HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2021 June 30.

Published in final edited form as:

Science. 2020 October 09; 370(6513): 208–214. doi:10.1126/science.aaz2582.

The mole genome reveals regulatory rearrangements associated with adaptive intersexuality

Francisca M. Real[1,2], **Stefan A. Haas**[3], **Paolo Franchini**[4], **Peiwen Xiong**[4], **Oleg Simakov**[5], **Heiner Kuhl**[6], **Robert Schöpflin**[1,2], **David Heller**[3], **M-Hossein Moeinzadeh**[3], **Verena Heinrich**[3], **Thomas Krannich**[3], **Annkatrin Bressin**[3], **Michaela F. Hartmann**[7], **Stefan A. Wudy**[7], **Dina K. N. Dechmann**[8,9], **Alicia Hurtado**[10,11], **Francisco Barrionuevo**[10,11], **Magdalena Schindler**[1], **Izabela Harabula**[1], **Marco Osterwalder**[12,13], **Michael Hiller**[14,15,16], **Lars Wittler**[17], **Axel Visel**[12,18,19], **Bernd Timmermann**[1], **Axel Meyer**[4], **Martin Vingron**[3], **Rafael Jiménez**[10,11], **Stefan Mundlos**[1,2,20,*], **Darío G. Lupiáñez**[1,2,20,21,*]

[1.]RG Development & Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany. [2.]Institute for Medical and Human Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany. [3.]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [4.]Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, 78457 Konstanz, Germany. [5.]Department of Molecular Evolution and Development, University of Vienna, 1090 Vienna, Austria. [6.]Department of Ecophysiology and Aquaculture, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [7.]Steroid Research & Mass Spectrometry Unit, Laboratory for Translational Hormone Analytics in Paediatric Endocrinology, Division of Paediatric Endocrinology & Diabetology, Center of Child and Adolescent Medicine, Justus Liebig University, Giessen, Germany. [8.]Department of Migration and Immuno-Ecology, Max Planck Institute for Animal Behavior, Radolfzell, Germany. [9.]Department of Biology, University of Konstanz, Konstanz, Germany. [10.]Departamento de Genética, Universidad de Granada, Granada, Spain. [11.]Instituto de Biotecnología, Centro de Investigación Biomédica, Universidad de Granada, Armilla, Granada, Spain. [12.]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. [13.]Department for BioMedical Research (DBMR), University of Bern, Murtenstrasse 35, 3008 Bern, Switzerland. [14.]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany. [15.]Max Planck Institute for the Physics of

[*]corresponding authors: Stefan Mundlos: mundlos@molgen.mpg.de, Darío G. Lupiáñez: Dario.Lupianez@mdc-berlin.de.

Author contributions: F.M.R., S.M. and D.G.L. designed the experimental approach. F.M.R., D.G.L., F.B, and R.J. captured moles and prepared the samples. F.M.R., M.S. and D.G.L. processed the samples for DNA, Hi-C, RNA, ChIP and ATAC sequencing. S.A.H., B.T., A.M., and M.V. supervised bioinformatic analyses. R.S., D.H., M-H.M., and H.K., assembled the genome. P.F., P.X. and O.S. performed the gene annotation and posterior genomic analyses. S.A.H. and A.B. analyzed the RNA-seq data. S.A.H. performed the epigenetic analysis, and the prediction of synteny breaks. M.H. performed multispecies alignments. R.S., V.H. and T.K., analyzed chromatin interaction datasets. M.F.H., S.A.W., performed hormone measurements. F.M.R., D.G.L. and A.H. performed the *CYP17A1* enhancer transgenic experiment. F.M.R., I.H. and M.O. performed the *FGF9* transgenic experiments. L.W. derived the XX mESC line and performed blastocyst aggregations. F.M.R., M.S., S.M. and D.G.L. performed ELISA experiments and grip strength tests. F.M.R., A.C.T., D.D., A.V., A.M., S.M. and D.G.L. participated in data interpretation and discussion. F.M.R., S.M. and D.G.L. wrote the manuscript with the contribution from all other authors.

Competing interests: the authors declare non-competing interests.

Data and materials availability: The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession RCFO00000000 (BioProject number: PRJNA494291). All transcriptomes and epigenetic data have been deposited in GEO under this accession number: GSE120589. Alignments have been deposited at https://bds.mpi-cbg.de/hillerlab/IberianMole

Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany. [16.]Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany [17.]Department of Developmental Genetics, Transgenic Unit. Max Planck Institute for Molecular Genetics, Berlin, Germany. [18.]US Department of Energy Joint Genome Institute, Berkeley, California 94720, USA. [19.]School of Natural Sciences, University of California, Merced, California 95343, USA. [20.]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité - Universitätsmedizin Berlin, Berlin, Germany. [21.]Epigenetics and Sex Development Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Berlin, Germany.

Abstract

Linking genomic variation to phenotypical traits remains a major challenge in evolutionary genetics. Here, we employ phylogenomic strategies to investigate a unique trait among mammals: the development of masculinizing ovotestes in female moles. Combining a chromosome-scale genome assembly of the Iberian mole, *Talpa occidentalis*, with transcriptomic, epigenetic and chromatin interaction datasets, we identify rearrangements altering the regulatory landscape of genes with distinct gonadal expression patterns. These include a tandem-triplication involving *CYP17A1,* a gene controlling androgen synthesis, and an intra-chromosomal inversion involving the pro-testicular factor *FGF9,* which is heterochronically expressed in mole ovotestes. Transgenic mice with a knocked-in mole *CYP17A1* enhancer or overexpressing *FGF9* showed phenotypes recapitulating mole sexual features. Our results highlight how integrative genomic approaches can reveal the phenotypic impact of non-coding sequence changes.

One Sentence Summary:

Phylogenomic analyses identify altered regulation at the *CYP17A1* and *FGF9* loci, which contributes to mole ovotestis formation

---

Phenotypic diversity across species results from differences in genome sequence and structure, which provide the molecular foundation for environmental adaptation. In evolutionary genetics, linking genomic alterations to phenotypic traits has largely relied on candidate gene (1) or linkage-mapping analyses (2). However, the combination of next generation sequencing with proximity-ligation methods, in particular Hi-C, allows the generation of chromosome-scale genome assemblies (3), opening ample possibilities for comparative genomics. More importantly, Hi-C enables the integration of 3D genome structure with transcriptional control. Vertebrate genomes are spatially organized into regulatory units, termed topologically associating domains (TADs) (4, 5). Although TADs are generally preserved across species (6, 7), studies of human disease highlighted that alterations in TAD organization can cause changes in gene expression and developmental phenotypes by re-wiring enhancer-promoter contacts (8-10). Thus, analytical strategies that consider 3D organization and regulatory potential become essential for a comprehensive annotation of genomes. Here, we introduce a phylogenomic strategy that combines comparative whole-genome, epigenomic, transcriptomic and chromatin interaction data to identify phenotype-associated genomic changes. We demonstrate the power of this approach

by elucidating the molecular underpinnings of generalized intersexuality in female moles, an evolutionary trait unique among mammals.

In mammals, sex is determined genetically and results in the differentiation of the bipotential gonad into either testicular or ovarian tissue, originating sex-specific anatomical, hormonal, and behavioral differences (11). A striking exception to this paradigm occurs in moles (Family Talpidae) where, in at least eight species, XX-genotypic females have an intersex phenotype (12, 13). While male moles have normal testes, genotypic females develop ovotestes instead of ovaries (Fig. 1A). These unusual gonads are composed of an ovarian part (OP) that fully supports sexual reproduction, and a testicular part (TP) that lacks fertile germ cells but contains typical male cell populations, such as androgen-producing Leydig cells (12) (fig. S1). As a consequence of increased androgen synthesis, female moles develop masculinized external genitalia, as well as prominent muscles and aggressive behavior (14), traits that likely represent adaptations to a subterranean lifestyle.

The Iberian mole genome

To investigate the molecular origins of mole ovotestes, we generated a chromosome-scale genome assembly for *Talpa occidentalis* based on long- and short-read sequencing, and scaffolded using Hi-C data (Fig. 1B; fig. S2). The assembly comprises 2.099 Gigabases (Gb) and up to 30% is made of transposable elements, with a different repeat profile to closely-related mammals (fig. S3, Supplementary Text). Combining RNA-seq datasets and homology-based predictions, we identified 18,751 genes including 2,370 single-copy orthologues. This gene subset was used to determine 1,580 one-to-one orthologous genes in nine species and build a phylogenetic tree (Fig. 1C). Our analysis confirm moles as a distinct family in the Order Eulipotyphla, with shrews and hedgehogs being the most closely related species (15) (Supplementary text).

Epigenetic and transcriptional landscape of mole gonadal development

We generated epigenetic and transcriptomic profiles of mole gonads at 7 days *post-partum* (P7), processing TP and OP separately (fig. S2). Specifically, we produced ChIP-seq datasets against histone modifications (H3K4me1, H3K4me3, H3K27ac and H3K27me3) to segment the mole genome into functional states for each tissue. Additionally, we performed assays for transposase-accessible chromatin using sequencing (ATAC-seq; fig. S2) and intersected both datasets to predict active enhancers in each tissue (22,105 in total; Fig 1D, data S1). While the TP and testis shared a higher number of putative enhancers than TP and OP, the large number of TP-specific putative enhancers (2,726) indicates a distinct molecular profile from testis and OP. Principal component analysis of RNA-seq data further confirmed the unique nature of the TP (Fig. 1E).

An analytical framework for evolutionary analyses

We combined our functional datasets with comparative genomics analyses and focused on three distinct layers: genes, regulatory elements and 3D chromatin organization (fig. S4). We reasoned that phenotype-relevant mutations affecting these layers should be shared by the

Iberian mole and the American star-nosed mole (*Condylura cristata*), whose females also develop ovotestes. We first searched for gene families that underwent expansion/contraction in the mole lineage, as well as genes under positive selection (Supplementary Text, data S2-S3). GO enrichment analyses revealed signatures in metabolic, immunological processes and the olfactory receptor repertoire. By filtering with GO terms related to sex differentiation ("sex, gonad", data S4), we identified eight positively-selected genes that could affect mole gonadal development (table S1). To gain a functional insight into these genes, we searched for mouse- and human-reported mutations. Although some mutations led to reduced fertility, none was reported to induce sex reversal, thus making a contribution to the mole intersex phenotype unlikely (data S3).

Next, we focused on regulatory elements by identifying mole-accelerated regions, defined as genomic segments that are highly conserved during mammalian evolution but strikingly diverged in moles (Supplementary Text, data S5). We identified 3,560 mole-accelerated regions that were subsequently filtered for overlap with our predicted gonadal enhancers (129 regions). Instead of associating the accelerated enhancers to the nearest gene, we employed TAD predictions from Hi-C data to delimit a genomic range of interaction for each element and assign them to the genes within such regions. Of note, TADs are well conserved across tissues (fig. S5), consistent with previous findings (4). The assigned genes were further filtered with GO terms related to sex differentiation, which revealed two candidates: the transcription factor *Osr1* and the cell-cycle regulator *Cdk2* (table S1). Although both genes are essential for gonadogenesis (16, 17), they show a similar expression pattern in moles and mice and are likely not relevant for mole ovotestis formation (fig. S6).

Since no relevant candidates were found, we searched for changes in 3D chromatin organization, based on the hypothesis that rearrangements can alter regulatory domains and affect gene expression (8, 9). Rearrangements can be identified in genome comparisons as synteny breaks, defined here as alterations on the conserved co-linear order of loci between species. To identify mole-specific rearrangements, we compared the mole genomes (*T. occidentalis* and *C. cristata*) with full-chromosome assemblies from human, mouse and from shrew, as the closest taxonomical outgroup with normal ovarian development (Fig. 2A). We employed our Hi-C domain predictions to identify genes located within TADs affected by a synteny break, for a total of 2,595 candidate genes considered to be susceptible to altered regulation (data S6). We filtered these candidates according to GO terms related to sex differentiation, restricting the list to 39 genes (table S1, data S4).

Using our functional datasets, we then searched for footprints of altered gene regulation that might be the consequence of mole-specific rearrangements. We considered the nature of each rearrangement (Fig. 2B, Supplementary text) and our TAD predictions, to delimit a potential region of novel interactions for each candidate gene and determine the number of active regulatory elements contained within. Of the 39 candidate genes, only 17 were predicted to gain *de novo* interaction with regions containing active enhancers. Furthermore, we ranked the candidate genes by correlating the number of active elements with the expression levels for each tissue, as an indicative parameter of potential effects of the rearrangement on transcription (Fig. 2C, fig. S7, data S7). A positive correlation between

active enhancers and gene expression was found for 10 genes. Among the top-ranked candidates, we selected those displaying higher expression for subsequent functional validation: the androgen-related gene *CYP17A1* and the pro-testicular factor *FGF9*.

A tandem triplication at the *CYP17A1* locus is linked to increased androgen production and strength

We detected an intra-TAD tandem triplication at the mole *CYP17A1* locus that creates two additional copies of the gene. Through comparative genomics analysis, we confirmed the exclusive presence of the rearrangement in the mole lineage, and its absence in other mammals (Fig. 3A, fig. S8). *CYP17A1* encodes a key enzyme controlling androgen synthesis (18), suggesting a role in female mole masculinization. The triplication was associated with high *CYP17A1* expression in testis and TP, both substantially exceeding the expression levels in mice (Fig. 3B, data S8). Searching for other genes of the steroidogenic pathway, we observed that *CYP19A1*, located downstream of *CYP17A1*, is not expressed in the TP (fig. S9). *CYP19A1* encodes for aromatase, an enzyme that converts androgens to estrogens (18) and is expressed exclusively in the OP (fig. S9). It is thus expected that the high levels of *CYP17A1* in the OP do not impede estrogen production and reproductive function, due to the protective effect of aromatase in degrading androgens locally. The absence of *CYP19A1* expression in the TP, in combination with high *CYP17A1* expression, provide a plausible explanation for the masculinization observed in female moles. We used gas chromatography-mass-spectrometry to quantify serum levels of male hormones and found high levels of circulating androgens in female moles. Androgen levels were similar as in male individuals (fig. S9), contrary to the general pattern among mammals where males display higher levels than females.

To examine the contributions of the additional *CYP17A1* alleles to increased androgen production, we analyzed RNA-seq data from mole testes and TP. The three *CYP17A1* paralogues have sufficiently diverged to enable unambiguous mapping of RNA-seq reads. Unexpectedly, the two newly emerged *CYP17A1-2* and *CYP17A1-3* jointly contribute less than 5% of *CYP17A1* transcript (Fig. 3C). Furthermore, sequence conservation analyses revealed that they diverge more from the human sequence than *CYP17A1-1* (fig. S10). These findings suggest that the triplication of the *CYP17A1* gene itself does not explain the increased androgen levels in moles. Instead, the triplication also caused the duplication and fusion of two predicted enhancer elements, termed “enhancer A” and “enhancer B” (Fig. 3C). As the triplication does not affect any TAD boundary, the duplicated enhancers and *CYP17A1* genes locate within the same regulatory domain (fig. S11). The duplicated fusion element, “enhancer A-B”, shows a high degree of sequence conservation with the original elements (84 and 90%, respectively) and high abundance of active enhancer marks (Fig. 3C). Computational predictions of binding affinities showed that enhancer A-B maintains significant binding affinities for transcription factors (TFs) found in the original enhancer A, as well as novel binding affinities for TFs that are expressed in the TP (data S9). Together, these observations suggest that duplication and functional changes in regulatory sequences, rather than amplification of coding sequence, cause the observed phenotypic adaptations.

To confirm this hypothesis, we inserted the mole enhancer A-B sequence into the *Cyp17a1* murine locus (Fig. 3D). Adult knock-in mice carrying the additional mole enhancer displayed a three and two-fold upregulation of *Cyp17a1* expression in females and males, respectively. The increased expression occurs in the same cell type (steroidogenic cells) as in wildtype controls (fig. S12). A similar effect was observed at embryonic stages (fig. S12), thus confirming the increased *in vivo* activity of the mole-specific fusion enhancer in gonadal tissue. The upregulation of *Cyp17a1* expression was accompanied by an increase in circulating testosterone in females and males (two and three-fold respectively, Fig. 3E). Since androgens have potent anabolic effects in muscle, we performed a strength grip test that revealed a significant increase in muscle strength in mutants compared to wildtype controls (Fig. 3F). Thus, the regulatory nature of the *CYP17A1* rearrangement offers a plausible molecular mechanism for the observed shift in hormone levels and the corresponding phenotype.

An inversion at the *FGF9* locus is associated to delayed meiosis and gonadal masculinization

We identified a rearrangement at the *FGF9* locus which is exclusive to the mole lineage and not present in any other mammals examined (Fig. 4A, fig. S13-S14). *FGF9* is a known testis-determining gene that functions in positive feedback with *SOX9* and inhibits the ovary-determining *WNT4*/β-catenin pathway (19) (fig. S15). Consequently, loss of *Fgf9* in XY gonads results in downregulation of testicular markers and male-to-female sex reversal. Comparative analyses against human, representative of the ancestral organization at the locus, revealed a large inversion that relocates a distant genomic region (26 Mb away in the human genome) to the mole *FGF9* locus (Fig. 4B). Hi-C data showed that the synteny break occurs within the human *FGF9* TAD, disrupting its 3D organization. Thus, the mole locus is reorganized, with most of the *FGF9* TAD remaining conserved, but extending beyond the synteny breakpoint on the centromeric side. This new interaction domain is delimited by the presence of two CTCF binding sites with divergent orientation, a genomic signature associated to TAD boundaries (7) (Fig. 4B). A closer examination of the newly interacting region revealed several elements enriched for active epigenetic marks, some of them specific for the mole ovotestis (fig. S16). This novel interaction pattern was also validated through independent 4C-seq experiments (Fig. 4B), which confirmed the association of the *FGF9* gene with the region containing the putative tissue-specific enhancers. We tested one element (fig. S16) in mouse LacZ reporter assays. Although this element displayed enhancer activity in tissues like the eye (fig. S17), we did not observe gonadal staining, which could be due to specific requirements for additional trans-acting factors in the mole. Alternatively, it may reflect known limitations of the reporter assay that might be intensified by the inter-species nature of the experiment (20).

We then explored possible alterations on the dimorphic *FGF9* expression pattern observed in other mammals, which is essential for suppressing germ cell meiosis (21). In mice, *Fgf9* is first expressed in the bipotential gonad of both sexes and becomes progressively restricted to the testis and turned off in ovaries, allowing the initiation of female meiosis at E13.5. In mole gonads, however, *FGF9* expression is maintained after sex determination in both sexes

and becomes restricted to the OP at later stages (Fig. 4C). Immunostaining analyses showed that FGF9 expression persists in the OP across the entire prenatal period, and becomes confined to a thin rim at postnatal stages (P7). The spatial reduction in FGF9 expression is concomitant with the initiation of meiosis (Fig. 5A and fig. S18) which is considerably delayed in female moles until birth, an exceptional feature among mammals (22). Consequently, the observed heterochrony on mole *FGF9* expression, compared to mouse, and its potential effects are suggestive of a contribution to the masculinization of female mole gonads.

We hypothesized that *FGF9* expression during early XX mole gonadogenesis might prevent germ cells from entering meiosis in the OP, allowing the TP to develop further. To test this hypothesis, we engineered a BAC construct to overexpress *Fgf9* in somatic ovarian populations and generated transgenic mice through *PiggyBac* transgenesis and morula aggregation. Highly-chimeric animals displayed early embryonic lethality, likely due to *Fgf9* misexpression in other organs, an effect that precluded their study in later stages. Nevertheless, RNA-seq analysis of E13.5 ovaries demonstrated an inhibition of the meiotic process, manifested by downregulation of meiosis markers (Fig. 5B, data S10). In contrast, low-chimaera individuals composed of XX wild-type and XX mutant cells were viable and displayed *Fgf9* expression during the entire ovarian development. These animals showed a complete female-to-male sex reversal, defined by gonadal morphology and expression of male-specific genes like SOX9 (Figure 5C). These results directly confirm the potential of altered *FGF9* expression to induce masculinization in mammalian XX females.

Discussion

Vertebrate sex determination is controlled by a limited set of key regulators whose hierarchy has evolved dynamically (11). Most of these genes display pleiotropic effects, controlling regulatory networks in several tissues, often making them indispensable for embryonic viability (23, 24). It is thus plausible that variations in sex determination derive from regulatory changes that alter gene expression patterns but preserve essential functions, as suggested for other evolutionary adaptations (1, 2). These genomic changes appear to be linked to the evolutionary success of moles and demonstrate that regulatory innovation can overcome a priori seemingly incompatible situations such as female fertility in the presence of high androgen levels. But how could an intersex phenotype evolve and what is its adaptive significance? One potential explanation is the anabolic effects of androgens on muscle mass. Mole ovotestes may have evolved to equalize muscular strength among sexes, by increasing androgens in females. Such an advantage, combined with other androgen-derived effects such as aggressive behavior, might have been key for the adaptation to the demanding requirements of a burrowing underground lifestyle (14).

This study also highlights the evolutionary importance of genomic rearrangements, and their potential to modulate developmental gene expression. In most cases, genomic rearrangements would have limited effects as they preserve whole regulatory units and do not disrupt enhancer-promoter interactions (6, 25). However, as shown here, they can also alter the regulatory potential within the local environment of a TAD, as for the *CYP17A1* locus, or shuffle the functional content of distant TAD units, exemplified by the *FGF9* locus.

Interestingly, similar effects have been observed in human genetic diseases for enhancer duplications (26) and for inversions (8, 9). Therefore, regulation of genes by enhancer elements and their organization in TADs might constitute a mechanism of “evolvability” resulting from a modular system with vast flexibility and enormous evolutionary potential for the origin of novelties. Based on this modularity, genomic rearrangements can easily change and reconstitute complex expression patterns, thereby contributing to the saltatory nature of phenotypic innovation observed in many phylogenetic lineages. We expect that approaches considering these important aspects will eventually reveal the evolutionary basis of many other traits and substantially enhance the toolbox for unlocking the secrets of phenotypic variation and adaptation across the animal kingdom.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank Asita Stiege, Norbert Brieske, Ute Fisher, Karol Macura, Judith Fiedler, Katja Zill, Christin Franke and Guido Mastrobuoni for technical support. We thank Peter Koopman for kindly providing us with the Wt1-BAC vector. We thank the sequencing and genotyping facility of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden for helping with the initial long-read sequencing project. We thank The Animal Outcome Core Facility of the NeuroCure center, Charité University, Berlin, for helping with strength tests. We also thank members of the Lupiáñez and Mundlos labs for fruitful discussions.

Funding:

This research was supported by grants from the Deutsche Forschungsgemeinschaft (grant numbers MU 880/15-1 and MU 880/16-1) and Max Planck Society. D.G.L. was supported by the Fundación Alfonso Martín Escudero and by a Helmholtz ERC Recognition Award grant from the Helmholtz-Gemeinschaft (ERC-RA-0033). Work conducted at the E. O. Lawrence Berkeley National Laboratory was performed under Department of Energy contract DE-AC02-05CH11231, University of California. A.V. and M.O. were supported by NIH grant R01HG003988. M.O. was supported by Swiss National Science Foundation grant PCEFP3_186993. O.S. was supported by the Austrian Science Fund FWF grant P32190.

References and Notes

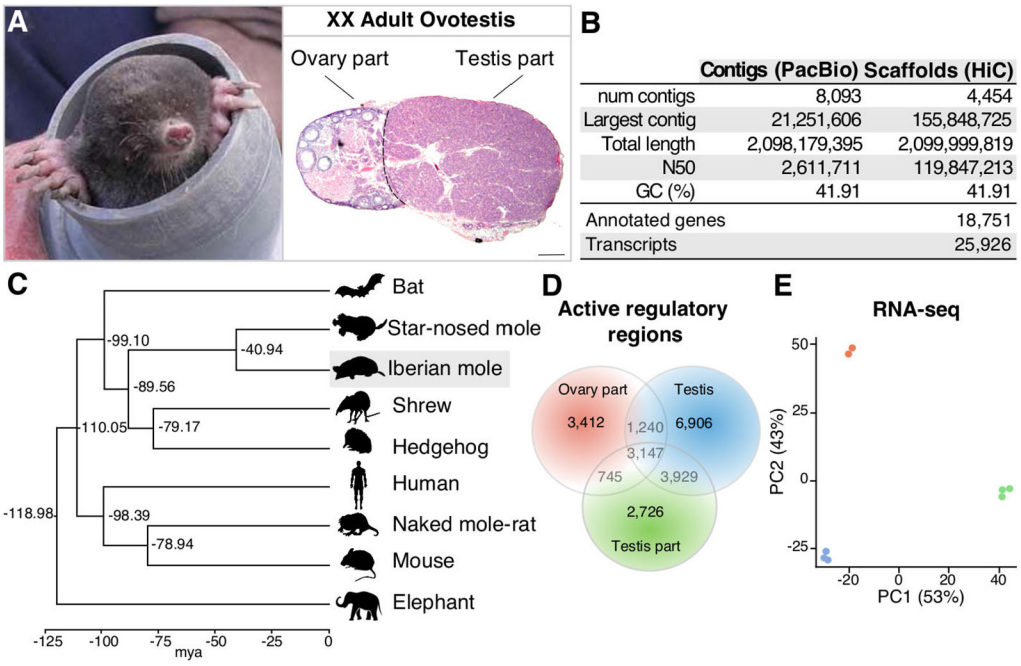
1. Lopez-Rios J et al., Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. Nature. 511, 46–51 (2014). [PubMed: 24990743]
2. Chan YF et al., Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 327, 302–305 (2010). [PubMed: 20007865]
3. Dudchenko O et al., De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, eaal3327 (2017).
4. Dixon JR et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485, 376–380 (2012). [PubMed: 22495300]
5. Nora EP et al., Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 485, 381–385 (2012). [PubMed: 22495304]
6. Harmston N et al., Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun. 8, 441 (2017). [PubMed: 28874668]
7. Gómez-Marín C et al., Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. PNAS. 112, 7542–7547 (2015). [PubMed: 26034287]
8. Spielmann M, Lupiáñez DG, Mundlos S, Structural variation in the 3D genome. Nat. Rev. Genet (2018), doi:10.1038/s41576-018-0007-0.

9. Lupiáñez DG et al., Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 161, 1012–1025 (2015). [PubMed: 25959774]
10. Franke M et al., Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 538, 265–269 (2016). [PubMed: 27706140]
11. Capel B, Vertebrate sex determination: evolutionary plasticity of a fundamental switch. Nat. Rev. Genet 18, 675–689 (2017). [PubMed: 28804140]
12. Barrionuevo FJ, Zurita F, Burgos M, Jiménez R, Testis-like development of gonads in female moles. New insights on mammalian gonad organogenesis. Dev. Biol 268, 39–52 (2004). [PubMed: 15031103]
13. Carmona FD et al., The evolution of female mole ovotestes evidences high plasticity of mammalian gonad development. J. Exp. Zool. B Mol. Dev. Evol 310, 259–266 (2008). [PubMed: 18085526]
14. Haeck J, Colonization of the mole (Talpa europea L.) in the ijsselmeer polders. Netherlands Journal of Zoology. 19 (2), 145–248 (1969).
15. Douady CJ et al., Molecular phylogenetic evidence confirming the Eulipotyphla concept and in support of hedgehogs as the sister group to shrews. Molecular Phylogenetics and Evolution. 25, 200–209 (2002). [PubMed: 12383761]
16. Wang Q, Lan Y, Cho E-S, Maltby KM, Jiang R, Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. Dev Biol. 288 (2005), doi:10.1016/j.ydbio.2005.09.024.
17. Berthet C, Aleem E, Coppola V, Tessarollo L, Kaldis P, Cdk2 Knockout Mice Are Viable. Current Biology. 13, 1775–1785 (2003). [PubMed: 14561402]
18. Hanukoglu I, Steroidogenic enzymes: structure, function, and role in regulation of steroid hormone biosynthesis. J. Steroid Biochem. Mol. Biol 43, 779–804 (1992). [PubMed: 22217824]
19. Capel B, Vertebrate sex determination: evolutionary plasticity of a fundamental switch. Nat. Rev. Genet 18, 675–689 (2017). [PubMed: 28804140]
20. Suryamohan K, Halfon MS, Overview Article: Identifying transcriptional cis-regulatory modules in animal genomes. Wiley Interdiscip Rev Dev Biol. 4, 59–84 (2015). [PubMed: 25704908]
21. Bowles J et al., FGF9 suppresses meiosis and promotes male germ cell fate in mice. Dev. Cell. 19, 440–449 (2010). [PubMed: 20833365]
22. Zurita F et al., Meiosis onset is postponed to postnatal stages during ovotestis development in female moles. Sex Dev. 1, 66–76 (2007). [PubMed: 18391517]
23. Colvin JS, White AC, Pratt SJ, Ornitz DM, Lung hypoplasia and neonatal death in Fgf9-null mice identify this gene as an essential regulator of lung mesenchyme. Development. 128, 2095–2106 (2001). [PubMed: 11493531]
24. Bi W et al., Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc. Natl. Acad. Sci. U.S.A 98, 6698–6703 (2001). [PubMed: 11371614]
25. Lazar NH et al., Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. Genome Res. (2018), doi:10.1101/gr.233874.117.
26. Will AJ et al., Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat. Genet 49, 1539–1545 (2017). [PubMed: 28846100]
27. Peterson JK, Moran F, Conley AJ, Bird IM, Zonal expression of endothelial nitric oxide synthase in sheep and rhesus adrenal cortex. Endocrinology. 142, 5351–5363 (2001). [PubMed: 11713235]
28. Rao SSP et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159, 1665–1680 (2014). [PubMed: 25497547]
29. Durand NC et al., Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3, 95–98 (2016). [PubMed: 27467249]
30. Li H, Durbin R, Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26, 589–595 (2010). [PubMed: 20080505]
31. Dobin A et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29, 15–21 (2013). [PubMed: 23104886]
32. Love MI, Huber W, Anders S, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014). [PubMed: 25516281]

33. Thomas PD et al., PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141 (2003). [PubMed: 12952881]
34. Mammana A, Chung H-R, Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. Genome Biol. 16, 151 (2015). [PubMed: 26206277]
35. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ, Curr Protoc Mol Biol, in press, doi:10.1002/0471142727.mb2129s109.
36. Langhammer M et al., Reproductive performance primarily depends on the female genotype in a two-factorial breeding experiment using high-fertility mouse lines. Reproduction. 153, 361–368 (2017). [PubMed: 28096494]
37. Bakhaus K et al., Sodium-dependent organic anion transporter (Slc10a6−/−) knockout mice show normal spermatogenesis and reproduction, but elevated serum levels for cholesterol sulfate. J. Steroid Biochem. Mol. Biol 179, 45–54 (2018). [PubMed: 28743544]
38. Kraft K et al., Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Rep (2015), doi:10.1016/j.celrep.2015.01.016.
39. Zhao L, Ng ET, Koopman P, A piggyBac transposon- and gateway-enhanced system for efficient BAC transgenesis. Dev. Dyn 243, 1086–1094 (2014). [PubMed: 24924516]
40. Wang J et al., An improved recombineering approach by adding RecA to lambda Red recombination. Mol. Biotechnol 32, 43–53 (2006). [PubMed: 16382181]
41. Rostovskaya M et al., Transposon-mediated BAC transgenesis in human ES cells. Nucleic Acids Res. 40, e150 (2012). [PubMed: 22753106]
42. van de Werken HJG et al., 4C technology: protocols and data analysis. Meth. Enzymol 513, 89–112 (2012).
43. Marçais G, Kingsford C, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27, 764–770 (2011). [PubMed: 21217122]
44. Vurture GW et al., GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 33, 2202–2204 (2017). [PubMed: 28369201]
45. Sánchez A, Stamatopoulos C, Redi CA, Descriptive kinetics of the seminiferous epithelium cycle and genome size in the mole Talpa occidentalis (Insectivora). J. Exp. Zool 273, 51–58 (1995). [PubMed: 7561724]
46. Dolezel J, Bartos J, Voglmayr H, Greilhuber J, Nuclear DNA content and genome size of trout and human. Cytometry A. 51, 127–128; author reply 129 (2003). [PubMed: 12541287]
47. Koren S et al., Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736 (2017). [PubMed: 28298431]
48. Walker BJ et al., Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE. 9, e112963 (2014). [PubMed: 25409509]
49. Jiménez R, Burgos M, Guardia RDDL, Karyotype and Chromosome Banding in the Mole (Talpa Occidentalis) from the South-East of the Iberian Peninsula. Implications on its Taxonomic Position. Caryologia. 37, 253–258 (1984).
50. Bao Z, Eddy SR, Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–1276 (2002). [PubMed: 12176934]
51. Price AL, Jones NC, Pevzner PA, De novo identification of repeat families in large genomes. Bioinformatics. 21 Suppl 1, i351–358 (2005). [PubMed: 15961478]
52. Jurka J et al., Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467 (2005). [PubMed: 16093699]
53. Haas BJ et al., Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9, R7 (2008). [PubMed: 18190707]
54. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515 (2019). [PubMed: 30395287]
55. Slater GSC, Birney E, Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6, 31 (2005). [PubMed: 15713233]
56. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 11, 1650–1667 (2016). [PubMed: 27560171]

57. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M, BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32, 767–769 (2016). [PubMed: 26559507]
58. Haas BJ et al., Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666 (2003). [PubMed: 14500829]
59. Grabherr MG et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol 29, 644–652 (2011). [PubMed: 21572440]
60. Kim D, Langmead B, Salzberg SL, HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]
61. Pearson WR, Curr Protoc Bioinformatics, in press, doi:10.1002/0471250953.bi0309s53.
62. Drăgan M-A, Moghul I, Priyam A, Bustos C, Wurm Y, GeneValidator: identify problems with protein-coding gene predictions. Bioinformatics. 32, 1559–1561 (2016). [PubMed: 26787666]
63. Chan PP, Lowe TM, tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol. Biol 1962, 1–14 (2019). [PubMed: 31020551]
64. Chan PP, Lowe TM, GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 44, D184–D189 (2016). [PubMed: 26673694]
65. Kalvari I et al., Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46, D335–D342 (2018). [PubMed: 29112718]
66. Kalvari I et al., Non-Coding RNA Analysis Using the Rfam Database. Curr Protoc Bioinformatics. 62, e51 (2018). [PubMed: 29927072]
67. Nawrocki EP, Eddy SR, Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29, 2933–2935 (2013). [PubMed: 24008419]
68. Li H et al., TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. 34, D572–580 (2006). [PubMed: 16381935]
69. Lin Q et al., The seahorse genome and the evolution of its specialized morphology. Nature. 540, 395–399 (2016). [PubMed: 27974754]
70. Katoh K, Standley DM, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol 30, 772–780 (2013). [PubMed: 23329690]
71. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25, 1972–1973 (2009). [PubMed: 19505945]
72. Stamatakis A, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312–1313 (2014). [PubMed: 24451623]
73. Yang Z, PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol 24, 1586–1591 (2007). [PubMed: 17483113]
74. Kumar S, Stecher G, Suleski M, Hedges SB, TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol 34, 1812–1819 (2017). [PubMed: 28387841]
75. De Bie T, Cristianini N, Demuth JP, Hahn MW, CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 22, 1269–1271 (2006). [PubMed: 16543274]
76. Zhang J, Nielsen R, Yang Z, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol 22, 2472–2479 (2005). [PubMed: 16107592]
77. Benjamini Y, Hochberg Y, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 57, 289–300 (1995).
78. Reimand J et al., g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 44, W83–89 (2016). [PubMed: 27098042]
79. Harris RS, thesis, Pennsylvania State University, USA (2007).
80. Kent WJ et al., The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002). [PubMed: 12045153]
81. Blanchette M et al., Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14, 708–715 (2004). [PubMed: 15060014]

82. Siepel A et al., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005). [PubMed: 16024819]
83. Quinlan AR, Hall IM, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841–842 (2010). [PubMed: 20110278]
84. Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, E. D. Green, S. Batzoglou, A. Sidow, Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15, 901–913 (2005). [PubMed: 15965027]
85. Siepel A, Pollard KS, Haussler D, in Proceedings of the 10th annual international conference on Research in Computational Molecular Biology (Springer-Verlag, Venice, Italy, 2006; 10.1007/11732990_17), RECOMB'06, pp. 190–205.
86. Dutheil JY, Gaillard S, Stukenbrock EH, MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics. 15, 53 (2014). [PubMed: 24447531]
87. Raudvere U, et al., g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198 (2019). [PubMed: 31066453]
88. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC, Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487–493 (2011). [PubMed: 21209072]
89. McKay SJ, Vergara IA, Stajich JE, Curr Protoc Bioinformatics, in press, doi:10.1002/0471250953.bi0912s31.
90. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D, Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. U.S.A. 100, 11484–11489 (2003). [PubMed: 14500911]
91. Osipova E, Hecker N, Hiller M, RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. Gigascience. 8 (2019), doi:10.1093/gigascience/giz132.
92. Suarez HG, Langer BE, Ladde P, Hiller M, chainCleaner improves genome alignment specificity and sensitivity. Bioinformatics. 33, 1596–1603 (2017). [PubMed: 28108446]
93. Haeussler M et al., The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 47, D853–D858 (2019). [PubMed: 30407534]
94. Shin H et al., TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res. 44, e70 (2016). [PubMed: 26704975]
95. Thomas-Chollier M et al., Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat Protoc. 6, 1860–1869 (2011). [PubMed: 22051799]
96. MANFREDI RMG, The nuclear content of desoxyribonucleic acid and some problems of Mammalian phylogenesis. Mammalia. 49, 369–386 (2009).
97. Vinogradov AE, Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry. 31, 100–109 (1998). [PubMed: 9482279]
98. Smith JDL, Bickham JW, Gregory TR, Patterns of genome size diversity in bats (order Chiroptera). Genome. 56, 457–472 (2013). [PubMed: 24168629]
99. Gallardo MH, Bickham JW, Kausel G, Köhler N, Honeycutt RL, Gradual and quantum genome size shifts in the hystricognath rodents. J. Evol. Biol 16, 163–169 (2003). [PubMed: 14635891]
100. Capparelli R et al., DNA content differences in laboratory mouse strains determined by flow cytometry. Cytometry. 29, 261–266 (1997). [PubMed: 9389443]
101. Kasai F, O'Brien PCM, Ferguson-Smith MA, Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping. Genomics. 102, 468–471 (2013). [PubMed: 24055950]



	Contigs (PacBio)	Scaffolds (HiC)
num contigs	8,093	4,454
Largest contig	21,251,606	155,848,725
Total length	2,098,179,395	2,099,999,819
N50	2,611,711	119,847,213
GC (%)	41.91	41.91
Annotated genes		18,751
Transcripts		25,926

Fig. 1. Mole genome and epigenetic and transcriptional study of ovotestis development.

(A) An Iberian mole (*Talpa occidentalis*) and an adult mole ovotestis. Scale bar represents 500 μm. **(B)** Genome assembly of *T. occidentalis* and gene annotation statistics. **(C)** Phylogenetic tree, based on four-fold degenerate sites, between selected species. **(D)** Venn diagram of active enhancers (data S1). **(E)** PC analysis of RNA-seq datasets of P7 mole gonads.

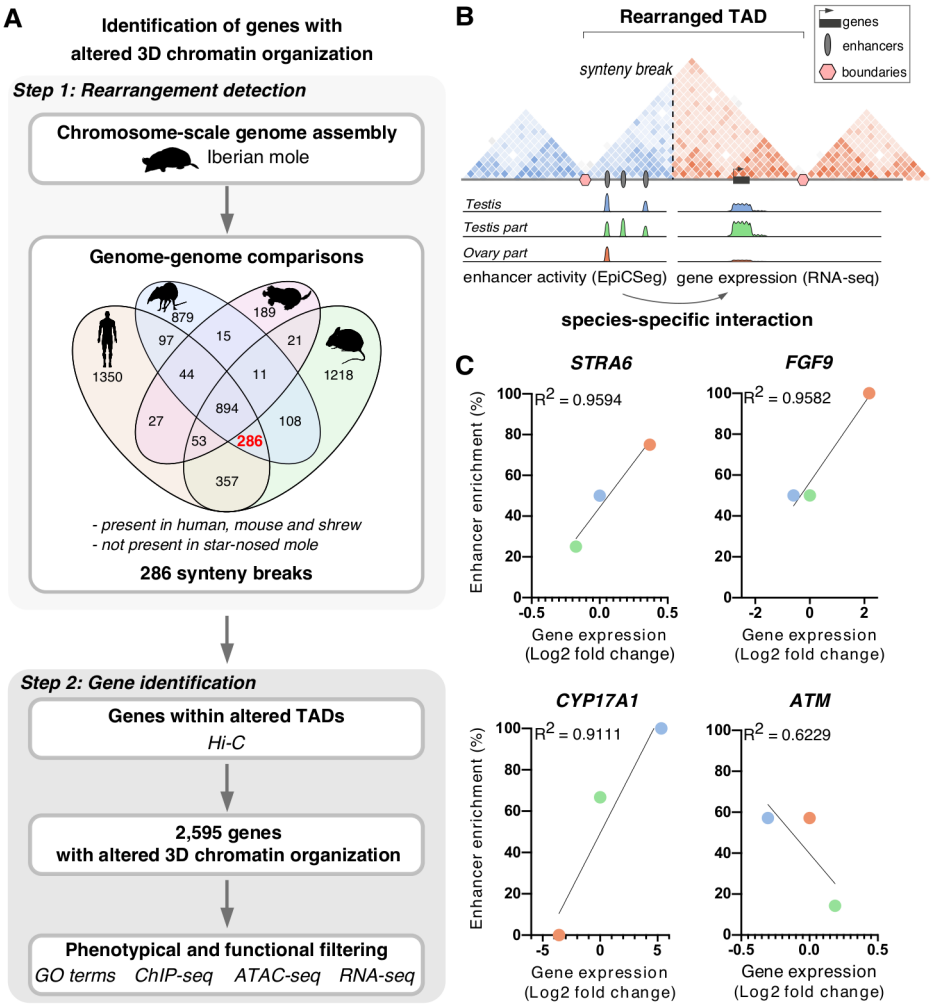


Fig. 2. Identification of genes with altered 3D chromatin regulatory landscapes.

(A) Strategy used to identify genes with altered 3D chromatin organization as a result of species-specific rearrangements. **(B)** Strategy used to assign regulatory elements to candidate genes. Number of active enhancers are correlated to gene expression levels for each tissue. **(C)** Correlation between the percentage of active enhancers and gene expression per tissue (orange=ovary part, green=testis part, blue=male testis) for selected candidates (full gene dataset in fig. S4 and data S7). *STRA6, FGF9* and *CYP17A1* display the highest positive correlation, *ATM* shows negative correlation.

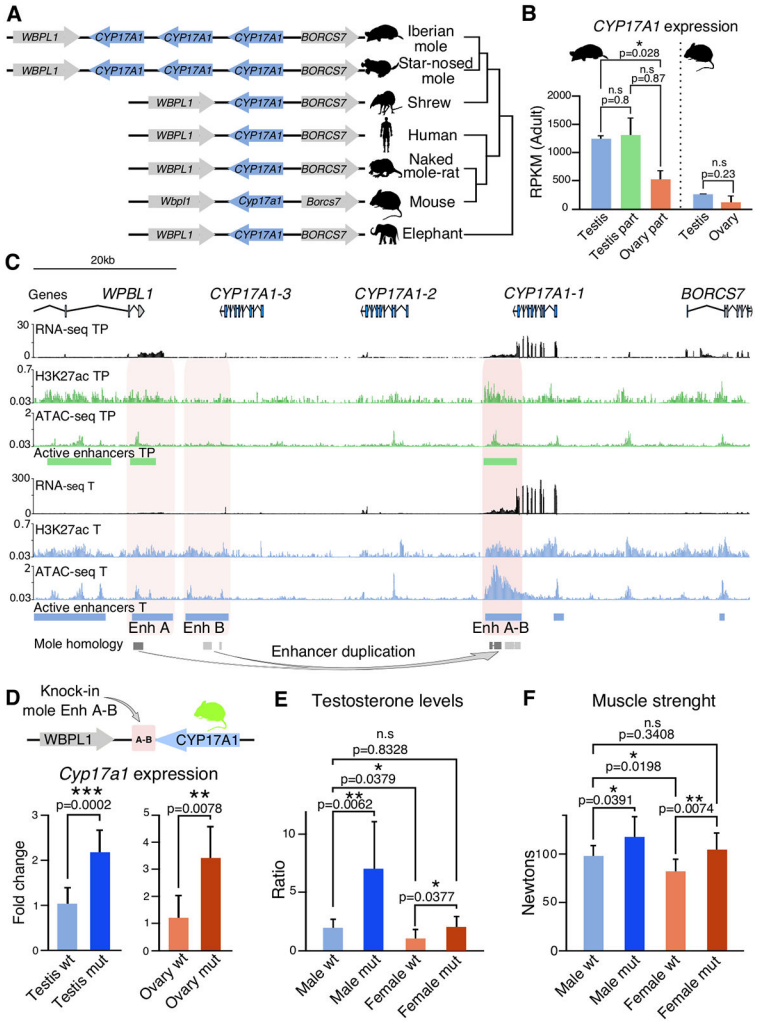


Fig. 3. Duplication of regulatory elements at the *CYP17A1* locus and associated increase in androgen production and strength.

(A) Comparative genomics at the *CYP17A1* locus. **(B)** *CYP17A1* expression (RNA-seq) in mole and mouse adult gonads (n=2). **(C)** Expression profile (RNA-seq, top), enhancer marks (H3K27Ac, center) and open chromatin (ATAC-seq, bottom) for testis part and testis at P7 gonads. Segmentation for active enhancers for testis part (green bars) and testis (blue bars). BLAT Sequence homology is represented in gray boxes. Duplicated enhancer (A-B) results from fusion of enhancer A and B. **(D)** Up: Integration of the mole *CYP17A1* duplicated enhancer into the mouse *Cyp17a1* locus. Down: Expression analysis of *Cyp17a1* (RT-qPCR) in adult mouse mutant gonads and wild-type controls (n ≥ 5). **(E)** Circulating testosterone levels in adult mouse mutants and wild-type controls (n=7). **(F)** Grip strength test in adult mouse mutants and wild-type controls (n=7). Bars represent mean and SD. Two-sided Student's t-test, n.s = non-significant, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.

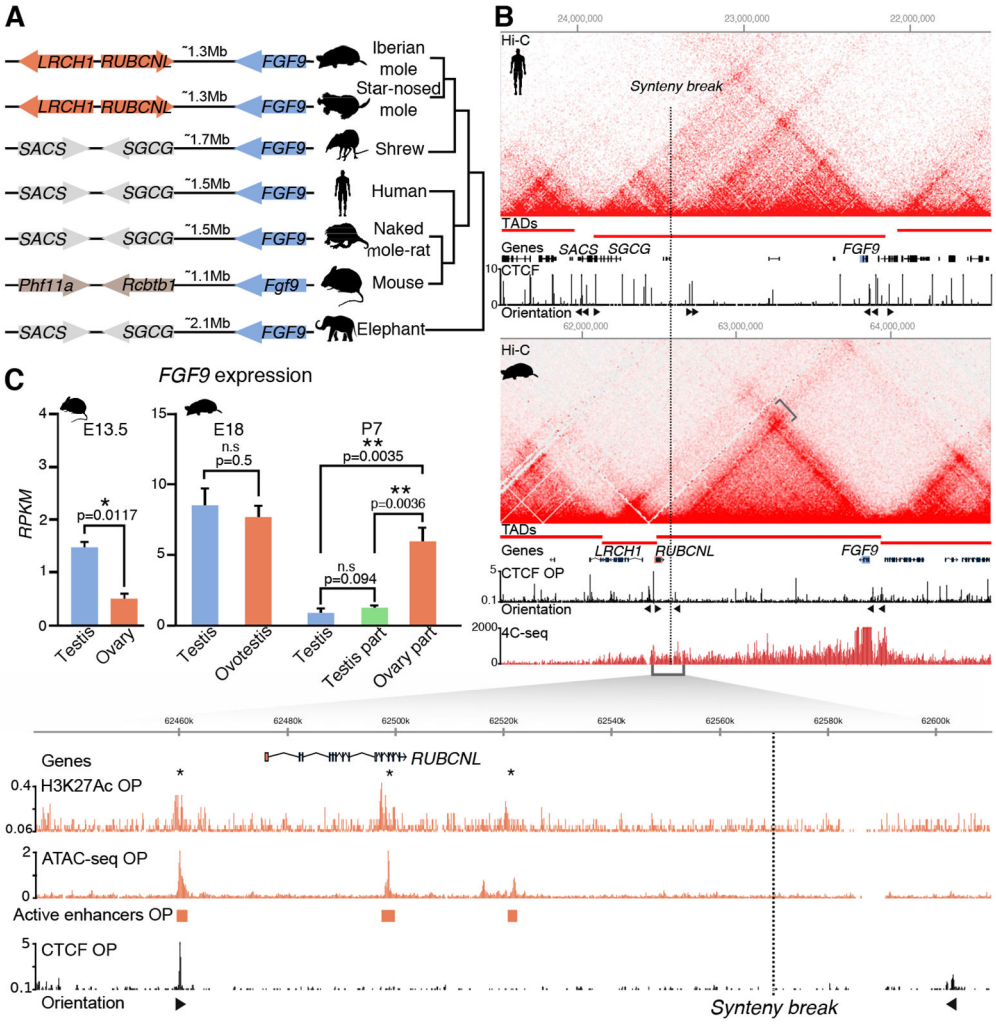


Fig. 4. An inversion altering the regulatory landscape of the *FGF9* mole locus.

(A) Comparative genomics at the *FGF9* locus. **(B)** Hi-C maps for human and mole displaying synteny break (discontinuous line) and TAD prediction. Below, CTCF ChIP-seq (mole P7 gonads) with peak orientation. 4C-seq using mole *FGF9* promoter as viewpoint shows contact extension beyond synteny break. Zoom of *FGF9* interacting region shows active ovarian enhancers (asterisks). H3K27Ac, ATAC-seq and segmentation for active enhancer tracks are displayed in orange. **(C)** *FGF9* expression (RNA-seq) in mice and moles at different timepoints. Bars represent mean and SD ($n \geq 2$). Two-sided Student's t-test, n.s = non-significant, *$P \leq 0.05$, **$P \leq 0.01$.

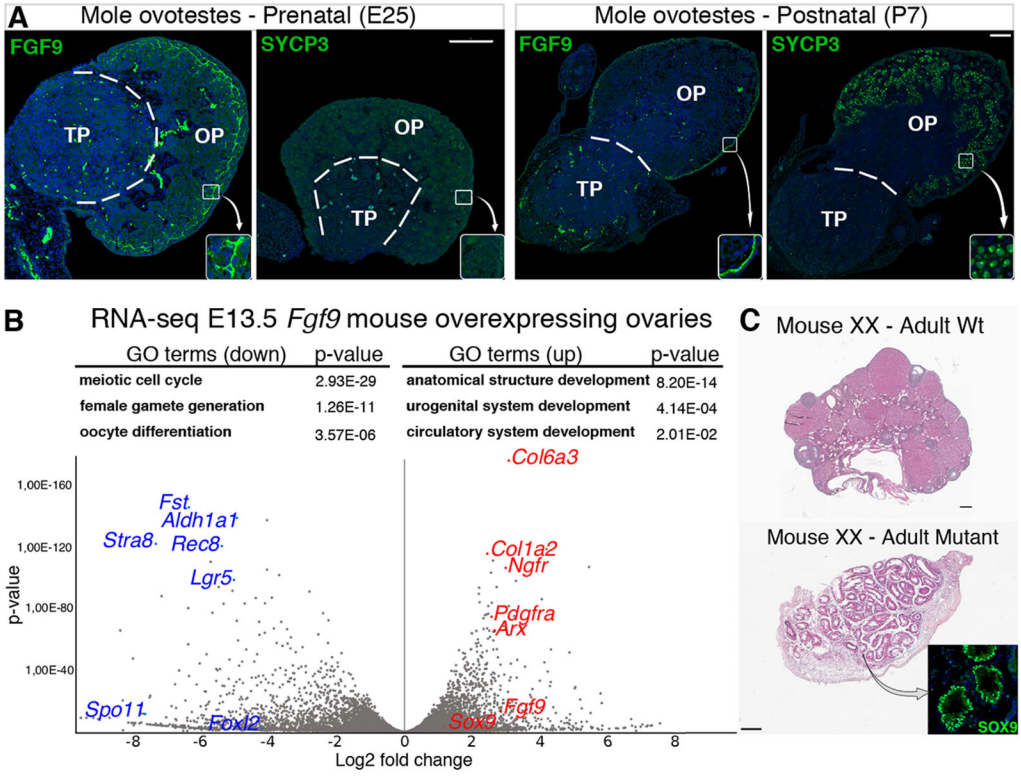


Fig. 5. *FGF9* sustained expression delays meiosis and promotes XX gonadal masculinization.

(A) Spatio-temporal expression of FGF9 and the meiotic marker SYCP3 (immunostaining, green, DAPI in blue). Insets display zoomed regions from OP. Scale bars represent 100 μm. **(B)** Volcano plot from RNA-seq of XX mutant versus wild-type gonads (E13.5) and gene ontology analysis. **(C)** Hematoxylin and Eosin staining of XX gonads of adult mutants and wild-type controls. Cord-like structures in mutants denote XX-to-XY sex reversal. Inset shows SOX9 expression (immunostaining, green, DAPI in blue). Scale bars represent 200 μm.