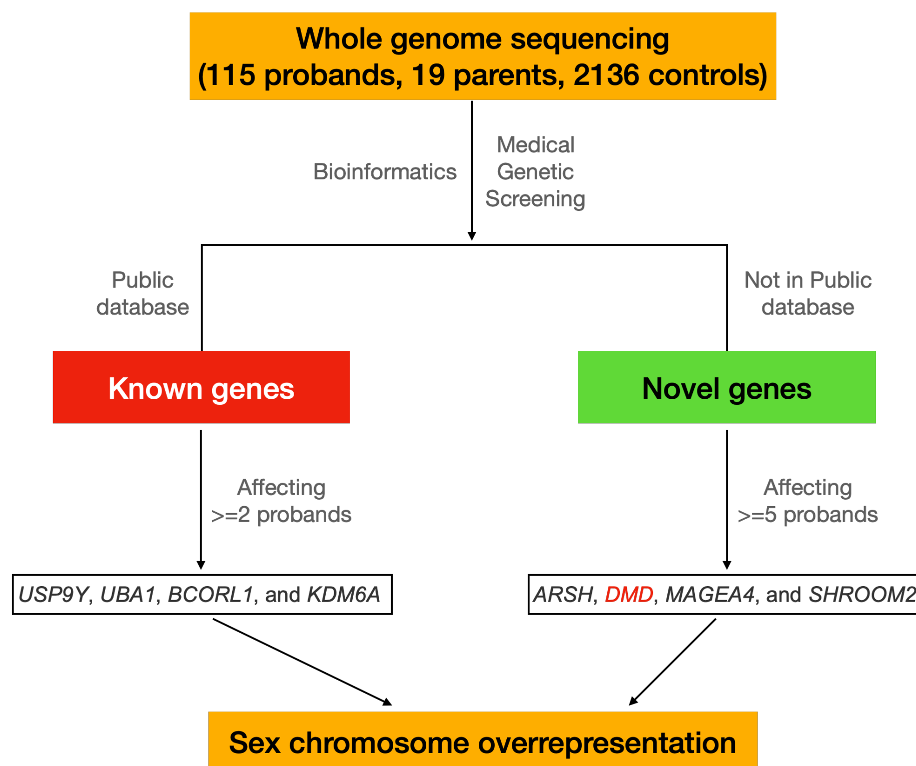
Supplementary note

---Bioinformatic pipeline and mice movement tests

1. Pathogenic variants identification in known and novel genes

The general framework analyzing data flows this.



Scheme figure 1. The analyzing processes and major results for boys' cryptorchidism based on whole genome sequencing data.

In detail, to determine rare pathogenic variants for both known and novel genes, we used a pipeline comprised of the co-segregation analysis, variant impact prediction, and common variants filtering by incorporating multiple software and databases, including SnpEff v5.1 [1], gnomAD v3.1.2 [2], 1000 genomes [3], PLINK v1.9 [4], VEP [5], SIFT [6],

PolyPhen [7], LoFtools [8], BLOSUM [9], and dbNSFP [10]. Only the variants with genetic consequence categories “HIGH” and “MODERATE” based on the VEP were included. The "HIGH" impact variants indicate disruptive effects on genes or chromosomes, such as the chromosome number variation, exon loss variant, frameshift variant, rare amino acid variant, splice acceptor variant, splice donor variant, start lost, stop gained, stop lost, and transcript ablation. The "MODERATE" impacts, such as inframe insertion, disruptive inframe insertion, inframe deletion, disruptive inframe deletion, missense variant, splice region variant, 3 prime UTR truncation, and 5 prime UTR truncation, may result in a non-disruptive but still primary protein sequence-altering effects.

Specifically, variant filtering and prioritization were performed with the following criteria for all variants per sample obtained from HaplotypeCaller in the GVCF mode: (a) The homozygous and heterozygous sites were adjusted with a threshold of reads proportions (RP). To increase the reliability of genotyping, only the alternative variants with RP scores over 85% were assigned as homozygous sites. The sites with RP scores between 15% and 85% were assigned as heterozygous. (b) To exclude sites with a high mutation rate, only biallelic sites were kept. (c) Only the variants with ten or more supporting reads were allowed. (d) For variants with "MODERATE" impacts, only those predicted to be deleterious or damaging by at least one algorithm implemented in the dbNSFP database were included. (e) Since the undescended testes naturally exist in certain species of Boreoeutheria and Afrotheria, for missense variants, we kept only the sites conserved at least in primate species. (f) Considering the phenotypic complexity of cryptorchidism, genes without expression in any reproductive system, at levels of tissue RNA, single-cell RNA, and protein, were screened using the HPA

database [11], and removed. (g) Variants with allele frequencies higher than 0.01 registered in any population database were removed. (h) The candidate variants' quality was finally manually visualized with IGV to inspect the mapping quality, and variants within highly variable regions were removed. (i) The co-segregation between genotype and affected status was based on autosomal recessive, compound heterozygous, and X-linked recessive inheritance mode. (j) All variants in the final list were confirmed using the Sanger sequencing. (k) The significance of identified novel genes was further confirmed using multiple algorithms of burden tests in the RVTESTS package by incorporating variants from the “1000 genomes” database [12].

Following the annotation of candidate variants, the known and novel candidate genes can be recognized. The known genes are within a local “gene pool” related to cryptorchidism reported in PubMed literature (Supplementary Table 2) and the HPO database [13], which integrates three databases: Orphanet [14], DECIPHER [15], and OMIM [16]. The cases with candidate rare pathogenic variants in known genes were filtered out during the identification of novel candidate genes. To make a rigorous identification, only the known genes affecting at least two cases and novel genes affecting at least five cases were included to estimate the diagnostic rate. To statistically test our newly identified candidate genes, we performed burden tests for rare variants with the RVTEST package [12], based on the methods of CMC test [17], CMC Fisher’s Exact test [17], CMC Wald test [17], Fp test [18], and Zeggini test [19].

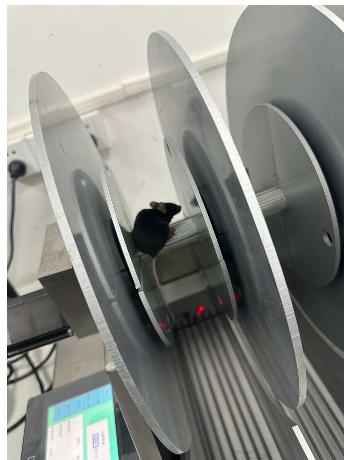
2. Mouse Behavioral Experiment Report

Grouping

Five groups were formed based on lineage and age: the control group, dmd-8 weeks group, dmd-12 weeks group, dmd-16 weeks group, and dmd-20 weeks group, with three mice per group. Due to the small number of subjects (n=3), multiple rounds of each test were performed to ensure a sufficient amount of data. The data recorded (Supplementary Table 6 and Figure 3e) have been cleaned of outliers.

Rotarod Test

Mice were trained two days prior to testing with a rotation speed of 10 rpm for 180 seconds. During the official test, the acceleration of the rotarod device was set to 2 rpm every 10 seconds, and the time the mice remained on the rotating rod was recorded in seconds. Each animal underwent five rounds of testing with a day's interval between each round, and three attempts per round with a 60-minute interval between each attempt.



Note figure 2. Rotarod test

Climbing Test

Custom climbing equipment consisting of a 50 cm long, 1 cm diameter rough pole was used, with the top end having a 3 cm diameter sphere. Medical tape was wrapped around the surface to prevent slipping. Mice were trained a day before the test to climb from the top to the bottom of the pole. On the test day, mice were acclimatized to the environment for at least half an hour. The time it took for the mice to climb from the top to the bottom was recorded in seconds. The same testing interval and round structure as the rotarod test were followed.

Note figure 2. Climbing test

Grip Strength Test

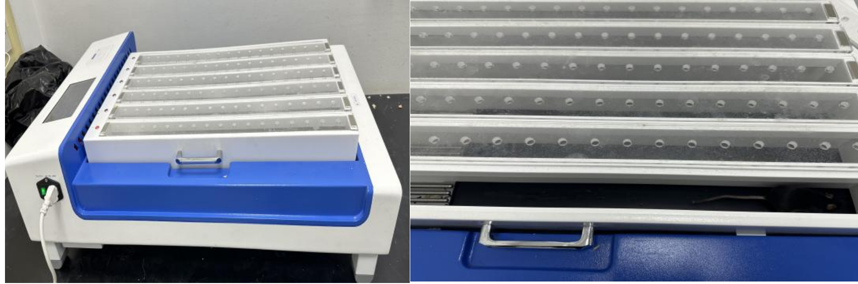
A grip strength meter was used to measure the forelimb grip strength of the mice. The meter was fixed horizontally on a table, and the mouse was placed on the pulling rod, gripping it with both forelimbs. The measurement at the moment the mouse released

the rod due to the tail being pulled at a constant speed was recorded. Each mouse underwent three rounds of testing with a day's interval between rounds, with eight measurements per round and a 10-second interval between each measurement. The maximum and minimum values were discarded, and the average of the six remaining values was taken. The order of the mice tested was randomized to reduce error.

Note figure 3. Grip strength test

Treadmill Test

Before starting the test, mice were placed on a treadmill for 5 minutes to acclimatize to the environment and reduce stress. The treadmill test involved running at set speeds, times, and inclines, with the mouse's running distance being recorded. The test followed a four-round structure similar to the other tests but with a 3-hour interval between each attempt.



Note figure 4. Treadmill test

Note: During all tests, we took great care to minimize stress in the mice. Any test was promptly halted upon signs of fatigue. No abnormal behavior or signs of distress were noted.

Reference

1. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 2012;6(2):80-92.
2. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434-43.
3. Clarke L, Zheng-Bradley X, Smith R, et al. The 1000 Genomes Project: data management and community access. *Nature Methods* 2012;9(5):459-62. doi: 10.1038/nmeth.1974
4. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* 2007;81(3):559-75.
5. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome biology* 2016;17(1):1-14.
6. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 2003;31(13):3812-14.
7. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen‐2. *Current protocols in human genetics* 2013;76(1):7.20. 1-7.20. 41.
8. Fadista J, Oskolkov N, Hansson O, et al. LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals. *Bioinformatics* 2017;33(4):471-74.
9. Mount DW. Using BLOSUM in sequence alignments. *Cold Spring Harbor Protocols* 2008;2008(6):pdb. top39.

10. Liu X, Li C, Mou C, et al. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome medicine* 2020;12(1):1-8.
11. Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Science* 2018;27(1):233-44.
12. Zhan X, Hu Y, Li B, et al. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016;32(9):1423-26.
13. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research* 2018;47(D1):D1018-D27. doi: 10.1093/nar/gky1105
14. Weinreich SS, Mangon R, Sikkens JJ, et al. [Orphanet: a European database for rare diseases]. *Ned Tijdschr Geneeskd* 2008;152(9):518-9.
15. Wright ES. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 2015;16(1):322. doi: 10.1186/s12859-015-0749-z
16. Hamosh A, Scott AF, Amberger J, et al. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* 2000;15(1):57-61. doi: 10.1002/(sici)1098-1004(200001)15:1<57::Aid-humu12>3.0.Co;2-g
17. Li B, Leal SM. Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *The American Journal of Human Genetics* 2008;83(3):311-21. doi: 10.1016/j.ajhg.2008.06.024
18. Lin D-Y, Tang Z-Z. A General Framework for Detecting Disease Associations with Rare Variants in Sequencing Studies. *The American Journal of Human Genetics* 2011;89(3):354-67. doi: 10.1016/j.ajhg.2011.07.015
19. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genetic Epidemiology* 2010;34(2):188-93. doi: https://doi.org/10.1002/gepi.20450