

# Androgens mediate sexual dimorphism in Pilarowski-Bjornsson Syndrome

Kimberley Jade Anderson<sup>1</sup>, Eirny Tholl Thorolfsson<sup>1</sup>, Ilana M. Nodelman<sup>2</sup>, Sara Tholl Halldorsdottir<sup>3</sup>, Stefania Benonisdottir<sup>4,5</sup>, Malak Alghamdi<sup>6</sup>, Naif Almontashiri<sup>7,8</sup>, Brenda J. Barry<sup>9,10,11</sup>, Matthias Begemann<sup>12</sup>, Jacquelyn F. Britton<sup>13</sup>, Sarah Burke<sup>14</sup>, Benjamin Cogne<sup>15,16</sup>, Ana S.A. Cohen<sup>17,18</sup>, Carles de Diego Boguñá<sup>19</sup>, Evan E. Eichler<sup>20,21</sup>, Elizabeth C. Engle<sup>9,10,11,22,23,24,25</sup>, Jill A. Fahrner<sup>13,26</sup>, Laurence Faivre<sup>27,28</sup>, Mélanie Fradin<sup>29</sup>, Nico Fuhrmann<sup>30</sup>, Christine W. Gao<sup>13,31</sup>, Gunjan Garg<sup>32,33,34</sup>, Dagmar Grečmalová<sup>35</sup>, Mina Grippa<sup>36</sup>, Jacqueline R. Harris<sup>13,37</sup>, Kendra Hoekzema<sup>21</sup>, Tova HersHKovitz<sup>38</sup>, Sydney Hubbard<sup>14</sup>, Katrien Janssens<sup>39</sup>, Julie A. Jurgens<sup>9,10,22,23</sup>, Stanislav Kmoch<sup>40</sup>, Cordula Knopp<sup>12</sup>, Meral Aktas Koptagel<sup>12</sup>, Farah A. Ladha<sup>41</sup>, Pablo Lapunzina<sup>42</sup>, Tobias Lindau<sup>43</sup>, Marije Meuwissen<sup>39</sup>, Andreina Minicucci<sup>44</sup>, Emily Neuhaus<sup>45</sup>, Mathilde Nizon<sup>46</sup>, Lenka Nosková<sup>40</sup>, Kristen Park<sup>47</sup>, Chirag Patel<sup>48</sup>, Rolph Pfundt<sup>49</sup>, Pankaj Prasun<sup>50</sup>, Nils Rahner<sup>51</sup>, Nathaniel H. Robin<sup>52</sup>, Carey Ronspies<sup>53</sup>, Jasmin Roohi<sup>54</sup>, Jill Rosenfeld<sup>55,41</sup>, Margarita Saenz<sup>47</sup>, Carol Saunders<sup>17,18</sup>, Zornitza Stark<sup>56</sup>, Isabelle Thiffault<sup>17,18</sup>, Sarah Thull<sup>30</sup>, Danita Velasco<sup>53</sup>, Clara Velmans<sup>30</sup>, Jolijn Verseput<sup>57</sup>, Antonio Vitobello<sup>27</sup>, Tianyun Wang<sup>58,59,60</sup>, Karin Weiss<sup>61</sup>, Ingrid M. Wentzensen<sup>62</sup>, Genay Pilarowski<sup>63</sup>, Thor Eysteinnsson<sup>64,65</sup>, Madelyn Gillentine<sup>66</sup>, Kári Stefánsson<sup>67,68</sup>, Agnar Helgason<sup>67,69</sup>, Gregory D. Bowman<sup>2</sup>, Hans Tomas Bjornsson<sup>1,3,13,26 \*</sup>

\*Corresponding author. Email: [htb@hi.is](mailto:htb@hi.is)

<sup>1</sup>Department of Genetics and Molecular Medicine, Landspítali University Hospital, Reykjavik, Iceland.

<sup>2</sup>T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA.

<sup>3</sup>The Louma G. Laboratory of Epigenetic Research, Faculty of Medicine, University of Iceland, Reykjavik, Iceland.

<sup>4</sup>Institute of Physical Sciences, University of Iceland, Reykjavik, Iceland.

<sup>5</sup>Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford and Nuffield College, Oxford, UK.

<sup>6</sup>Medical Genetics Division, Pediatric Department, College of Medicine, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia.

<sup>7</sup>College of Applied Medical Sciences and Center for Genetics and Inherited Diseases, Taibah University, Madinah, Kingdom of Saudi Arabia.

<sup>8</sup>Research Department, King Khaled Eye Specialist Hospital, Riyadh, Saudi Arabia.

<sup>9</sup>F.M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA, USA.

<sup>10</sup>Department of Neurology, Boston Children's Hospital, Boston, MA, USA.

<sup>11</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA.

- <sup>12</sup>Institute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University, Aachen, Germany.
- <sup>13</sup>McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
- <sup>14</sup>Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, Nebraska, USA.
- <sup>15</sup>Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France.
- <sup>16</sup>Nantes Université, CHU de Nantes, Service de Génétique médicale, F-44000 Nantes, France.
- <sup>17</sup>Department of Pathology and Laboratory Medicine, Genomic Medicine Center, Children's Mercy- Kansas City, Kansas City, MO, USA
- <sup>18</sup>The University of Missouri-Kansas City, School of Medicine, Kansas City, MO, USA.
- <sup>19</sup>Servicio de Genética, Hospital Universitario de Toledo, Toledo, Spain.
- <sup>20</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
- <sup>21</sup>Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA.
- <sup>22</sup>Department of Neurology, Harvard Medical School, Boston, MA, USA.
- <sup>23</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA.
- <sup>24</sup>Department of Ophthalmology, Boston Children's Hospital, Boston, MA, USA.
- <sup>25</sup>Department of Ophthalmology, Harvard Medical School, Boston, MA, USA.
- <sup>26</sup>Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
- <sup>27</sup>INSERM UMR1231, GAD team, Université de Bourgogne Europe, Dijon, France.
- <sup>28</sup>Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, de l'Interrégion Est, Centre Hospitalier Universitaire Dijon, 21079 Dijon, France.
- <sup>29</sup>Service de Genetique Medicale, Centre Labellisé Anomalies du Développement de l'Ouest, CHU Rennes, Rennes, France.
- <sup>30</sup>Institute of Human Genetics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.
- <sup>31</sup>Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.
- <sup>32</sup>Department of Clinical Genetics, Liverpool Hospital, Sydney, New South Wales, Australia.
- <sup>33</sup>Hunter Genetics, Waratah, New South Wales, Australia.
- <sup>34</sup>School of Women's and Children's Health, University of New South Wales, Sydney, New South Wales, Australia.
- <sup>35</sup>Institute of Molecular and Clinical Pathology and Medical Genetics, University Hospital Ostrava, Czech Republic.
- <sup>36</sup>SSD Medical Genetics, Maternal and Child Department, AOU Policlinico Modena, Modena, Italy.
- <sup>37</sup>Kennedy Krieger Institute, Department of Neurology, Baltimore, Maryland, USA.
- <sup>38</sup>The Genetics Institute, Galilee Medical Center, Nahriya, Israel.
- <sup>39</sup>Center of Medical Genetics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium.
- <sup>40</sup>Research Unit for Rare Diseases, Department of Pediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University, Prague, Czech Republic.
- <sup>41</sup>Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX, USA.

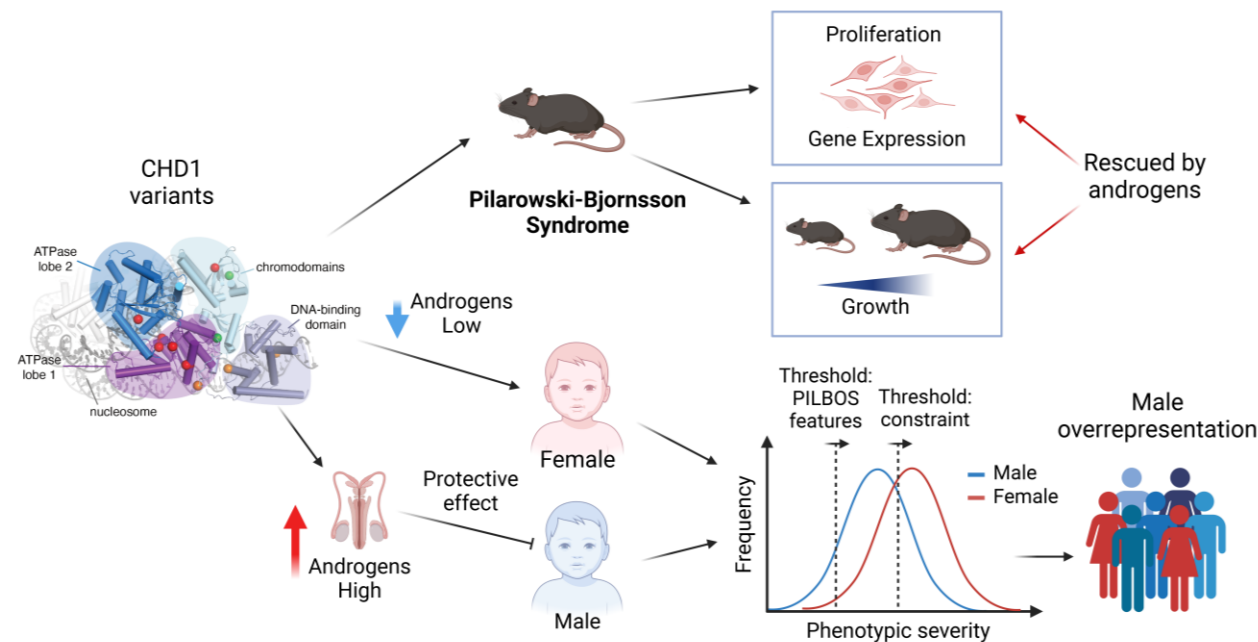
- 
- <sup>42</sup>INGEMM-Institute of Medical and Molecular Genetics, IdiPAZ- CIBERER- Hospital Universitario La Paz, Madrid, Spain and ERNITHACA, Madrid, Spain.
- <sup>43</sup>Department of Pediatrics, Gemeinschaftsklinikum Mittelrhein Kemperhof, Koblenzer Straße 115-155, 56073 Koblenz, Germany.
- <sup>44</sup>IRCCS Azienda Ospedaliero-Universitaria di Bologna, U.O. Genetica Medica, 40138 Bologna, Italy.
- <sup>45</sup>Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA.
- <sup>46</sup>Service de Génétique Médicale, Unité de Génétique Clinique, Nantes, France.
- <sup>47</sup>Departments of Pediatrics and Neurology, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO, USA.
- <sup>48</sup>Genetic Health Queensland, Royal Brisbane & Women's Hospital Campus, Herston, Brisbane, Australia.
- <sup>49</sup>Department of Human Genetics, Donders Institute, Radboud University Medical Center, Nijmegen, The Netherlands.
- <sup>50</sup>Division of Genetics, Department of Pediatrics West Virginia School of Medicine, Morgantown, USA.
- <sup>51</sup>MVZ Institute for Clinical Genetics and Tumor Genetics, Bonn, Germany.
- <sup>52</sup>Department of Genetics, UAB Heersink School of Medicine, Birmingham AL, USA.
- <sup>53</sup>Department of Pediatrics, University of Nebraska Medical Center, Omaha, NE, USA.
- <sup>54</sup>Department of Genetics, Mid-Atlantic Permanente Medical Group, Washington, DC, USA.
- <sup>55</sup>Baylor Genetics, Houston, TX, USA.
- <sup>56</sup>Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Flemington Road, Parkville, Victoria, Australia.
- <sup>57</sup>Human Genetics Department, Radboud University Medical Center, Nijmegen, the Netherlands.
- <sup>58</sup>Department of Medical Genetics, Center for Medical Genetics, School of Basic Medical Sciences, Peking University, Beijing 100191, China.
- <sup>59</sup>Neuroscience Research Institute, Peking University; Key Laboratory for Neuroscience, Ministry of Education of China & National Health Commission of China, Beijing 100191, China.
- <sup>60</sup>Autism Research Center, Peking University Health Science Center, Beijing 100191, China.
- <sup>61</sup>The Genetics Institute Rambam Health Care Campus Haifa Israel.
- <sup>62</sup>GeneDx, LLC, Gaithersburg MD 20877 USA.
- <sup>63</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA.
- <sup>64</sup>Department of Physiology, Faculty of Medicine, University of Iceland.
- <sup>65</sup>Department of Ophthalmology, Landspítali University Hospital, Reykjavik, Iceland.
- <sup>66</sup>Independent Scientist
- <sup>67</sup>deCODE Genetics/Amgen Inc., Reykjavik, Iceland.
- <sup>68</sup>Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
- <sup>69</sup>Department of Anthropology, University of Iceland, Reykjavik, Iceland.

## Abstract

Sex-specific penetrance in autosomal dominant Mendelian conditions is largely understudied. The neurodevelopmental disorder Pilarowski-Bjornsson syndrome (PILBOS) was initially described in females. Here, we describe the clinical and genetic characteristics of the largest PILBOS cohort to date, showing that both sexes can exhibit PILBOS features, although males are overrepresented. A mouse model carrying a human-derived *Chd1* missense variant (*Chd1*<sup>R616Q/+</sup>) displays female-restricted phenotypes, including growth deficiency, anxiety and hypotonia. Orchiectomy unmasks a growth deficiency phenotype in male *Chd1*<sup>R616Q/+</sup> mice, while testosterone rescues the phenotype in females, implicating androgens in phenotype modulation. In the gnomAD and UK Biobank databases, rare missense variants in *CHD1* are overrepresented in males, supporting a male protective effect. We identify 33 additional highly constrained autosomal genes with missense variant overrepresentation in males. Our results support androgen-regulated sexual dimorphism in PILBOS and open novel avenues to understand the mechanistic basis of sexual dimorphism in other autosomal Mendelian disorders.

**Keywords:** CHD1, Mendelian disease, Sex differences, Neurodevelopmental disorder

# Graphical Abstract



## Introduction

Traditionally, autosomal dominant disorders have been considered to have limited sex differences, unlike X-linked disorders defined by this feature. Pilarowski-Björnsson syndrome (PILBOS, MIM#617682) is an autosomal dominant disorder first described by our group in 2018 in six affected girls with developmental delay, hypotonia, speech apraxia, autism, and seizures<sup>1</sup>. Each individual carried a heterozygous missense variant in the chromatin remodeler gene *CHD1*, located on chromosome 5, and four were de novo variants. The absence of males led to the hypothesis of male lethality, consistent with the strong male overrepresentation in neurodevelopmental disorders (NDDs)<sup>2,3</sup>. Since then, two case reports have been published including individuals of either sex<sup>4,5</sup>. Thus, the question arises: does this disorder affect both sexes equally and does the disease mechanism of these variants involve gain-of-function or loss-of-function? To date no large scale PILBOS study has been performed and therefore the full phenotypic details and mechanistic basis remains elusive.

*CHD1* is a highly conserved chromatin remodeler that was first characterized for its role at sites of active transcription<sup>6,7</sup>, performing ATP-dependent chromatin remodeling of nucleosomes<sup>8,9</sup>. *CHD1* is known to recognize and directly bind trimethylated histone H3 lysine 4 (H3K4me3)<sup>10</sup> and is thought to drive global transcriptional output<sup>11</sup>, co-operating with elongation factors<sup>12</sup> and shifting the +1 nucleosome to allow RNA polymerase II to escape from promoters<sup>13</sup>. *CHD1* also has important roles in DNA replication, negatively regulating S-phase in yeast<sup>14</sup>, remodeling the nucleosome behind the replication fork<sup>15</sup>, and remodeling chromatin at stalled replication forks<sup>16</sup>. Furthermore, *CHD1* is important for the homologous recombination and nucleotide excision repair DNA damage repair pathways and is known to directly interact with DNA repair factors<sup>17-21</sup>.

The investigation of sex differences in mouse models of neurological conditions is critically understudied<sup>22</sup>. The careful identification of robust examples of sexually dimorphic conditions could help identify additional such disorders and lead to therapeutic strategies for those sharing the same causal modifier. An obvious difference between the two sexes involves the sex hormones, which play key roles in neurodevelopment as the main drivers of sex differences in the brain<sup>23</sup>. Interestingly, in prostate cancer, *CHD1* is known to directly interact with the androgen receptor (AR), binding to AR-occupied enhancers, and loss of *CHD1* leads to genome-wide reorganization of AR binding<sup>24,25</sup>. The *CHD1*-AR interaction may offer a unique window into a potential mechanism for sexual dimorphism in PILBOS and other related NDD phenotypes.

Here, we provide a comprehensive characterization of a large set of individuals with a diagnosis of PILBOS and their corresponding *CHD1* genetic variants. Using a novel mouse model (*Chd1*<sup>R616Q/+</sup>) carrying a variant identified in one of these individuals, we demonstrate sexual dimorphism of core PILBOS phenotypes. We further show that exposure to androgens mediates the protective effect observed in males. Finally, we utilize human population data to demonstrate differential frequencies of *CHD1* variants between the sexes and show that our findings of sexual dimorphism may extend to other autosomal dominant Mendelian disorders.

## Results

### PILBOS affects both sexes through distinct mutational mechanisms

To further explore the phenotypic presentation of PILBOS, we identified additional individuals using GeneMatcher<sup>26</sup> (<https://genematcher.org/>), through direct communication with individuals, families and providers, and through an online PILBOS support group. This resulted in the identification of 48 unrelated individuals from 14 countries. Together with the eight individuals previously reported in the literature<sup>1,4,5,27-29</sup>, this amounted to 52 separate variants in 56 individuals: 39 males and 17 females. Notably, males were significantly overrepresented in this cohort ( $p = 0.0046$ , binomial test). A total of 33 individuals carried 29 missense variants, of which: 17 were *de novo* (52%), 5 inherited (15%), and 11 unknown (33%). Twenty individuals carried truncating loss of function (LOF) variants: 11 were *de novo* (55%), 4 inherited (20%), and 5 unknown (25%); one carried a stop-loss variant (*de novo*), and two carried possible non-canonical splice-site variants (one *de novo*, one of unknown inheritance). Three of the missense variants were recurring: firstly, a variant identified in two unrelated males (c.2090G>A, p.Arg697Gln), secondly, a variant identified in two unrelated females (c.3335G>T, p.Arg1112Leu), and thirdly, a variant identified in two sisters and an unrelated male (c.5123G>A, p.Arg1708Gln). This third variant (c.5123G>A, p.Arg1708Gln) has been reported at a low frequency ( $3.73 \times 10^{-6}$ ) in the gnomAD population database (v4.1.0)<sup>30</sup> and has also been identified in unaffected parents from clinical cohorts, raising the possibility of this being a benign variant. However, since we are interested in penetrance, here we include information on all variants.

Of all the variants, 24 (46%) were classified as being likely pathogenic using the American College of Medical Genetics and Genomics (ACMG) guidelines. Most of the missense variants (83%) were classified as variants of unknown significance, with five of them (17%) being classified as likely

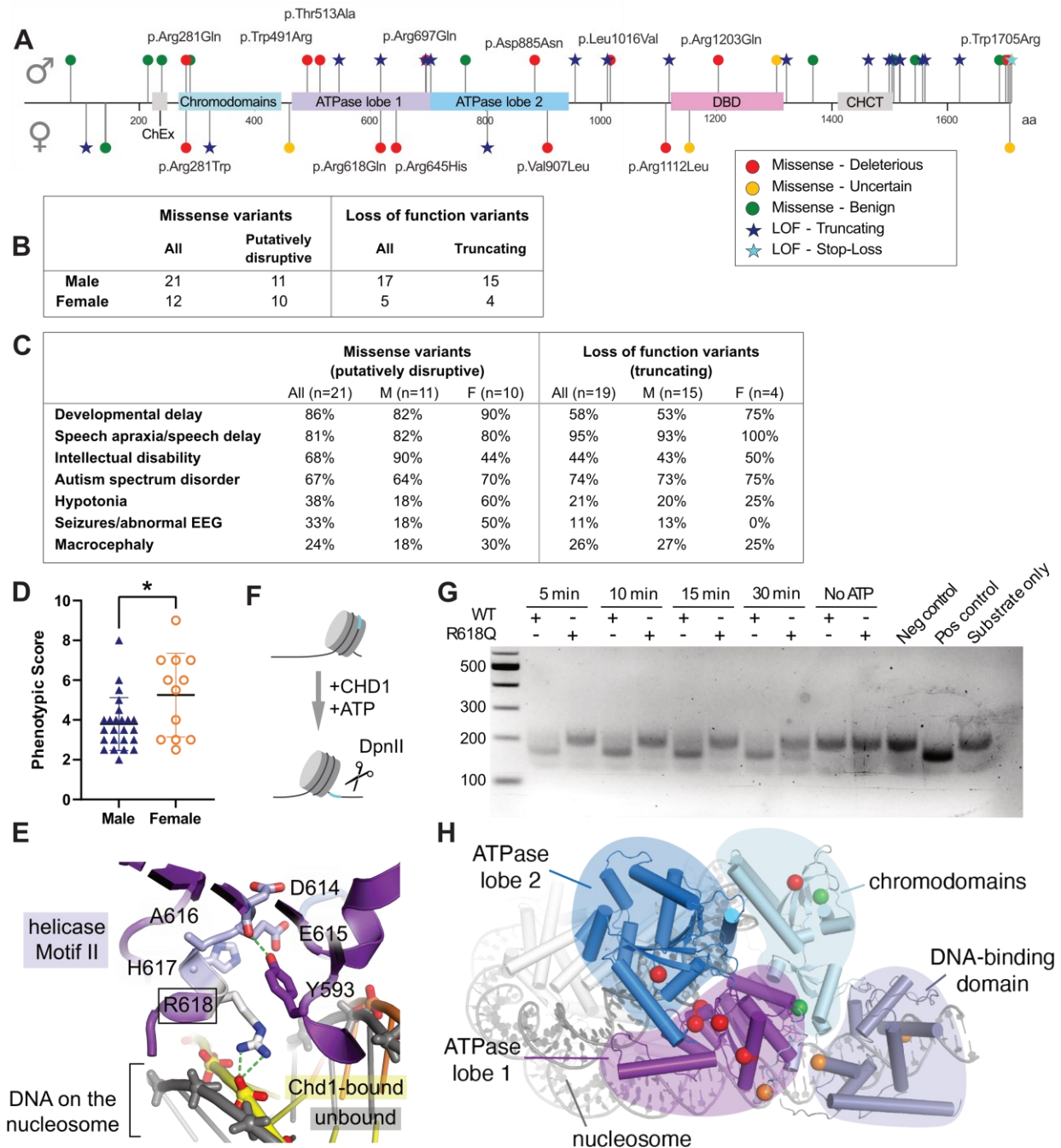
pathogenic. We used AlphaMissense<sup>31</sup> to perform structural predictions, identifying 12 additional missense variants that may disrupt structure and function (deleterious or uncertain classification, see **Supplementary Table 1**), and created a color-coded visual representation of LOF and missense variants for either sex (**Figure 1A**). Twelve of the 29 missense variants (41%) are categorized as deleterious and appear to lead to loss of protein function based on AlphaMissense structural predictions, and a further four missense variants (14%) are categorized as uncertain, while 13 are categorized as benign (45%). Comparing the sexes, 21 missense variants were found in males (64%) compared to 12 in females (36%). For the 21 individuals carrying putatively disruptive missense variants (categorized as deleterious or uncertain), there were 11 males and 10 females, whereas 15 males carried LOF truncating variants categorized as likely pathogenic (79%) compared to 4 females (21%) (**Fig. 1B**). Subsequently, our collaborators at GeneDx analyzed a separate cohort of individuals from their database who underwent exome sequencing-based testing and were found to have pathogenic or likely pathogenic *CHD1* variants ( $n < 15$ ). Notably, no females in this cohort had frameshift or nonsense variants. Instead, females had missense (67%) or splice site variants (33%), while males had LOF (33%) or missense variants (67%).

Collectively, we identified substantial evidence for 36 variants, observed across 40 individuals, which we used to define the core PILBOS phenotype. Key phenotypic characteristics of the 21 individuals carrying putatively disruptive missense variants included developmental delay (86%), speech apraxia/speech delay (81%), and autism spectrum disorder (67%). Two of these individuals were too young to be assessed for intellectual disability, which was present in 68% of the individuals for whom cognitive assessment was possible. For the 19 patients carrying truncating LOF variants, the key phenotypic aspects were speech apraxia/speech delay (95%), autism spectrum disorder (74%), developmental delay (58%) and intellectual disability (44%, with one individual from this group being too young for assessment) (**Figure 1C, Supplementary Tables 1&2**). In order to investigate this further, we scored the probands with curated variants according to the number and severity of core phenotypes. When grouping individuals with missense and LOF variants, females exhibited significantly higher phenotypic scores than males (**Fig. 1D**,  $p = 0.0440$ , Welch's t-test), suggesting that females may be more sensitive to *CHD1* variants than males. Together, these data demonstrate that males are overrepresented in the context of PILBOS, whereas females tend to exhibit more severe phenotypic manifestations, suggesting that male sex gives some protection against *CHD1* variants in humans.

## Loss of function of *CHD1* caused by specific missense variants



The present cohort carries a high proportion of missense variants in *CHD1* compared to LOF variants, which were not found in the original study<sup>1</sup>. Our computational prediction suggests that many of these missense variants lead to loss of function (**Supplementary Table 1**). To gather more evidence supporting a loss of function mechanism, we performed functional and predictive analyses. First, we selected one representative variant (P\_32, **Supplementary Table 1**), a heterozygous variant leading to an amino acid change from a positively charged arginine to neutral glutamine at position 618 in CHD1 which corresponds to the R616Q in mice. This arginine immediately follows the Walker B DExx Box, also known as helicase Motif II (DEAHR), and is strictly conserved in CHD1 orthologs. Since it directly contacts nucleosomal DNA that is actively distorted by CHD1<sup>32</sup> (**Fig. 1E**), we expected that this arginine-to-glutamine substitution would likely interfere with CHD1 remodeling activity. To assess this, we used a human *CHD1* cDNA construct with an N-terminal deletion, as the N-terminus has been described to inhibit the DNA-binding and ATPase activity of CHD1<sup>19</sup>. We generated the p.R618Q mutation using site-directed mutagenesis, then produced both the wild type (WT) and mutant protein. Using a recombinant DNA-wrapped nucleosome with a DpnII restriction site, we incubated it with DpnII, ATP and either CHD1-WT or CHD1-R618Q at 37°C for multiple different exposure times (**Fig. 1F-G**). While CHD1-WT was rapidly able to shift the nucleosome along the DNA and allow access for DpnII activity in the presence of ATP, CHD1-R618Q was not (**Fig. 1G**, lower band). Thus, the p.R618Q variant causes deficient CHD1 chromatin remodeling activity. Finally, we analyzed the locations of the other missense variants in high resolution structures of human and yeast CHD1 (overview in **Fig. 1H**). Most variants flagged as severe are themselves highly conserved and/or interact with highly conserved residues and motifs and thus would be expected to interfere with enzymatic function (**Fig. S1A**).



**Figure 1: PILBOS missense variants lead to loss of function of *CHD1***

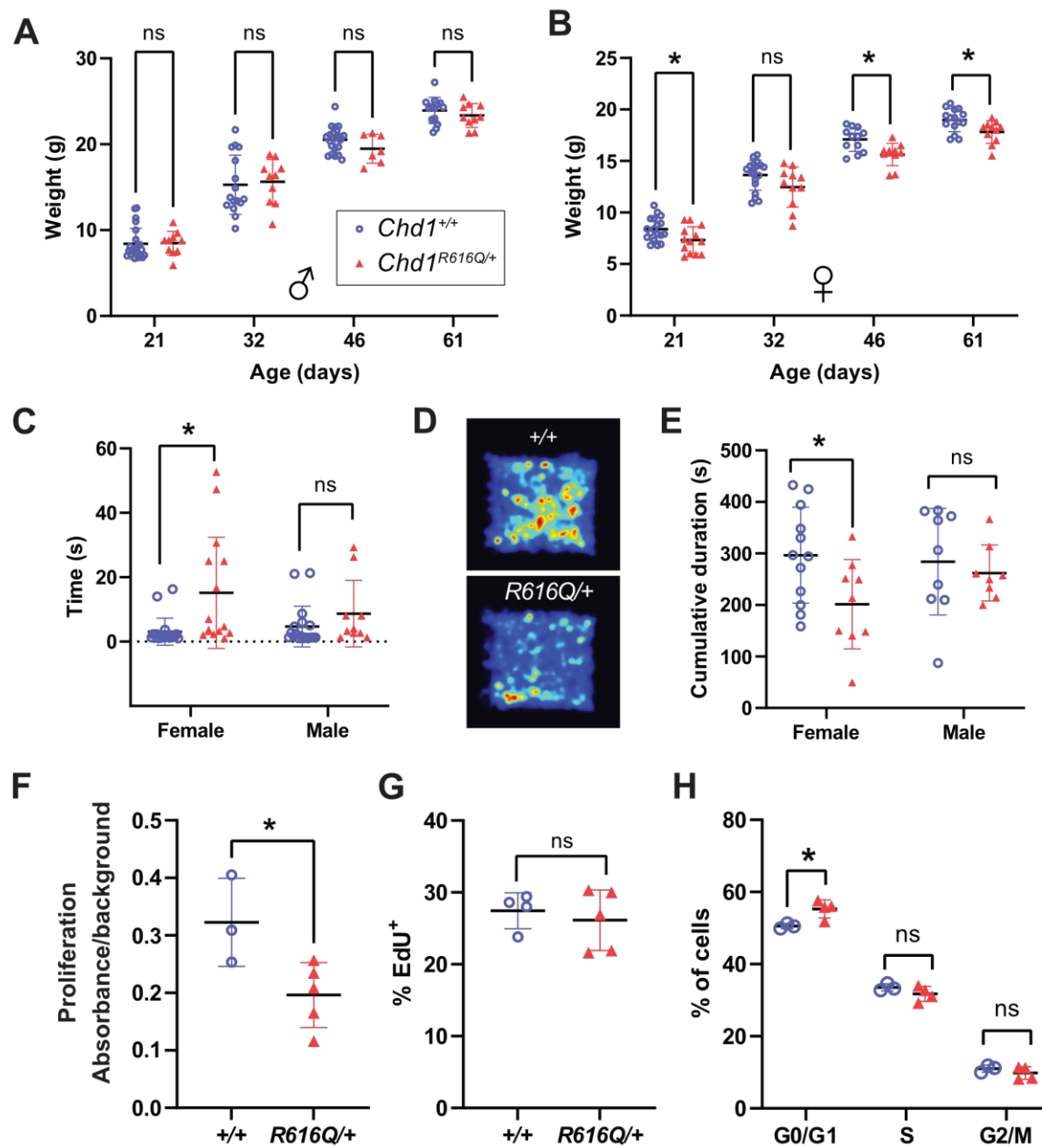
(A) A schematic representation of the CHD1 protein. Location of variants shown as stars (LOF variants) and circles (missense variants). Male variants are above and female variants below the baseline. Color coding of missense variants represents likelihood of the variant leading to loss of function according to AlphaMissense. ChEx: Chd1 Exit-side binding domain, DBD: DNA-binding domain, CHCT: CHD C-terminal domain. (B) The distribution of missense and LOF variants in both sexes. (C) Key phenotypic aspects of

individuals carrying missense variants predicted to cause loss of protein function and individuals carrying loss of function variants. (D) Phenotypic scores for females and males with curated missense and LOF variants (male, n = 24, female, n = 12). (E) Expected interactions for human CHD1 R618, based on a yeast Chd1-nucleosome structure<sup>32</sup>, which would be potentially disrupted by the missense variant p.R618Q. (F) A schematic of the *in vitro* nucleosome remodeling assay. (G) A representative image from three biological replicates of the chromatin remodeling assay. The CHD1-WT and CHD1-R618Q proteins are denoted by “+” and “-” symbols. The upper band ~220bp represents the uncut nucleosome-wrapped DNA substrate. The lower band at ~180bp represents remodeled DpnII-digested DNA-nucleosome substrate. The negative control comprised the DNA alone, with no DpnII restriction site, and the positive control comprised DNA alone with the DpnII restriction site. DpnII was added to all conditions. (H) An overview of the CHD1 protein based on a yeast Chd1-nucleosome structure<sup>32</sup>, with highlighted residues harboring missense variants. \**p* < 0.05, Welch’s two-tailed t-test, ns = not significant.

## Characterization of *Chd1*<sup>R616Q</sup> mice

In order to examine whether PILBOS phenotypes are observed in a murine model and explore their malleability, we used CRISPR-Cas9 to generate a mouse model harboring the equivalent of the *CHD1*<sup>R618Q</sup> variant (*Chd1*<sup>R616Q</sup>). Given previous reports of homozygous lethality in a *Chd1* knockout mouse model<sup>11</sup> but not in a mouse model carrying a hypomorphic variant<sup>33</sup>, we sought to investigate the effects of *Chd1*<sup>R616Q</sup> in homozygosity. Over 5 generations of heterozygote-heterozygote breeding, we never observed a homozygous mouse out of a total of 96 pups from 18 litters. Furthermore, we observed decreased Mendelian ratios for mice carrying the variant in heterozygosity: given the number of 39 *Chd1*<sup>+/+</sup> mice, we would expect 78 heterozygotes according to normal Mendelian ratios, but instead observed 57 (**Supplementary Table 3**, *p* = 2.72 × 10<sup>-9</sup>, Fisher’s exact test). These data suggest functional importance of this variant on fitness either at conception or *in utero*. We next assessed differences in gene expression in cells with the *Chd1*<sup>R616Q</sup> variant compared to cells from *Chd1*<sup>+/+</sup> littermates. Using RNA extracted from the cortex of *Chd1*<sup>R616Q/+</sup> mice, we performed RT-qPCR and observed no significant difference in *Chd1* mRNA levels compared to WT littermates (**Fig. S1B**). We then derived, cultured and extracted protein from mouse embryonic fibroblasts (MEFs) from both genotypes. A western blot demonstrated no significant difference in total CHD1 protein quantity (**Fig. S1C-D**). Thus, although the R616Q variant affects chromatin remodeling activity, it does not appear to do so by reducing the stability of the transcript or the protein.

As growth retardation is a common feature of PILBOS, we examined the weight of the mice over time, starting at postnatal day 21. While the males did not show any significant difference in weight over time (**Fig. 2A**), the females showed a significant weight deficit at postnatal days 21, 46 and 61 ( $p < 0.05$ , mixed effects analysis, **Fig. 2B**). Using linear regression analysis, we observed a significant difference in the elevation of the regression line between *Chd1*<sup>R616Q/+</sup> mice and WT littermates for females ( $p = 0.0004$ ) but not males ( $p = 0.4883$ ). Since hypotonia is a common PILBOS phenotype, we assessed this feature in the mouse model. Indeed, female *Chd1*<sup>R616Q/+</sup> mice displayed an impaired righting reflex at postnatal day 6 ( $p = 0.0181$ , Welch's t-test), which was not observed in males (**Fig. 2C**). In contrast, when using the hindlimb suspension test, which measures strength and endurance, we did not observe a difference in either females (**Fig. S2A**) or males (**Fig. S2B**). To further assess behavioral features in our mouse model, we performed an open field test at 3 months of age, in which the mice were allowed to freely explore an open space for 10 minutes. The female *Chd1*<sup>R616Q/+</sup> mice spent significantly less time in the center of the open field compared to their WT littermates ( $p = 0.0267$ , Welch's t-test, **Fig. 2D-E**), a known anxiety-like behavior in mice<sup>34</sup>. Again, no abnormality was observed in male *Chd1*<sup>R616Q/+</sup> mice (**Fig. 2E**). In contrast, there was no difference in the overall distance travelled for either females (**Fig. S2C**) or males (**Fig. S2D**). Intellectual disability is another common PILBOS feature, however female *Chd1*<sup>R616Q/+</sup> mice did not display memory deficits in a Morris water maze test, an established test of spatial learning and memory<sup>35</sup> (**Fig S2E-H**). However, they travelled less distance ( $p = 0.0119$ , Welch's t-test, **Fig. S2I**) and swam with decreased velocity in the water maze ( $p = 0.0148$ , Welch's t-test, **Fig. S2J**) compared to WT littermates, concordant with a potential adult motor deficit. Given that speech apraxia and autism are also frequent features of PILBOS, we assessed the ultrasonic vocalizations (UVs) of pups, a measure which can be used to assess differences in social communication, an autism-like phenotype<sup>36</sup>. Measuring every two days from postnatal day 3-9, we did not observe any major differences (**Fig. S2K-L**). Since abnormal eye morphology appears as one of the most significant phenotypes identified in *Chd1* KO mice in data from the International Mouse Phenotyping Consortium ([www.mousephenotype.org](http://www.mousephenotype.org))<sup>37</sup>, we also examined the structure and electrophysiological responses of the posterior chamber (retina, retinal pigment epithelium and choroid) of the eyes of *Chd1*<sup>R616Q/+</sup> mice. However, we observed no significant differences from their WT littermates (**Fig. S2M-N**). Taken together, *Chd1*<sup>R616Q/+</sup> mice display female-limited weight, motor and behavioral phenotypes.



**Figure 2:  $Chd1^{R616Q/+}$  mice exhibit female-limited growth, motor and behavioral phenotypes**

(A) Weight of male  $Chd1^{R616Q/+}$  mice ( $n = 7-10$ ) and WT littermates ( $n = 14-23$ ). (B) Weight of female  $Chd1^{R616Q/+}$  mice ( $n = 11-13$ ) and WT littermates ( $n = 12-18$ ). (C) Righting reflex times in female (WT:  $n = 20$ ,  $R616Q/+$ :  $n = 15$ ) and male (WT:  $n = 18$ ,  $R616Q/+$ :  $n = 10$ )  $Chd1^{R616Q/+}$  mice and WT littermates at postnatal day 6. (D) A representative heatmap depicting the arena area covered by female mice of both genotypes during a 10-minute open field test. (E) Cumulative duration of time spent in the center area of the open field for female (WT:  $n = 12$ ,  $R616Q/+$ :  $n = 9$ ) and male (WT:  $n = 9$ ,  $R616Q/+$ :  $n = 8$ )  $Chd1^{R616Q/+}$

mice and WT littermates during a 10-minute test. (F) Proliferation of female NPCs in culture as measured by absorbance over background in a MTT assay (WT:  $n = 3$ ,  $R616Q/+$ :  $n = 5$ ). (G) Flow cytometry measurements of EdU incorporation into female NPC DNA over a 2h period in culture (WT:  $n = 3$ ,  $R616Q/+$ :  $n = 5$ ). (H) DNA content in female NPCs as measured by flow cytometry following DAPI staining of ethanol-permeabilized cells (WT:  $n = 3$ ,  $R616Q/+$ :  $n = 4$ ). \* $p < 0.05$ , ns: not significant.

### ***Chd1*<sup>R616Q/+</sup> mice demonstrate cellular phenotypes in neuronal progenitors**

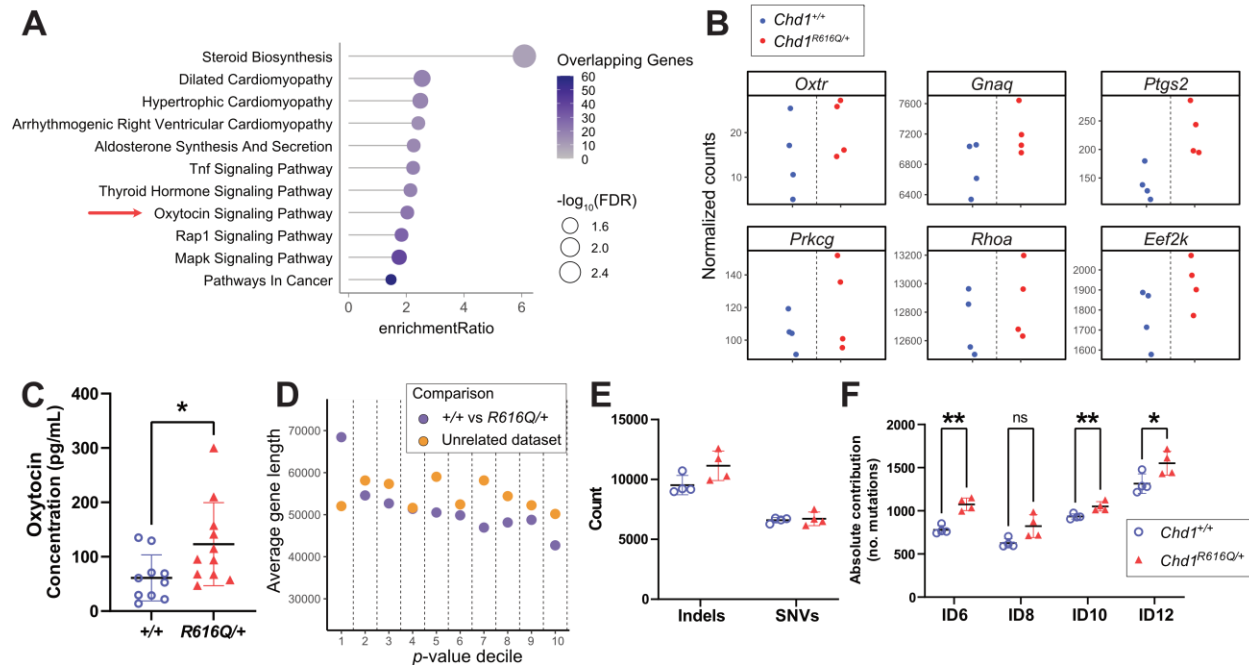
As key PILBOS phenotypes are linked to neuronal function, we next investigated cellular phenotypes in cortical neuronal progenitor cells (NPCs) derived from E18.5 female mouse embryos. Using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay as a measure of proliferation, we observed significantly decreased proliferation in the *Chd1*<sup>R616Q/+</sup> NPCs compared to NPCs from WT littermates ( $p = 0.0354$ , Student's two-tailed t-test, **Fig. 2F**). We then assessed whether the proliferation defect was due to differences in the speed of progression through S-phase by using a 5-ethynyl-2'-deoxyuridine (EdU) assay. Using a 2-hour (h) pulse, we did not observe a difference in the short-term rate of replication (**Fig. 2G**). However, when we assessed the cellular DNA content using DAPI staining, we observed a higher proportion of *Chd1*<sup>R616Q/+</sup> NPCs in the G0/G1 phase of the cell cycle, suggesting a possible delayed entry to S phase, a phenomenon known to occur during neuronal differentiation<sup>38</sup> ( $p = 0.025$ , Student's two-tailed t-test, **Fig. 2H**). *Chd1*<sup>R616Q/+</sup> NPCs did not show any change in the proportion of apoptotic cells as measured by Annexin V assay (**Fig. S2O**). In summary, the R616Q mutation causes decreased proliferation likely through lengthening of the G1 phase of the cell cycle in NPCs, a possible mechanism for the cellular phenotypes observed with disruption of CHD1 in NPCs.

In order to further examine the functional impact of the p.R616Q variant, we performed bulk RNA-sequencing (RNAseq) using female NPCs derived from both genotypes. Modest transcriptomic changes were observed, with 44 significantly differentially expressed genes identified between WT and *Chd1*<sup>R616Q/+</sup> (**Fig. S3A, Supplementary Table 4**). In order to not miss any relevant changes, we expanded to a list of genes with the top 10% most significant p-values ( $n = 1809$ ;  $p$  ranging from  $3.33 \times 10^{-15}$  to 0.08). Using this list, we performed overrepresentation analysis and observed "microfibril", "neuron-to-neuron synapse" and "centrosome" amongst the most significantly enriched cellular component terms (**Fig. S3B**). Next, using molecular pathway enrichment, we observed "oxytocin signaling pathway" amongst



the significantly overrepresented pathways (**Fig. 3A**). This finding is particularly noteworthy as oxytocin is known to exhibit sexually dimorphic effects on weight and anxiety<sup>39-44</sup>. Genes encoding the oxytocin receptor (*Oxtr*) and downstream effectors including *Gnaq*, *Ptgs2*, *Prkcg*, *Rhoa* and *Eef2k* were upregulated in the *Chd1*<sup>R616Q/+</sup> cells compared to wildtype (**Fig. 3B**). To investigate the functional implications of this observation, we conducted an enzyme-linked immunoassay (ELISA) on protein extracted from hypothalamic tissue of 3-month-old female mice and observed significantly elevated levels of oxytocin in the *Chd1*<sup>R616Q/+</sup> mice compared to WT littermates ( $p = 0.0337$ , Welch's t-test, **Fig. 3C**). These results suggest that even genes that did not meet the more stringent significance threshold may still hold biological relevance, justifying our approach of considering genes in the top 10% of p-values.

As CHD1 is known to play roles in DNA damage repair<sup>17-20</sup>, and long genes are more susceptible to DNA damage which can affect their transcription<sup>45</sup>, we next assessed the gene-lengths of detected genes in our RNAseq datasets ranked according to p-value. Here we observed that the genes with the top 10% most significant p-values (decile 1), were significantly longer than all other detected genes ( $p = 2.973 \times 10^{-6}$ , Welch's t-test, **Fig. 3D**). In contrast, when looking at an unrelated WT RNAseq dataset also generated from E18.5 WT mouse NPCs<sup>46</sup>, there was no significant difference in gene length for the most significantly differentially expressed genes ( $p = 0.3084$ , Welch's t-test **Fig. 3D**). Next, we assessed the number of indels and single nucleotide variants (SNVs) in the RNAseq data. We observed a slight but non-significant increase in the number of indels in the *Chd1*<sup>R616Q/+</sup> NPCs compared to WT ( $p = 0.07$ , Welch's t-test), whereas the number of SNVs was unchanged (**Fig. 3E**). Further analysis of the COSMIC mutational signatures<sup>47</sup> associated with the indels revealed the signature "ID1" as most enriched, which consists primarily of T-base insertions. Given that this pattern could reflect poly(A) tails, which are expected in RNA-seq data, we excluded ID1 from further analysis to avoid potential confounding. Among the remaining signatures, "ID6", "ID8", "ID10" and "ID12" were notably enriched and three were significantly increased in *Chd1*<sup>R616Q/+</sup> NPCs relative to WT (ID6:  $p = 0.0008$ , ID8:  $p = 0.0517$ , ID10:  $p = 0.0144$ , ID12:  $p = 0.0399$ , Welch's t-test, **Fig. 3F**). While ID10 and ID12 are not associated with any specific etiology, ID6 is associated with defective homologous recombination DNA damage repair and ID8 has a proposed etiology linked to repair of double-strand breaks by non-homologous end-joining (NHEJ). These findings suggest that *Chd1*<sup>R616Q/+</sup> NPCs may be deficient in homologous recombination activity, which is known to be dependent on CHD1<sup>17,19,20</sup>, and instead rely more heavily on the error-prone NHEJ pathway, potentially implicating DNA damage in the etiology of the PILBOS phenotype.



**Figure 3: Gene expression in *Chd1*<sup>R616Q/+</sup> NPCs shows changes in oxytocin signaling, enrichment of long genes and a distinct mutational signature**

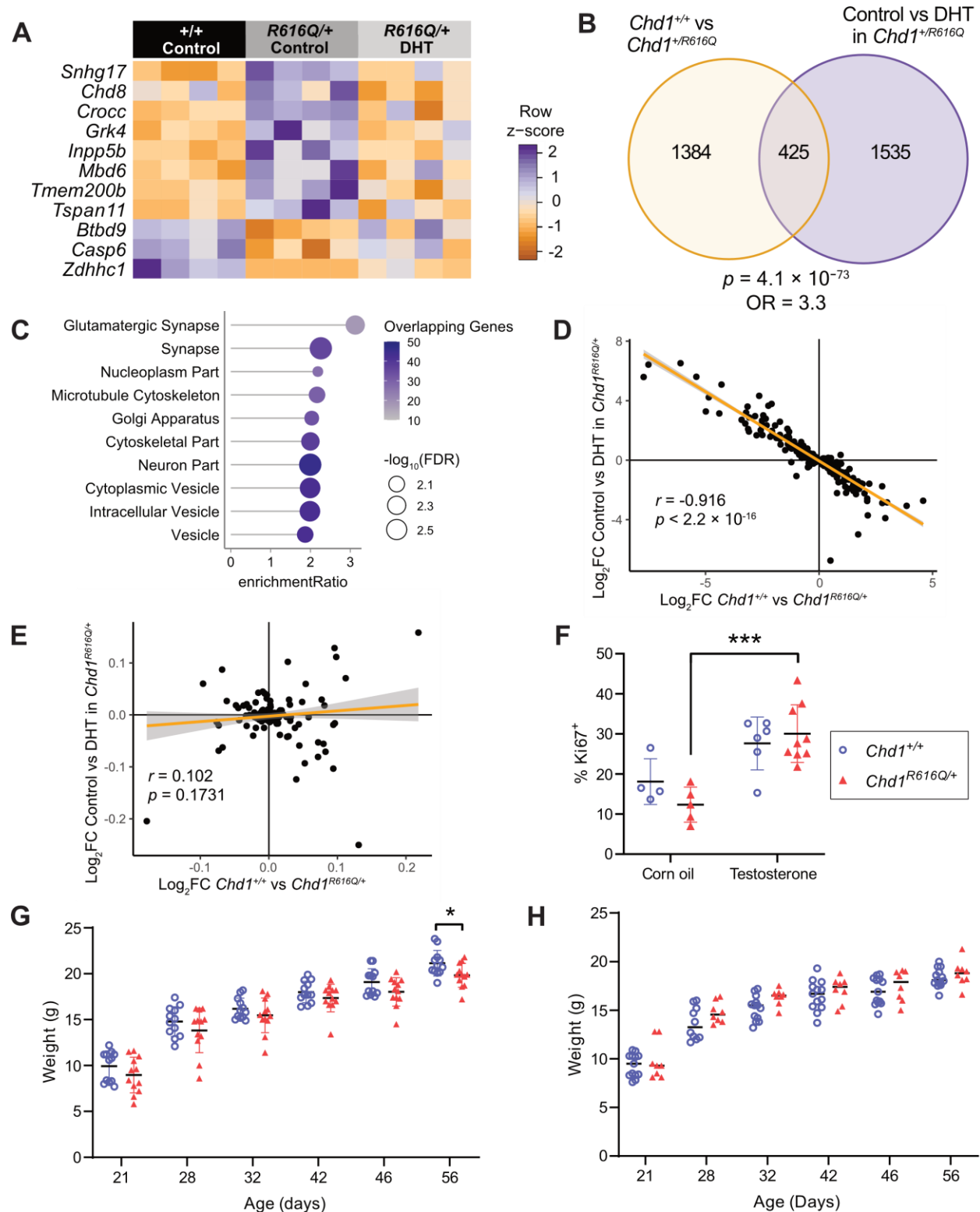
(A) KEGG pathway gene ontology overrepresentation analysis for the comparison of gene expression results from *Chd1*<sup>+/+</sup> vs *Chd1*<sup>R616Q/+</sup>. (B) Normalized RNAseq counts for genes involved in the oxytocin signaling pathway in *Chd1*<sup>+/+</sup> and *Chd1*<sup>R616Q/+</sup> NPCs, selected from the gene ontology overrepresentation analysis. (C) Oxytocin concentration in hypothalamic lysates from 3-month-old female *Chd1*<sup>R616Q/+</sup> mice and WT littermates as measured by ELISA (WT: *n* = 10, *R616Q/+*: *n* = 11). (D) Average gene lengths for each *p*-value decile, with 1 being the most significant *p*-values for the WT vs *Chd1*<sup>R616Q/+</sup> comparison and an unrelated wild-type NPC dataset. (E) Counts of indels and SNVs in RNAseq data. (F) Counts of mutations associated with COSMIC indel mutational signatures. \**p* < 0.05, \*\**p* < 0.01.

### Androgen exposure rescues *Chd1*<sup>R616Q/+</sup> transcriptional abnormalities

Previous studies have shown direct protein interactions between CHD1 and the AR<sup>24,25</sup>. We hypothesized that the greater androgen exposure that naturally occurs in males could be a potential mechanism mediating the male protective effect in our *Chd1*<sup>R616Q/+</sup> mice and human clinical cohort. To test this, we performed RNAseq using NPCs treated with 10nM dihydrotestosterone (DHT) or with DMSO vehicle control (**Supplementary Table 5, Fig. S4A**). As before, we collected the genes with the top 10%



most significant  $p$ -values from the comparison of DHT vs DMSO control treatment in *Chd1*<sup>R616Q/+</sup> cells ( $n = 1960$ ,  $p$  ranging from  $6.66 \times 10^{-16}$  to  $0.08$ ). The most significantly enriched cellular component gene ontology terms in this list also included “neuron-to-neuron synapse” and “centrosome”, as well as “cytosolic ribosome” (**Fig. S4B**). KEGG pathway analysis also revealed enrichment of the term “ribosome”, which may imply differences in regulation of protein synthesis, a known effect of androgens<sup>48-50</sup> (**Fig. S4C**). Interestingly, for several genes whose expression was the most significantly changed in the *Chd1*<sup>R616Q/+</sup> cells compared to WT, DHT treatment of the *Chd1*<sup>R616Q/+</sup> cells restored the expression towards baseline WT levels (**Fig. 4A**). We next compared these two lists (WT vs *Chd1*<sup>R616Q/+</sup> and control vs DHT treatment in *Chd1*<sup>R616Q/+</sup> cells) and observed a highly significant overlap of 425 genes ( $p = 4.1 \times 10^{-73}$ , OR = 3.3, Fisher’s exact test, **Fig. 4B**). Among the most significantly enriched gene ontology terms for this overlapping list of genes were “glutamnergic synapse”, “cytoskeletal part” and “neuron part”, with genes including *Agrn*, *Bcan*, *Syt1*, *Dab1* and *Ntrk3*, suggesting changes in neuronal function and differentiation (**Fig. 4C**). We then sought to understand the relationship between the changes caused by the R616Q variant and DHT. When we plotted the log<sub>2</sub> fold change for the overlapping genes from each comparison against one another, we observed a strong negative correlation ( $r = -0.916$ ,  $p < 2.2 \times 10^{-16}$ , **Fig. 4D**). This indicates that genes which were over-expressed in *Chd1*<sup>R616Q/+</sup> cells exhibit lower expression following DHT treatment and vice versa, suggesting that DHT treatment reverses the transcriptional dysregulation caused by the R616Q variant in NPCs. In contrast, when we looked at the lists of genes with the 10% least significant  $p$ -values from the two comparisons, there was an overlap of 181 genes, which was not statistically significant. Furthermore, when we plotted the log<sub>2</sub> fold change for these genes, we did not observe a correlation ( $r = 0.102$ ,  $p = 0.1731$ , **Fig. 4E**), demonstrating that this effect is specific to the most highly differentially expressed genes. As testosterone is converted to estradiol by the aromatase cytochrome P450 enzyme<sup>51</sup>, we examined the expression of this gene (*Cyp19a1*) and the estrogen receptor genes in the NPCs and observed relatively low levels of expression compared to the androgen receptor (**Fig. S4D**). Together with the fact that the non-aromatizable androgen, DHT restores normal expression in mutant cells, this suggests the effects of testosterone treatment may be more likely to be driven by AR activation rather than the estrogen receptor, at least in neural progenitor cells.



**Figure 4: Androgens rescue *Chd1*<sup>R616Q/+</sup> transcriptomic and *in vivo* phenotypes**

(A) A heatmap depicting mRNA expression in NPCs, either WT or *Chd1*<sup>R616Q/+</sup> treated with DMSO vehicle control, or *Chd1*<sup>R616Q/+</sup> treated with 10nM DHT for 6h. (B) A Venn diagram displaying overlapping genes with the top 10% most significant p-values from the comparison between *Chd1*<sup>+/+</sup> vs *Chd1*<sup>+/R616Q</sup> and DMSO vehicle control vs DHT in *Chd1*<sup>+/R616Q</sup> NPCs. The overlap was tested using Fisher's exact test with a background of 19980 genes (number of protein coding genes in GENCODE hg38). (C) Overrepresentation analysis of overlapping gene list from B. (D) Plot of log<sub>2</sub> fold changes for the genes with the 10% lowest p-values from the overlap in B. (E) Plot of log<sub>2</sub> fold changes for the overlapping genes with the 10% highest p-values in the "WT vs *Chd1*<sup>R616Q/+</sup>" and "DMSO vs DHT in *Chd1*<sup>R616Q/+</sup>" comparisons. (F) Percent Ki-67<sup>+</sup> cells in the telencephalon of embryonic day 18.5 embryos, as measured by antibody staining and flow cytometry (WT-corn oil: *n* = 4, *R616Q/+*-corn oil: *n* = 5, WT-testosterone: *n* = 6, *R616Q/+*-testosterone: *n* = 9). Embryos were obtained from 3-4 dams for each treatment. (G) Weight measurements over time in orchiectomized male *Chd1*<sup>R616Q/+</sup> mice and WT littermates (*n* = 12 per genotype). (H) Weight measurements over time in female mice with a testosterone implant placed under skin at P15 (WT: *n* = 10-13, *R616Q/+*: *n* = 8). \**p* < 0.05, \*\*\**p* < 0.001.

### Androgens ameliorate *in vivo* *Chd1*<sup>R616Q/+</sup> phenotypes

Given the reversal of *Chd1*<sup>R616Q/+</sup>-driven transcriptional abnormalities with androgen exposure in NPCs *in vitro*, we next sought to investigate if androgen exposure could rescue *in vivo* phenotypes in our mouse model. As we had previously observed decreased proliferation of NPCs from the *Chd1*<sup>R616Q/+</sup> model (Fig. 2F), we sought to test whether testosterone treatment would ameliorate this defect. Indeed, 10 days of testosterone injection of pregnant dams led to increased proliferation in the telencephalon of *Chd1*<sup>R616Q/+</sup> embryos at E18.5 (*p* = 0.0001, Welch's t-test, Fig. 4F). We hypothesized that if androgens were protecting males from phenotypes driven by CHD1-deficiency, then removal of androgens might unmask these same phenotypes in *Chd1*<sup>R616Q/+</sup> males. To test this, we performed orchiectomy of both WT and *Chd1*<sup>R616Q/+</sup> male mice at postnatal day 15 (P15) and then followed their weight prospectively. Although there was no significant weight difference between the genotypes prior to orchiectomy (Fig. S4E), after the P15 procedure a weight deficit emerged in the orchiectomized *Chd1*<sup>R616Q/+</sup> mice compared to their orchiectomized WT littermates at P56 (*p* = 0.0255, mixed effects analysis, Fig. 4G). Using linear regression analysis, we also identified a significant difference in the elevation of the regression line for *Chd1*<sup>R616Q/+</sup> males compared to their WT littermates (*p* = 0.0037). Next, to test if the female weight deficit could be rescued through androgen treatment, we surgically implanted slow-release testosterone silastic

implants under the skin of both WT and *Chd1*<sup>R616Q/+</sup> female mice at P15 (**Fig. S4F**). Following their weight over time, we observed amelioration of the previously observed weight deficit in the *Chd1*<sup>R616Q/+</sup> females compared to their WT littermates, with no significant decrease in their weights (mixed effects analysis, **Fig. 4H**). Furthermore, using linear regression analysis, we observed less difference in the elevation of the regression line for *Chd1*<sup>R616Q/+</sup> females compared to their WT littermates ( $p = 0.0254$ , previously  $0.0004$ ). We also tested if orchietomy would uncover the female-limited anxiety-like behavior in males using an open field test. However, there was no difference in the time spent in the center of the open field following orchietomy (**Fig. S4G**), suggesting that the sequence of events underlying postnatal anxiety-like behavior may occur prior to P15 (time of orchietomy). Taken together, our data demonstrate a protective effect of androgens against weight and neuronal proliferation deficits in *Chd1*<sup>R616Q/+</sup> mice.

## Rare *CHD1* missense alleles are overrepresented in healthy males

Given the overrepresentation of males and the increased phenotypic scores of females in our PILBOS cohort, along with the protective effect of androgens in our mouse model, we sought to explore if this could be attributed to a general sex-specific susceptibility to *CHD1* variants. In exome sequencing data from the Genome Aggregation Database (gnomAD, v4.1.0)<sup>30</sup>, which mostly excludes individuals with pediatric disease, we assessed the frequency of alternate alleles in *CHD1* in female (XX) and male (XY) populations. As previously reported<sup>1</sup>, *CHD1* is highly intolerant to loss of function variants ( $pLI = 1$ ) so there are few LOF variants to evaluate. Considering all missense variants, we decided to filter out those with a minor allele frequency (MAF) of  $\geq 1\%$  as these are much less likely to have functional effects. Using the remaining rare missense variants, we calculated the aggregate sum of alternate allele counts (AC) divided by the aggregate sum of total allele number (AN) for each sex and compared these frequencies (AF) between the sexes using a two-proportion z-test. We observed a significant overrepresentation of missense alleles in males ( $AF = 1.35 \times 10^{-5}$ ) compared to females ( $AF = 1.24 \times 10^{-5}$ ) ( $p = 3.65 \times 10^{-10}$ ,  $\chi^2 = 39.29$ ,  $95\% \text{ CI} = 5.73 \times 10^{-7} - 1.37 \times 10^{-6}$ , **Fig. 5A**). To test the validity of our findings, we next repeated this analysis in whole genome sequencing (WGS) data from the UK biobank (UKB)<sup>52,53</sup>. Here, rare missense variants in *CHD1* were also overrepresented in males ( $AF = 1.53 \times 10^{-5}$ ) compared to females ( $AF = 1.47 \times 10^{-5}$ ) ( $p = 0.019$ ,  $\chi^2 = 5.48$ ,  $95\% \text{ CI} = 4.46 \times 10^{-8} - 5.10 \times 10^{-7}$ , **Fig. 5A**). Together, these data suggest that healthy males (i.e., those without rare pediatric conditions) are more likely than females to carry rare missense variants in *CHD1*, indicating a sex-differential susceptibility to these

variants. In the All of Us dataset<sup>54</sup>, sex at birth was not associated with being a carrier of at least one rare (MAF < 1%) missense variant in *CHD1* ( $p = 0.15$ , two-sided z-test, logistic regression, Methods), potentially reflecting differences in recruitment practices, demographic composition, or variant calling pipelines across cohorts.

## Autosomal genes show sex-biased allele frequencies in healthy populations

Next, we extended our analysis to all autosomal genes with a probability of loss of function intolerance (pLI) score of > 0.9, i.e. highly constrained genes. Using the gnomAD exome sequencing database excluding UK biobank samples, out of a total of 3140 genes, we observed 663 genes with a significant overrepresentation of rare (MAF < 1 %) missense variants in males and 1060 genes with female overrepresentation after Benjamini-Hochberg multiple-testing correction ( $p\text{-adj} < 0.05$ , **Supplementary Table 6**). To assess the impact of variant frequencies on sex-specific overrepresentation, we next compared gene lists obtained using lower MAF thresholds (< 0.5% and < 0.1%) and observed similar results (**Fig. S5A-D**), suggesting that the observed sex bias is largely driven by very rare variants.

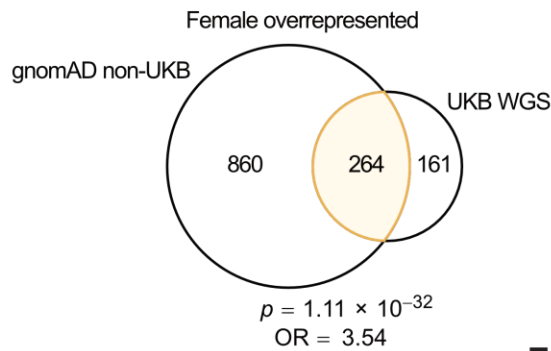
We repeated the sex overrepresentation analysis for rare missense variants (MAF < 1%) in the UKB WGS dataset, and also in the total gnomAD exomes dataset and the UK Biobank-only exomes dataset in gnomAD, observing positive correlations between frequency difference values (male AF – female AF, **Fig. S5E-G**). Using the UKB WGS dataset analysis, we identified 6 genes with male overrepresentation and 124 genes with female overrepresentation after Benjamini-Hochberg multiple-testing correction ( $p\text{-adj} < 0.05$ , **Supplementary Table 7**). To expand upon the analysis, we then examined all genes with  $p$ -values < 0.05 from the gnomAD non-UKB and UKB WGS datasets. Of the 425 genes with nominally significant female overrepresentation in the UKB WGS dataset, 264 genes overlapped with nominally significant female-biased genes in the non-UKB gnomAD dataset ( $p = 1.11 \times 10^{-32}$ , OR = 3.54, **Fig. 5B**). Of the 68 genes with nominally significant male overrepresentation in UKB, 34 overlapped with genes with male overrepresentation in gnomAD ( $p = 5.63 \times 10^{-7}$ , OR = 3.53, **Fig. 5C**). This list included *CHD1* and *MAP3K1*, where gain-of-function variants in *MAP3K1* are known to cause gonadal dysgenesis in 46,XY individuals<sup>55,56</sup>, while loss-of-function variants impair female reproductive development<sup>57</sup>. The top 10 overlapping genes with male overrepresentation included *SEZ6L*, *HLX* and *ALKBH5*, while female-biased genes included *PDZD2*, *CASR* and *CFH* (**Fig. 5D**). The results were similar using lower MAF thresholds (< 0.5% and < 0.1%, **Fig. S5H-K**).

Overall, fewer genes exhibited sex bias in the UKB WGS dataset compared to the gnomAD non-UKB exomes dataset, with the most pronounced difference observed among genes with male overrepresentation. Notably, the UKB-only exome data from gnomAD showed a similar number of genes with male or female sex bias as the UKB WGS dataset using the MAF < 1% threshold (**Fig. S5L-M**). Differences between the gnomAD and UKB datasets may therefore reflect disparities in cohort compositions. The UK Biobank is known to exhibit ascertainment bias toward healthier individuals with higher income and educational attainment<sup>58,59</sup>, 88.26% of whom self-identify as British/White.<sup>60</sup> The gnomAD non-UKB v4 exomes dataset is drawn from case-control studies of adult-onset disorders as well as diverse biobanks from around the world and includes a higher proportion of individuals with non-European genetic ancestries<sup>30</sup> (35.87%, <https://gnomad.broadinstitute.org/stats>). Thus, using two independent population datasets, we have identified 298 genes that display sex-biased allele frequencies of rare missense variants (**Fig. 5E, Supplementary Table 8**).

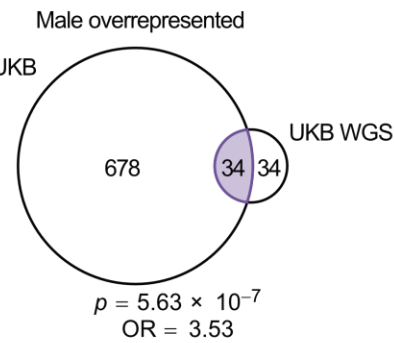
A

	Agg. alt. allele count		Agg. total allele count		Agg. alt. allele freq.		N loci	FrqDiff	p value	$\chi^2$	95% CI
	Male	Female	Male	Female	Male	Female					
gnomAD exomes	11975	11170	$8.89 \times 10^8$	$9.00 \times 10^8$	$1.35 \times 10^{-5}$	$1.24 \times 10^{-5}$	1261	$1.07 \times 10^{-6}$	$3.65 \times 10^{-10}$	39.29	$3.66 \times 10^{-7} - 7.01 \times 10^{-7}$
UK Biobank WGS	7565	8643	$4.96 \times 10^8$	$5.87 \times 10^8$	$1.53 \times 10^{-5}$	$1.47 \times 10^{-5}$	1104	$5.54 \times 10^{-7}$	0.0192	5.48	$4.46 \times 10^{-8} - 5.10 \times 10^{-7}$

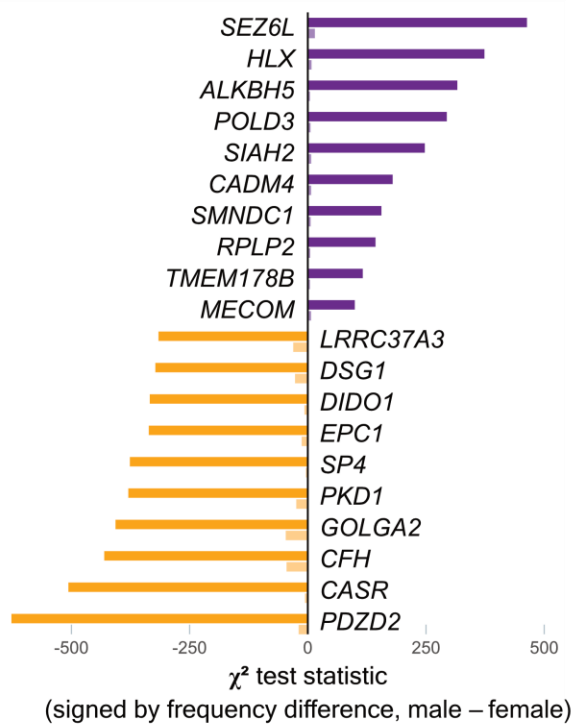
B



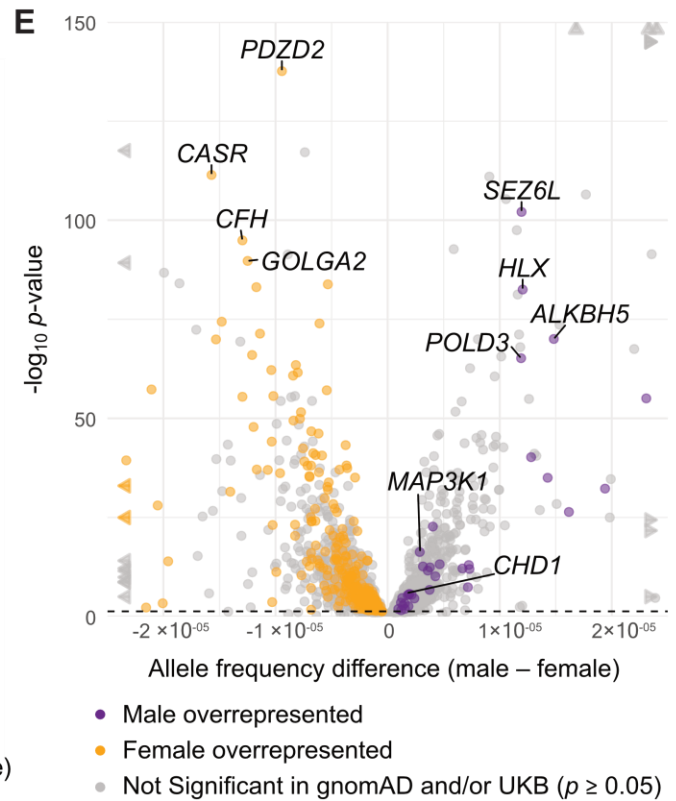
C



D



E



**Figure 5: Sex-bias of rare missense variants in *CHD1* and other highly constrained genes**

(A) Comparison of aggregated alternate allele frequencies of rare (MAF < 1%) missense variants across males and females in the *CHD1* gene in the total gnomAD exomes and UK Biobank WGS datasets. The two-proportions z-test was used to compare frequencies between sexes. The overlap between genes with nominal (B) female and (C) male overrepresentation ( $p < 0.05$ ) of rare (MAF < 1%) missense variants



for highly constrained ( $pLI > 0.9$ ) autosomal genes in the gnomAD non-UKB exomes and UKB WGS datasets. The  $p$  values and odds ratios for the overlaps were calculated using Fisher's exact test using a background of 19980 genes (number of protein coding genes in GENCODE hg38). (D) Chi-squared test statistic multiplied by the sign of the aggregated allele frequency differences for the top 10 genes from the overlapping gene lists. The sign corresponds to the direction of the aggregated allele frequency difference (male – female). (E) Sex differences in rare ( $MAF < 1\%$ ) missense variant frequencies for highly constrained ( $pLI > 0.9$ ) autosomal genes. Each dot represents a gene, showing the aggregated alternate allele frequency difference (male – female) against the  $-\log_{10} p$ -value from the gnomAD non-UKB exomes dataset. Statistics were computed using two-proportions z-test. Arrows represent genes which exceed the plot range. The dashed line represents the  $p = 0.05$  threshold for the gnomAD dataset analysis. Colored genes are those with  $p$ -values  $< 0.05$  in both the gnomAD non-UKB exomes and UKB WGS datasets.

## Neurodevelopmental functions are enriched in autosomal sex-biased genes

A subset of our sex-biased genes may be true examples of disorders with male or female protective effects. Any sexual dimorphisms may however hinder the ability to define them as disease-causing genes. When we assessed the overlap of our sex-biased gene list with a published list of genes with sex-biased expression in human organs<sup>61</sup>, *CHD1* (overrepresented in males), *VPS54*, *TTC7B* and *LIMK1* (overrepresented in females) appeared in both lists, with differential expression in brain tissue.

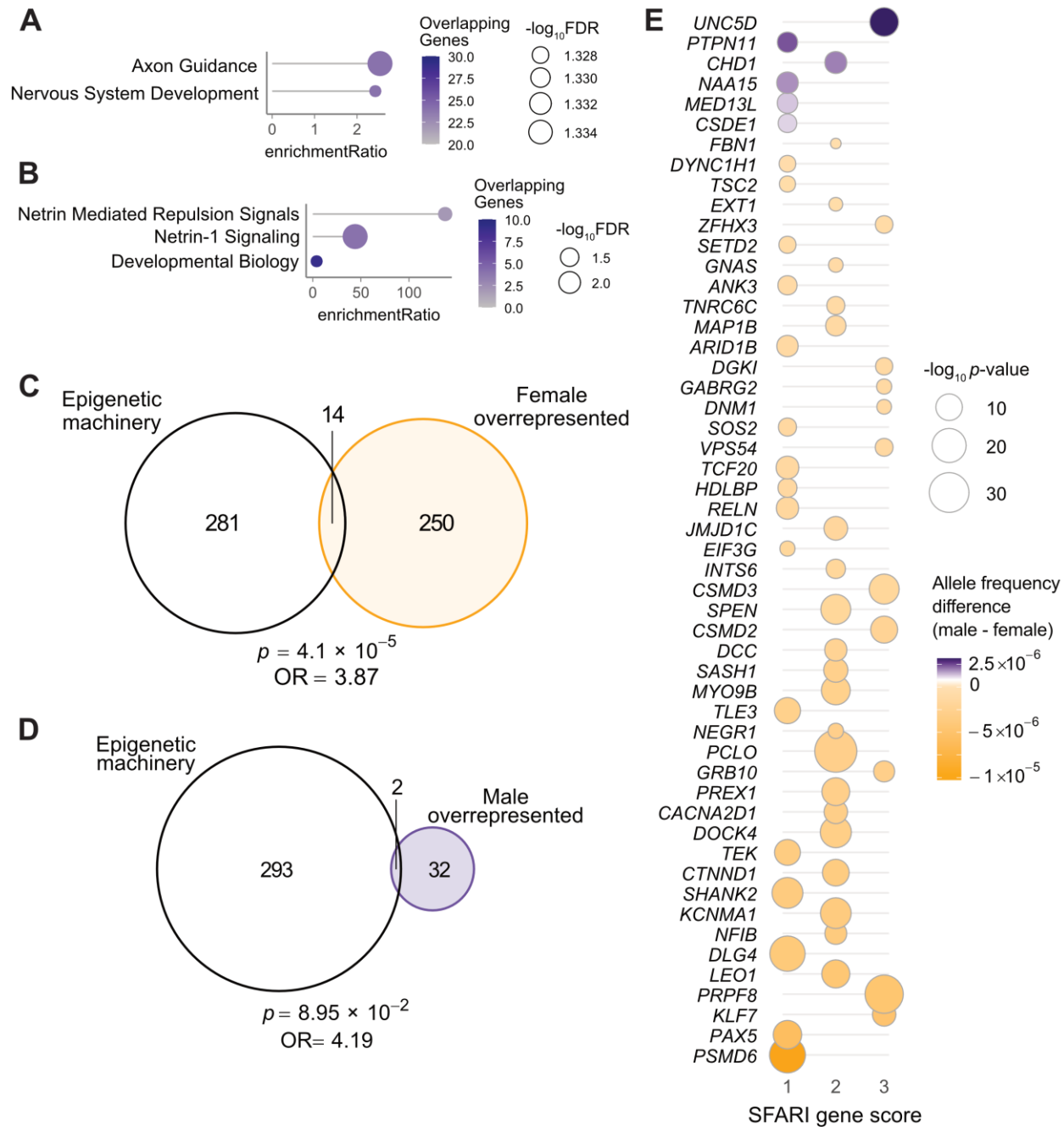
We then looked deeper into the molecular functions of the genes in each list by performing overrepresentation analysis. Genes with female overrepresentation in both gnomAD and UKB were significantly enriched for molecular function gene ontology terms relating to “vascular endothelial growth factor receptor activity”, “histone deacetylase activity”, “transmembrane receptor protein kinase activity” and transcriptional regulation (**Fig. S6A**,  $FDR < 0.1$ ), and for Reactome pathway terms “axon guidance” and “nervous system development” (**Fig. 6A**,  $FDR < 0.1$ ). In contrast, while genes with male overrepresentation were not significantly enriched for molecular function gene ontology terms, they were significantly enriched for the Reactome pathway term “netrin mediated repulsion signals”, “netrin-1 signaling” and “developmental biology” (**Fig. 6B**,  $FDR < 0.1$ ).

Given that terms relating to histone deacetylation, transcriptional regulation and neurological development were amongst the most significantly enriched for the list of genes with female



overrepresentation, we next examined overlap with the epigenetic machinery (EM). EM genes are a known class of genes that are highly intolerant to loss of function variation and frequently linked to NDD, with *CHD1* being a part of this group<sup>62</sup>. Using a pre-defined EM gene list<sup>62</sup>, we observed a significant overlap of 14 EM genes with female overrepresentation in gnomAD and UKB including *DNMT1*, *HDAC3* and *JMJD1C* ( $p = 4.1 \times 10^{-5}$ , OR = 3.87, **Fig. 6C**), and a non-significant overlap of 2 EM genes with male overrepresentation, *CHD1* and *SMNDC1* ( $p > 0.05$ , **Fig. 6D**, epigenetic function denoted in **Supplementary Table 8**). Notably, the *SMNDC1* transcript is a known target of repression by the fragile X mental retardation protein (FMRP)<sup>63,64</sup>.

As autism is an NDD with known sex-bias, we next assessed the overlap of the genes we identified with female or male overrepresentation with the SFARI Gene autism candidate list<sup>65</sup>. We identified a significant overlap of 48 SFARI genes with female overrepresentation ( $p = 5.1 \times 10^{-12}$ , OR = 3.57, **Fig. S6B**) and a significant overlap of 6 SFARI genes with male overrepresentation ( $p = 0.015$ , OR = 3.36, **Fig. S6C**). Among these, we observed 2 with a “Syndromic” classification but no score, 20 with a SFARI gene score of 1 (high confidence), 21 with score 2 (strong candidate), and 11 with score 3 (suggestive evidence) (**Fig. 6E, Supplementary Table 8**). Although only 6 of the genes with male overrepresentation overlapped with SFARI autism genes, additional genes are associated with neurodevelopmental and psychiatric disorders including *ADD3*<sup>66</sup>, *FBXW7*<sup>67</sup>, *POLD3*<sup>68</sup> and *SEZ6L*<sup>69</sup>. These findings highlight a notable enrichment of neurodevelopment-associated genes with significant differences in the frequency of rare missense variants between males and females, suggesting that genetic factors contributing to neurodevelopment may be particularly prone to sex-biased penetrance.



**Figure 6: Overrepresentation of neurodevelopmental, epigenetic and autism-related genes among those showing sex bias in variant frequencies**

Overrepresentation analysis of genes found in both gnomAD and UKB with (A) female overrepresentation and (B) male overrepresentation using the reactome pathways gene sets. (C) Overlap of genes that show female overrepresentation in gnomAD and UKB datasets with epigenetic machinery genes. (D) Overlap of genes that show male overrepresentation in gnomAD and UKB datasets with

epigenetic machinery genes. The  $p$  values and odds ratios for the overlaps were calculated using Fisher's exact test with a background of 19980 genes (number of protein coding genes in GENCODE hg38). (E) Genes with male or female overrepresentation in gnomAD and UKB that overlap with SFARI autism candidate genes. SFARI gene score is shown on the x-axis, 1: High confidence, 2: Strong candidate, or 3: Suggestive evidence. The male - female aggregated alternate allele frequency difference is depicted by the color of the circles and  $-\log_{10} p$ -value is depicted by the size of the circles.

## Discussion

In the present study, we have extended the understanding of PILBOS, highlighting key clinical information from the largest PILBOS cohort assembled to date. We have demonstrated validity of a new PILBOS mouse model, which may serve as a valuable reagent to the research community in the future. Most importantly we have identified a protective role of androgens against two of the disorder phenotypes, growth retardation and the neuronal proliferation defect in mice, and demonstrated a sex-bias in *CHD1* variants on the human population level. Furthermore, we have identified additional highly constrained genes with a similar sex bias, suggesting that other autosomal Mendelian disorders may show similar sexual dimorphism, and this may be worth further study in murine models and clinical settings.

As no cohort studies of PILBOS have been published since the original description of the disorder, the data presented here will likely prove valuable to clinicians in placing diagnoses and counseling patients as well as to DNA diagnostic laboratories in interpreting variants. Notably, males were overrepresented amongst PILBOS probands and females had more severe phenotypic scores, suggesting that males may have increased tolerance for disruption of the CHD1 protein compared to females. This may be a useful rule that can be further fleshed out as additional individuals are diagnosed.

Our mouse model may be useful in the future to test potential treatments for PILBOS, or to gain deeper insights into the role of CHD1 in development. Although a prior study observed no phenotypic abnormalities in *Chd1*<sup>+/-</sup> mice<sup>41</sup>, this may be caused by not keeping track of sex of individual animals or the use of different assays to evaluate phenotypes. Another possibility is that missense variants in *Chd1* can have more severe dominant-negative effects, but given the paucity of heterozygous nonsense or null

*CHD1* variants in the human population and our functional data in this study, this seems less likely. A second study identified spatial memory deficits in a homozygous mouse model with an N-terminal deletion in the gene encoding the CHD1 protein<sup>70</sup>, which contrasts with our findings in the *Chd1*<sup>R616Q/+</sup> mice. This discrepancy may be due to the presence of a functional N-terminus in our model, or the fact that our mice have one fully functional allele. It could also relate to the absence of intellectual disability in the subject from which the original human p.R618Q variant was derived. Potentially, variants with more profound effects on the CHD1 protein may lead to learning deficits in mouse models that are easier to discriminate from WT littermates.

Notably, oxytocin levels were elevated in the *Chd1*<sup>R616Q/+</sup> mouse model. This may suggest abnormal hypothalamic function, which by itself could lead to both weight and anxiety phenotypes, as increased oxytocin is known to reduce weight gain by decreasing food intake, increasing energy expenditure and promoting lipolysis<sup>39-41</sup>, and is also associated with anxiety<sup>42-44</sup>. Oxytocin itself is also known to drive sexually dimorphic circuits in the brain<sup>71</sup>, and its receptor is expressed in a sexually dimorphic manner<sup>72</sup>. *PTEN* variants, a common cause of ASD, also lead to elevated levels of oxytocin in mouse models<sup>73</sup>, and our finding is thus a second example of elevated oxytocin levels in the hypothalamus in the context of ASD. This link opens avenues to study elevated oxytocin as a novel biomarker of a subset of ASD, which may suggest shared pathogenesis and a potential for shared therapeutic strategies.

The protective role of androgens we demonstrate in this study opens a new path for investigation into therapeutic strategies for PILBOS treatment. Indeed, how do androgens protect against CHD1 deficiency? Testosterone is known to increase survival of neurons in the rodent hippocampus<sup>74,75</sup> and androgen signaling has been reported to promote proliferation in the mouse cortex during development<sup>76</sup>. In human cerebral organoids, androgens promote excitatory neurogenesis<sup>77</sup> and studies indicate their neuroprotective effects across multiple brain regions, including the rat hypothalamus and visual cortex<sup>78</sup>, as well as in promoting motor neuron survival in injury models<sup>79</sup>. Thus, androgens may be protective against PILBOS through counteracting the neural proliferation deficit and promoting neuronal survival. At the molecular level, the AR may facilitate chromatin remodeling at sites typically regulated by CHD1, potentially compensating for the deficiency in its normal function. Upon androgen stimulation, the AR is known to recruit pioneer factors, which promote chromatin opening and enhance access to target gene loci<sup>80</sup>. Furthermore, AR activation is protective against genomic instability and DNA damage accumulation, driving transcription of non-homologous end-joining pathway genes and the AR and TOP2B are co-localized to genomic breakpoints following androgen stimulation<sup>81</sup>. Interestingly, the

effects of CHD1 deficiency on the transcription of long genes mirrors that of another chromatin remodeler, CHD7, whose interaction with TOP2B maintains transcription of long neuronal genes<sup>82</sup>. Notably, TOP2B is also a CHD1 interaction partner<sup>21</sup>, indicating a partially shared function of the two remodelers. Another possible driver of PILBOS phenotypes could be reduced androgen levels in PILBOS-affected individuals of either sex, a phenotype which is common with CHD7 deficiency<sup>83,84</sup>. In line with this hypothesis, genes involved in the “steroid biosynthesis” pathway were significantly enriched amongst the most significantly altered genes in NPCs from our mouse model, with almost all being downregulated.

It is interesting to speculate that altered expression of many androgen-responsive genes could give rise to the female-specific defects in PILBOS, which are prevented in the majority of males by their naturally higher levels of androgen signaling, bringing the expression of these genes closer to the baseline level. If this effect occurs widely *in utero*, as is the case for neuronal proliferation in our mouse model, it could lead to increased female embryonic lethality, resulting in fewer females with PILBOS diagnoses. Indeed, combining the clinical data with our analysis of the gnomAD and UK Biobank population data, our findings suggest increased embryonic lethality amongst female carriers of pathogenic *CHD1* variants. In males with the most severe defects in CHD1, normal androgen signaling may be insufficient to overcome the induced changes in gene expression, resulting in the development of PILBOS features. A consequence of this line of thought would be that PILBOS features may improve in males upon puberty. However, since our cohort in this report consists almost entirely of young children, we cannot draw conclusions on this matter at present, but it would be of great interest to follow these individuals prospectively and measure hormone levels such as androgens and hypothalamic hormones. This could open up a potential treatment right away by providing testosterone treatment to those males that have the lowest values.

Overrepresentation of a particular sex in a clinical cohort of a genetic disorder can indicate one of two possibilities: greater susceptibility of that sex to variants in a given gene, or protection against them. Traditionally, the former has been assumed, as seen in arguments for male susceptibility in autism being characterized by male overrepresentation. However, population-level data can challenge this assumption. Using PILBOS—a genetic disorder with male overrepresentation—we illustrate that when a sex is overrepresented both in affected individuals and in the healthy population, this pattern may instead reflect a protective effect in that sex. A similar pattern is seen for the gene *PTPN11*, whose variants we observe to be overrepresented amongst males in the healthy population. *PTPN11* is a cause

of Noonan syndrome, which is known to exhibit a male-biased sex ratio, both for *de novo* and familial cases<sup>85</sup>. Furthermore, our analysis revealed variants in the gene *HDLBP* to be overrepresented in females. Notably, *HDLBP* resides within the locus for 2q37.3 microdeletion syndrome, a NDD with a female-biased sex ratio<sup>86</sup>.

Conversely, some genes do support the traditional interpretation of increased susceptibility for the sex that is overrepresented in clinical cohorts. For example, *SEZ6L* was the most significant high-confidence male-overrepresented gene in our population dataset analysis. Variants in *SEZ6L* have been associated with bipolar disorder specifically in females<sup>69</sup>, and *Sez6l* knockout mice show female-limited weight differences and motor-coordination deficits<sup>87</sup>. Another example is *NRG1*, which was overrepresented in females in both healthy population datasets and is linked to sex-specific neurological and behavioral phenotypes that are more severe in male mice<sup>88</sup> and rats<sup>89</sup>. In humans, decreased circulating NRG1 levels correlate with increased severity of early-onset schizophrenia specifically in males<sup>90</sup> and variants in *NRG1* are associated with elevated schizophrenia risk in males<sup>91</sup>.

One caveat of our analysis is that the same individual could be counted more than once if they have multiple variants in our gene list. However, the effect of this potential bias is likely minimal given the rarity of the variants and the stringent filtering criteria applied.

Overall, the results of our population database analyses are congruent with the concept of a female-protective effect being more common in genetic disorders, with many more genes showing female rather than male overrepresentation. However, the notable number of genes with male overrepresentation suggests that several disorders may involve a male-protective effect, raising the possibility that androgens may play a protective role in more contexts. Supporting this idea, both testosterone administration and sex reversal using an *Sry* transgene have been shown to reduce female embryonic lethality in a mouse model of MCM deficiency<sup>92</sup>. It is compelling to speculate that identifying a targetable downstream effector of androgen signaling could provide a potential therapeutic applicable to many NDDs.

In summary, we have identified a non-X-linked NDD where males are protected by their naturally higher androgen levels and demonstrate that this principle may be more widely applicable. These findings offer new avenues for developing targeted therapeutic interventions for PILBOS and potentially other NDDs with sex-based differences, and for exploring the role of sex-mediated penetrance in other autosomal dominant disorders.

## Materials and methods

### GeneMatcher and clinical data collection

We used the GeneMatcher platform<sup>26</sup> (<https://genematcher.org>) to identify additional individuals carrying *CHD1* variants. Our study was specifically approved by the Icelandic Ethics Committee (Vísindasiðanefnd) and the Ethics Committee of Landspítali University Hospital. All work was completed in compliance with the Institutional Review Board of each institution and proper written consent was obtained from each patient/patient family included in the study.

### Variant Classification

Variant classification was performed by applying American College of Medical Genetics and Genomics (ACMG) guidelines using Franklin variant interpretation platform by Genoox<sup>93</sup> (<https://franklin.genoox.com>). Missense variants were further interpreted using AlphaMissense<sup>31</sup> by Google DeepMind, a model built on artificial intelligence designed to predict the effects of missense variants scoring them on a scale between 0 and 1. Missense variants were classified based on these AlphaMissense scores, with high-scoring variants flagged as deleterious (>0.761), low-scoring variants as benign (<0.331), and those in between uncertain (0.331-0.761).

### Visualization of variants

A lollipop plot was generated using Python with the Matplotlib library<sup>94</sup> to visualize the distribution and frequency of variants in carried by individuals in the study.

### Phenotype scoring

A custom phenotype scoring system was designed, where those individuals carrying truncating LOF variants or putatively disruptive missense variants (categorized as deleterious or unknown by AlphaMissense) were scored based on the phenotype. One point was assigned for each of the following core phenotypic characteristics; developmental delay, speech apraxia/speech delay, hypotonia, seizures/abnormal EEG, macrocephaly, almond shaped eyes and being below 1% ile or above 99% ile in height or weight (1 point each). One point was given for intellectual disability and autism spectrum disorder, but if stated that symptoms were mild, half a point was assigned. Individuals with scarce



phenotype information were excluded, i.e. less than three phenotype entries (1/20 in the truncating LOF group, 3/20 in the putatively disruptive missense group). The maximum possible score was 10 points.

### **CHD1 protein production**

The human *CHD1* cDNA sequence was cloned from a plasmid (Sino Biological), and the p.R618Q variant was introduced using site-directed mutagenesis (New England Biolabs). N-terminal deletions of *CHD1*<sup>WT</sup> and *CHD1*<sup>R618Q</sup> were generated by PCR and cloned into the pCoofy51-TwinStrep tag expression plasmid using Sequence and Ligation Independent Cloning (primer sequences provided in **Supplementary Table 9**). Protein production and purification was performed at the EMBL Protein Expression and Purification Core Facility. The proteins were produced using High Five insect cell culture, then purified using StrepTactin affinity purification and size-exclusion chromatography. The presence of CHD1 protein was confirmed by mass spectrometry.

### **Nucleosome remodeling assay**

The nucleosome remodeling assay was modified from the Epicypher Nucleosome Remodeling Assay protocol (<https://www.epicypher.com/content/documents/protocols/epidyne-guide.pdf>). Reactions were prepared by combining the previously synthesized human CHD1 proteins with N-terminal deletions (either WT or R618Q) at a concentration of 20 nM with 40 nM Epidyne Nucleosome Remodeling Assay Substrate (Epicypher, 16-4101) and 25 units of DpnII enzyme (New England Biolabs) in assay buffer (20 mM Tris pH 7.5, 50 mM KCl, 3 mM MgCl<sub>2</sub>, 0.1 mg/mL bovine serum albumin) to a final volume of 20 µL. Reactions were initiated by addition of 1 mM ATP, then incubated for different lengths of times at 37°C. Reactions were quenched by the addition of 20 µL assay quench buffer (10 mM Tris pH 7.5, 40 mM EDTA, 0.60% SDS, 50 µg/mL Proteinase K), then incubated at 50°C for 20 min and allowed to cool to room temperature before loading on gel. Negative controls without exposure to ATP were included for both CHD1 proteins, as well as a substrate-only control, and a positive and negative digest control (Epicypher, 23618-4101 & 23618-4100).

### **Mice**

The *Chd1*<sup>R616Q/+</sup> mouse line were created by Wellcome Sanger (Hinxton, Cambridge) through a grant from INFRAFRONTIER 2020 to H.T.B and G.P. Mice were created by CRISPR-Cas9 to generate a mutant allele containing the missense mutation Chr17:15738534 G>A [R616Q]. In an effort to create the desired variant (R616Q), a synonymous mutation was created: Chr17:15738553 C>A (guide RNA and mutant oligo sequences in **Supplementary Table 9**). The mice carrying the synonymous variant demonstrated no



obvious phenotypic difference from WT littermates, as an example we show the righting assay (**Fig. S7A**). The mice were bred and maintained according to the Institute's standard protocols before being shipped to our facility. All animal procedures were conducted in accordance with the guidelines and regulations approved by the University of Iceland and the Icelandic Food and Veterinary Authority (license no. 2208602) and accepted by the National Expert Advisory Board on Animal Welfare. The animals were maintained in a 12 h light-dark cycle in a temperature- and humidity-controlled environment (21-23°C, ~40% humidity), and provided with food (Altromin NIH#31 M (breeding) or Altromin 1324 (maintenance), Brogaarden) and water *ad libitum*. Cages were equipped with appropriate environmental enrichment to promote natural behaviors and well-being. Enrichment included paper nesting materials, cardboard tunnels for burrowing and exploration, and plastic shelters to provide hiding spaces. Additionally, mice were provided with wooden chew sticks to support dental health and prevent overgrowth of teeth. Enrichment materials were replenished as needed. Mice were housed with 2-5 animals per cage. The mice were backcrossed to the C57BL/6N strain (Taconic Biosciences), then regularly crossed with WT C57BL/6N females. Genotyping was performed using tissue from ear clips with the following primers: Chd1\_fwd (5'-CAGCAGGCCAGTCAACTGAGGC-3'), Chd1\_rev (5'-TCTCCCAACTGCATCTGAGCACA-3'). The presence of the mutant allele was confirmed either by Sanger sequencing or by restriction digest with ClaI, which recognizes the WT but not the Chr17:15738534 G>A sequence.

#### **RNA extraction, cDNA synthesis and RT-qPCR**

RNA was extracted by homogenization in Tri-reagent (Thermo Fisher Scientific, 15596026), then isolated using the Direct-zol RNA isolation kit (Zymo, R2062) according to the manufacturer's instructions. cDNA was synthesized from 1 µg of RNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, 4368814) according to the manufacturer's instructions. RT-qPCR was performed using the Luna Universal qPCR Master Mix (NEB, M3003E) according to manufacturer's instructions on the CFX384™ Real-Time PCR Detection System (Bio-Rad). Relative quantitation was calculated according to the Pfaffl method<sup>95</sup>, with *Actb* used as a reference gene. To determine primer efficiencies, standard curves were prepared using 10-fold dilutions of mouse cDNA. Primer amplification factors were calculated based on  $10^{-1/\text{slope of standard curve}}$ . Each biological replicate represented an individual mouse and was performed in technical triplicate.

#### **Primary cell isolation and culture**

Mouse embryonic fibroblasts (MEFs) were isolated from day 14.5 embryos as previously described<sup>96</sup>. MEFs were cultured in DMEM-F12 (Gibco, 31331093) with 10% fetal bovine serum (Gibco, A5256701), and sub-cultured when confluency reached 70-80%. Mouse cortical neuronal progenitor cells (NPCs) were isolated from day 18.5 embryos as previously described<sup>46</sup>. NPCs were cultured in neurobasal growth medium containing 1X B27 supplement (Thermo Fisher Scientific, 17504044), 1X Penicillin/Streptomycin (Thermo Fisher Scientific, 15140122), 1X Glutamax (Thermo Fisher Scientific, 35050038), 20 ng/mL FGF-2 (Peprotech, 100-18B), 20 ng/mL EFG (Peprotech, AF-100-15), and 2 µg/mL Heparin (MP Biomedicals, 210193125). NPCs were either sub-cultured or media was replaced every 2 days.

### **Protein extraction and western blot**

Cells ( $1 \times 10^6$ ) were lysed in 50 µL RIPA buffer (50 mM Tris pH 8, 150 mM NaCl, 1% NP-40, 1% Sodium Deoxycholate, 0.1% SDS, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride with 1% protease inhibitor cocktail, Sigma, P8340) on ice for 30 min, then centrifuged at 13000 rpm for 15 min at 4°C. Supernatants were mixed with 4X sample buffer (240 mM Tris pH 6.8, 8% SDS, 40% glycerol, 0.04% bromophenol blue, 5% beta-mercaptoethanol) to a 1X final concentration, then heated at 95°C for 5 min and allowed to cool to room temperature before being loaded on 8% polyacrylamide gel and electrophoresis for western blot<sup>97</sup>. Proteins were transferred to PVDF membranes for 2 h at 200 mA. Membranes were blocked for 1 h at room temperature in 5% bovine serum albumin (BSA, Applichem, A1391), then incubated overnight with primary antibodies (anti-CHD1, 1:1000 dilution, Cell Signaling Technologies, 4351; anti-beta-Actin, 1:1000 dilution, Abcam, ab82224). Membranes were washed 3X with tris-buffered saline with 0.1% tween-20 (TBST), then incubated 1 h with secondary antibodies (IRDye® 800CW Donkey anti-Rabbit IgG (H + L) & IRDye® 680RD Donkey anti-Mouse IgG (H + L), Licor, 926-32213 & 926-68072). Membranes were again washed 3X with TBST before visualization using an ODYSSEY® DLx (Licor). Band intensities were quantified using ImageJ Fiji software<sup>98</sup>.

### **Mouse motor/behavioral testing**

The mouse righting reflex and hindlimb strength was assessed in postnatal day 6 pups using the surface righting and hindlimb suspension tests as previously described<sup>99</sup>. The Morris water maze test was performed as previously described<sup>100</sup>. The data were analyzed using the EthoVision XT software (Noldus). For the open field test, each mouse was placed in an open field consisting of a 32 x 32 cm square enclosure open on top and allowed to roam freely while being filmed for 10 minutes. The open field

enclosure was cleaned thoroughly with soap and water, then with 70% ethanol and allowed to dry completely between each test. The video data were analyzed using EthoVision XT software (Noldus), with the center of the floor area defined as an 18 x 18 cm area in the center of the open field.

### **Mouse vocalization measurements**

Pups at postnatal day 3-11 were temporarily removed from their dams and placed inside a soundproof box. Vocalizations were recorded over a four-minute period using the Petterson M500-384kHz USB Ultrasound Microphone with gain set to 97. Pups were then returned to their home cage. Audio recordings were processed with UltraVox software (Noldus) to count the number of vocalizations and measure the latency until first vocalization.

### **Electroretinography (ERG) and Optical Coherence Tomography (OCT)**

Corneal ERG responses were obtained from mice anesthetized with an IP administered mixture of ketamine (40 mg/kg) and xylazine (4 mg/kg) after 24 hours dark adaptation and then after 10 minutes light adaptation with a light-adapting background of 9 cd/m<sup>2</sup>. ERG recordings were performed with a Celeris D430 rodent recording system (Diagnosys LLC, Lowell, MA, USA) equipped with “bright standard” stimulators and light guide electrodes, following procedures described previously<sup>101</sup>. OCT scans from the eyes of anesthetized mice were obtained with a Micron IV SD-OCT system (Phoenix-Micron, Inc, Bend, OR, USA) on a rodent fundus camera. Pupils were dilated with mydriatic eye drops (phenylephrine and tropicamide) and corneas locally anesthetized with tetracaine eyedrops for both ERG and OCT examinations.

### **Oxytocin ELISA**

Hypothalamus tissue was manually dissected from mouse brains and lysed in RIPA buffer as above. Protein concentrations were quantified using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, 23225) according to the manufacturer's instructions. Oxytocin levels were quantified using the Oxytocin ELISA kit (Enzo, ADI-901-153A-0001). Protein lysates were normalized to a concentration of 0.4 µg/µL by dilution in the kit assay buffer and then the assay was performed according to the manufacturer's instructions.

### **MTT proliferation assay**

Proliferation was assessed using the CellTiter 96® Non-Radioactive Cell Proliferation Assay MTT (Promega, G4000) according to the manufacturer's instructions. MEFs and NPCs were seeded at a

density of  $1 \times 10^3$  cells per well in a 96-well flat-bottomed plate. After 24 h, media was replaced with 100  $\mu$ L fresh media plus 15  $\mu$ L kit dye solution and plates were incubated at 37°C (MEFs for 3 h, NPCs for 4 h). The reaction was stopped by the addition of 100  $\mu$ L stop solution and incubated for 1 h at room temperature protected from light. Absorbance was read at 570 nm and results were normalized by subtracting the background signal.

#### **Cell cycle analysis**

Cells were harvested and pelleted, then pellet was resuspended in 400  $\mu$ L ice-cold PBS. While gently vortexing, 800  $\mu$ L ice-cold 100% ethanol was added to a final concentration of 66% for fixation. After washing in PBS, cells were stained with 0.2  $\mu$ g/mL DAPI (AAT Bioquest, ABD-17511) diluted in PBS, with 100  $\mu$ g/mL RNase A (Abcam, ab139418) and incubated at 37°C for 30 min protected from light. Staining was measured by flow cytometry using the FACS Calibur (BD, for MEFs) or Attune NxT (Thermo Fisher Scientific, for NPCs). Data were analyzed with FlowJo software (v.10.8.1) with gating to include only live single cells (**Fig. S7B**).

#### **EdU DNA synthesis assay**

DNA synthesis was measured using the Click-iT™ Plus EdU Alexa Fluor™ 647 Flow Cytometry Assay Kit (Thermo Fisher Scientific, C10634) according to the manufacturer's instructions. Cells were incubated with 10  $\mu$ M EdU for 2h before harvesting, fixation and Click-iT reaction. EdU incorporation was assessed by flow cytometry using Attune NxT (Thermo Fisher Scientific).

#### **Annexin V apoptosis assay**

Apoptosis was quantified using the FITC Annexin V Apoptosis Detection Kit I (BD, 556547) according to the manufacturer's instructions. Staining was measured by flow cytometry using the Attune NxT (Thermo Fisher Scientific) and FlowJo software (v.10.8.1) was used for analysis.

#### **RNA sequencing**

NPCs obtained from 4 individual embryos per genotype from 2 separate litters were seeded (passage 3)  $1.5 \times 10^5$  cells per well in a 6-well plate coated with poly-D-lysine and laminin. After 3 days of culture, cells were treated with DMSO vehicle control or 10 nM dihydrotestosterone (Selleckchem, S4757) for 6 h, then RNA was extracted as described above. RNA quality and integrity was assessed using a Nanodrop spectrophotometer and a Bioanalyzer instrument with the RNA 6000 Nano Kit (Agilent, 5067-1511). Only samples with RNA integrity (RIN) scores above 8 were used for library preparation. PolyA-enriched

RNAseq library construction and Illumina paired-end sequencing was performed by Novogene, to obtain 30 million paired-end 150 bp reads for each sample. Reads were pseudoaligned to the GRCm39 transcriptome using Kallisto using paired-end mode and 100 bootstraps, then prepared for differential expression analysis using the Tximport package in R, using the TxDb.Mmusculus.UCSC.mm10.ensGene package to assign reads to genes. Differential expression analysis was performed using DESeq2, keeping only transcripts with at least 10 counts. For the DMSO vs DHT treatment comparison, we used a design matrix that accounted for embryo as well as genotype and treatment, since both treatments were each applied to the same embryo. Since the p-value histograms displayed a right-skewed distribution, which can indicate non-random noise or inappropriate assumptions, we utilized the DESeq2 z-score data as input for the FDRtool package to achieve more accurate false discovery rate estimations, and used these corrected p-values and q-values for further analysis. Overlaps were calculated by Fisher's exact test using the GeneOverlap package in R. Overrepresentation analysis was performed using WebGestalt<sup>102</sup> with the functional database settings, "GeneOntology: CellularComponent" and "Pathway: KEGG".

### **Mutational signature analysis**

The method used for variant discovery in the RNAseq data was adapted from the GATK<sup>103</sup> Best Practice Workflow for RNAseq short variant discovery (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels>). Raw RNAseq data were aligned to the GRCm39 v. 112 mouse genome using STAR<sup>104</sup> with two-pass mode, with bam output files sorted by coordinate. Duplicates were marked in bam files using Picard (<https://broadinstitute.github.io/picard/>) then with GATK tools<sup>103</sup>, SplitNCigarReads was used to re-format intron-spanning alignments, BaseRecalibrator, ApplyBQSR, and AnalyzeCovariates were used for base quality recalibration. For variant discovery, first variants were called on the WT samples using GATK Mutect2 in tumor-only mode to obtain variants that were then used to assemble a panel of normals using the CreateSomaticPanelOfNormals command. Variants were then called in each sample against the panel of normals using Mutect2. Variants common to at least 12 samples out of 16 were identified using the BCFtools<sup>105</sup> -iSec command, then filtered out of the VCF files. VCF files were read into R using the MutationalPatterns package<sup>106</sup> and limited to variants found within exons using the TxDb.Mmusculus.UCSC.mm39.refGene package. Indels were assessed using the get\_indel\_context() command. Mutational signatures were fit to known COSMIC signatures<sup>47</sup> using the fit\_to\_signatures() command.

### **Orchiectomy and testosterone implant surgeries**

Surgeries were performed on mice at postnatal day 15. Mice were anesthetized with 40 mg/kg ketamine and 4 mg/kg xylazine intra-peritoneally, as well as 7 mg/kg lidocaine subcutaneously at the incision site. In male mice, two abdominal incisions were made, one for each testis. Testicular tissue was pulled through the incision by the epididymal fat to expose the testis, and then the testicular blood vessels were crushed using hemostats with heat applied to prevent hemorrhage. The testes were then excised, and the incision closed with absorbable sutures (Vicryl, 4-0, Ethicol, V310H). For post-operative analgesia, male mice were treated with 0.1 mg/kg buprenorphine. In female mice, a small incision was made in the back of the neck and the testosterone 60-day slow-release silastic implant (Belma Technologies, T-M/60, Batch No. M60T1711-03NB) was placed under the skin. The incision was closed with an absorbable suture.

### **Testosterone treatment of dams**

Testosterone propionate (T, MedChemExpress, HY-B1269) was dissolved in corn oil to a stock concentration of 5 mg/mL by heating for 1 h at 50°C with frequent vortexing. Pregnant dams were injected with 20 mg/kg T or 4 µL corn oil per gram body weight subcutaneously in the hindleg once per day from embryonic day 8.5-17.5. On embryonic day 18.5, dams were euthanized by cervical dislocation and embryo telencephalon tissue was used to isolate nuclei. Embryos were obtained from 3-4 litters for each treatment.

### **Nuclear isolation for flow cytometry**

The telencephalon was dissected from embryo brains and placed in ice-cold homogenization buffer (5 mM calcium chloride, 3 mM magnesium acetate, 10 mM Tris pH 7.8, 0.17% beta-mercaptoethanol, 320 mM sucrose, 1 mM EDTA, 0.1% IGEPAL CA-630) and homogenized using 2 mL Dounce homogenizer. Homogenate was transferred to 1.5 mL centrifuge tube and centrifuged 8000 rpm for 30 seconds to pellet nuclei. Nuclei were resuspended in 200 µL FACS buffer and stained with anti-Ki67 (1:500, Abcam, ab16667) for 20 min at room temperature. Samples were stained with Alexa fluor-647 anti-rabbit secondary (1:1000, Thermo Fisher Scientific, A-21244) and 0.2 µg/mL DAPI (AAT Bioquest, ABD-17511) for 30 min at room temperature protected from light. Nuclei were passed through a pre-wet 100 µm filter, then fluorescence was measured by FACS using the SH800S flow cytometer (Sony). Data were analyzed using FlowJo software (v.10.8.1).

### **Human population variant analysis**

The list of all genes with available pLI scores was downloaded from the gnomAD v4.1 website<sup>30</sup> (gnomad.v4.1.constraint\_metrics.tsv). This list was filtered to keep only genes with pLI score > 0.9. The gnomAD v4.1 exome hail table data ("gnomad.exomes.v4.1.sites.ht") was downloaded from the gnomAD website<sup>30</sup>. Variants assigned to genes of interest were then selected, and the data were filtered to exclude sex chromosomes. Then the variant filters "AC0" (variants whose allele count is zero after filtering out low-confidence genotypes) and "AS\_VSQR" (variants that failed allele-specific Variant Quality Score Recalibration filtering thresholds of -4.0598 for SNPs and 0.1078 for indels) were applied. We then limited to missense variants that occurred at < 1% adjusted allele frequency at sites with at least 100 000 allele number (AN) counts.

Analysis of data from the UK Biobank resource was performed at deCODE Genetics using the UK Biobank resource, application number 56270. The filtered list of genes with pLI score > 0.9 previously obtained from gnomAD was used to extract rare (MAF < 1%) missense variants from the UK Biobank WGS dataset. For data extracted from all datasets, we calculated the aggregated alternate allele frequency per gene as the sum of the alternate allele counts (AC) divided by the sum of the allele number (AN) counts for each sex. We then used the two-proportions z-test using the prop.test() function in R to compare the aggregated alternate allele frequencies for each sex. The *p*-values were adjusted using the Benjamini-Hochberg correction. The frequency difference (FrqDiff) was calculated by subtracting the allele frequency for females from that of males (male – female). Overrepresentation analysis was performed using WebGestalt<sup>102</sup> with the functional database settings "Gene ontology: Molecular Function" and "Pathway: Reactome". Overlaps were calculated by Fisher's exact test using the GeneOverlap package in R.

Analysis in the All of Us Research Program's Controlled Tier Dataset version 8, were based on the exome sequencing data in a maximal set of unrelated samples<sup>54</sup>. We identified biallelic missense variants in *CHD1* from the columns "consequence" and "gene\_symbol" in the variant annotation table provided in the All of Us data release. Allele frequencies were computed for these variants in the exome sequencing data and variants with MAF ≥ 1% were subsequently filtered out. After identifying individuals carrying at least one rare missense variant in *CHD1*, we performed logistic regression in R where sex at birth (male or female) was the dependent variable and *CHD1* rare missense variant carrier status (carrier or non-carrier) was the explanatory variable (n = 380,281, male-fraction = 39.8%, carrier-fraction = 2.3%).



## List of Supplementary Tables

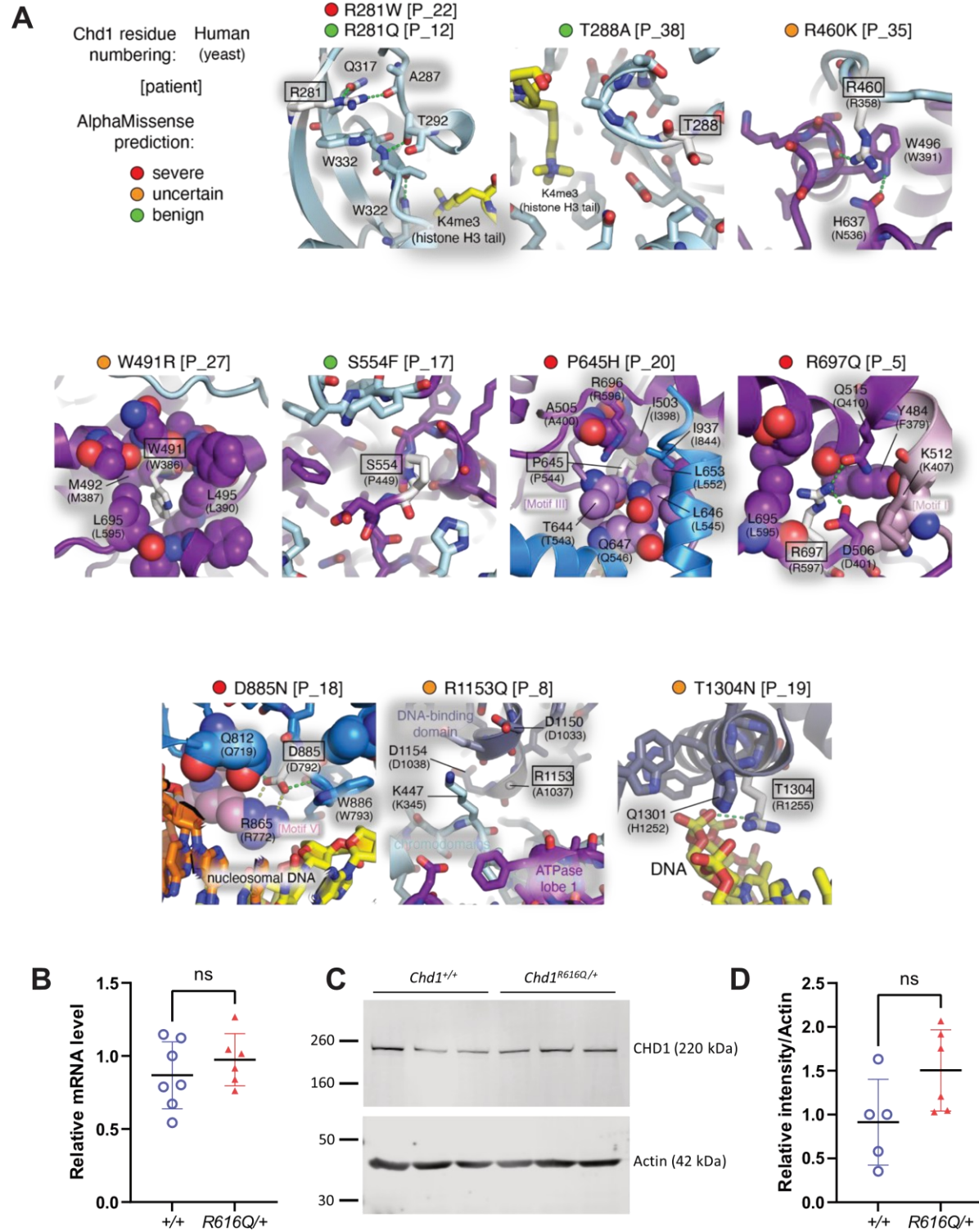
- Supplementary Table 1. Phenotypes observed in the PILBOS cohort with missense variants.
- Supplementary Table 2. Phenotypes observed in the PILBOS cohort with LOF and potential splice-site variants.
- Supplementary Table 3. Mendelian ratios of *Chd1*<sup>R616Q/+</sup> mice and WT littermates.
- Supplementary Table 4. Differentially expressed genes in NPCs from *Chd1*<sup>R616Q/+</sup> and wildtype mouse embryos.
- Supplementary Table 5. Differentially expressed genes in NPCs from *Chd1*<sup>R616Q/+</sup> mouse embryos, treated with DMSO vehicle control or dihydrotestosterone.
- Supplementary Table 6. A list of genes from gnomAD non-UKB dataset - all autosomal genes with pLI > 0.9 with test for male or female overrepresentation in rare missense variants
- Supplementary Table 7. A list of genes from UK Biobank WGS dataset - all autosomal genes with pLI > 0.9 with test for male or female overrepresentation in rare missense variants
- Supplementary Table 8. A list of genes that show sex-specific skewing of missense variants in gnomAD and UK Biobank.
- Supplementary Table 9. Guide RNA and oligo sequences used to create the mouse model, primer sequences from the current study.

## Data availability

- Raw sequencing data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE281601. Data will be made publicly available upon peer-reviewed publication of the manuscript.
- This study used data from the All of Us Research Program's Controlled Tier Dataset version 8, available to authorized users on the Researcher Workbench.

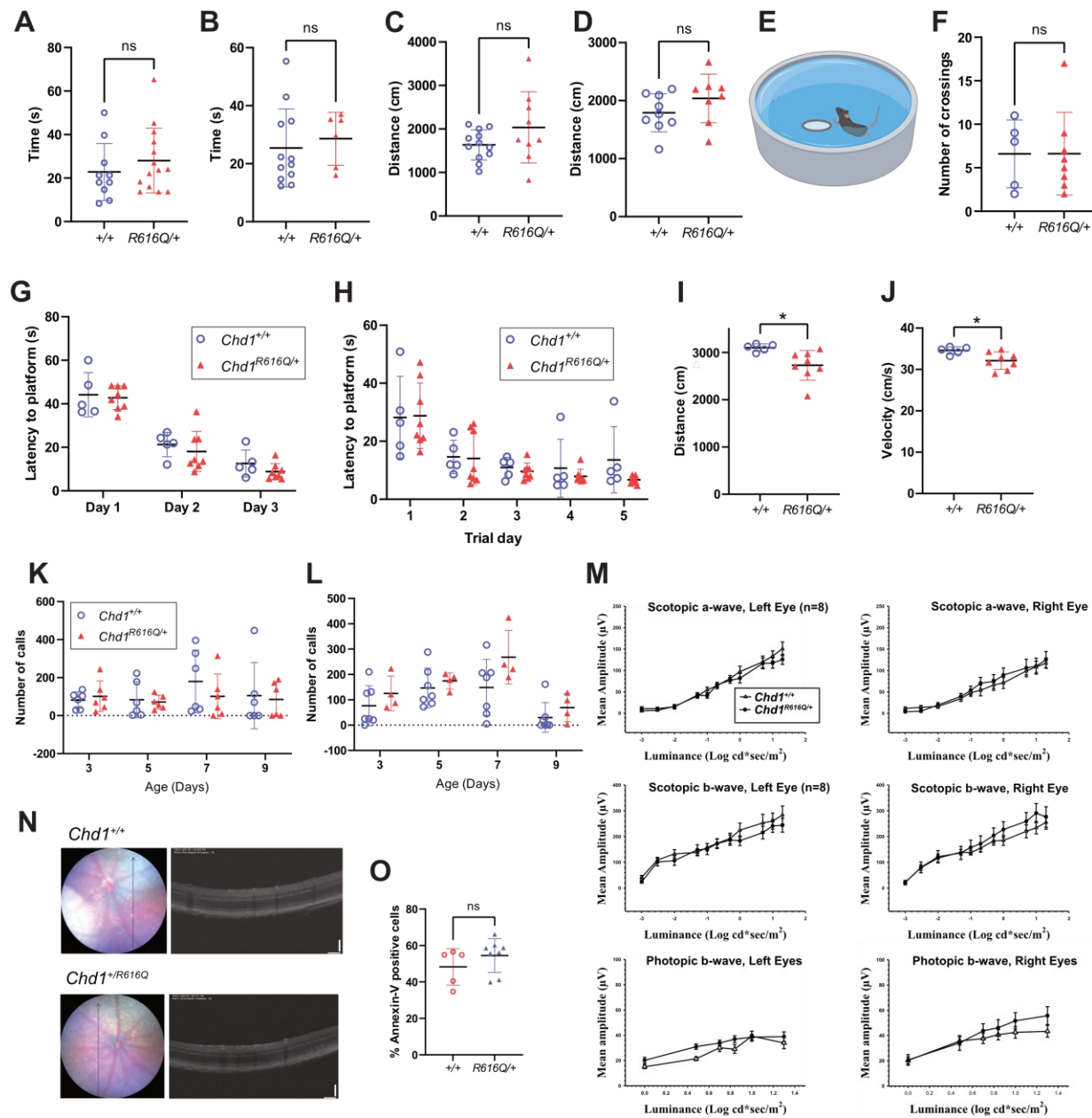


## Supplementary figures



Supplementary Figure 1: The effects of CHD1 variants

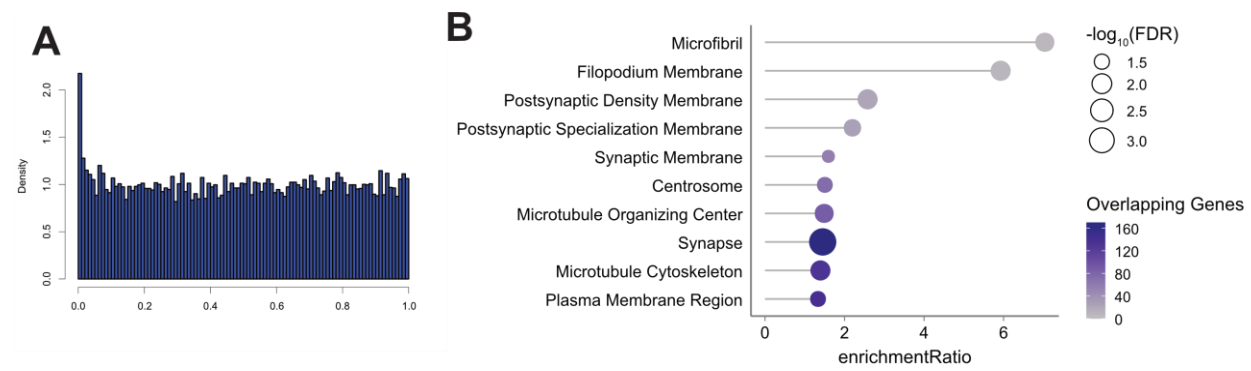
(A) Variant positions (white, boxed) were analyzed using Chd1 crystal and cryoEM structures from yeast (PDB code 7TN2, 3TED; with yeast amino acid positions shown in parentheses beneath corresponding human positions) and human (2B2W). In each case, neighboring residues that may be affected by the missense change are highlighted as well. Those considered to be deleterious are often part of, contacting, or near conserved helicase motifs. (B RT-qPCR results showing *Chd1* mRNA levels in tissue obtained from the adult cortex of *Chd1*<sup>R616Q/+</sup> mice and WT littermates. (C) A representative western blot showing CHD1 protein levels in mouse embryonic fibroblasts (MEFs), quantified in (D). ns: not significant.



**Supplementary Figure 2: Additional phenotyping of the  $Chd1^{R616Q/+}$  mice.**

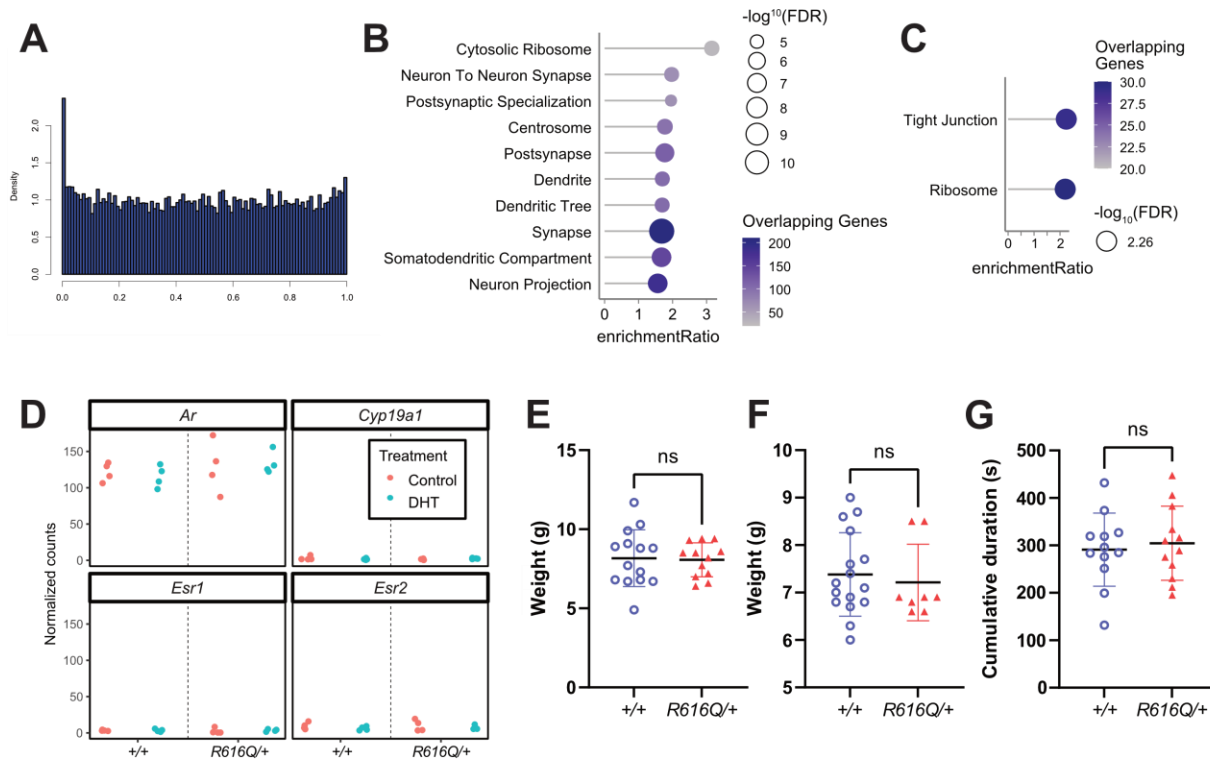
Time spent in suspension during the hindlimb suspension test in (A) females (WT:  $n = 10$ ,  $R616Q/+$ :  $n = 14$ ) and (B) males (WT:  $n = 12$ ,  $R616Q/+$ :  $n = 6$ ). (C) Distance travelled in an open field during the 10-minute period for (C) females (WT:  $n = 12$ ,  $R616Q/+$ :  $n = 9$ ) and (D) males (WT:  $n = 9$ ,  $R616Q/+$ :  $n = 8$ ). (E) A diagram illustrating the basis of MWM swimming task, females only (WT:  $n = 5$ ,  $R616Q/+$ :  $n = 8$ ). (F) Quantification of the number of platform crossings of female  $Chd1^{R616Q/+}$  mice and WT littermates during the hidden platform test on the final day of the MWM task, used as a measure of visuospatial learning and memory. (G) Time latency to visible platform. (H) Time latency to hidden platform. Quantification of

(I) distance travelled and (J) swimming velocity of female *Chd1*<sup>R616Q/+</sup> mice and WT littermates during the hidden platform test of the MWM. Quantification of the total number of pup calls over a 4-minute period upon separation from their dams at postnatal day 3, 5, 7 and 9 in (K) females (WT: *n* = 6, *R616Q/+*: *n* = 6) and (L) males (WT: *n* = 7, *R616Q/+*: *n* = 4). (M) Mean amplitudes of dark-adapted ERG a- and b-waves and light-adapted b-waves of *Chd1*<sup>R616Q/+</sup> mice and WT littermates. (N) Sample OCT scans from the eyes of a *Chd1*<sup>R616Q/+</sup> mouse and a WT littermate. Locations of the scans on the fundi are indicated by the black vertical arrows in the adjacent fundus images. (O) The percentage of annexin V positive NPCs in culture as measured by flow cytometry (WT: *n* = 5, *R616Q/+*: *n* = 8). \**p* < 0.05, ns = not significant.



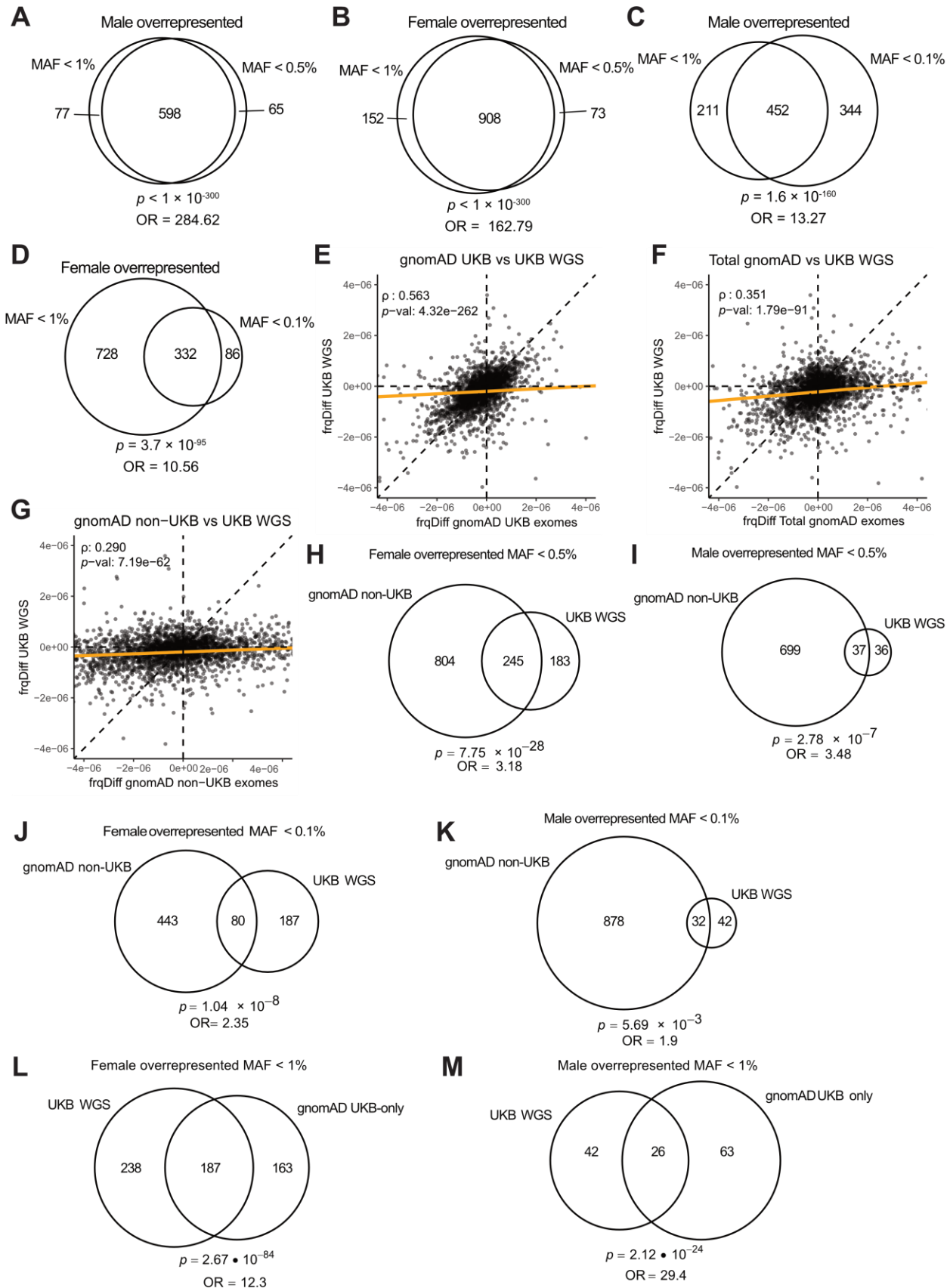
**Supplementary Figure 3: Quality measures and enriched pathways from RNA-Seq.**

(A) Histogram of  $p$ -values for the comparison  $Chd1^{+/+}$  vs  $Chd1^{R616Q/+}$  bulk RNAseq in NPCs ( $n = 4$  per genotype). (B) Cellular component gene ontology overrepresentation analysis using genes with the top 10% most significant  $p$ -values for the comparison of  $Chd1^{+/+}$  vs  $Chd1^{R616Q/+}$ .



**Supplementary Figure 4: Additional data regarding the effect of DHT on gene expression in the two genotypes.**

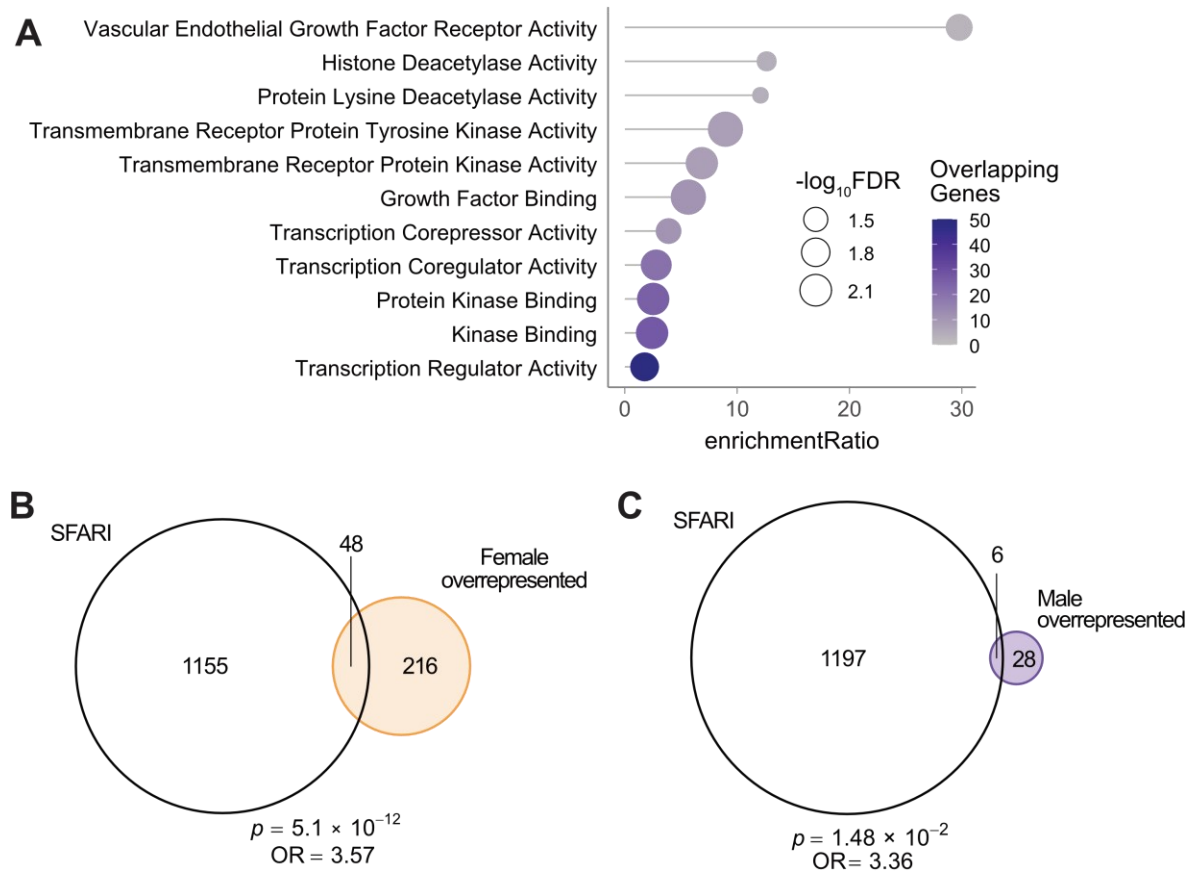
(A) A histogram of  $p$ -values for the comparison of DMSO control vs DHT treatment only in *Chd1*<sup>R616Q/+</sup> NPCs ( $n = 4$  per treatment). (B) Cellular component gene ontology overrepresentation analysis using genes with the top 10% most significant  $p$ -values for the comparison of DMSO control vs DHT treatment only in *Chd1*<sup>R616Q/+</sup> NPCs. (C) KEGG pathway gene ontology overrepresentation analysis for the comparison of DMSO control vs DHT treatment only in *Chd1*<sup>R616Q/+</sup> NPCs. (D) Normalized count data for the genes *Ar*, *Cyp19a1*, *Esr1* and *Esr2*. (E) Weights of male mice at P15 prior to orchiectomy. (WT:  $n = 14$ , *R616Q/+*:  $n = 12$ ). (F) Weights of female mice at P15 prior to testosterone implantation. (WT:  $n = 16$ , *R616Q/+*:  $n = 8$ ). (G) Cumulative duration of time spent in the center of the open field for orchiectomized male mice at 3 months of age ( $n = 12$  per genotype). ns = not significant.





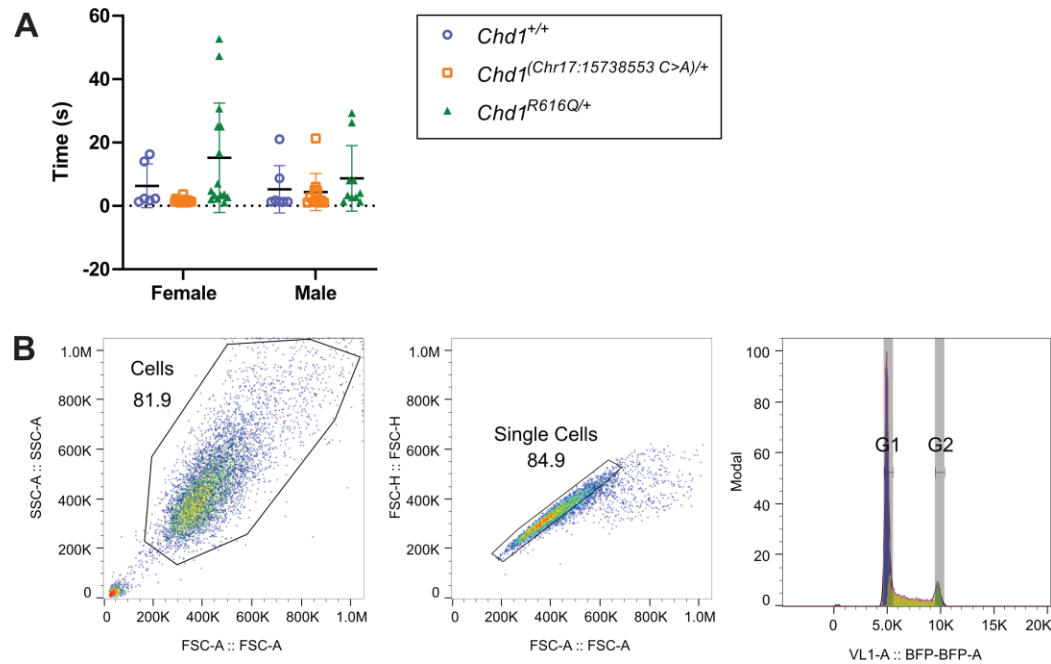
## Supplementary Figure 5: Quality control for sex differences in rare missense variants in gnomAD and UK Biobank

Overlapping genes with significant sex overrepresentation of missense variants in gnomAD exome data identified using MAF cutoffs of  $< 0.5\%$  and  $< 1\%$  in (A) males and (B) females. Overlapping genes with significant sex overrepresentation of missense variants identified using MAF cutoffs of  $< 1\%$  and  $< 0.1\%$  in (C) males and (D) females. Male – female aggregated allele frequency differences (frqDiff) for rare ( $MAF < 1\%$ ) missense variants in (E) gnomAD UK biobank exome sequencing samples compared to UK Biobank WGS samples, (F) total gnomAD exomes dataset compared to UK Biobank WGS samples, and (G) gnomAD non-UK Biobank exomes samples compared to UK Biobank WGS samples. Spearman's correlation ( $\rho$ ) and  $p$ -value shown for each plot. Each dot represents one gene. Overlapping genes with sex overrepresentation for missense variants with  $MAF < 0.5\%$  in gnomAD non-UKB and UKB WGS datasets in (H) females and (I) males. Overlapping genes with sex overrepresentation for missense variants with  $MAF < 0.1\%$  in gnomAD non-UKB and UKB WGS datasets in (J) females and (K) males. The overlap of sex-biased genes identified from variants with  $MAF < 1\%$  and  $p$ -value  $< 0.05$  in the gnomAD UKB-only dataset compared to the UKB WGS dataset for genes with (L) male overrepresentation and (M) female overrepresentation. All  $p$ -values and odds ratios calculated using Fisher's exact test, with a background of 3140 genes (the number of autosomal genes with  $pLI > 0.9$ ).



## Supplementary Figure 6: Functional analysis of genes that show a sex bias in our study

(A) Molecular function gene ontology overrepresentation analysis of genes with female overrepresentation in gnomAD and UKB. Venn diagrams depicting the overlap of SFARI genes with our identified lists of genes with (B) female or (C) male overrepresentation in gnomAD and UKB. The  $p$  values and odds ratios for the overlaps were calculated using Fisher's exact test using a background of 19980 genes (number of protein coding genes in GENCODE hg38).



**Supplementary Figure 7: *Chd1* synonymous mutation and flow cytometry gating for single cells using NPCs for cell cycle analysis.**

(A) Quantification of righting reflex times in female (WT:  $n = 6$ , *Chr17:15738553 C>A*/+:  $n = 14$ , *R616Q*/+:  $n = 15$ ) and male (WT:  $n = 7$ , *Chr17:15738553 C>A*/+:  $n = 11$ , *R616Q*/+:  $n = 10$ ) *Chd1*<sup>R616Q/+</sup>, *Chd1*<sup>(Chr17:15738553 C>A)/+</sup>, and WT littermates at postnatal day 6. The plot includes data depicted in Fig. 2C).

(B) Flow cytometry gating for single cells and cell cycle analysis using DAPI staining of NPCs.

## Supplementary Tables

**Supplementary Table 3: Altered Mendelian ratios of *Chd1*<sup>R616Q</sup> mice**

Genotype	Expected	Observed	Females	Males
+/+	39	39	21	18
R616Q/+	78	57	30	27
R616Q/R616Q	39	0	0	0

**Supplementary Table 9: Guide RNAs, oligos and primer sequences from the study.**

<b>gRNAs/Oligos/Primers</b>	
<b>Mouse model guides and oligos</b>	
<b>DMAA Synonymous mutation: g17: 15738553 C&gt;A</b>	
Guide	GATTAAAGTTTTGTACAGAA(GGG)
Oligo	CATTTTAAGGCATTCTTGGTGGTCTGAATTGGGCATTCATAGGTGTTGATGAAGCG CATCAATTAAGAATGATGATTCACTTCTGTACAAAACTTAATCGACTTTAAATCTAA CCATCGC
Mutant allele sequence	>DMAA AATATACTAACTGTTTTTTTATTCATTTTAAGGCATTCTTGGTGGTCTGAATTGGGCAT TCATAGGTGTTGATGAAGCGCATCGATTAAAGAATGATGATTCACTTCTGTACAAAAC TTAATCGACTTTAAATCTAACCATCGCCTTCTGATCACTGGAACCCCTCTACAGAA CTCCCTGAAGGAGCTCTGGTCACTGCTGCA
<b>DMON SNP: g17: 15738534 G&gt;A [R616Q] with synonymous mutation: g17: 15738553 C&gt;A</b>	
Guide	GATTAAAGTTTTGTACAGAA(GGG)
Oligo	CATTTTAAGGCATTCTTGGTGGTCTGAATTGGGCATTCATAGGTGTTGATGAAGCG CATCAATTAAGAATGATGATTCACTTCTGTACAAAACTTAATCGACTTTAAATCTAA CCATCGC
Mutant allele sequence	>DMON CTCTCTAAATATACTAACTGTTTTTTTATTCATTTTAAGGCATTCTTGGTGGTCTGAAT TGGGCATTCATAGGTGTTGATGAAGCGCATCAATTAAGAATGATGATTCACTTCTGT ACAAAAC
<b>Cloning Primers</b>	
<b>CHD1 mutagenesis primers</b>	
R618Q_mut_F	TGAAGCACACcaaTTAAAGAATG
R618Q_mut_R	TCAACACCTATAAATGCC
<b>pCoofy51 backbone primers</b>	
3C_fwd	GGGCCCCTGGAACAGAACTTCCAG
ccdB_rev	CGCCATTAACCTGATGTTCTGGGG
<b>CHD1 N-terminal deletion insert primers</b>	
3C_CHD1Ndel_fwd	AAGTTCTGTTCCAGGGGCCCATGAAACCATAGAAAGATTATGG
ccdB_CHD1_rev	CCCCAGAACATCAGGTTAATGGCGTAGAGCCTCCACCCCCTGTTT
<b>RT-qPCR primers</b>	
mChd1_Fwd_qPCR	AGGAAACGGCCGAAGAAACG
mChd1_Rev_qPCR	AGGGGGCCACCAATTTCTTA
ms-Actb_qPCR_F	GTGGGAATGGGTCAGAAGGA
ms-Actb_qPCR_R	CTTCTCCATGTCGTCCCAGT
<b>Genotyping PCR primers</b>	
Chd1_long_fwd	CAGCAGGCCAGTCAACTGAGGC
Chd1_long_rev	TCTCCCAACTGCATCTGAGCACA

## Acknowledgements

We would like to thank all the individuals with PILBOS diagnosis and their families that participated in this study. We would like to thank the European mouse project. We thank the Wellcome Trust Sanger Institute Mouse Genetics Project (Sanger MGP) and its funders for providing the mutant mouse line *Chd1*<sup>R616Q/+</sup> (15738534 G>A [R616Q] with synonymous mutation: 15738553 C>A). Funding information may be found at [www.sanger.ac.uk/mouseportal](http://www.sanger.ac.uk/mouseportal) and associated primary phenotypic information at [www.mousephenotype.org](http://www.mousephenotype.org). This research has been conducted using the UK Biobank Resource under Application Number 56270. We gratefully acknowledge All of Us participants for their contributions. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study. We thank the National Center for Medical Genomics, Czech Republic (LM2023067) for WES analyses.

## Funding

H.T.B. is funded by the Louma G. Foundation. This project was also specifically funded by a 3-year grant from Landspítali Science Fund (H.T.B.). The mouse model (*Chd1*<sup>R616Q/+</sup>) creation was funded by a grant from INFRAFRONTIER 2020. The behavioral core equipment used for behavioral testing was funded by a grant from the Icelandic Infrastructure Fund (191708-003). E.E.E. is funded by the Simons Foundation (RFA 810018EE). L.N., D.G. and S.K. were supported by grant NU23-07-00281 from the Ministry of Health of the Czech Republic. L.N. and S.K. were supported from the project MULTIOMICS\_CZ (Programme Johannes Amos Comenius, Ministry of Education, Youth and Sports of the Czech Republic/ID Project CZ.02.01.01/00/23\_020/0008540) – Co-funded by the European Union. This work was also supported, in part, by the National Natural Science Foundation of China (82201314 and 82471194) and the “Fundamental Research Funds for the Central Universities” starting fund (BMU2022RCZX038) to T.W. G.D.B. is funded by NIH grant R01-GM084192.

## Author contributions

H.T.B., G.P. and K.J.A. conceived the study. H.T.B. acquired funding for the study. K.J.A., S.T.H., and T.E. performed all the experimental work. E.T.T. and H.T.B. performed clinical collection and characterization of clinical variants. G.D.B. and I.M.N. did computational modelling of missense variants. K.J.A., S.B., A.H., and K.S. performed human population variant analysis. M.A., N.A., B.J.B., M.B., J.F.B., S.B., B.C., A.S.A.C., C.D.B.B., E.E., E.C.E., J.A.F., L.F., M.F., N.F., C.W.G., G.G., D.G., M.G., J.R.H., K.H., T.H., S.H., K.J., J.A.J., S.K., C.K., M.A.K., P.L., T.L., M.M., A.M., E.N., M.N., L.N., K.P., C.P., R.P., P.P., N.R., N.H.R., C.R., J.Roohi., J.Rosenfeld., M.S., C.S., Z.S., I.T., S.T., D.V., C.V., J.V., A.V., T.W., K.W., I.M.W. and M.G. provided clinical data.

## Competing interests

Dr. Bjornsson is Founder of KALDUR therapeutics and a consultant for Mahzi therapeutics. S.B. is a part of a working group called Alzheimer diagnostics that has received a grant from the Icelandic Technology Development Fund. I.M.W. is an employee of and may own stock in GeneDx. E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc. K.S. and A.H. are employees of deCODE genetics, a subsidiary of Amgen. The Department of Medical and Molecular Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics Laboratories.

## Declaration of generative AI and AI-assisted technologies in the writing process

1046 During the preparation of this work the authors used Chat GPT-4 in order to improve language and  
1047 readability, and for troubleshooting and refining R and python scripts. After using this tool/service, the  
1048 authors reviewed and edited the content as needed and take full responsibility for the content of the  
1049 publication.

## References

- 1 Pilarowski, G. O. *et al.* Missense variants in the chromatin remodeler CHD1 are associated with neurodevelopmental disability. *J Med Genet* **55**, 561-566 (2018).  
<https://doi.org/10.1136/jmedgenet-2017-104759>
- 2 Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry* **56**, 466-474 (2017).  
<https://doi.org/https://doi.org/10.1016/j.jaac.2017.03.013>
- 3 Bölte, S. *et al.* Sex and gender in neurodevelopmental conditions. *Nat Rev Neurol* **19**, 136-159 (2023). <https://doi.org/10.1038/s41582-023-00774-6>
- 4 Sunwoo, Y., Seo, S. H., Kim, H.-J., Park, M. S. & Cho, A. Variant of CHD1 gene resulting in a Korean case of Pilarowski-Bjornsson syndrome. *J Genet Med* **19**, 111-114 (2022).  
<https://doi.org/10.5734/JGM.2022.19.2.111>
- 5 Al-Aamri, M., Alshaqaaq, M. & Al-Abdi, S. Y. A Saudi Girl With Co-occurring CHD1 (Pilarowski-Bjornsson Syndrome) and ASH1L Gene Variants. *Cureus* **15**, e49905 (2023).  
<https://doi.org/10.7759/cureus.49905>
- 6 Stokes, D. G., Tartof, K. D. & Perry, R. P. CHD1 is concentrated in interbands and puffed regions of *Drosophila* polytene chromosomes. *Proceedings of the National Academy of Sciences* **93**, 7137-7142 (1996). <https://doi.org/doi:10.1073/pnas.93.14.7137>
- 7 Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. & Collins, F. S. Characterization of the CHD family of proteins. *Proceedings of the National Academy of Sciences* **94**, 11472-11477 (1997).  
<https://doi.org/doi:10.1073/pnas.94.21.11472>
- 8 Tran, H. G., Steger, D. J., Iyer, V. R. & Johnson, A. D. The chromo domain protein Chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *The EMBO Journal* **19**, 2323-2331 (2000). <https://doi.org/https://doi.org/10.1093/emboj/19.10.2323>
- 9 Lusser, A., Urwin, D. L. & Kadonaga, J. T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nature Structural & Molecular Biology* **12**, 160-166 (2005).  
<https://doi.org/10.1038/nsmb884>
- 10 Sims, R. J. *et al.* Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell* **28**, 665-676 (2007).  
<https://doi.org/https://doi.org/10.1016/j.molcel.2007.11.010>
- 11 Guzman-Ayala, M. *et al.* Chd1 is essential for the high transcriptional output and rapid growth of the mouse epiblast. *Development* **142**, 118 (2015). <https://doi.org/10.1242/dev.114843>
- 12 Simic, R. *et al.* Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *The EMBO Journal* **22**, 1846-1856 (2003).  
<https://doi.org/https://doi.org/10.1093/emboj/cdg179>
- 13 Skene, P. J., Hernandez, A. E., Groudine, M. & Henikoff, S. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife* **3**, e02042 (2014). <https://doi.org/10.7554/eLife.02042>
- 14 Biswas, D. *et al.* A Role for Chd1 and Set2 in Negatively Regulating DNA Replication in *Saccharomyces cerevisiae*. *Genetics* **178**, 649-659 (2008).  
<https://doi.org/10.1534/genetics.107.084202>
- 15 Yadav, T. & Whitehouse, I. Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. *Cell Reports* **15**, 715-723 (2016).  
<https://doi.org/https://doi.org/10.1016/j.celrep.2016.03.059>



1095 16 Delamarre, A. *et al.* MRX Increases Chromatin Accessibility at Stalled Replication Forks to  
1096 Promote Nascent DNA Resection and Cohesin Loading. *Molecular Cell* **77**, 395-410.e393 (2020).  
1097 <https://doi.org/https://doi.org/10.1016/j.molcel.2019.10.029>  
1098 17 Kari, V. *et al.* Loss of CHD1 causes DNA repair defects and enhances prostate cancer therapeutic  
1099 responsiveness. *EMBO reports* **17**, 1609-1623 (2016).  
1100 <https://doi.org/https://doi.org/10.15252/embr.201642352>  
1101 18 R  themann, P., Balbo Pogliano, C., Codilupi, T., Garajov  , Z. & Naegeli, H. Chromatin remodeler  
1102 CHD1 promotes XPC to TFIIH handover of nucleosomal UV lesions in nucleotide excision repair.  
1103 *The EMBO Journal* **36**, 3372-3386-3386 (2017).  
1104 <https://doi.org/https://doi.org/10.15252/embj.201695742>  
1105 19 Zhou, J. *et al.* Human CHD1 is required for early DNA-damage signaling and is uniquely regulated  
1106 by its N terminus. *Nucleic Acids Research* **46**, 3891-3905 (2018).  
1107 <https://doi.org/10.1093/nar/gky128>  
1108 20 Shenoy, T. *et al.* CHD1 loss sensitizes prostate cancer to DNA damaging therapy by promoting  
1109 error-prone double-strand break repair. *Annals of Oncology* **28**, 1495-1507 (2017).  
1110 21 Bulut-Karslioglu, A. *et al.* Chd1 protects genome integrity at promoters to sustain  
1111 hypertranscription in embryonic stem cells. *Nature Communications* **12**, 4859 (2021).  
1112 <https://doi.org/10.1038/s41467-021-25088-3>  
1113 22 Rechlin, R. K., Splinter, T. F. L., Hodges, T. E., Albert, A. Y. & Galea, L. A. M. An analysis of  
1114 neuroscience and psychiatry papers published from 2009 and 2019 outlines opportunities for  
1115 increasing discovery of sex differences. *Nature Communications* **13**, 2137 (2022).  
1116 <https://doi.org/10.1038/s41467-022-29903-3>  
1117 23 McEwen, B. S. & Milner, T. A. Understanding the broad influence of sex hormones and sex  
1118 differences in the brain. *Journal of Neuroscience Research* **95**, 24-39 (2017).  
1119 <https://doi.org/https://doi.org/10.1002/jnr.23809>  
1120 24 Augello, M. A. *et al.* CHD1 Loss Alters AR Binding at Lineage-Specific Enhancers and Modulates  
1121 Distinct Transcriptional Programs to Drive Prostate Tumorigenesis. *Cancer Cell* **35**, 603-617.e608  
1122 (2019). <https://doi.org/https://doi.org/10.1016/j.ccell.2019.03.001>  
1123 25 Metzger, E. *et al.* Assembly of methylated KDM1A and CHD1 drives androgen receptor–  
1124 dependent transcription and translocation. *Nature Structural & Molecular Biology* **23**, 132-139  
1125 (2016). <https://doi.org/10.1038/nsmb.3153>  
1126 26 Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for  
1127 connecting investigators with an interest in the same gene. *Hum Mutat* **36**, 928-930 (2015).  
1128 <https://doi.org/10.1002/humu.22844>  
1129 27 Stessman, H. A. F. *et al.* Targeted sequencing identifies 91 neurodevelopmental-disorder risk  
1130 genes with autism and developmental-disability biases. *Nature Genetics* **49**, 515-526 (2017).  
1131 <https://doi.org/10.1038/ng.3792>  
1132 28 Zhou, X. *et al.* Integrating de novo and inherited variants in 42,607 autism cases identifies  
1133 mutations in new moderate-risk genes. *Nature Genetics* **54**, 1305-1319 (2022).  
1134 <https://doi.org/10.1038/s41588-022-01148-2>  
1135 29 Li, S. *et al.* Identification of novel Mendelian disorders of the epigenetic machinery (MDEMs)-  
1136 associated functional mutations and neurodevelopmental disorders. *QJM: An International*  
1137 *Journal of Medicine* **116**, 355-364 (2023). <https://doi.org/10.1093/qjmed/hcad005>  
1138 30 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456  
1139 humans. *Nature* **581**, 434-443 (2020). <https://doi.org/10.1038/s41586-020-2308-7>  
1140 31 Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense.  
1141 *Science* **381**, eadg7492 (2023). <https://doi.org/10.1126/science.adg7492>

1142 32 Nodelman, I. M. *et al.* Nucleosome recognition and DNA distortion by the Chd1 remodeler in a  
1143 nucleotide-free state. *Nature structural & molecular biology* **29**, 121-129 (2022).  
1144 33 Piatti, P. *et al.* Embryonic stem cell differentiation requires full length Chd1. *Scientific Reports* **5**,  
1145 8007 (2015). <https://doi.org/10.1038/srep08007>  
1146 34 Seibenhener, M. L. & Wooten, M. C. Use of the Open Field Maze to measure locomotor and  
1147 anxiety-like behavior in mice. *J Vis Exp*, e52434 (2015). <https://doi.org/10.3791/52434>  
1148 35 Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related  
1149 forms of learning and memory. *Nature Protocols* **1**, 848-858 (2006).  
1150 <https://doi.org/10.1038/nprot.2006.116>  
1151 36 Caruso, A., Ricceri, L. & Scattoni, M. L. Ultrasonic vocalizations as a fundamental tool for early  
1152 and adult behavioral phenotyping of Autism Spectrum Disorder rodent models. *Neuroscience &*  
1153 *Biobehavioral Reviews* **116**, 31-43 (2020).  
1154 <https://doi.org/https://doi.org/10.1016/j.neubiorev.2020.06.011>  
1155 37 Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout  
1156 phenotyping underpinning the study of human disease. *Nucleic Acids Research* **51**, D1038-D1045  
1157 (2023). <https://doi.org/10.1093/nar/gkac972>  
1158 38 Salomoni, P. & Calegari, F. Cell cycle control of mammalian neural stem cells: putting a speed  
1159 limit on G1. *Trends in Cell Biology* **20**, 233-243 (2010). <https://doi.org/10.1016/j.tcb.2010.01.006>  
1160 39 Arletti, R., Benelli, A. & Bertolini, A. Influence of oxytocin on feeding behavior in the rat. *Peptides*  
1161 **10**, 89-93 (1989). [https://doi.org/https://doi.org/10.1016/0196-9781\(89\)90082-X](https://doi.org/https://doi.org/10.1016/0196-9781(89)90082-X)  
1162 40 Noble, E. E., Billington, C. J., Kotz, C. M. & Wang, C. Oxytocin in the ventromedial hypothalamic  
1163 nucleus reduces feeding and acutely increases energy expenditure. *American Journal of*  
1164 *Physiology-Regulatory, Integrative and Comparative Physiology* **307**, R737-R745 (2014).  
1165 <https://doi.org/10.1152/ajpregu.00118.2014>  
1166 41 Li, E. *et al.* Control of lipolysis by a population of oxytocinergic sympathetic neurons. *Nature* **625**,  
1167 175-180 (2024). <https://doi.org/10.1038/s41586-023-06830-x>  
1168 42 Peters, S., Slattery, D. A., Uschold-Schmidt, N., Reber, S. O. & Neumann, I. D. Dose-dependent  
1169 effects of chronic central infusion of oxytocin on anxiety, oxytocin receptor binding and stress-  
1170 related parameters in mice. *Psychoneuroendocrinology* **42**, 225-236 (2014).  
1171 <https://doi.org/https://doi.org/10.1016/j.psyneuen.2014.01.021>  
1172 43 Grillon, C. *et al.* Oxytocin increases anxiety to unpredictable threat. *Mol Psychiatry* **18**, 958-960  
1173 (2013). <https://doi.org/10.1038/mp.2012.156>  
1174 44 Winter, J. *et al.* Chronic oxytocin-driven alternative splicing of Crfr2 $\alpha$  induces anxiety. *Molecular*  
1175 *Psychiatry* **28**, 4742-4755 (2023). <https://doi.org/10.1038/s41380-021-01141-x>  
1176 45 Wei, P.-C. *et al.* Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural  
1177 Stem/Progenitor Cells. *Cell* **164**, 644-655 (2016).  
1178 <https://doi.org/https://doi.org/10.1016/j.cell.2015.12.039>  
1179 46 Rafnsdottir, S. *et al.* SMYD5 is a regulator of the mild hypothermia response. *Cell Reports* **43**  
1180 (2024). <https://doi.org/10.1016/j.celrep.2024.114554>  
1181 47 Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer.  
1182 *Nucleic Acids Research* **52**, D1210-D1217 (2024). <https://doi.org/10.1093/nar/gkad986>  
1183 48 Liang, T. & Liao, S. A very rapid effect of androgen on initiation of protein synthesis in prostate.  
1184 *Proceedings of the National Academy of Sciences* **72**, 706-709 (1975).  
1185 <https://doi.org/10.1073/pnas.72.2.706>  
1186 49 Ferrando, A. A. *et al.* Testosterone injection stimulates net protein synthesis but not tissue amino  
1187 acid transport. *Am J Physiol* **275**, E864-871 (1998).  
1188 <https://doi.org/10.1152/ajpendo.1998.275.5.E864>

1189 50 Rizk, J., Sahu, R. & Duteil, D. An overview on androgen-mediated actions in skeletal muscle and  
1190 adipose tissue. *Steroids* **199**, 109306 (2023).  
1191 <https://doi.org/https://doi.org/10.1016/j.steroids.2023.109306>  
1192 51 Handelsman, D. J. Androgen physiology, pharmacology, use and misuse. *Endotext [Internet]*  
1193 (2020).  
1194 52 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,  
1195 203-209 (2018). <https://doi.org/10.1038/s41586-018-0579-z>  
1196 53 Li, S., Carss, K. J., Halldorsson, B. V., Cortes, A. & Consortium, U. B. W.-G. S. Whole-genome  
1197 sequencing of half-a-million UK Biobank participants. *medRxiv*, 2023.2012.2006.23299426  
1198 (2023). <https://doi.org/10.1101/2023.12.06.23299426>  
1199 54 Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340-346 (2024).  
1200 <https://doi.org/10.1038/s41586-023-06957-x>  
1201 55 Granados, A. *et al.* MAP3K1-related gonadal dysgenesis: Six new cases and review of the  
1202 literature. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* **175**, 253-  
1203 259 (2017). <https://doi.org/https://doi.org/10.1002/ajmg.c.31559>  
1204 56 Ostrer, H. Pathogenic Variants in MAP3K1 Cause 46,XY Gonadal Dysgenesis: A Review. *Sexual*  
1205 *Development* **16**, 92-97 (2022). <https://doi.org/10.1159/000522428>  
1206 57 Kimura, E. *et al.* MAP3K1 regulates female reproductive tract development. *Disease Models &*  
1207 *Mechanisms* **17** (2024). <https://doi.org/10.1242/dmm.050669>  
1208 58 Schoeler, T. *et al.* Participation bias in the UK Biobank distorts genetic associations and  
1209 downstream analyses. *Nature Human Behaviour* **7**, 1216-1227 (2023).  
1210 <https://doi.org/10.1038/s41562-023-01579-9>  
1211 59 Benonisdottir, S. & Kong, A. Studying the genetics of participation using footprints left on the  
1212 ascertained genotypes. *Nature Genetics* **55**, 1413-1420 (2023). [https://doi.org/10.1038/s41588-](https://doi.org/10.1038/s41588-023-01439-2)  
1213 [023-01439-2](https://doi.org/10.1038/s41588-023-01439-2)  
1214 60 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,  
1215 203-209 (2018). <https://doi.org/10.1038/s41586-018-0579-z>  
1216 61 Rodríguez-Montes, L. *et al.* Sex-biased gene expression across mammalian organ development  
1217 and evolution. *Science* **382**, eadf1046 (2023). <https://doi.org/10.1126/science.adf1046>  
1218 62 Boukas, L. *et al.* Coexpression patterns define epigenetic regulators associated with neurological  
1219 dysfunction. *Genome Research* **29**, 532-542 (2019). <https://doi.org/10.1101/gr.239442.118>  
1220 63 Ascano, M., Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein  
1221 expression. *Nature* **492**, 382-386 (2012). <https://doi.org/10.1038/nature11737>  
1222 64 McAninch, D. S. *et al.* Fragile X mental retardation protein recognizes a G quadruplex structure  
1223 within the survival motor neuron domain containing 1 mRNA 5'-UTR. *Molecular BioSystems* **13**,  
1224 1448-1457 (2017). <https://doi.org/10.1039/C7MB00070G>  
1225 65 Banerjee-Basu, S. & Packer, A. SFARI Gene: an evolving database for the autism research  
1226 community. *Disease Models & Mechanisms* **3**, 133-135 (2010).  
1227 <https://doi.org/10.1242/dmm.005439>  
1228 66 Gonçalves, S. *et al.* A homozygous KAT2B variant modulates the clinical phenotype of ADD3  
1229 deficiency in humans and flies. *PLOS Genetics* **14**, e1007386 (2018).  
1230 <https://doi.org/10.1371/journal.pgen.1007386>  
1231 67 Stephenson, S. E. M. *et al.* Germline variants in tumor suppressor FBXW7 lead to impaired  
1232 ubiquitination and a neurodevelopmental syndrome. *The American Journal of Human Genetics*  
1233 **109**, 601-617 (2022). <https://doi.org/https://doi.org/10.1016/j.ajhg.2022.03.002>  
1234 68 Riestra, M. R. *et al.* Human Autosomal Recessive DNA Polymerase Delta 3 Deficiency Presenting  
1235 as Omenn Syndrome. *Journal of Clinical Immunology* **44**, 2 (2023).  
1236 <https://doi.org/10.1007/s10875-023-01627-z>

1237 69 Xu, C. *et al.* Polymorphisms in seizure 6-like gene are associated with bipolar disorder I: Evidence  
1238 of genexgender interaction. *Journal of Affective Disorders* **145**, 95-99 (2013).  
1239 <https://doi.org/https://doi.org/10.1016/j.jad.2012.07.017>

1240 70 Schoberleitner, I. *et al.* Role for Chromatin Remodeling Factor Chd1 in Learning and Memory.  
1241 *Frontiers in Molecular Neuroscience* **12** (2019). <https://doi.org/10.3389/fnmol.2019.00003>

1242 71 Sofer, Y. *et al.* Sexually dimorphic oxytocin circuits drive intragroup social conflict and aggression  
1243 in wild house mice. *Nature Neuroscience* **27**, 1565-1573 (2024). [https://doi.org/10.1038/s41593-](https://doi.org/10.1038/s41593-024-01685-5)  
1244 [024-01685-5](https://doi.org/10.1038/s41593-024-01685-5)

1245 72 Uhl-Bronner, S., Waltisperger, E., Martínez-Lorenzana, G., Condes Lara, M. & Freund-Mercier, M.  
1246 J. Sexually dimorphic expression of oxytocin binding sites in forebrain and spinal cord of the rat.  
1247 *Neuroscience* **135**, 147-154 (2005).  
1248 <https://doi.org/https://doi.org/10.1016/j.neuroscience.2005.05.025>

1249 73 Sarn, N., Thacker, S., Lee, H. & Eng, C. Germline nuclear-predominant Pten murine model  
1250 exhibits impaired social and perseverative behavior, microglial activation, and increased  
1251 oxytocinergic activity. *Molecular Autism* **12**, 41 (2021). [https://doi.org/10.1186/s13229-021-](https://doi.org/10.1186/s13229-021-00448-4)  
1252 [00448-4](https://doi.org/10.1186/s13229-021-00448-4)

1253 74 Spritzer, M. D. & Galea, L. A. Testosterone and dihydrotestosterone, but not estradiol, enhance  
1254 survival of new hippocampal neurons in adult male rats. *Developmental neurobiology* **67**, 1321-  
1255 1333 (2007).

1256 75 Hamson, D. K. *et al.* Androgens increase survival of adult-born neurons in the dentate gyrus by  
1257 an androgen receptor-dependent mechanism in male rats. *Endocrinology* **154**, 3294-3304  
1258 (2013).

1259 76 La Rosa, P. *et al.* Androgen Receptor signaling promotes the neural progenitor cell pool in the  
1260 developing cortex. *Journal of Neurochemistry* **157**, 1153-1166 (2021).  
1261 <https://doi.org/https://doi.org/10.1111/jnc.15192>

1262 77 Kelava, I., Chiaradia, I., Pellegrini, L., Kalinka, A. T. & Lancaster, M. A. Androgens increase  
1263 excitatory neurogenic potential in human brain organoids. *Nature* **602**, 112-116 (2022).  
1264 <https://doi.org/10.1038/s41586-021-04330-4>

1265 78 Nuñez, J. L., Jurgens, H. A. & Juraska, J. M. Androgens reduce cell death in the developing rat  
1266 visual cortex. *Developmental Brain Research* **125**, 83-88 (2000).  
1267 [https://doi.org/https://doi.org/10.1016/S0165-3806\(00\)00126-7](https://doi.org/https://doi.org/10.1016/S0165-3806(00)00126-7)

1268 79 Tetzlaff, J. E., Huppenbauer, C. B., Tanzer, L., Alexander, T. D. & Jones, K. J. Motoneuron injury and  
1269 repair. *Journal of Molecular Neuroscience* **28**, 53-64 (2006).  
1270 <https://doi.org/10.1385/JMN:28:1:53>

1271 80 Pihlajamaa, P. *et al.* Tissue-specific pioneer factors associate with androgen receptor cistromes  
1272 and transcription programs. *The EMBO Journal* **33**, 312-326-326 (2014).  
1273 <https://doi.org/https://doi.org/10.1002/emboj.201385895>

1274 81 Haffner, M. C. *et al.* Androgen-induced TOP2B-mediated double-strand breaks and prostate  
1275 cancer gene rearrangements. *Nature Genetics* **42**, 668-675 (2010).  
1276 <https://doi.org/10.1038/ng.613>

1277 82 Feng, W. *et al.* Chd7 is indispensable for mammalian brain development through activation of a  
1278 neuronal differentiation programme. *Nature Communications* **8**, 14758 (2017).  
1279 <https://doi.org/10.1038/ncomms14758>

1280 83 Gonçalves, C. I. *et al.* High frequency of CHD7 mutations in congenital hypogonadotropic  
1281 hypogonadism. *Sci Rep* **9**, 1597 (2019). <https://doi.org/10.1038/s41598-018-38178-y>

1282 84 Zhou, Q., Sheng, W., Yang, S. & Zou, C. The Clinical and Genetic Characteristics in Children with  
1283 Idiopathic Hypogonadotropic Hypogonadism. *Journal of Oncology* **2022**, 7973726 (2022).  
1284 <https://doi.org/https://doi.org/10.1155/2022/7973726>



1285 85 Tartaglia, M. *et al.* Paternal Germline Origin and Sex-Ratio Distortion in Transmission of  
1286 <em>PTPN11</em> Mutations in Noonan Syndrome. *The American Journal of Human Genetics*  
1287 **75**, 492-497 (2004). <https://doi.org/10.1086/423493>  
1288 86 Leroy, C. *et al.* The 2q37-deletion syndrome: an update of the clinical spectrum including  
1289 overweight, brachydactyly and behavioural features in 14 new patients. *Eur J Hum Genet* **21**,  
1290 602-612 (2013). <https://doi.org/10.1038/ejhg.2012.230>  
1291 87 Ong-Pålsson, E. *et al.* The  $\beta$ -Secretase Substrate Seizure 6-Like Protein (SEZ6L) Controls Motor  
1292 Functions in Mice. *Molecular Neurobiology* **59**, 1183-1198 (2022).  
1293 <https://doi.org/10.1007/s12035-021-02660-y>  
1294 88 O'Tuathaigh, C. M. *et al.* Sexually dimorphic changes in the exploratory and habituation profiles  
1295 of heterozygous neuregulin-1 knockout mice. *NeuroReport* **17** (2006).  
1296 89 Taylor, S. B., Markham, J. A., Taylor, A. R., Kanaskie, B. Z. & Koenig, J. I. Sex-specific  
1297 neuroendocrine and behavioral phenotypes in hypomorphic Type II Neuregulin 1 rats.  
1298 *Behavioural Brain Research* **224**, 223-232 (2011).  
1299 <https://doi.org/https://doi.org/10.1016/j.bbr.2011.05.008>  
1300 90 Zhang, Z. *et al.* Sex differences in circulating neuregulin1- $\beta$ 1 and  $\beta$ -secretase 1 expression in  
1301 childhood-onset schizophrenia. *Comprehensive Psychiatry* **100**, 152176 (2020).  
1302 <https://doi.org/https://doi.org/10.1016/j.comppsy.2020.152176>  
1303 91 Ananloo, E. S., Yoosefee, S. & Karimipour, M. Neuregulin1 gene variants as a biomarker for  
1304 cognitive impairments in patients with schizophrenia. *European Journal of Psychiatry* **34**, 11-19  
1305 (2020). <https://doi.org/10.1016/j.ejpsy.2019.12.004>  
1306 92 McNairn, A. J., Chuang, C. H., Bloom, J. C., Wallace, M. D. & Schimenti, J. C. Female-biased  
1307 embryonic death from inflammation induced by genomic instability. *Nature* **567**, 105-108 (2019).  
1308 <https://doi.org/10.1038/s41586-019-0936-6>  
1309 93 Rodrigues, E. D. S. *et al.* Variant-level matching for diagnosis and discovery: Challenges and  
1310 opportunities. *Hum Mutat* **43**, 782-790 (2022). <https://doi.org/10.1002/humu.24359>  
1311 94 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-  
1312 95 (2007). <https://doi.org/10.1109/MCSE.2007.55>  
1313 95 Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic*  
1314 *Acids Res* **29**, e45 (2001). <https://doi.org/10.1093/nar/29.9.e45>  
1315 96 Durkin, M. E., Qian, X., Popescu, N. C. & Lowy, D. R. Isolation of Mouse Embryo Fibroblasts. *Bio*  
1316 *Protoc* **3** (2013). <https://doi.org/10.21769/bioprotoc.908>  
1317 97 Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide  
1318 gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National*  
1319 *Academy of Sciences* **76**, 4350-4354 (1979). <https://doi.org/doi:10.1073/pnas.76.9.4350>  
1320 98 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**,  
1321 676-682 (2012). <https://doi.org/10.1038/nmeth.2019>  
1322 99 Feather-Schussler, D. N. & Ferguson, T. S. A Battery of Motor Tests in a Neonatal Mouse Model of  
1323 Cerebral Palsy. *J Vis Exp* (2016). <https://doi.org/10.3791/53569>  
1324 100 Björnsson, H. T. *et al.* Histone deacetylase inhibition rescues structural and functional brain  
1325 deficits in a mouse model of Kabuki syndrome. *Sci Transl Med* **6**, 256ra135 (2014).  
1326 <https://doi.org/10.1126/scitranslmed.3009278>  
1327 101 García-Llorca, A., Ólafsson, K. H., Sigurdsson, A. T. & Eysteinnsson, T. Progressive Cone-Rod  
1328 Dystrophy and RPE Dysfunction in Mitfmi/+ Mice. *Genes* **14**, 1458 (2023).  
1329 102 Elizarraras, J. M. *et al.* WebGestalt 2024: faster gene set analysis and new support for  
1330 metabolomics and multi-omics. *Nucleic Acids Research* **52**, W415-W421 (2024).  
1331 <https://doi.org/10.1093/nar/gkae456>

1332 103 BD, V. d. A. G. O. C. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st edition)*.  
1333 (O'Reilly Media, 2020).  
1334 104 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).  
1335 <https://doi.org/10.1093/bioinformatics/bts635>  
1336 105 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).  
1337 <https://doi.org/10.1093/gigascience/giab008>  
1338 106 Manders, F. *et al.* MutationalPatterns: the one stop shop for the analysis of mutational  
1339 processes. *BMC Genomics* **23**, 134 (2022). <https://doi.org/10.1186/s12864-022-08357-3>  
1340