REPORT

# *TAF1* Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations

Jason A. O'Rawe,[1,2] Yiyang Wu,[1,2] Max J. Dörfel,[1] Alan F. Rope,[3] P.Y. Billie Au,[4] Jillian S. Parboosingh,[4] Sungjin Moon,[5] Maria Kousi,[5] Konstantina Kosma,[6,7] Christopher S. Smith,[4] Maria Tzetis,[6,7] Jane L. Schuette,[8,9] Robert B. Hufnagel,[10,11] Carlos E. Prada,[10,12] Francisco Martinez,[13] Carmen Orellana,[13] Jonathan Crain,[1] Alfonso Caro-Llopis,[13] Silvestre Oltra,[13] Sandra Monfort,[13] Laura T. Jiménez-Barrón,[1,14] Jeffrey Swensen,[15] Sara Ellingwood,[16] Rosemarie Smith,[17] Han Fang,[1] Sandra Ospina,[18] Sander Stegmann,[19] Nicolette Den Hollander,[20] David Mittelman,[21] Gareth Highnam,[21] Reid Robison,[22,23] Edward Yang,[24] Laurence Faivre,[25,26] Agathe Roubertie,[27,28] Jean-Baptiste Rivière,[25] Kristin G. Monaghan,[29] Kai Wang,[30] Erica E. Davis,[5] Nicholas Katsanis,[5] Vera M. Kalscheuer,[31] Edith H. Wang,[32] Kay Metcalfe,[33] Tjitske Kleefstra,[34] A. Micheil Innes,[4] Sophia Kitsiou-Tzeli,[6] Monica Rosello,[13] Catherine E. Keegan,[8,9] and Gholson J. Lyon[1,2,22,*]

We describe an X-linked genetic syndrome associated with mutations in *TAF1* and manifesting with global developmental delay, intellectual disability (ID), characteristic facial dysmorphology, generalized hypotonia, and variable neurologic features, all in male individuals. Simultaneous studies using diverse strategies led to the identification of nine families with overlapping clinical presentations and affected by de novo or maternally inherited single-nucleotide changes. Two additional families harboring large duplications involving *TAF1* were also found to share phenotypic overlap with the probands harboring single-nucleotide changes, but they also demonstrated a severe neurodegeneration phenotype. Functional analysis with RNA-seq for one of the families suggested that the phenotype is associated with downregulation of a set of genes notably enriched with genes regulated by E-box proteins. In addition, knockdown and mutant studies of this gene in zebrafish have shown a quantifiable, albeit small, effect on a neuronal phenotype. Our results suggest that mutations in *TAF1* play a critical role in the development of this X-linked ID syndrome.

Transcription factor II D (TFIID) consists of TATA binding protein (TBP) and 12–14 TBP-associated factors (TAFs). TFIID promotes transcriptional initiation by recognizing promoter DNA and facilitating the nucleation of other general transcription factors for assembly into a functional pre-initiation complex,[1–5] and it also functions as a co-activator by interacting with transcriptional activators.[1] Subunits of TFIID have been suggested to play a possible role in neurodegenerative diseases and developmental delay when they are disrupted.[6–12] Indeed, variants in *TAF2* (MIM: 604912) and *TBP* (MIM: 600075) have been implicated in intellectual disability (ID) and developmental delay with or without corpus callosum hypoplasia.[10,11,13] Recent work toward understanding the molecular basis of Cornelia de Lange syndrome (CdLS [MIM: 122470, 300590, 300882, 614701, 610759]) has also suggested

[1]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA; [2]Graduate Program in Genetics, Stony Brook University, Stony Brook, NY 11794, USA; [3]Department of Medical Genetics, Northwest Kaiser Permanente, Portland, OR 97227, USA; [4]Department of Medical Genetics and Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T3B 6A8, Canada; [5]Center for Human Disease Modeling, Duke University, Durham, NC 27708, USA; [6]Department of Medical Genetics, Medical School, University of Athens, Athens 11527, Greece; [7]Research Institute for the Study of Genetic and Malignant Disorders in Childhood, Aghia Sophia, Children's Hospital, Athens 11527, Greece; [8]Department of Pediatrics, Division of Genetics, University of Michigan, Ann Arbor, MI 48109, USA; [9]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA; [10]Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA; [11]Unit on Pediatric, Developmental, and Genetic Ophthalmology, Ophthalmic Genetics and Visual Function Branch, National Eye Institute, NIH, Bethesda, MD 20892, USA; [12]Centro de Medicina Genómica y Metabolismo, Fundación Cardiovascular de Colombia, Bucaramanga, Santander 681004, Colombia; [13]Unidad de Genetica, Hospital Universitario y Politecnico La Fe, Valencia 46026, Spain; [14]Centro de Ciencias Genomicas, Universidad Nacional Autonoma de México, Cuernavaca, Morelos 62210, Mexico; [15]Caris Life Sciences, Phoenix, AZ 85040, USA; [16]Division of Genetics, Pediatric Specialty Care, Maine Medical Partners, ME 04102, USA; [17]Department of Pediatrics, Division of Genetics, Barbara Bush Children's Hospital, Maine Medical Center, Portland, ME 04102, USA; [18]Fundacion Cardioinfantil and Universidad del Rosario, Bogota 681004, Colombia; [19]Deptartment of Clinical Genetics, 6202 AZ Maastricht, the Netherlands; [20]Department of Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, the Netherlands; [21]Gene by Gene Ltd., Houston, TX 77008, USA; [22]Utah Foundation for Biomedical Research, Salt Lake City, UT 84107, USA; [23]Tute Genomics, 150 South 100 West Street, Provo, UT 84601, USA; [24]Department of Radiology, Boston Children's Hospital, Boston, MA 02115, USA; [25]Génétique des Anomalies du Développement EA4271, Université Bourgogne Franche-Comté, 21000 Dijon, France; [26]Centre de Génétique, Centre de Référence Maladies Rares "Anomalies du Développement et Syndromes Malformatifs," Centre Hospitalier Universitaire de Dijon, 21000 Dijon, France; [27]Institut des Neurosciences de Montpellier, Université de Montpellier, BP 74103, 34950 Montpellier Cedex 5, France; [28]Service de Neuropédiatrie, Centre Hospitalier Universitaire Gui de Chauliac, 34000 Montpellier, France; [29]GeneDx, Gaithersburg, MD 20877, USA; [30]Zilkha Neurogenetic Institute, Department of Psychiatry and Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA; [31]Research Group Development and Disease, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [32]Department of Pharmacology, University of Washington, Seattle, WA 98195, USA; [33]Clinical Genetics Service, Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester M13 9WL, UK; [34]Department of Human Genetics, Radboud University Medical Center, 6500 HB Nijmegen, the Netherlands

*Correspondence: gholsonjlyon@gmail.com

http://dx.doi.org/10.1016/j.ajhg.2015.11.005. 

that mutations in *TAF6* (MIM: 602955), encoding a component of TFIID, play an important role in the pathogenesis of this syndrome.[14] CdLS is a phenotypically and genetically heterogeneous syndrome characterized by distinct facial features, hirsutism, developmental delay, ID, and limb abnormalities,[15] and mutations in several different genes have been implicated in contributing to this heterogeneous clinical presentation.[16–19] Two mutations in *TAF6* were implicated in disease pathogenesis in some individuals with phenotypic features of CdLS and were shown, through biochemical investigations, to reduce binding of TAF6 to other core components of TFIID.[14]

A recent paper nominated *TAF1* (MIM: 313650) as a candidate gene for ID on the basis of segregation of missense variants in two different pedigrees; however, no clinical information other than ID was provided.[20] *TAF1* encodes the largest subunit of the TFIID complex and has been ranked 53$^{rd}$ among the top 1,003 constrained human genes in a recent population-scale study,[21] suggesting that TAF1 plays a critical role in normal cellular function. Previous work in *Drosophila* cells has shown that TAF1 depletion increases the magnitude of the initial transcription burst and causes delay in the shutoff of transcription upon removal of the stimulus.[22] The authors showed that the magnitude of the transcriptional response to the same signaling event, even at the same promoter, can vary greatly depending on the composition of the TFIID complex in the cell. In addition, and consistent with the notion that TAF1 is important in controlling the binding patterns of TFIID to specific promoter regions, this study showed that the set of genes conferring increased expression were enriched with TATA-containing promoters, suggesting an association between the depletion of TAF1 and increased expression of genes with the TATA motif. The genomic region containing *TAF1* has also been suggested to play an important role in X-linked dystonia-parkinsonism (XDP [MIM: 314250]), although the exact mechanism remains undetermined[6,8,23] (Dy et al., 2015, Movement Disorders, abstract). XDP is an X-linked recessive movement disorder characterized by adult-onset dystonia and parkinsonism, which lead to eventual death as a result of oropharyngeal dystonia or secondary infections.[24] Studies investigating the molecular basis of XDP have demonstrated aberrant neuron-specific expression levels of *TAF1* isoforms in neuronal tissue containing *TAF1* variants. Herzfeld et al. corroborated previous reports suggesting that a reduction in *TAF1* expression is associated with large-scale expression differences across hundreds of genes,[6] and studies in rat and mouse brain have also corroborated the importance and relevance of *TAF1* expression patterns specific to neuronal tissues.[7,9]

In this study, we describe a recognizable syndrome attributed to mutations in *TAF1*. This work represents a collaborative research effort between independent groups engaged in studying the molecular basis of human disease. A "genotype-first" approach[25] was used for finding families with variants in TAF1. This approach included phenotypic evaluations and screening of families with individuals harboring mutations in *TAF1*. This process was facilitated by the use of databases such as DECIPHER and reporting initial findings on the BioRxiv preprint server.[26] These efforts culminated in a study cohort of 14 affected individuals from 11 unrelated families. Twelve of these individuals (from nine unrelated families) contain single-nucleotide changes in *TAF1*, and two (from two unrelated families) have large duplications involving *TAF1*. Written informed consent was obtained from all study participants for all families, except in those instances when exomes were ordered through GeneDx on clinical grounds, and research was carried out in compliance with the Declaration of Helsinki.

Shared phenotypic features representing the cardinal characteristcs of this syndrome (see Figures 1 and 2, Tables 1, 2, and S5, and Movies S1, S2, S3, S4, S5, S6, and S7) include global developmental delay, ID, characteristic facial dysmorphologies, and generalized hypotonia. Shared facial dysmorphologies include prominent supraorbital ridges, down-slanted palpebral fissures, sagging cheeks, a long philtrum, low-set and protruding ears, a long face, a high palate, a pointed chin, and anteverted nares. The probands also share a characteristic gluteal crease with a sacral caudal remnant (Figure S13), although spine MRI on two probands did not show any major underlying defect (Figure S8). The affected individuals have generalized hypotonia, as well as joint hypermobility. Other widely shared features include hearing impairment, microcephaly, and hypoplasia of the corpus callosum (Figure S7 and Table S5). Interestingly, probands 8A (II-1 in family 8; Figure 2) and 11A (II-1 in family 11; Figure 3) are also affected by abnormal development of the thoracic cage. Some additional neurological features include spastic diplegia, dystonic movements, and tremors. Individuals 8A, 10A (IV-3 in family 10; Figure 3), and 11A have progressive symptoms, and one individual (11A) died of severe cardiopulmonary insufficiency attributed to an infection. Importantly, probands bearing duplications of *TAF1* (10A and 11A) not only demonstrate severe and progressive neurodegeneration but also do not share some of the more common features of the rest of the probands (see Table 1 and Figures 1 and 3 for comparison). Detailed clinical information is included in the Supplemental Note.

Several strategies were used for identifying candidate disease-related sequence variation. These included whole-genome sequencing, exome sequencing, targeted gene-panel sequencing, and microarray-based strategies (Figures S1–S5 and Tables S1–S4). Sanger sequencing was used for validating sequence variations. Many of the families studied here underwent genotyping via clinical microarrays or gene-specific sequencing for a small number of genomic regions. See the Supplemental Data for more-detailed descriptions of the sequencing and analysis methods used for each family. All 14 affected individuals were found to contain *TAF1* sequence variants, the majority of which (11 out of 14) are missense variants. One proband (II-1 in

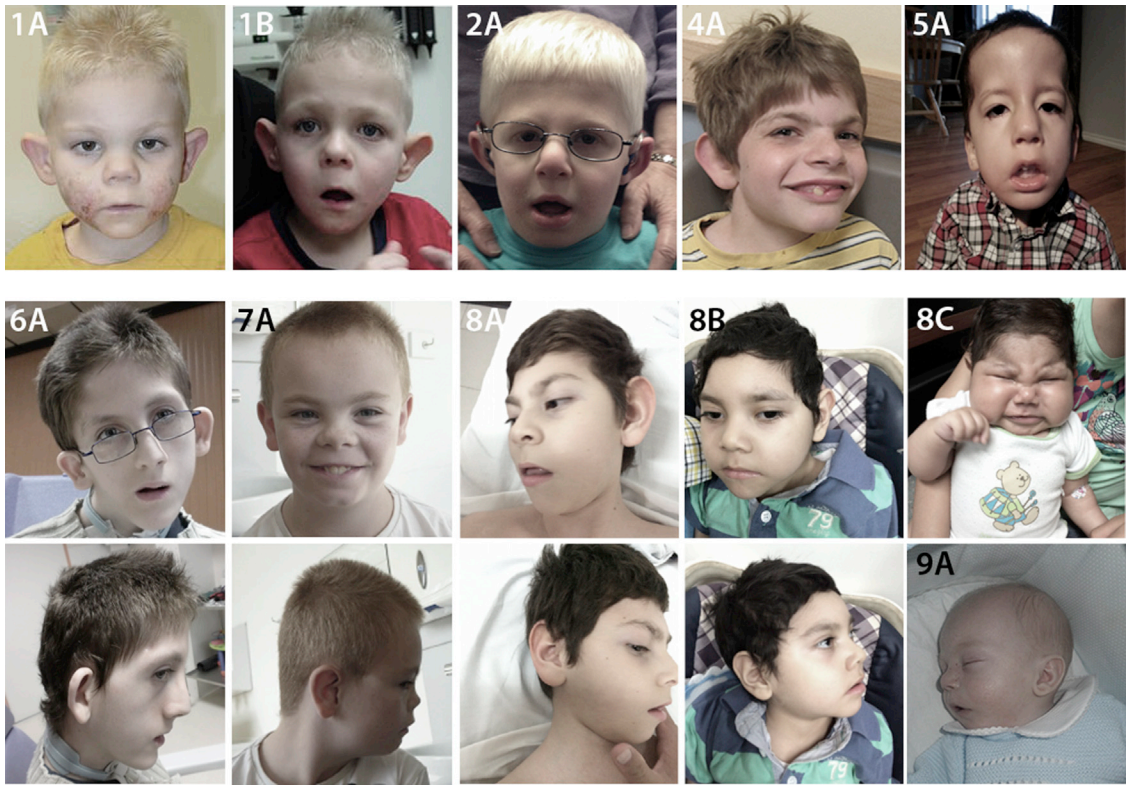


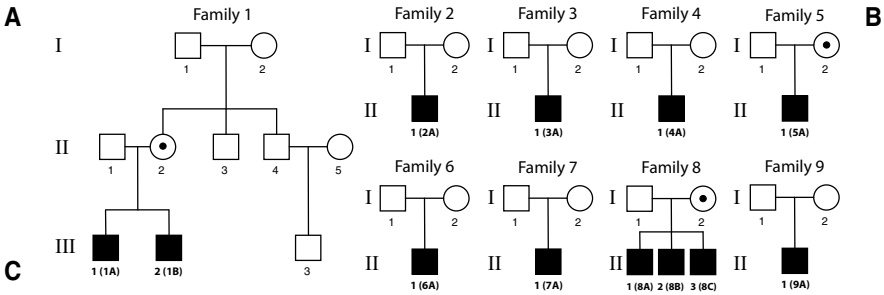

**Figure 1. Images of the Facial Phenotype from Families 1, 2, and 4–9**
Cardinal facial dysmorphologies include prominent supraorbital ridges (seen in 1A, 1B, 2A, 6A, 8A–8C, and 9A), down-slanted palpebral fissures (1A, 1B, 2A, 4A, 5A, 6A, 8A–8C, and 9A), sagging cheeks (1A, 1B, 5A, 8A–8C, and 9A), a long philtrum (1A, 1B, 2A, 4A, 5A, 8A–8C, and 9A), low-set and protruding ears (1A, 1B, 2A, 4A, 5A, 6A, 8A–8C, and 9A), a long face (1A, 1B, 2A, 5A, 6A, 8A–8C, and 9A), a high palate (5A, 6A, 8A–8C, and 9A), a pointed chin (1A, 1B, 2A, 4A, 5A, 6A, and 8A–8C), and anteverted nares (2A, 4A, 5A, 7A, 8A–8C, and 9A).

family 5; 5A in Figure 2) was found to have a variant that influences *TAF1* splicing, and two probands (10A and 11A) were found to have a duplication that involves *TAF1*. As stated above, the two probands with duplications exhibit less overlap with those who harbor single-nucleotide changes (see Table 1) and exhibit severe progressive neurologic impairment. All of the mutations reported here, including the duplications, are de novo or co-segregate with the phenotype in other affected male individuals (see Figures 2 and 3).

Families 1 and 10 were tested for X chromosome inactivation, which showed that female carriers of *TAF1* mutations and duplications demonstrate highly skewed inactivation (99:1). The female carriers in families 5 and 11 were not informative for the polymorphic CAG repeat in the human androgen-receptor gene.

All missense variants were found to affect evolutionarily conserved residues (Figure 2B) and were not present at any frequency in public databases such as dbSNP137, 1000 Genomes phase 1, the NHLBI Exome Sequencing Project Exome Variant Server (ESP6500), or ExAC Browser version 0.2, which contains allelic information derived from ~60,000 exome sequences. The *TAF1* missense variants were also predicted to be deleterious by a range of prediction software (CADD, SIFT, GERP++, and phyloP) (Figure 2B). The splice-site variant discovered in family 5 was not predicted to be deleterious by the prediction software listed. However, this variant does not change the amino acid content of the predicted protein; instead, it affects *TAF1* splicing in both the mother and the proband (Figure S11). In addition, the single-nucleotide variants (SNVs) described here fall within *TAF1* regions that are relatively sparsely covered by non-synonymous sequence variations reported in the ExAC Browser (Figure 2C). Indeed, four of these missense variants (c.2419T>C [p.Cys807Arg], c.2926G>C [p.Asp976His], c.3736C>T [p.Arg1246Trp], and c.3708A>G [p.Arg1228Ilefs*16]) were found to be within the two *TAF1* regions that are significantly underrepresented by non-synonymous sequence variation reported in the ExAC Browser when restricted to European and Latin populations (Figure 2C). Additional variants found in families 6 and 7 (c.4549A>C [p.Asn1517His] and c.4355G>A [p.Arg1431His]) also fall within a region that is underrepresented by non-synonymous sequence variation, although this last cluster is not statistically significant (p value = 0.29) (see Supplemental Data for further details).

Recent structural work in yeast points to an epigenetic role of the TAF1-TAF7 complex in general TFIID function and/or assembly of the pre-initiation complex (PIC).[4,5]




| Variant | Family | CADD | SIFT | GERP++ | phyloP |
|---|---|---|---|---|---|
| c.4010T>C | 1 | 27 | 0.003 | 5.6 | 7.695 |
| c.2419T>C | 2 | 23.7 | 0.004 | 4.7 | 7.691 |
| c.3736C>T | 3 | 35 | 0.000 | 5.54 | 2.216 |
| c.1514T>A | 4 | 29.9 | 0.000 | 5.51 | 7.655 |
| c.3708A>G | 5 | 6.17 | 1.000 | -0.087 | 1.681 |
| c.4549A>C | 6 | 20.8 | 0.000 | 4.65 | 8.910 |
| c.4355G>A | 7 | 29.9 | 0.001 | 4.17 | 9.259 |
| c.1786C>T | 8 | 29.3 | 0.008 | 5.97 | 7.408 |
| c.2926G>C | 9 | 30 | 0.000 | 4.81 | 9.412 |

**Figure 2. TAF1 Domains, Variant Scores, and ExAC Sequence Variation Plot**
(A) Pedigree drawings of the nine families who were found to harbor *TAF1* SNVs (NCBI Gene ID: 6872 according to the GRCh37.p13 assembly). Black dots indicate maternal carriers.
(B) All nine SNVs are listed and annotated with CADD, SIFT, GERP++, and phyloP scores (which indicate conservation across 99 vertebrate genomes and humans). All of the SNVs are considered to be potentially deleterious by all of the listed annotations, except for c.3708A>G, which is a splice-site variant and as a consequence is not necessarily expected to be categorized as deleterious by any of the listed scores, because it does not affect amino acid composition of the predicted protein.
(C) Known TAF1 domains are shown with respect to their corresponding genic positions. All but non-synonymous variants reported in the ExAC Browser for *TAF1* are plotted below as lines; white and gray indicate exon boundaries. Red lines indicate the relative positions of the eight missense variants described in this paper (see Table 2). Numerals link the sequence variants shown on the ExAC plot to their familial origin, and those noted with a star fall within *TAF1* regions that are significantly underrepresented by non-synonymous sequence variation in the ExAC Browser in European and Latin populations (p values of 0.032 and 0.037 for the first [c.2419T>C, c.2926G>C, and c.3736C>T] and second [c.3708A>G ] clusters, according to Cucala's hypothesis-free multiple scan statistic with a variable window[27]).

Human TAF1 is a 1,893 amino acid multifunctional protein that has been reported to possess activities related to DNA promoter binding, histone acetylation, and protein phosphorylation.[28–31] The histone acetyltransferase (HAT) activity of TAF1 can be blocked by TAF7 binding.[32,33] Studies suggest that phosphorylation of TAF7 at Ser264 causes release from TFIID and thus alleviates its inhibitor effect on TAF1.[34] Therefore, mutations that disrupt this inhibitory protein-protein interaction could have devastating effects on gene-expression profiles during human development. Intriguingly, four of the eight missense variants in *TAF1* change residues that are conserved in higher eukaryotes and map to domains important for TAF7 binding. Substitutions p.Cys807Arg, p.Pro596Ser, and p.Asp976His from families 2, 8, and 9, respectively, fall within an evolutionarily conserved central domain (DUF3591) that spans residues 586–1,049. DUF3591 encompasses the TAF1 HAT domain and numerous points of contact with TAF7.[35] The recently reported human TAF1-TAF7 crystal structure reveals that Cys807 is buried in the center of a heterodimeric triple barrel formed by segments of TAF1 and TAF7.[35] Replacement of the cysteine, a polar amino acid capable of disulfide-bond formation, with the large basic amino acid arginine (p.Cys807Arg) is predicted to destabilize the triple-barrel fold and potentially interfere with the interaction between TAF1 and TAF7. The p.Asp976His substitution also has the potential to disrupt TAF1-TAF7 binding. The acidic Asp976 maps to a separate TAF1-TAF7 protein interface within DUF3591 and undergoes intermolecular hydrogen binding in the highly conserved glycine-rich loop of TAF1. This loop interacts extensively with a highly conserved Arg-rich motif in TAF7. The acidic-to-basic amino acid change (p.Asp976His) has the potential to disrupt the architecture of the glycine-rich motif and its ability to effectively bind to TAF7. The p.Arg1246Trp substitution described in this work and the published p.Arg1190Cys TAF1 variant[20] reside in the RAP74-interacting domain (RAPiD) of TAF1 (residues 1,120–1,279), which also has been shown to be important for TAF7 binding.[35,36] The number of *TAF1* de novo missense mutations co-segregating with ID syndromes and predicted to affect TAF7 binding is quite striking, further strengthening the importance of TAF7 in the regulation of TAF1 function. TAF1 is a difficult protein to work with in isolation, and it is likely to be unstable in the absence of a binding partner, such as TAF7.[4,32,35] Thus, we speculate that variants that are not within any known protein domain (i.e., p.Ile1337Thr) might affect domain packing of TAF1, and this might interfere with the TAF1-TAF7 interacting surface[4,35] or mark the protein for proteolytic degradation.

Bromodomains are a common feature of transcription factors that possess HAT activity. TAF1 contains two bromodomains (Bromo1: 1,397–1,467; Bromo2: 1,520–1,590),

**Table 1. Summary of the Clinical Features of *TAF1* ID Syndrome**

| Features (Human Phenotype Ontology Nos.) | Proband | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1A | 1B | 2A | 3A | 4A | 5A | 6A | 7A | 8A | 8B | 8C | 9A | 10A[a] | 11A[a] |
| Sex | M | M | M | M | M | M | M | M | M | M | M | M | M | M |
| Age (years) | 15 | 13 | 5 | 6 | 9 | 3 | 22 | 11 | 9 | 4 | 1 | 3 | 16 | 8 |
| Postnatal growth retardation (HP: 0008897) | + | + | + | + | + | + | − | − | + | + | + | + | UK | + |
| Delayed gross motor development (HP: 0002194) | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Delayed speech and language development (HP: 0000750) | + | + | + | + | + | + | + | + | + | + | + | UK | + | + |
| Oral-pharyngeal dysphagia (HP: 0200136) | + | + | + | + | UK | + | UK | − | + | + | + | UK | + | UK |
| Prominent supraorbital ridges (HP: 0000336) | + | + | − | + | UK | − | + | − | + | + | + | + | + | + |
| Downslanted palpebral fissures (HP: 0000494) | + | + | + | − | + | + | + | − | + | + | + | − | + | UK |
| Sagging cheeks | + | + | − | − | − | + | − | − | + | + | + | + | + | + |
| Long philtrum (HP: 0000343) | + | + | + | + | + | + | − | + | + | + | + | + | − | − |
| Low-set ears (HP: 0000369) | + | + | + | + | + | + | + | − | + | + | + | + | − | + |
| Protruding ears (HP: 0000411) | + | + | + | + | + | + | + | − | + | + | + | − | − | + |
| Long face (HP: 0000276) | + | + | − | − | UK | + | + | − | + | + | + | + | + | + |
| High palate (HP: 0000218) | UK | UK | + | + | − | + | + | − | + | + | + | + | + | + |
| Pointed chin (HP: 0000307) | + | + | − | − | + | + | + | − | + | + | + | − | + | + |
| Anteverted nares (HP: 0000463) | − | − | + | + | + | + | − | + | + | + | + | + | − | + |
| Hearing impairment (HP: 0000365) | + | + | + | + | UK | + | − | − | + | + | + | − | UK | − |
| Chromic otitis media (HP: 0000389) | + | + | + | − | + | + | − | + | + | + | + | − | UK | − |
| Strabismus (HP: 0000486) | + | + | + | + | UK | + | + | − | + | − | − | + | + | − |
| Microcephaly (HP: 0000252) | + | + | + | + | + | − | − | + | + | + | + | + | − | − |
| Hypoplasia of the corpus callosum (HP: 0002079) | + | + | + | UK | + | + | + | UK | + | + | + | + | UK | − |
| Generalized hypotonia (HP: 0001290) | + | + | + | + | + | + | + | − | + | + | + | + | − | + |
| Unusual gluteal crease with sacral caudal remnant and sacral dimple (abnormal sacral segmentation [HP: 0008468] and prominent protruding coccyx [HP: 0008472]) | + | + | + | + | + | + | + | + | + | + | + | + | UK | − |
| Joint hypermobility (HP: 0001382) | + | + | − | + | UK | + | − | − | + | + | + | + | − | UK |
| Autistic behaviors (HP: 0000729) | + | + | + | − | UK | UK | + | + | + | + | + | − | + | + |
| Intellectual disability (HP: 0001249) | + | + | + | + | + | UK | + | + | + | + | + | + | + | + |

This table demonstrates clinical features shared by eight or more probands across all affected individuals in the families. See Table S5 for a more comprehensive phenotypic table that includes phenotypic and clinical idiosyncrasies. Abbreviations are as follows: M, male; and UK, unknown.
[a]Probands containing duplications; they are generally less similar to the probands containing SNVs and share fewer common clinical features.

each of which consists of a bundle of four alpha helices that form a hydrophobic pocket that can recognize acetylated lysines found on the N-terminal tails of histones.[37,38] A close examination of the published TAF1 bromodomain structure revealed that the p.Arg1431His substitution in family 7 is a surface residue on one face of the alpha helix. It forms hydrogen bonds with two residues on a nearby helix, most likely playing a supporting role in maintaining the bromodomain fold. Substitution p.Arg1431His could affect the stability of the acetyl-lysine ligand binding site in TAF1 and alter promoter recognition.

As an initial attempt to explore whether a biological signature could be seen on the level of the transcriptome for at least one of the mutations in the above families, we conducted RNA sequencing (RNA-seq) with RNA isolated from the blood of family 1. Blood was collected from the two probands, their parents, and the maternal grandparents. Except for the single blood draws for the grandparents, two blood draws were taken on separate days for each subject. RNA libraries were generated, and the final pooled library was sequenced on a HiSeq 2000 across three lanes (paired-end 100 bp). A mean of

**Table 2. Summary of *TAF1* Variants across All Affected Individuals in This Study**

| Proband | Inheritance | Genetic Background | *TAF1* Mutation (hg19) |
|---|---|---|---|
| 1A | maternal | European decent | chrX: g.70621541T>C (c.4010T>C; p.Ile1337Thr) |
| 1B | maternal | European decent | chrX: g.70621541T>C (c.4010T>C; p.Ile1337Thr) |
| 2A | de novo | European decent | chrX: g.70607243T>C (c.2419T>C; p.Cys807Arg) |
| 3A | de novo | European decent | chrX: g.70618477C>T (c.3736C>T; p.Arg1246Trp) |
| 4A | de novo | European decent | chrX: g.70601686T>A (c.1514T>A; p.Ile505Asn) |
| 5A | maternal | Ecuadorian | chrX: g.70618449A>G (c.3708A>G; r.[3708a>g; 3681_3708del28]; p.Arg1228Ilefs*16) |
| 6A | de novo | European decent | chrX: g.70643003A>C (c.4549A>C; p.Asn1517His) |
| 7A | de novo | British | chrX: g.70627912G>A (c.4355G>A; p.Arg1431His) |
| 8A | maternal | Colombian | chrX: g.70602671C>T (c.1786C>T; p.Pro596Ser) |
| 8B | maternal | Colombian | chrX: g.70602671C>T (c.1786C>T; p.Pro596Ser) |
| 8C | maternal | Colombian | chrX: g.70602671C>T (c.1786C>T; p.Pro596Ser) |
| 9A | de novo | Spanish | chrX: g.70612503G>C (c.2926G>C; p.Asp976His) |
| 10A | maternal | Albanian | 0.423 Mb duplication including *TAF1* and other genes at Xq13.1(70,370,794–70,794,385); deletion containing *KANSL1* and other genes at 17q21.31 (0.63 Mb) |
| 11A | de novo | Greek | 0.42 Mb duplication including *TAF1* and other genes: arr Xq13.1(70,287,519–70,711,110)×2 |

49,792,652 (SD = 11,666,119) properly paired reads were generated, and a mean spliced mapping percentage of 85.41% (SD = 7.1%) per sample was observed (Table S6). HISAT[39] was used for spliced alignment to UCSC Human Genome Browser hg19, and Stringtie[40] was used for quantifying known transcripts. Cuffdiff[41] was used for performing differential expression analysis, and CummeRbund[42] was used for analyzing, filtering, and visualizing the results from the Cuffdiff differential-expression analysis. Gene-set enrichment analyses were performed with WebGestalt,[43] which uses Molecular Signatures Database (MSigDB) for transcription-factor targets,[44] Kyoto Encyclopedia of Genes and Genomes (KEGG),[45] Gene Ontology (GO),[46] and Human Phenotype Ontology (HPO)[47] to obtain gene annotations and set-inclusion information. We used various analysis tools in the CummeRbund package[42] to evaluate the quality of our RNA-seq data and analysis for family 1 (Figure S10).

We found 213 genes (Table S7) to be differentially expressed between the two affected male probands from family 1 and their unaffected parents and grandparents. Of these 213 genes, 179 were expressed less (downregulated) and 34 were expressed more (upregulated) in the affected individuals. Among those genes that were upregulated in the affected individuals, 24 out of the 34 were expressed with a $\log_2$ fold-change value greater than 1, and no genes were found to be upregulated with a $\log_2$ fold-change value greater than 2. Among those genes that were downregulated in the affected individuals, 161 out of 179 were expressed with a $\log_2$ fold-change value less than −1, and 41 out of 179 were expressed with a $\log_2$ fold-change value less than −2.

Among the genes that were upregulated in the affected boys, we found no obvious interpretable biological signal. We present the results from the gene-set enrichment analysis performed on the set of genes that were downregulated in the affected individuals with a $\log_2$ fold-change value less than −1, because this set represents the genes that were significantly downregulated and to a potentially biologically relevant level. Transcription-factor-target enrichment analysis revealed a significant enrichment of genes regulated by E-box proteins (CANNTG promoter motifs, Benjamini-Hochberg [BH]-corrected p value of 0.0052). KEGG-pathway enrichment analysis revealed an enrichment of genes involved in Parkinson, Alzheimer, and Huntington diseases and cardiac-muscle contraction (BH-corrected p values of 2.95e−7, 1.54e−6, 2.13e−6, and 2.13e−6, respectively). Enrichments in genes associated with HPO annotations included type 1 muscle-fiber predominance, and laxity of ankles, fingers, and wrists (all four phenotypes were reported with a BH-corrected p value of 0.0481).

It is important to note that the RNA-seq results are preliminary and potentially confounded by age- or sex-specific expression differences between the affected and unaffected groups, and the results here were derived from a single family (because these were the samples that we were able to collect to date). Complete blood counts were not performed on the blood samples used for RNA-seq; thus, our results could also be confounded by secondary differences in mRNA abundances. The RNA-seq and whole-genome sequencing data have been deposited to the Sequence Read Archive (SRA) under BioProject ID PRJNA301337.

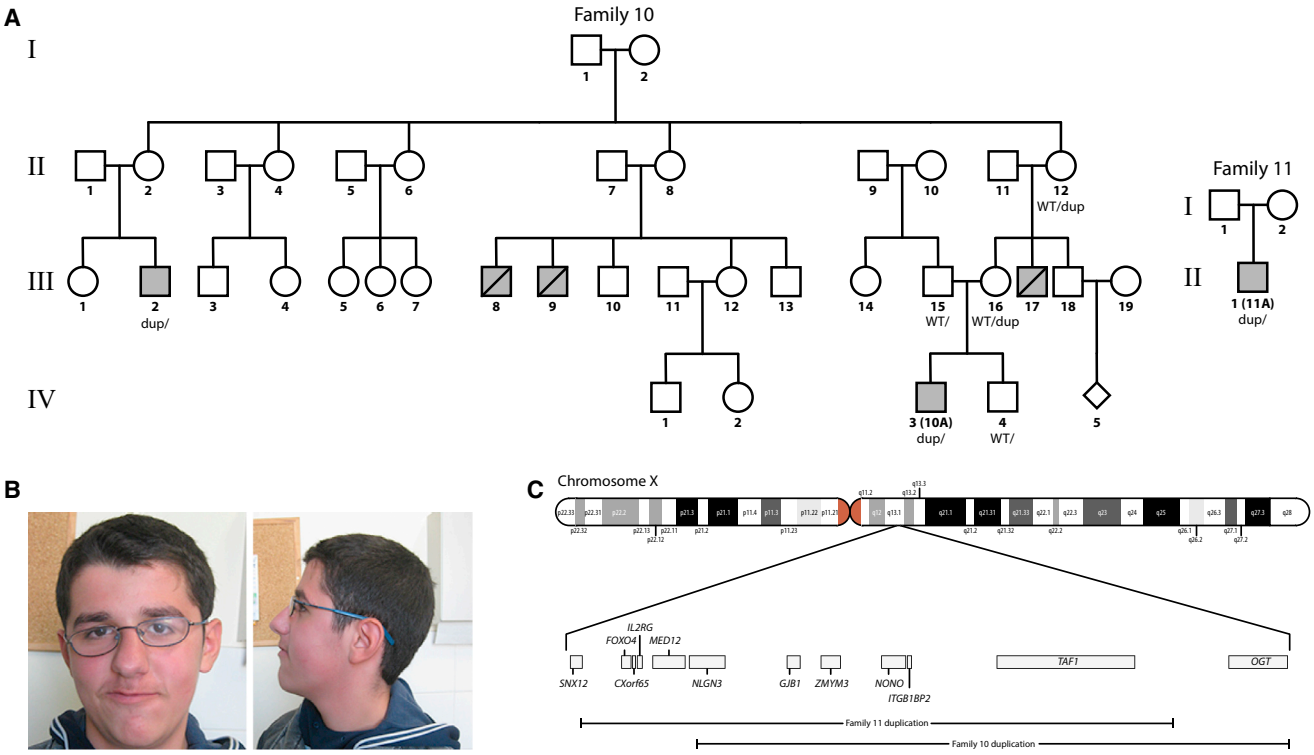



**Figure 3. Duplications Involving *TAF1* from Families 10 and 11**
(A) Pedigree drawings of families 10 and 11.
(B) The facial phenotype of proband 10A is notable for prominent supraorbital ridges, down-slanted palpebral fissures, sagging cheeks, a long face, a high palate, and a pointed chin.
(C) Chromosome X cytobands are plotted above a more focused view of the region containing duplications that involve *TAF1* in families 10 and 11. UCSC refGenes (from the UCSC Genome Browser tables) whose canonical transcript start or stop sites overlap either of the two duplications are plotted.

To evaluate structural CNS defects, we performed in vivo functional modeling of TAF1 by using transient knockdown in developing zebrafish embryos. The optic tectum is a neuroanatomical structure that occupies the majority of the space within the midbrain, and its size is a proxy for head size (e.g., microcephaly,[48] which is one of the most robustly observed phenotypes).

We designed a splice-blocking morpholino (MO) targeting the donor site of exon 9 of the sole zebrafish *taf1* ortholog (Ensembl: ENSDART00000051196; Figure S12) and performed stainings for acetylated tubulin to evaluate the structure of the optic tectum. In total, we analyzed 134 control embryos, 133 embryos injected with *taf1* MO, 109 embryos injected with *taf1* MO and wild-type (WT) *TAF1* RNA, and 78 embryos injected with WT *TAF1* mRNA by measuring the area of the optic tectum 3 days after fertilization. We found that the relative area of the optic tectum was reduced by approximately 10% in embryos injected with a MO targeting the donor site of exon 9 of *D. rerio taf1* ($p < 0.0001$; Figures 4A and 4B and S12). The effect was specific to the MO knockdown, given that the phenotype could be reliably restored by co-injection of MO and WT human *TAF1* mRNA ($p < 0.0001$). Notably, overexpression of WT human *TAF1* mRNA did not result in a phenotype that was significantly different from that of the controls ($p = 0.79$).

To confirm these findings, we also used CRISPR/Cas9 to disrupt *taf1*. For the CRISPR experiments, *taf1* guide RNA was generated as described.[49] In agreement with our observation that the area of the optic tectum was smaller in embryos injected with *taf1* MO than in control embryos, F0 embryos with CRISPR-disrupted *taf1* showed a smaller area of this structure than did un-injected controls ($p < 0.001$) (Figures 4C and S12). Unfortunately, the relatively small effects seen with MO knockdown and CRISPR-mediated mosaicism in F0 embryos precluded us from being able to show differences between WT and mutated *TAF1* constructs in terms of rescuing the neuroanatomical defect (data not shown).

In conclusion, we have presented evidence showing that mutations involving *TAF1* contribute to the phenotype described here. Differences in genetic background and the environment can certainly account for the phenotypic differences among the various males in these families.[50–56] Other studies also suggest a functional link between developmental delay and the TFIID multi-protein complex,[10–12] although the phenotypic variability and expression of other variants in *TAF1* and in other TAFs remains to be determined. We have also provided the results of initial studies

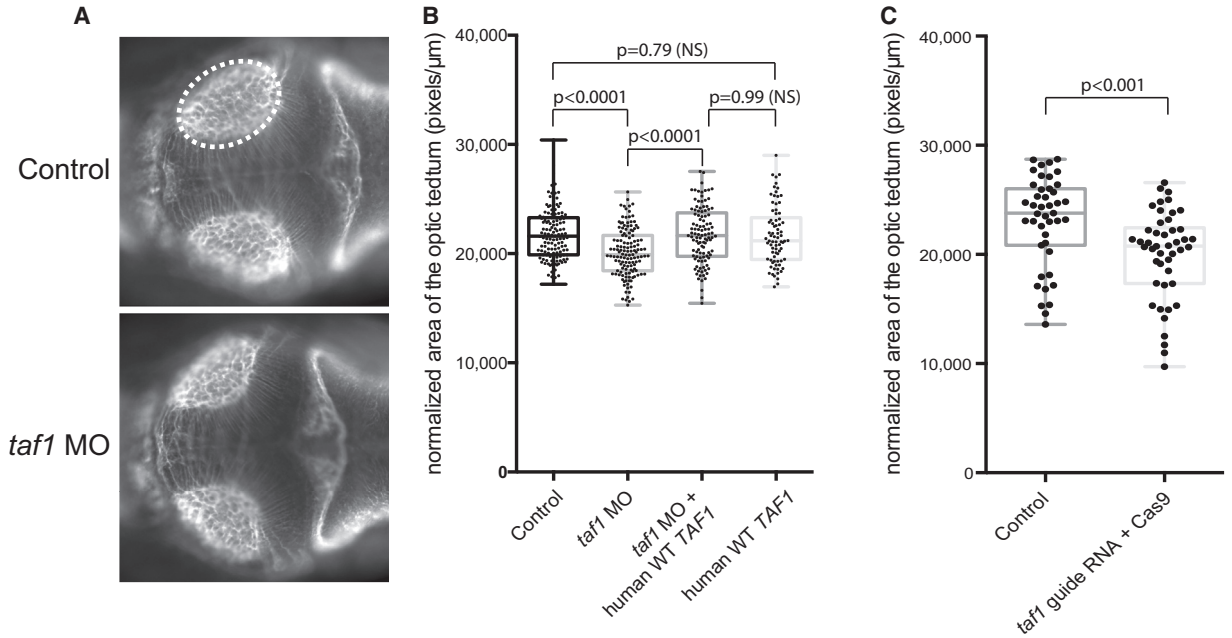



**Figure 4. Suppression or Genetic Mutation of Endogenous *taf1* Induces Decreased Size of the Optic Tectum In Vivo**
(A) Dorsal view of a control embryo (top) and an embryo injected with a morpholino (MO) targeting the donor site of exon 9 of *D. rerio taf1* 3 days after fertilization. An antibody against α-acetylated tubulin was used for visualizing the axon tracts in the brain of evaluated embryos. The assay consisted of measuring the area of the optic tectum (highlighted with the dashed ellipse), a neuroanatomical structure that occupies the majority of the space within the midbrain.
(B) A boxplot shows quantitative differences in the size of the optic tectum for each condition tested across three biological replicates. Suppression of *taf1* consistently induced a decrease of ~10% in the relative area of the optic tectum ($p < 0.0001$). The MO phenotype could be restored by co-injection of MO and wild-type (WT) human *TAF1* mRNA ($p < 0.0001$), denoting the specificity of the phenotype due to *taf1* suppression. Overexpression of WT human *TAF1* mRNA alone did not induce a phenotype that was significantly different from that of controls ($p = 0.79$). The numbers of embryos evaluated per condition were as follows: control, 134; *taf1* MO, 133; *taf1* MO + WT *TAF1* RNA, 109; and WT *TAF1* RNA, 78.
(C) A boxplot shows quantitative differences in the size of the optic tectum between uninjected controls and F0 embryos with CRISPR-disrupted *taf1*. The phenotype observed for both MO-injected embryos and embryos with CRISPR-disrupted *taf1* was concordant and reproducible across different experiments and across the two different methodologies. The p values were calculated with a Student's t test.

that suggests a possible regulatory role for at least one variant presented here and have shown that zebrafish knockdown and mutant studies for this gene have demonstrated a quantifiable, albeit small, effect on a neuronal phenotype. In our current study, we did not see a clear mechanistic link between the missense and splice-site variants in some of the probands and the duplications in two of the probands, and the probands with duplications share fewer phenotypic features with the rest of the disease cohort. There are many possible explanations for why a duplication of a gene might mimic the effect of a disruptive missense or splice-site variant, and our current study lacks the necessary experimental evidence to point to any particular scenario with certainty. Further work is needed for teasing apart the contributions of these duplications and their constituent genes to this more complex phenotype. It should also be noted that exome and/or whole-genome sequencing has not yet been performed in family 10 or 11, so other mutations could be contributing to the phenotypes of the probands.

## Accession Numbers

The RNA-seq and whole-genome sequencing data reported in this paper have been deposited to the Sequence Read Archive (SRA) under BioProject ID PRJNA301337. The study accession number is SRA: SRP065848, and the sample accession numbers are SRA: SRS1153874, SRS1153892, SRS1153893, SRS1153896, SRS1153897, SRS1153898, SRS1153899, SRS1153900, SRS1153901, SRS1153902.

## Supplemental Data

Supplemental Data include Supplemental Material and Methods, a Supplemental Note, 13 figures, 7 tables, and 8 movies and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.11.005.

## Conflicts of Interest

G.J.L. serves on advisory boards for GenePeeks Inc. and Omicia Inc. and is a consultant to Good Start Genetics. D.M. was until recently an employee and shareholder of Gene By Gene Inc. and is currently chief scientific officer of Tute Genomics Inc. K.W. is a board member and shareholder of Tute Genomics Inc., and R.R. is an employee, chief executive officer, and shareholder of Tute Genomics Inc. This study did not involve these companies and did not use products from these companies. The sequencing service by Complete Genomics was provided via a data-analysis grant, but the company was not involved in any subsequent analyses.

## Acknowledgments

G.J.L. is supported by funds from the Stanley Institute for Cognitive Genomics at Cold Spring Harbor Laboratory. K.W. is supported by NIH grant HG006465. The sequencing by Complete Genomics was provided by a data-analysis grant to K.W. We would like to thank Megan Cho for facilitating discovery of affected individuals through GeneDx. Margaret Yoon and David Tegay assisted with Human Phenotype Ontology annotation. A.M.I. and J.S.P. were supported by a grant from the Alberta Children's Hospital Foundation and would like to thank Francois Bernier and Ryan Lamont for helpful discussions and contributions. F.M., M.R., and colleagues are supported by grant PI14/00350 (Instituto de Salud Carlos III Acción Estratégica en Salud 2013–2016; Fondo Europeo de Desarrollo Regional). This work was supported in part by funding from grants from the Netherlands Organization for Health Research and Development, ZonMw (grant 907-00-365 to T.K.), and an institutional award from Cincinnati Children's Hospital (R.B.H.). N.K. is a distinguished George W. Brumley Professor. The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about.

Received: July 27, 2015
Accepted: November 5, 2015
Published: December 3, 2015

## Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/data
ANNOVAR, http://annovar.openbioinformatics.org/en/latest/
dbSNP, http://www.ncbi.nlm.nih.gov/snp
DECIPHER, http://decipher.sanger.ac.uk/
Ensembl, http://www.ensembl.org
Exome Aggregation Consortium (ExAC) Browser, http://exac.broadinstitute.org/
GeneSplicer, https://ccb.jhu.edu/software/genesplicer/
Golden Helix GenomeBrowse, http://www.goldenhelix.com/GenomeBrowse/index.html
HomoloGene, http://www.ncbi.nlm.nih.gov/homologene
Human Gene Mutation Database, http://www.hgmd.org/
Human Splicing Finder, http://www.umd.be/HSF/
MaxEntScan, http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.ncbi.nlm.nih.gov/omim
PennCNV, http://penncnv.openbioinformatics.org/en/latest/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
SIFT, http://sift.jcvi.org/

## References

1. Papai, G., Weil, P.A., and Schultz, P. (2011). New insights into the function of transcription factor TFIID from recent structural studies. Curr. Opin. Genet. Dev. *21*, 219–224.
2. Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. Cell *152*, 120–131.
3. Kotani, T., Miyake, T., Tsukihashi, Y., Hinnebusch, A.G., Nakatani, Y., Kawaichi, M., and Kokubo, T. (1998). Identification of highly conserved amino-terminal segments of dTAFII230 and yTAFII145 that are functionally interchangeable for inhibiting TBP-DNA interactions in vitro and in promoting yeast cell growth in vivo. J. Biol. Chem. *273*, 32254–32264.
4. Bhattacharya, S., Lou, X., Hwang, P., Rajashankar, K.R., Wang, X., Gustafsson, J.A., Fletterick, R.J., Jacobson, R.H., and Webb, P. (2014). Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D. Proc. Natl. Acad. Sci. USA *111*, 9103–9108.
5. Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. Nature *493*, 699–702.
6. Herzfeld, T., Nolte, D., Grznarova, M., Hofmann, A., Schultze, J.L., and Müller, U. (2013). X-linked dystonia parkinsonism syndrome (XDP, lubag): disease-specific sequence change DSC3 in TAF1/DYT3 affects genes in vesicular transport and dopamine metabolism. Hum. Mol. Genet. *22*, 941–951.
7. Jambaldorj, J., Makino, S., Munkhbat, B., and Tamiya, G. (2012). Sustained expression of a neuron-specific isoform of the Taf1 gene in development stages and aging in mice. Biochem. Biophys. Res. Commun. *425*, 273–277.
8. Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., Matsumoto, S., Tabuena, M.D., Maranon, E., Dantes, M., et al. (2007). Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am. J. Hum. Genet. *80*, 393–406.
9. Sako, W., Morigaki, R., Kaji, R., Tooyama, I., Okita, S., Kitazato, K., Nagahiro, S., Graybiel, A.M., and Goto, S. (2011). Identification and localization of a neuron-specific isoform of TAF1 in rat brain: implications for neuropathology of DYT3 dystonia. Neuroscience *189*, 100–107.
10. Rooms, L., Reyniers, E., Scheers, S., van Luijk, R., Wauters, J., Van Aerschot, L., Callaerts-Vegh, Z., D'Hooge, R., Mengus, G., Davidson, I., et al. (2006). TBP as a candidate gene for mental retardation in patients with subtelomeric 6q deletions. Eur. J. Hum. Genet. *14*, 1090–1096.
11. Hellman-Aharony, S., Smirin-Yosef, P., Halevy, A., Pasmanik-Chor, M., Yeheskel, A., Har-Zahav, A., Maya, I., Straussberg, R., Dahary, D., Haviv, A., et al. (2013). Microcephaly thin corpus callosum intellectual disability syndrome caused by mutated TAF2. Pediatr. Neurol. *49*, 411–416.e1.
12. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature *478*, 57–63.
13. Abu-Amero, K.K., Hellani, A., Salih, M.A., Al Hussain, A., al Obailan, M., Zidan, G., Alorainy, I.A., and Bosley, T.M. (2010). Ophthalmologic abnormalities in a de novo terminal 6q deletion. Ophthalmic Genet. *31*, 1–11.
14. Yuan, B., Pehlivan, D., Karaca, E., Patel, N., Charng, W.-L., Gambin, T., Gonzaga-Jauregui, C., Sutton, V.R., Yesil, G., Bozdogan, S.T., et al. (2015). Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. J. Clin. Invest. *125*, 636–651.
15. Deardorff, M.A., and Krantz, I.D. (2008). NIPBL and SMC1L1 (now SCM1A) and the Cornelia de Lange Syndrome. In Inborn Errors of Development, C.J. Epstein, R.P. Erickson, and A. Wynshao-Boris, eds. (Oxford University Press), pp. 1020–1031.

16. Gil-Rodríguez, M.C., Deardorff, M.A., Ansari, M., Tan, C.A., Parenti, I., Baquero-Montoya, C., Ousager, L.B., Puisac, B., Hernández-Marcos, M., Teresa-Rodrigo, M.E., et al. (2015). De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. Hum. Mutat. *36*, 454–462.
17. Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., et al. (2015). Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat. Genet. *47*, 338–344.
18. Kline, A.D., Calof, A.L., Schaaf, C.A., Krantz, I.D., Jyonouchi, S., Yokomori, K., Gauze, M., Carrico, C.S., Woodman, J., Gerton, J.L., et al. (2014). Cornelia de Lange syndrome: further delineation of phenotype, cohesin biology and educational focus, 5th Biennial Scientific and Educational Symposium abstracts. Am. J. Med. Genet. A. *164A*, 1384–1393.
19. Mannini, L., Cucco, F., Quarantotti, V., Krantz, I.D., and Musio, A. (2013). Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. Hum. Mutat. *34*, 1589–1596.
20. Hu, H., Haas, S.A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A.P., Weinert, S., Froyen, G., Frints, S.G., Laumonnier, F., et al. (2015). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. Mol. Psychiatry. Published online February 3, 2015. http://dx.doi.org/10.1038/mp.2014.193.
21. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950.
22. Pennington, K.L., Marr, S.K., Chirn, G.W., and Marr, M.T., 2nd. (2013). Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. Proc. Natl. Acad. Sci. USA *110*, 7678–7683.
23. Domingo, A., Westenberger, A., Lee, L.V., Brænne, I., Liu, T., Vater, I., Rosales, R., Jamora, R.D., Pasco, P.M., Cutiongco-Dela Paz, E.M., et al. (2015). New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3). Eur. J. Hum. Genet. *23*, 1334–1340.
24. Nolte, D., Niemann, S., and Müller, U. (2003). Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism. Proc. Natl. Acad. Sci. USA *100*, 10347–10352.
25. Stessman, H.A., Bernier, R., and Eichler, E.E. (2014). A genotype-first approach to defining the subtypes of a complex disease. Cell *156*, 872–877.
26. O'Rawe, J., Wu, Y., Rope, A., Barrón, L.T.J., Swensen, J., Fang, H., Mittelman, D., Highnam, G., Robison, R., Yang, E., et al. (2015). A variant in TAF1 is associated with a new syndrome with severe intellectual disability and characteristic dysmorphic features. bioRxiv, http://dx.doi.org/10.1101/014050.
27. Cucala, L. (2008). A hypothesis-free multiple scan statistic with variable window. Biom. J. *50*, 299–310.
28. Mizzen, C.A., Yang, X.-J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261–1270.
29. Lee, D.-H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol. Cell. Biol. *25*, 9674–9686.
30. Dikstein, R., Ruppert, S., and Tjian, R. (1996). TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. Cell *84*, 781–790.
31. Chalkley, G.E., and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II) 150 complex recognizes the initiator. EMBO J. *18*, 4835–4845.
32. Gegonne, A., Weissman, J.D., and Singer, D.S. (2001). TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. Proc. Natl. Acad. Sci. USA *98*, 12432–12437.
33. Chiang, C.M., and Roeder, R.G. (1995). Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science *267*, 531–536.
34. Kloet, S.L., Whiting, J.L., Gafken, P., Ranish, J., and Wang, E.H. (2012). Phosphorylation-dependent regulation of cyclin D1 and cyclin A gene transcription by TFIID subunits TAF1 and TAF7. Mol. Cell. Biol. *32*, 3358–3369.
35. Wang, H., Curran, E.C., Hinds, T.R., Wang, E.H., and Zheng, N. (2014). Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. Cell Res. *24*, 1433–1444.
36. Ruppert, S., and Tjian, R. (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. Genes Dev. *9*, 2747–2755.
37. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., et al. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell *149*, 214–231.
38. Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422–1425.
39. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.
40. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.
41. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.
42. L. Goff, C.T., D. Kelley. (2013). cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2100, http://bioconductor.org/packages/release/bioc/html/cummeRbund.html.
43. Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. *41*, W77–W83.
44. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

45. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–D205.
46. Consortium, T.G.O.; Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056.
47. Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. *42*, D966–D974.
48. Borck, G., Hög, F., Dentici, M.L., Tan, P.L., Sowada, N., Medeira, A., Gueneau, L., Thiele, H., Kousi, M., Lepri, F., et al. (2015). BRF1 mutations alter RNA polymerase III-dependent transcription and cause neurodevelopmental anomalies. Genome Res. *25*, 155–166.
49. Jao, L.-E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc. Natl. Acad. Sci. USA *110*, 13904–13909.
50. Hogben, L.T. (1933). Nature and nurture (W.W. Norton Company).
51. Weldon, W.F.R. (1902). Mendel's laws of alternative inheritance in peas. Biometrika *1*, 228–254.
52. Carayol, J., Schellenberg, G.D., Dombroski, B., Amiet, C., Génin, B., Fontaine, K., Rousseau, F., Vazart, C., Cohen, D., Frazier, T.W., et al. (2014). A scoring strategy combining statistics and functional genomics supports a possible role for common polygenic variation in autism. Front. Genet. *5*, 33.
53. Lyon, G.J., and O'Rawe, J. (2015). Human genetics and clinical aspects of neurodevelopmental disorders. In The Genetics of Neurodevelopmental Disorders, K. Mitchell, ed. (Wiley).
54. Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R., Ayroles, J.F., Duncan, L., Jordan, K.W., Lawrence, F., Magwire, M.M., et al. (2012). Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc. Natl. Acad. Sci. USA *109*, 15553–15559.
55. Mackay, T.F., and Moore, J.H. (2014). Why epistasis is important for tackling complex human disease genetics. Genome Med. *6*, 124.
56. Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. Nature *494*, 234–237.