin M, Lavillaureix A, Afenjar A,
4 Burglen L, Whalen S, Charles P, Marey I, Heide S, Jacqueline A, Heron D,
5 Doummar D, Rodriguez D, Billette de Villemeur T, Moutard ML, Guet A, Xavier J,
6 Perisse D, Cohen D, Demurger F, Quelin C, Depienne C, Odent S, Nava C,
7 David V, Pasquier L, Mignot C. Using medical exome sequencing to identify the
8 causes of neurodevelopmental disorders: Experience of 2 clinical units and 216
9 patients. *Clin Genet* 2018;**93**:567-76 doi: 10.1111/cge.13102published Online
10 First: 2017/07/15].
- 11
12 32. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-
13 Djebbar S, St-Onge J, Steinmetz A, Huet F, Chouchane M, Darmency-Stamboul
14 V, Callier P, Thauvin-Robinet C, Faivre L, Riviere JB. Diagnostic odyssey in
15 severe neurodevelopmental disorders: toward clinical whole-exome sequencing
16 as a first-line diagnostic test. *Clin Genet* 2016;**89**:700-7 doi:
17 10.1111/cge.12732published Online First: 2016/01/13].
- 18
19 33. Zhang H, Hu J, Recce M, Tian B. PolyA_DB: a database for mammalian mRNA
20 polyadenylation. *Nucleic Acids Res* 2005;**33**:D116-20 doi:
21 10.1093/nar/gki055published Online First: 2004/12/21].
- 22
23 34. Sheets MD, Ogg SC, Wickens MP. Point mutations in AAUAAA and the poly (A)
24 addition site: effects on the accuracy and efficiency of cleavage and
25 polyadenylation in vitro. *Nucleic Acids Res* 1990;**18**:5799-805 Online First:
26 1990/10/11].
- 27
28 35. Jankovic L, Efremov GD, Petkov G, Kattamis C, George E, Yang KG, Stoming TA,
29 Huisman TH. Two novel polyadenylation mutations leading to beta(+)-
30 thalassemia. *Br J Haematol* 1990;**75**:122-6 Online First: 1990/05/01].
- 31
32 36. Bilgen T, Clark OA, Ozturk Z, Akif Yesilipek M, Keser I. Two novel mutations in the
33 3' untranslated region of the beta-globin gene that are associated with the mild
34 phenotype of beta thalassemia. *Int J Lab Hematol* 2013;**35**:26-30 doi:
35 10.1111/j.1751-553X.2012.01456.xpublished Online First: 2012/08/07].
- 36
37 37. Gieselmann V, Polten A, Kreysing J, von Figura K. Arylsulfatase A
38 pseudodeficiency: loss of a polyadenylation signal and N-glycosylation site.
39 *Proc Natl Acad Sci U S A* 1989;**86**:9436-40 Online First: 1989/12/01].
- 40
41 38. Gyawali S, Subaran R, Weissman MM, Hershkowitz D, McKenna MC, Talati A, Fyer
42 AJ, Wickramaratne P, Adams PB, Hodge SE, Schmidt CJ, Bannon MJ, Glatt CE.
43 Association of a polyadenylation polymorphism in the serotonin transporter and
44 panic disorder. *Biol Psychiatry* 2010;**67**:331-8 doi:
45 10.1016/j.biopsych.2009.10.015published Online First: 2009/12/09].
- 46
47 39. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes*
48 *Dev* 1997;**11**:2755-66 Online First: 1997/11/14].
- 49
50 40. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define
51 the mRNA polyadenylation site. *Genes Dev* 2015;**29**:889-97 doi:
52 10.1101/gad.261974.115published Online First: 2015/05/03].
- 53
54 41. Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, Tong L. Molecular basis for
55 the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad*
56 *Sci U S A* 2018;**115**:E1419-E28 doi: 10.1073/pnas.1718723115published Online
57 First: 2017/12/07].

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
42. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;**133**:1-9 doi: 10.1007/s00439-013-1358-4published Online First: 2013/10/01].
 43. Giordano PC, Bouva MJ, Van Delft P, Akkerman N, Kappers-Klunne MC, Hartevelde CL. A new polyadenylation site mutation associated with a mild beta-thalassemia phenotype. *Haematologica* 2005;**90**:551-2 Online First: 2005/04/12].
 44. Berkovits BD, Mayr C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 2015;**522**:363-7 doi: 10.1038/nature14321published Online First: 2015/04/22].
 45. Mayr C. Regulation by 3'-Untranslated Regions. *Annu Rev Genet* 2017;**51**:171-94 doi: 10.1146/annurev-genet-120116-024704published Online First: 2017/08/31].
 46. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405-24 doi: 10.1038/gim.2015.30published Online First: 2015/03/06].

1
2
3 Figure 1. Pedigrees for families 1 to 3. Clinically affected individuals are depicted by filled
4 symbols, black symbols depict individuals with eye findings, grey symbols depict individuals
5 without eye findings, genotypes of tested individuals are noted.
6
7
8
9

10 Figure 2. Reverse transcription and quantitative PCR analysis of *NAA10* expression in mRNA,
11 data were normalized to *ACTB* mRNA levels. Panel A, relative *NAA10* expression in whole blood
12 from affected individuals (family 1, individuals III-2, III-5 and IV-3), carrier females (family 1,
13 individuals II-2, II-6 and III-15) and male and female control individuals (C1-C4). All values are
14 shown relative to control C1. Panel B, *NAA10* expression levels in lymphoblasts from affected
15 male IV-6 shown relative to the expression level in unaffected male V-9, family 2.
16
17
18
19
20
21
22
23
24

25 Figure 3. RNA-seq data comparing reads from four affected individuals, two carrier females and
26 three control individuals, one female and two males, for *NAA10*. The variant positions in the
27 *NAA10* polyadenylation signal are marked by a filled red arrow. Transcription in affected
28 individuals (family 1, individuals III-2, III-5 and IV-3 and family 2, individual IV-6) and carrier
29 females (family 1, individuals II-2 and III-15) continues past the normal polyadenylation
30 cleavage site and uses a cryptic signal approximately 600 bp downstream at an alternate
31 polyadenylation cleavage site depicted by an open red arrow. *NAA10* gene models are from
32 GENCODE Version 19 as included in the UCSC genome browser (GRCh37/hg19).
33
34
35
36
37
38
39
40
41
42
43

44 Figure 4. Comparison of the effects of AAUAAA mutations on cleavage and poly (A) addition.
45 Variants identified in this study are noted with an asterisk. Figure is reproduced from Sheets et
46 al. 1990 with permission from Oxford University Press.[34]
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. *NAA10* polyadenylation signal variants identified in microphthalmia/anophthalmia and review of polyadenylation signal alterations as reported in HGMD.

<i>Gene</i>	WT ^a	Variant	HGMD ID	HGMD pathogenicity	ClinVar Descriptor	GRCh37 Position & Variation	ClinVar variation ID	ClinVar Assertion	PMID
<i>NAA10</i>		AATAGA				ChrX:153195397T>C			
<i>NAA10</i>		GATAAA				ChrX:153195401T>C			
<i>NAA10</i>		AGTAAA				ChrX:153195400T>C			
<i>BMP1</i>	AGTAAA	AGCAAA	CR150372	DM	NM_001199.3(BMP1):c.*241T>C	Chr8:22058957T>C	190231	Pathogenic	25214535
<i>F9</i>		AATGAA	CR005437	DM					11013449
<i>FOXP3</i>		AATGAA	CR014834	DM					11685453
<i>FOXP3</i>		AATAAG	CR097218	DM					19471859
<i>GLA</i>	ATTAA <u>A</u> ga	ATTAGA	CD031841	DM					12796853
<i>HBA2</i>		AATAAG	CR830007	DM					6646217
<i>HBA2</i>		AATGAA	CR920785	DM	NM_000517.4(HBA2):c.*92A>G	Chr16:223691A>G	15647	Pathogenic/Likely Pathogenic	1581238
<i>HBA2</i>		AATAAC	CR106042	DM					19912309
<i>HBA2</i>	AATAA <u>A</u> gt	AATAGT	CD941949	DM					7803252
<i>HBB</i>		AAAAAA	CR045224	DM					15481893
<i>HBB</i>		AAGAAA	CR014260	DM					11300343
<i>HBB</i>		AATATA	CR057232	DM					15820953
<i>HBB</i>		AACAAA	CR850010	DM	NM_000518.4(HBB):c.*110T>C	Chr11:5246718A>G	36332	Pathogenic	4018033
<i>HBB</i>		AATAAG	CR880076	DM	NM_000518.4(HBB):c.*113A>G	Chr11:5246715T>C	15473	Pathogenic	3048433

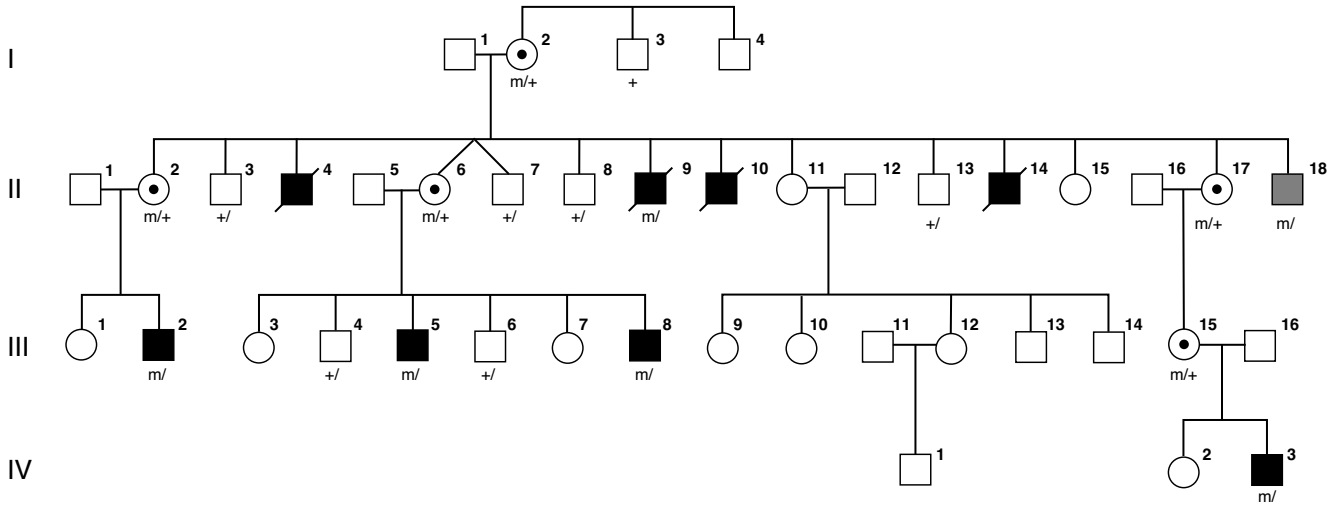
<i>HBB</i>		AATGAA	CR900265	DM	NM_000518.4(<i>HBB</i>):c.*111A>G	Chr11:5246717T>C	15488	Pathogenic	2375910
<i>HBB</i>		AATAGA	CR900266	DM	NM_000518.4(<i>HBB</i>):c.*112A>G	Chr11:5246716T>C	15476	Pathogenic	2375910
<i>HBB</i>		CATAAA	CR016252	DM					11722440
<i>HBB</i>		GATAAA	CR127145	DM					22862814
<i>HBD</i>		AATTAA	CR109506	DM					20854114
<i>IGF1</i>	AATATA	AAAATA	CR033689	DM					14684690
<i>IL2RG</i>		AATAAG	CR0910465	DM					19841577
<i>INS</i>		AATAAG	CR101141	DM	NM_000207.2(<i>INS</i>):c.*59A>G	Chr11:2181023T>C	65581	Pathogenic	20133622
<i>ITGA2B</i>		AACAAA	CR153724	DM					25728920
<i>ABCG2</i>	AAATGA	AAGTGA	CR1718077	DP					28930109
<i>ARSA</i>	AATAAC	AGTAAC	CR890137	DP	NM_000487.5(<i>ARSA</i>):c.*96A>G	Chr22:51063477T>C	3049	Conflicting	2574462
<i>HES7</i>	AAAATC	AAGATC	CR1510755	DM?					25928698
<i>NAT1</i>		AAAAAA	CR1111021	FP					7585580
<i>PEX6</i>	AATAAAcaca	AACACA	CD1716837	FP					29220678
<i>PTGFR</i>		AATGAA	CR156013	DFP					25977569
<i>SLC6A4</i>	ATTAAC	AGTAAC	CR102248	DFP					19969287
<i>TP53</i>		AATACA	CR118782	DFP					21946351

^aWild type is AATAAA unless otherwise specified. Underlined nucleotides are deleted and the nucleotides given in lower case are shifted into the variant hexamer. Thymidine is specified instead of uracil, referencing the genomic sequence, rather than the mRNA.

Figure 1

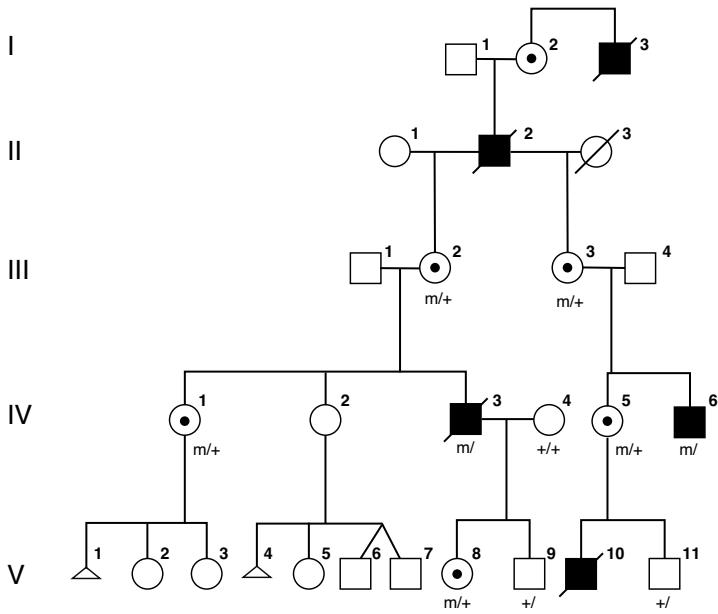
A

c.*43A>G



B

c.*39A>G



C

c.*40A>G

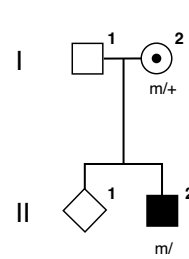


Figure 2

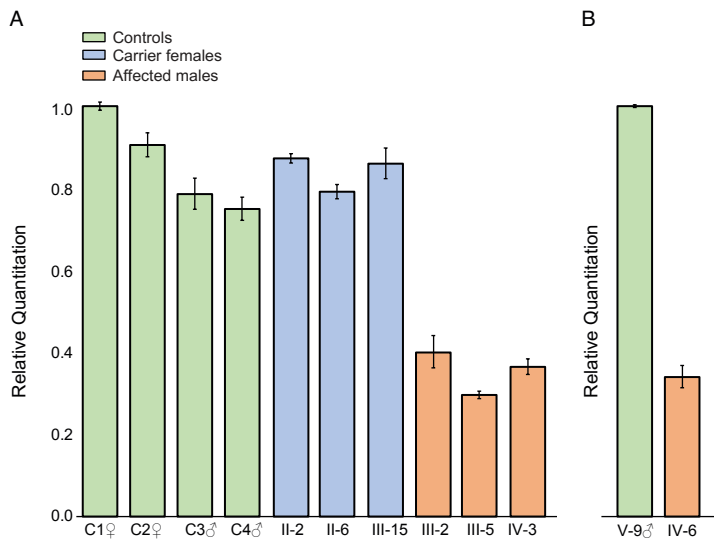


Figure 3

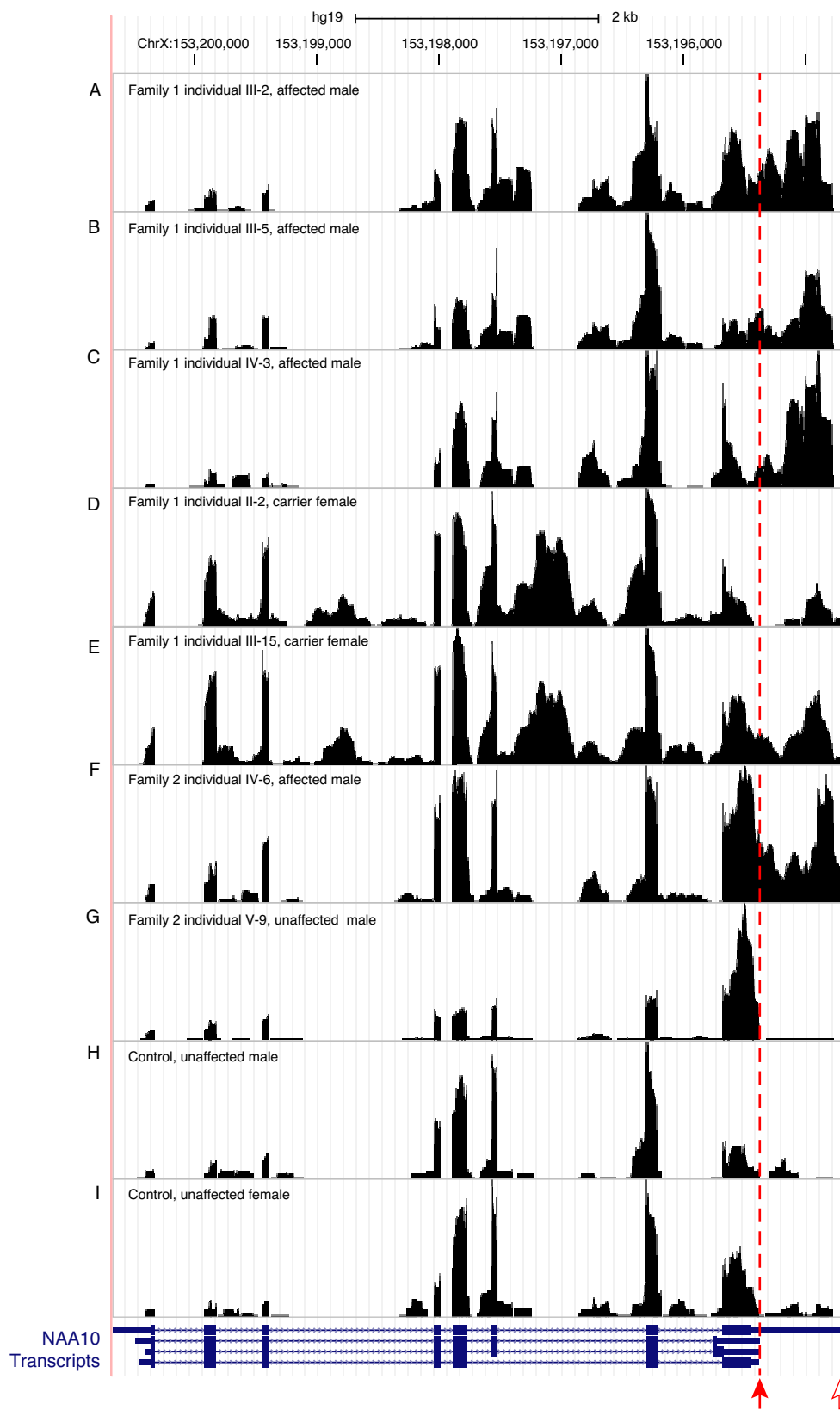


Figure 4

