## PROJECT NARRATIVE

### A. Background and Rationale

The **goal** of this proposal is to map genetic modifiers of tuberous sclerosis complex (TSC) outcomes in a genetically diverse mouse population (Fig. 1). TSC is a rare autosomal dominant disease caused by loss-of-function (LoF) mutations in either of the genes *TSC1* or *TSC2* (*TSC1/2*). These mutations cause aberrant up-regulation of mammalian target of rapamycin (mTOR) signaling, resulting in high rates of epilepsy, autism spectrum disorder (ASD), and tumor formation in multiple organ systems [1]. However, presentation of TSC in the patient population is highly heterogeneous ranging from subclinical to severe [1–7]. There are at least three distinct contributors to this phenotypic complexity: 1) a patient's specific TSC gene mutation; 2) the random occurrence of somatic second-hit *TSC1/2* mutations, and 3) genetic modifiers, i.e., alleles inherited separately from the *TSC1/2* genes that alter the severity of TSC-associated phenotypes. The **central hypothesis** of this project is that genetic modifiers are a major source of patient heterogeneity. Despite high heterogeneity within affected families sharing the same mutation [5–7], TSC outcomes are more similar in identical twins and within families than across families [8], supporting the existence of genetic modifiers in the human population.

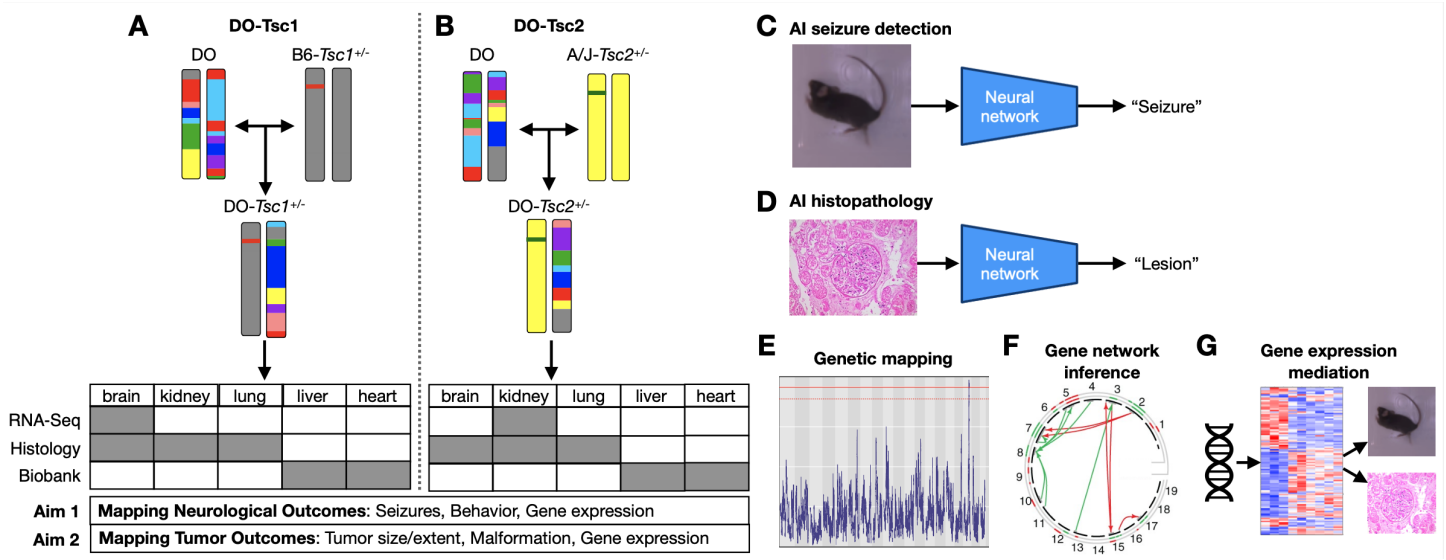


Figure 1: **Project overview.** Breeding schematics show generation of A) DO-*Tsc1* and B) DO-*Tsc2* mice by crossing a DO female to an inbred mouse carrying a *Tsc1* or *Tsc2* null allele. DO-*Tsc1* and DO-*Tsc2* mice will be used in both aims of this proposal. Both mouse subpopulations will be run through all phenotyping pipelines and will have brain, kidneys, lung, liver, and heart harvested. RNA sequencing will be performed in brain for the DO-*Tsc1* subpopulation and kidney for the DO-*Tsc2* subpopulation. We will use AI technology to quantify C) seizures and D) histological phenotypes in all mice. We will map TSC-related phenotypic outcomes to E) individual QTL and F) to genetic interaction networks. G) Gene expression will be used as a mediator of genetic effects on phenotypes to identify molecular pathways influencing tumor and neurological outcomes.

The **rationale** for this study is that by mapping genetic modifiers, we will identify novel genes and pathways that can be therapeutically targeted to mitigate poor outcomes. Although current therapies that directly target mTOR signaling show promise in treating TSC-associated conditions [9–12], significant challenges remain. Indeed, some TSC tumors are unaffected by treatment with rapamycin [12–16], and both seizures [17] and tumors [18, 19] develop resistance to these therapies with long-term treatment. Thus, the identification of alternative or adjunct therapies for TSC is critical for long-term disease management. Previous work supports strategies targeting mTOR-independent pathways. For example, genetic studies in patients with TSC led to work in a mouse model demonstrating that treatment with IFN-$\gamma$ in combination with rapamycin was more effective in reducing tumor growth and improving survival more than rapamycin alone [20]. Identification of this and other alternative pathways [15, 21–23] through genetic modifier screens, may be critical for treatment of refractory TSC.

Our preliminary data and prior reports definitively demonstrate the existence of genetic modifiers of both neurological and tumor outcomes in haploinsufficiency mouse models of TSC, providing a tractable experimental platform for genetic discovery. Here, we propose to use natural genetic variation in a population-based model of TSC to identify additional molecular pathways that drive variation in TSC outcomes.

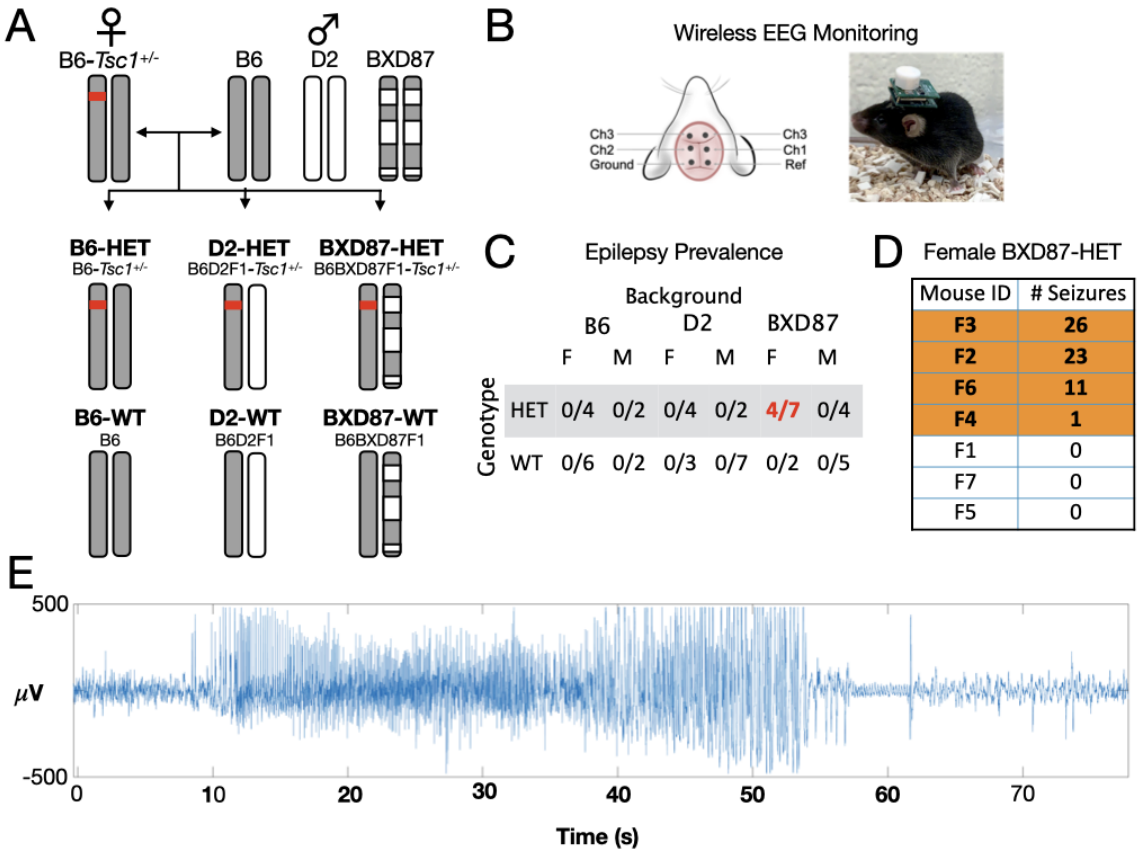


Figure 2: **Previous experiments demonstrated that genetic background modifies seizure outcomes in a haploinsufficiency mouse model of TSC.** A) Breeding schematic describing generation of *Tsc1*$^{+/-}$ mice on three different genetic backgrounds (B6, D2, and BXD87). B) Diagram and image of wireless EEG setup for seizure monitoring. C) Of all mice tested only female BXD87-HET- mice experienced seizures. D) Seizure counts for individual female BXD87-HET mice. E) Example EEG trace showing electrographic evidence of a seizure in a BXD87-HET mouse. All seizures were confirmed with concurrent video.

*Genetic modifiers of seizure outcomes exist in a haploinsufficiency mouse model of TSC*. In **Aim 1** we will genetically map modifiers of neurological outcomes in a mouse population model of *Tsc1/2* haploinsufficiency. Our recently completed study (DOD TS180087) tested whether epilepsy outcomes in *Tsc1*$^{+/-}$ mice were influenced by genetic background (Fig. 2). We crossed a C57BL/6J-*Tsc1*$^{+/-}$ mouse to multiple wild-type genetic backgrounds–C57BL/6J (B6), DBA/2J (D2), and BXD87/Rww (BXD87)–allowing us to test the relationship between strain background and disease outcomes (Fig. 2A). The BXD87 strain is a recombinant inbred strain whose genome is a mosaic of the B6 and D2 genomes (Fig. 2A), which we chose because wild-type BXD87 mice are outliers for both low hippocampal volume and poor cognitive performance [24–26], suggesting a developmentally susceptible state to poor TSC outcomes. Because the carrier parent was heterozygous for the *Tsc1* knockout, the F1 progeny could inherit either one knockout allele (HET) or the B6 wild-type allele (WT) (Fig. 2A). We performed five-day video-electroencephalogram (EEG) monitoring (starting around postnatal day 70) in N = 48 mice (Fig. 2B). We found that only BXD87-HET mice developed spontaneous, tonic-clonic seizures when carrying a single copy of the *Tsc1* null allele. Over the course of our experiment 4 out of 7 BXD87 females had a total of 61 seizures (Fig. 2C-D), averaging 1.7 seizures per BXD87 female mouse per day. This is the first time that a mouse model has been observed to develop chronic epilepsy without an induced second hit to *Tsc1/2*, demonstrating that a single germline

mutation of *Tsc1* can cause chronic seizures in mice in a manner that is modified by genetic background. We note that haploinsufficiency models of spontaneous seizures have been reported before; however, in these models early-life epilepsy resolved by post-natal day 19 [27], well before adulthood, and hence does not recapitulate the chronic epilepsy of the most severely affected human patients. We can make two important observations about the genetics of epilepsy in our BXD87-HET model. First, the appearance of seizures in BXD87, but not the parent strains B6 and D2, implies that there must be at least two genetic modifiers in BXD87 responsible for epilepsy; If there were only one, we would have observed seizures in either B6-HET or D2-HET mice. Second, we observed seizures only in around half of the female BXD87-HET mice. Based on our cross design with a B6-*Tsc1*$^{+/-}$ dam (Fig. 2A), we infer that one of the modifier alleles may be on the X chromosome and that incomplete penetrance may be due to random X inactivation [28, 29]. We are currently testing this hypothesis, which is out of the scope of this project. However, to maximize our chance of detecting an X-linked modifier in this study, we will reverse the cross so that the sire is the inbred carrier and the dam is the genetically diverse, non-carrier. In this design, progeny of all sexes will have genetic variation on the X chromosome.

Our data above demonstrate that the polygenic combination of variants inherited by BXD87 from the B6 and D2 strains cannot suppress TSC-associated epilepsy. Mapping these modifiers only three strains (B6, D2, and BXD87) is impossible due to insufficient genetic resolution. However, using existing data on the BXD mapping panel available in the Mouse Phenome Database [30], we performed a re-analysis of hippocampal volume and spatial memory performance phenotypes, for which BXD87 is an outlier strain (low hippocampal volume, poor spatial memory performance) [24]. Briefly, we performed SNP association mapping for each trait individually and combined them using a Bayesian mixture model [31] to identify SNPs explaining BXD87's outlier status (Fig. 3). The top-ranked SNP overall was on chromosome 4 and linked to the gene *Gabbr2*. The human ortholog, *GABBR2*, has recently been implicated as a significant epilepsy modifier gene in TSC (Farach et al., abstract to the International TSC Research Conference 2021). The unlinked locus with the next most significant SNP was on the X chromosome. The top SNP in this locus is linked to the gene *Ubqln2*, which encodes the protein Ubiquilin 2. Significantly, human Ubiquilin 2 has been shown to be an activator of mTORC1 signaling [32, 33]. This analysis strongly suggests that: 1) genetic modifiers of baseline phenotypes in mice can predispose strains to dramatically different TSC outcomes, and 2) mouse modifier genes may translate directly to human TSC. While this analysis is only suggestive, we will be able to validate the locations of TSC modifier alleles and potential specific roles for *Gabbr2* and *Ubqln2* through gene-expression and genetic mapping studies in this proposal.

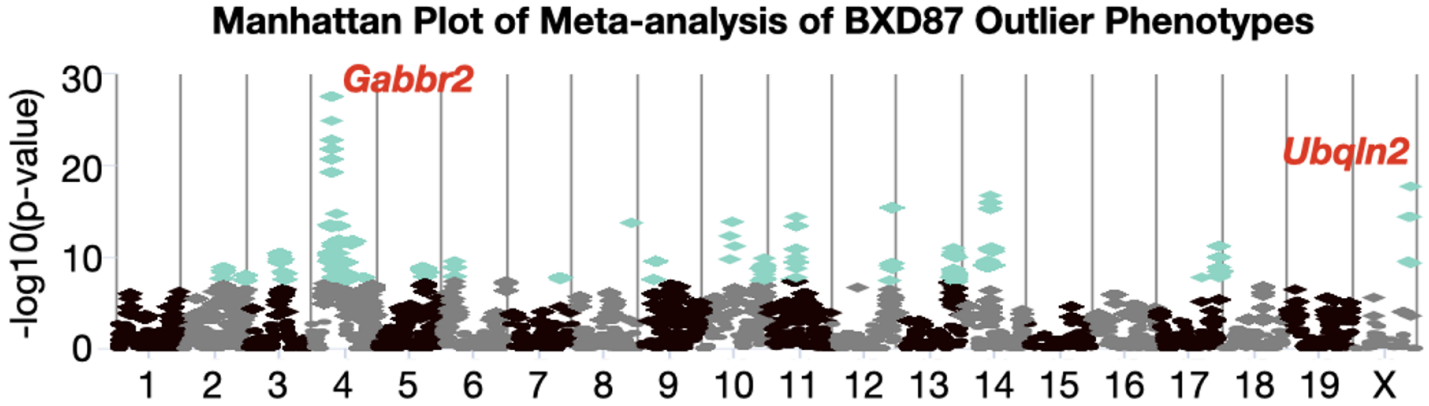


Figure 3: **Meta-analysis of outlier phenotypes in BXD87 mice** Using the Mouse Phenome Database, we re-analyzed neurological traits (hippocampal volume and radial arm maze performance) for which BXD87 is an outlier among all BXD strains. The top two loci were on chr 4 and chr X, and the top SNPs were linked to *Gabbr2* and *Ubqln2*, respectively. Both genes have a strongly plausible role in TSC-associated epilepsy.

*Genetic modifiers of tumor outcomes exist in haploinsufficiency mouse models of TSC.* In **Aim 2** we will genetically map modifiers of tumor outcomes in a population model of *Tsc1/2* haploinsufficiency. Unlike epilepsy, there is substantial prior work in rodents showing that genetic variation is an important driver of tumor outcomes in

*Tsc1/2* haploinsufficiency [12]. In Eker rats, which carry a null allele of *Tsc2*, variation in kidney tumor size and number have been mapped to alleles on rat chromosomes 3 [34] and 5 [35]. In mice carrying a null *Tsc2* allele, tumor number, extent, and location vary by strain background. Specifically, A/J-*Tsc2*$^{+/-}$ mice had a substantially higher burden of kidney tumors than 129S4/SvJae-*Tsc2*$^{+/-}$ mice [36], while 129S4/SvJae-*Tsc2*$^{+/-}$ mice had a higher burden of liver tumors compared to six other strains tested [12]. Similarly, a null allele of *Tsc1* caused high incidence of renal carcinoma with some lung metastases in BALB/c mice compared to other strains, while the same allele caused early-life mortality in one quarter of B6 pups [37]. Together these studies demonstrate that tumor outcomes in haploinsufficiency models of TSC are highly heritable.

*Rationale for a DO-TSC population model.* The heritability of TSC outcomes in inbred strains proves the existence of modifier alleles, but does not identify them. Here we propose to map these alleles using the Diversity Outbred (DO) mouse population. The DO mice are a highly recombined, outbred population derived from eight inbred founder strains [38], which capture 89% of the known genetic variation in laboratory mice [39]. Three of the DO founders–A/J, B6, and 129–have been shown to vary in TSC-related clinical outcomes in the studies cited above. Additionally, while they are not founder strains per se, nearly all variants in BALB/c and BXD87, which also carry modifiers of TSC outcomes, are present in the DO population. Thus, by generating F1 hybrids between carriers of *Tsc1/2* LoF mutations with DO mice (Fig. 1A-B), we can model both the germline haploinsufficiency of many human patients and the effects of background genetic diversity. We will make two subpopulations, DO-*Tsc1* with a B6-*Tsc1*$^{+/-}$ sire (Fig. 1A) and DO-*Tsc2* with an A/J-*Tsc2*$^{+/-}$ sire (Fig. 1B). Given our results above, the DO-*Tsc1* population is nearly certain to have significant variation in epilepsy outcomes, with both susceptible and resistant mice. Likewise, given prior reports on tumors in A/J-*Tsc2*$^{+/-}$ mice, the DO-*Tsc2* population is nearly certain to have significant variation in tumor outcomes. Thus, these subpopulations separately have power to detect strong modifier genes. However, because poor neurological and tumor outcomes often occur in the same patient, we will phenotype both subpopulations for both categories of outcome and perform pooled analyses for universal modifiers of *Tsc1/2* LoF. By pairing clinical phenotypes with high-throughput molecular data, a strategy called systems genetics, we can further identify the regulatory effects of modifier alleles on genes and pathways that drive variation [40]. Systems genetics in the DO population is a widely adopted design [41–74], with highly mature software tools for genetic mapping [75] (Fig. 1E), gene network inference [76] (Fig. 1F), and gene expression mediation analysis [77](Fig. 1G). We have successfully used DO mice to map individual QTLs and infer genetic networks [78, 79] (Fig. 1E-F). Others have successfully mapped tumor outcomes using modifier screens similar to this proposal [80, 81]. We have also thoroughly studied the effects of population structure on mapping studies in this population [82]. This prior work combined with our strong preliminary data demonstrate that the DO-TSC population will be a singularly powerful genetic discovery resource for TSC.

*Artificial intelligence (AI) approaches for scalable, automated phenotyping.* One of the persistent problems for genetic mapping has been the requirement for high-throughput phenotyping at the scale of hundreds of individuals. This has traditionally meant either simple summaries of complex observations (e.g., case vs. control) or surrogate biomarkers that may not capture the complete phenotype. However, recent advances in AI and machine learning have dramatically lowered the barriers to subtle phenotyping at scale, including scoring days-long home-cage behavior videos [83–91], non-invasive seizure monitoring [92], and analyzing whole-slide images of tissues [93, 94] (Fig. 1C-D). We have pioneered applications of AI-based image analysis for quantitative trait genetics. In a study of the *Far2* gene, which alters kidney aging phenotypes [95], we demonstrated that AI approaches could recapitulate known kidney phenotypes and detect novel phenotypes using only genotype as a training label [96]. In an extension of this work, Dr. Mahoney has an active R01 on histological imaging genetics to further develop these tools. Taken together, this prior work and our feasibility data (described below) demonstrate that we can use AI models to robustly score seizure, behavioral, and histopathological phenotypes at scale, and genetically map AI outputs as biologically meaningful traits. Thus, our innovative AI-based phenotyping pipeline overcomes one of the primary technical limitations of genetic mapping.

### B. Hypothesis and Objective

The **objective** of this project is to identify novel genetic modifiers of TSC outcomes and identify the gene expression

signatures that mediate these effects. To achieve this, we will perform a systems genetics analysis of a novel DO-TSC population model. We hypothesize that: 1) DO-TSC mice will vary significantly in seizure, behavioral, and tumor outcomes;2) each outcome will have modifiers mapping to specific locations in the genome, and 3) we will identify significant modifier pathways using gene-expression mediation analysis.

**C. Specific Aims**

**Aim 1: Identify genetic modifiers of epilepsy and TANDs in TSC** by: a) mapping genetic modifiers of epilepsy and TANDs in our DO-TSC model and b) performing mediation analysis with brain gene expression to identify the causal pathways driving poor neurological outcomes.

**Aim 2: Identify genetic modifiers of tumors in TSC** by: a) mapping genetic modifiers of tumor outcomes in brain, kidney, and lung in our DO-TSC model and b) performing mediation analysis with kidney gene expression to identify the causal pathways driving poor tumor outcomes.

**D. Research Strategy and Feasibility**

**Aim 1: Identify genetic modifiers of epilepsy and TANDs in TSC**

**Experimental approach:**

*Generation of experimental mice:* We will generate two subpopulations of DO-TSC mice, which we designate as DO-*Tsc1* and DO-*Tsc2* (Fig. 1A-B). We will phenotype both sub-populations in both aims (Fig. 1C-D). Each mouse will be an F1 hybrid between a genetically unique female DO mouse and a male inbred carrier mouse. Each F1 hybrid will also be genetically unique, because of recombination of the DO dam's chromosomes during meiosis. For the DO-*Tsc1* subpopulation, we will use B6-*Tsc1*$^{+/-}$ mice as the carrier sire. We will generate these mice as in our preliminary data, using a cross between our B6-*Tsc1*fl/fl mouse line and the CMV-Cre line, which expresses Cre recombinase at the germline, causing germline heterozygosity of *Tsc1* LoF. For the DO-*Tsc2* subpopulation, we will use A/J-*Tsc2*$^{+/-}$ mice as the carrier sire. These mice are not commercially available, so the JAX Genetic Engineering Technologies (GET) group will generate this line *de novo* using CRISPR/Cas9 to knock out exons 5 through 8 in the A/J background, matching the construct used in prior reports [12, 36]. In both subpopulations, we will genotype all pups for the TSC mutation before weaning using standard PCR and retain pups with a *Tsc1/2* mutation for phenotyping. We will also maintain detailed records of litter sizes and *Tsc1/2* mutation frequency as an additional phenotype. TSC mutations can cause fetal death in humans [97–99], and prenatal death would result in a significant deviation from the expected 50% frequency of *Tsc1/2* mutations among pups.

*Seizure and behavior monitoring.* For high-throughput seizure and behavioral phenotyping, we will use a novel home-cage behavior system developed by TLR Ventures. This home-cage behavior system was developed in collaboration with the Digital In Vivo Alliance (DIVA), a consortium that includes JAX and supports the development of automated home-cage technologies. DIVA cages have two cameras, one overhead and another at a side angle, that stream video directly to cloud-based computing centers. Once data are in the cloud, DIVA cage software automatically applies AI models to identify animal position (segmentation), identify key points on the body (key point detection), predict respiratory rate, and identify loss of righting reflex (LORR) during seizures. These data are stored for further processing in a cloud-based data science environment, and user-interface tools allow researchers to review videos, define segments of interest, and train models *de novo*. Starting around postnatal day 70, we will put up to three littermate DO-TSC mice into the DIVA cages for three consecutive days of continuous monitoring to measure the frequency and duration of seizures, as primary outcomes. We will also measure frequencies and durations of abnormal behaviors, such as backflipping, abnormal grooming, or other stereotypies that have been associated with ASD, as secondary outcomes. We will manually review periods of high predicted probability of LORR to confirm seizures and fine-tune the seizure detection model using manual annotations. Similarly, we will apply and fine-tune existing models to detect periods of highly stereotyped behavior, abnormal grooming activity, and identify backflipping [83–91].

*Tissue collection:* At 14 months of age, we will collect brain, kidney, lung, liver, and heart tissue from each

DO-TSC mouse. All tissues will be fixed with formalin and embedded in paraffin (FFPE) for long-term storage. We will separate brains along the corpus callosum and use one hemisphere for RNA sequencing and the other for histology.

*RNA sequencing:* To identify the gene expression signatures that mediate the effects of genetic modifiers on epilepsy and TANDs, we will perform bulk gene expression of brain tissue. Gene expression analysis of all mouse brains from this study is beyond the scope of this project. Thus, we will focus on DO-*Tsc1*, which has a high probability of manifesting heterogeneous neurological outcomes. We will deliver DO-*Tsc1* brain samples to the JAX Genome Technologies Service for 50 base pair, paired-end bulk RNA sequencing. Fastq files will be delivered to our group for analysis. We will perform read alignment and gene expression quantification using custom, publicly available RNA-seq pipelines developed by JAX Computational Sciences [100]. This pipeline performs read trimming to remove sequencing barcodes, quantifies gene expression using the RNA-Seq by Expectation Maximization (RSEM) algorithm [101], and delivers sample- and batch-level summary and quality-control statistics, including Picard alignment metrics [102] and a MultiQC [103] report. Gene expression levels will be normalized using a variance-stabilizing transformation [104, 105] for downstream processing.

*Genotyping and reconstruction of ancestral haplotypes:* For background genotyping, we will use miniMUGA SNP genotyping arrays [106] to generate SNP-level genotypes for each mouse. We will then compute the DO founder haplotypes using HaploQA [107]. We will use haplotypes for genetic mapping [108].

*Genetic mapping and power analyis:* We will perform two complementary forms of genetic mapping–linkage mapping and combined analysis of pleiotropy and epistasis (CAPE)–each of which has unique strengths. In linkage mapping, we correlate the haplotypes at individual genetic markers with variation in quantitative traits, such as seizure frequency. Linkage mapping seeks loci with strong marginal effects, i.e., strong effects across many genetic backgrounds. We will perform linkage mapping using R/qtl2 [75] and define statistical significance for QTLs using standard permutation-based statistical thresholds [109]. We will use sex and TSC-genotype as interacting covariates to identify potential sex- and genotype-specific modifiers. We will also use linear mixed models to account for the effects of variable relatedness among individuals [110]. Because the DO mice are highly recombined, we expect that mapping in this population will yield QTLs encompassing small genomic regions ($<1$ megabase (Mb)) with a small number of positional candidate genes. With $n = 400$ mice total, we will have 80% power to detect loci explaining 10% of the variance in any individual phenotype [108]. This population size conforms to prior successful mapping studies in DO-F1 hybrid mice [80]. Because linkage mapping is the most statistically stringent form of mapping, we have powered our study for this analysis. Using gene expression from RNA sequencing, we will also perform expression QTL (eQTL) analysis to identify genetic loci regulating gene expression levels. We will map eQTLs exactly as described above, once for each transcript. The effect sizes for eQTLs are generally much more significant than QTLs for clinical traits due to strong cis-regualtory effects. Thus, with $n = 200$ mice with gene expression, we expect large numbers of eQTLs. By identifying loci that influence both gene expression and traits, we can more easily link trait variation to molecular mechanisms [111]. In addition to mapping to individual QTLs, we will use CAPE, developed by Dr. Tyler [76], to map trait variation to epistatically interacting QTLs (genetic network inference). No gene acts in isolation, and many genetic modifiers of TSC may depend on interactions with the rest of the genome. CAPE leverages information across multiple traits to infer directed epistatic interactions affecting all traits simultaneously. The simultaneous analysis of multiple traits increases power to detect QTLs over classical linkage mapping [78]. Moreover, CAPE allows inference of directed interactions indicating which variant acts upstream of another variant in an interacting pair, thereby aiding in interpretation of the interaction and prioritization of causal positional candidate genes [112, 113]. Furthermore, the interactions identified are consistent across all analyzed traits, allowing us to relate clinical traits to genetic interactions through endophenotypes, such as the expression of genes in the mTOR signaling pathway, which we will include as traits in the analysis [78, 112, 114]. As with linkage mapping, we will use sex and TSC-genotype as interactive covariates to ascertain sex- and genotype-specific effects. We will also use linear mixed models to account for interindividual relatedness [82] and perform permutation testing to assess interaction significance [109]. We have experience using CAPE to analyze seizure outcomes in mouse models of absence epilepsy [115], and

complex traits in a DO population [78] comparable to the one proposed here. We expect that in this population we will be able to generate multiple testable molecular hypotheses about genetic variants that influence multiple clinical outcomes simultaneously.

*Transcriptomic mediation analysis:* One limitation to all forms of genetic mapping is that the effects of individual loci must be large enough to detect statistically. However, there is now extensive evidence that most complex traits are highly polygenic, and this polygenic effect is mediated through changes in transcriptional profiles [116, 117]. Thus, as a complement to genetic mapping per se, we will also use gene expression to identify the transcriptomic profiles driving heritable variation in neurological outcomes, independent of individual genomic loci. To do this, we will perform high-dimensional mediation analysis [118], which is an extension of canonical correlation analysis that, in our case, seeks a transcriptomic signature that is simultaneously correlated with underlying genetics and phenotypic outcomes. Formally, if G, T, and P, represent our genotype, transcriptome, and phenotype data matrices, we define composite genetic, transcriptomic, and phenotypic scores, $s_g = G \cdot l_g$, $s_t = T \cdot l_t$, and $s_p = P \cdot l_p$, respectively, through matrix multiplication of the data matrices with loading vectors lg, lt, and lp that define each variable's importance to the corresponding score. Using the theory of causal graphical models [119, 120], we maximize the likelihood function $L = -\log|S| - \mathrm{tr}(\Sigma S^{-1})$, where $\Sigma$ is the correlation matrix of the scores ($s_g$, $s_t$, and $s_p$), $S$ is the model-implied covariance matrix for the perfect mediation model, and $|.|$ and "tr" denote the determinant and trace of a matrix, respectively [121]. This model encodes the causal flow of information from genotype, through transcription, to phenotype. The outputs of this analysis are the optimal scores and loadings that define the causal influence of genetic background on TSC outcomes. We will rigorously define the significance of mediation by permuting the transcriptomes relative to the genotypes and phenotypes to generate a null distribution for the path coefficient from genotype to phenotype. To identify the significant pathways in the transcriptomic score, we will rank all individual transcripts by their loadings, lt, and perform gene set enrichment analysis (GSEA) [122]. In this way, the transcript rankings are a powerful complement to genetic mapping because they allow inference of the causal importance of genes and pathways, even when background modifier effects are highly polygenic.

**Feasibility:** As feasibility data for this proposal, we trained an AI model *de novo* to detect LORR in manually coded frames from videos of 25 mice during pentylenetetrazole-induced seizures (Fig. 4A). The model achieved greater than 95% accuracy detecting LORR in testing data (Fig. 4B). As an example readout, the model computes a time-varying probability (Fig. 4C, blue trace) that predicts when the mouse has fallen vs. is upright. Time intervals with high predicted probability were strongly correlated with expert annotations of LORR (Fig. 4C, pink boxes). These data demonstrate that the DIVA cages can robustly detect and quantify seizures at scale. The DIVA cages also compute animal segmentations and pose-estimation phenotypes from which we can score standard home cage behaviors, such as grooming.

**Potential problems and alternative strategies:** We anticipate three potential issues. First, we may not detect any QTLs at the genome-wide significance level using any of the proposed strategies, thereby limiting candidate gene prioritization from genetic mapping. While we stress that our combined approach is designed to mitigate this possibility, if it occurs, we anticipate that it will be due to both the presence of a large number of modifiers and limited statistical power to detect effects. If this transpires, we will use false discovery rate (FDR) approaches [123] to lower the burden of multiple testing for our associations. We also note that high-dimensional mediation analysis is an explicitly polygenic analysis that supports gene prioritization independent of QTLs. Second, we may not have sufficient genetic resolution to nominate candidate genes by mapping alone. While current DO generations have linkage blocks that are substantially below 1 Mb, it is possible that our analyses may not resolve individual candidate genes. If this transpires, we will pursue candidate gene prioritization within QTLs using network-based approaches that Drs. Mahoney and Tyler have extensively developed [124, 125]. Third, we may find that DO-TSC mice have heterogeneous seizure presentations that are difficult to track with our existing AI tools. Because we expect all animals having tonic-clonic seizures to have LORR, we anticipate that thresholding the AI-generated probabilities at a low level, e.g. 20%, will guide us to epochs that are enriched for seizure activity (see Fig. 4). If we identify misclassified epochs, we will first attempt to fine-tune the LORR model by manually

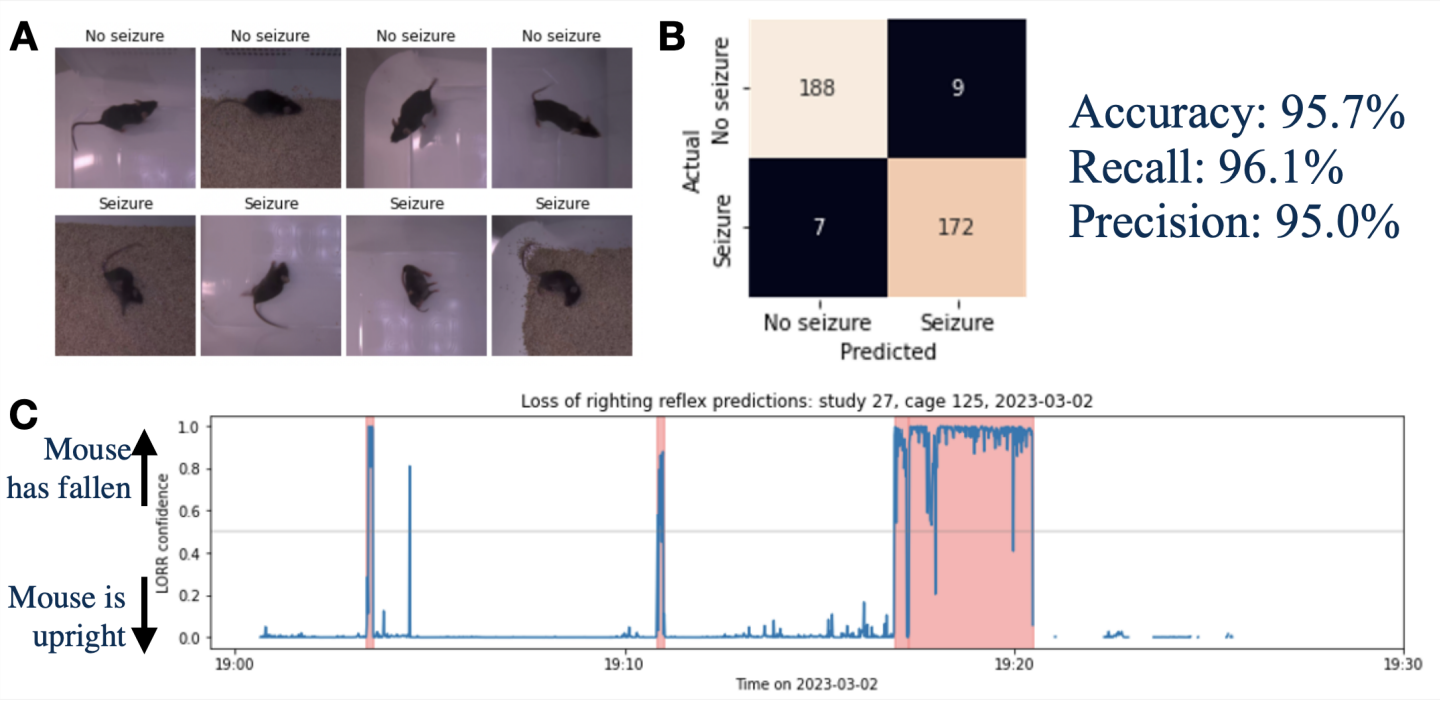


Figure 4: **AI classification of seizures.** A) Still images of normal behavior (top row) and mice experiencing seizures (bottom row). The bottom row shows the loss-of-righting reflex (LORR) that is classified by the AI model as a seizure. B) Confusion matrix showing the counts of correct and incorrect classifications by the AI along with calculated accuracy, recall, and precision of the model. C) A time-varying probability of LORR predicted by the AI model at each frame of a video (blue trace). High probabilities indicate frames for which the AI predicted that the mouse had fallen over. Pink areas highlight frames for which an expert reviewer indicated that the mouse had fallen over.

annotating a subset of these epochs as ground-truth labels. Barring that, following Geschwind et al. [92], we will perform unsupervised analyses that cluster animal behaviors into groups, independent of annotations, and manually annotate clusters corresponding to seizures and other abnormal behaviors. In the worst case, we will review all videos in their entirety, visually screening for seizures.

**Aim 2: Identify genetic modifiers of tumors in TSC.**

**Experimental approach:** For this aim we will use the same 400 DO-TSC mice developed in **Aim 1**. We will perform tissue pre-processing, genotyping, genetic mapping, and mediation analysis exactly as described above.

*Histological imaging:* We will deliver FFPE brain, kidney, and lung tissue blocks to the JAX Histology Service for sectioning and staining with hematoxylin and eosin (H&E). Stained slides will then be delivered to the JAX Microscopy Service for slide scanning with a 40X objective. Whole-slide image files will be delivered to our team for analysis. At the conclusion of this study, we will make all histological images available to the public.

*Image processing and genetic mapping:* To generate histological phenotypes at the scale required for genetic mapping, we will use an AI-based computer vision approach to automatically score the extent of tissue abnormalities without requiring extensive human annotation. From each tissue image, we will extract tiles, or sub-images, of 299x299 pixels, which we will input into the Inception v3 deep neural network [126]. The Inception v3 model is a top-performing network for natural image classification [127]. We will perform transfer learning [128], where we remove the last layers of the model and retain the numerical features at the second-to-last layer as a black-box feature vector encoding histological image structure. This transfer learning approach has been applied heavily in the cancer field with remarkable success [129, 130]. From the black-box feature vector, we will perform principal components analysis (PCA) to identify the dominant axes of variation across all images of the same tissue. For

each sample, we will average the PCA scores over tiles to compute mouse-level scores, which we will map as quantitative traits, using the techniques described in **Aim 1**. We will visualize the learned tissue features from our model using example montages, where we compare high- and low-scoring tiles to detect differences.

*RNA sequencing and analysis:* To identify the gene expression signatures that mediate the effects of genetic modifiers on tumors, we will perform bulk gene expression of kidney tissue. Gene expression analysis of all mouse tissues from this study is beyond the scope of this project. Thus, we will focus on DO-*Tsc2* kidneys, which have a high probability of manifesting heterogeneous tumor outcomes.

*Tissue biobanking:* We will biobank the brain, lung, and kidney tissue not used for histology or RNA sequencing, as well as all liver and heart samples, from each DO-TSC mouse. We expect our results from this project to provide a strong rationale to generate comprehensive histological and gene expression data for these remaining samples. We will also make the biobanked tissues available to the wider TSC research community upon reasonable request.

**Feasibility:** In ongoing work for Dr. Mahoney's R01, we have analyzed kidney aging in a population of 500 DO mice using the strategy described above (in preparation; Fig. 5). Using whole-slide images of periodic acid-Schiff-stained kidney sections, we extracted features from ~150k individual glomeruli and performed PCA to automatically identify age-related tissue features across the population. We genetically mapped these AI-derived scores and identified a significant QTL on chr. 17 (Fig. 5A). Examples of high- and low-scoring glomeruli show marked differences in mesangial matrix expansion, an age-associated histopathology (Fig. 5B), and cellularity, both of which we have validated quantitatively (data not shown). The allele effects for this AI-derived trait show that inheriting the A/J allele leads to worse outcomes while the NZO/LtJ allele leads to better outcomes (Fig. 5C). We have nominated high-confidence candidate genes within this locus–Xdh and Birc6–which have variants specific to the A/J and NZO/LtJ backgrounds, respectively, and strong evidence for association to kidney phenotypes [131, 132]. We have ongoing experiments with CRISPR knock-in mice to validate these effects. While this study does not speak to the hypotheses of this project, it demonstrates feasibility for mapping genetic modifiers of histological outcomes using AI-derived traits.

**Potential problems and alternative strategies:** We anticipate two potential problems. First, bulk RNA sequencing of kidney tissues necessarily mixes healthy tissue and tumor tissue, possibly limiting our ability to detect causally important transcriptomic changes relative to "passenger" changes. We can detect this by performing mediation analysis in reverse, with phenotypes as the mediator and gene expression as the outcome. Individual transcripts that are highly ranked in both models (i.e., with transcriptome as a mediator and an outcome) will be considered confounded. If this occurs, we will cross-reference our kidney expression to existing eQTL data from DO kidneys in mice without a TSC insult [42]. This will give added confidence in specific, genetically encoded transcript changes that cannot be due to tumor-composition effects. These data will also provide a strong rationale for follow-up proposals to generate single-cell and spatial transcriptomic atlases from our biobanked samples, from which we can deconvolve our bulk expression computationally [133]. A second potential problem is that we may fail to detect overt lesions using an unsupervised PCA approach. While we consider this extremely unlikely, if this occurs, we anticipate that it will be due to the heterogeneity of normal tissue structure in the DO obscuring the effects of relatively rarer lesional tissue. If our PCA montages (see Fig. 5B) do not contrast normal tissue with lesional tissue, then we will consult with JAX's on-site veterinary pathologist to both annotate the learned, non-lesional features and provide annotations of lesions that were missed by our analysis. With ground-truth labels of novel features, we can train models *de novo* to detect such features directly, which we will use as an additional phenotype. If lesions are extremely rare, we will apply outlier detection methodologies [134] to identify image tiles that exist far outside the distribution of healthy tissue. If all else fails, we will review each whole-slide image in its entirety, manually annotating abnormalities. These images will become public upon project completion for researchers in the wider community to apply their expertise and define novel phenotypes.

**Overall timeline**

Breeding of the DO-*Tsc1*$^{+/-}$ mice will begin immediately and the first round of experimental mice will be available around three months. Mice will be bred to generate waves of approximately 50 mice. Home-cage

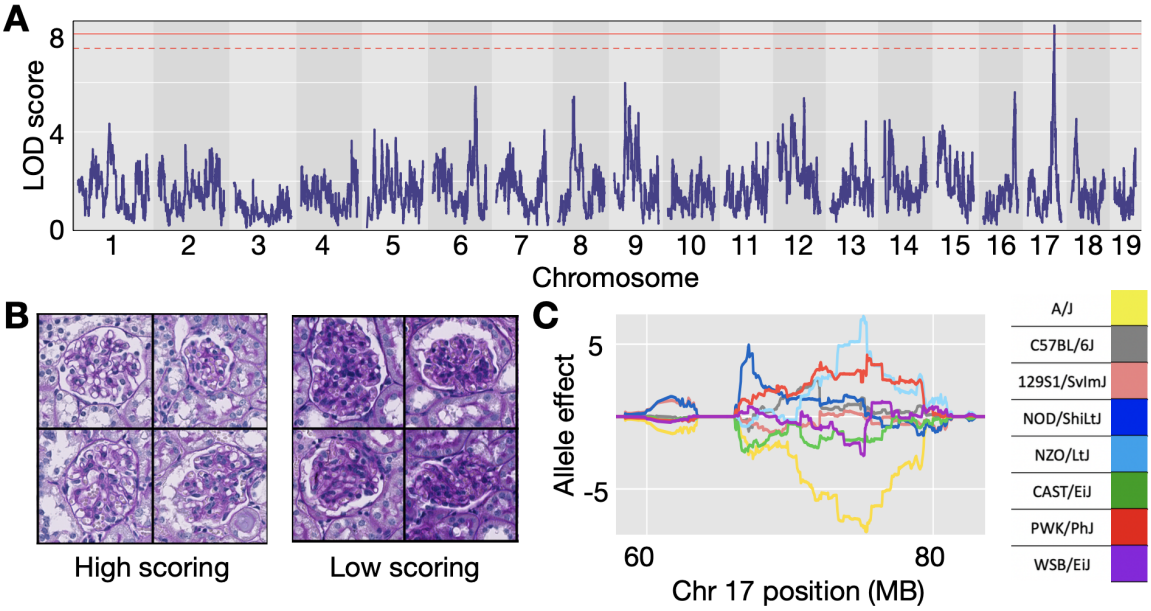


Figure 5: **Previous study demonstrating feasibility of mapping novel histological phenotypes generated by AI models.** A) QTL map for an age-related, AI-generated histological feature. This feature had a genome-wide significant QTL on mouse chromosome 17. The solid and dashed red lines indicate genome-wide significance of 1% and 5% respectively. B) Montage of images comparing glomeruli that scored high and low for this mapped feature. High- and low-scoring glomeruli differed in mesangial matrix expansion (darker purple staining on the right) and nucleus size (larger dark blue nuclei on the right). C) Founder allele effects at the QTL for this feature. Each founder is represented by one color. NZO alleles (light blue) were associated with better outcomes than A/J alleles (yellow).

behavior monitoring will occur continuously as new experimental waves arrive. After home-cage monitoring monitoring, mice will be housed until 14 months of age (17 months into the project), at which time they will sacrificed for tissue harvest. In parallel with the DO-*Tsc1*$^{+/-}$ experiment, we will generate the A/J-*Tsc2*$^{+/-}$ model using CRISPR. The full CRISPR pipeline, from generating the first knockouts and backcrossing to remove off-target effects, will take approximately 9 months. We will then begin breeding of the DO-*Tsc2*$^{+/-}$ mice, and experimental mice will be available by the end of Year 1. Thus, the DO-*Tsc2*$^{+/-}$ experiment will lag behind the DO-*Tsc1*$^{+/-}$ and be complete at around 29 months into the project. We will perform seizure and behavior analyses continuously throughout the study, monitoring model performance and fine-tuning as necessary. When all tissues from both subpopulations are collected, we will generate histological data and perform RNA sequencing in a single batch (around month 30). In the final 6 months, we will perform histological image analysis and genetic mapping, draft manuscripts, and prepare follow-up grant proposals.

**Integration of aims and project growth:** The aims of this project are conceptually and operationally distinct, but form an integrated whole. If both are successful, we will be positioned to perform a joint analysis of neurological and tumor outcomes. Indeed, we anticipate that many genetic modifiers will be pleiotropic across both classes of outcomes [78]. By mapping pleiotropic modifiers, we may identify universal disease-modifying targets. Going forward, we believe that the DO-TSC model from this Idea Development proposal will provide a platform for multiple, critical follow-up research programs, including an extended search for disease modifiers, pharmacogenetic studies of resistance to therapies (*e.g.*, rapalogs), and pre-clinical drug screening in genetically diverse animals, minimizing the probability of false-positive hits. Thus, by developing the DO-TSC population model, we will enable a significant new paradigm for model systems studies of TSC.

## REFERENCES CITED


[1] P. B. Crino, K. L. Nathanson, and E. P. Henske. The tuberous sclerosis complex. *N Engl J Med*, 355(13):1345–1356, Sep 2006.

[2] M.R. Gomez, J.R. Sampson, and V.H. Whittemore. *Tuberous Sclerosis Complex*. Developmental Perspectives in Psychiatry. Oxford University Press, 1999.

[3] H. Northrup, J. W. Wheless, T. K. Bertin, and R. A. Lewis. Variability of expression in tuberous sclerosis. *J Med Genet*, 30(1):41–43, Jan 1993.

[4] G. Matsuzaki, T. Lin, and K. Nomoto. Differentiation and function of intestinal intraepithelial lymphocytes. *Int Rev Immunol*, 11(1):47–60, 1994.

[5] T. Gilboa, R. Segel, S. Zeligson, G. Alterescu, and H. Ben-Pazi. Ganglioglioma, Epilepsy, and Intellectual Impairment due to Familial TSC1 Deletion. *J Child Neurol*, 33(7):482–486, Jun 2018.

[6] J. Fox, S. Ben-Shachar, S. Uliel, R. Svirsky, H. Saitsu, N. Matsumoto, and A. Fattal-Valevski. Rare familial TSC2 gene mutation associated with atypical phenotype presentation of Tuberous Sclerosis Complex. *Am J Med Genet A*, 173(3):744–748, Mar 2017.

[7] R. C. Caylor, L. Grote, I. Thiffault, E. G. Farrow, L. Willig, S. Soden, S. M. Amudhavalli, A. J. Nopper, K. A. Horii, E. Fleming, J. Jenkins, H. Welsh, M. Ilyas, K. Engleman, A. Abdelmoity, and C. J. Saunders. Incidental diagnosis of tuberous sclerosis complex by exome sequencing in three families with subclinical findings. *Neurogenetics*, 19(3):205–213, Aug 2018.

[8] D. A. Lyczkowski, K. D. Conant, M. B. Pulsifer, D. Y. Jarrett, P. E. Grant, D. J. Kwiatkowski, and E. A. Thiele. Intrafamilial phenotypic variability in tuberous sclerosis complex. *J Child Neurol*, 22(12):1348–1355, Dec 2007.

[9] M. Li, Y. Zhou, C. Chen, T. Yang, S. Zhou, S. Chen, Y. Wu, and Y. Cui. Efficacy and safety of mTOR inhibitors (rapamycin and its analogues) for tuberous sclerosis complex: a meta-analysis. *Orphanet J Rare Dis*, 14(1):39, Feb 2019.

[10] K. F. Xu, X. Tian, Y. Yang, and H. Zhang. Rapamycin for lymphangioleiomyomatosis: optimal timing and optimal dosage. *Thorax*, 73(4):308–310, Apr 2018.

[11] L. Lee, P. Sudentas, B. Donohue, K. Asrican, A. Worku, V. Walker, Y. Sun, K. Schmidt, M. S. Albert, N. El-Hashemite, A. S. Lader, H. Onda, H. Zhang, D. J. Kwiatkowski, and S. L. Dabora. Efficacy of a rapamycin analog (CCI-779) and IFN-gamma in tuberous sclerosis mouse models. *Genes Chromosomes Cancer*, 42(3):213–227, Mar 2005.

[12] D. J. Kwiatkowski. Animal models of lymphangioleiomyomatosis (LAM) and tuberous sclerosis complex (TSC). *Lymphat Res Biol*, 8(1):51–57, Mar 2010.

[13] H. Kenerson, T. A. Dundon, and R. S. Yeung. Effects of rapamycin in the Eker rat model of tuberous sclerosis complex. *Pediatr Res*, 57(1):67–75, Jan 2005.

[14] C. Li, E. Zhang, Y. Sun, P. S. Lee, Y. Zhan, Y. Guo, J. C. Osorio, I. O. Rosas, K. F. Xu, D. J. Kwiatkowski, and J. J. Yu. Rapamycin-insensitive up-regulation of adipocyte phospholipase A2 in tuberous sclerosis and lymphangioleiomyomatosis. *PLoS One*, 9(10):e104809, 2014.

[15] C. Li, P. S. Lee, Y. Sun, X. Gu, E. Zhang, Y. Guo, C. L. Wu, N. Auricchio, C. Priolo, J. Li, A. Csibi, A. Parkhitko, T. Morrison, A. Planaguma, S. Kazani, E. Israel, K. F. Xu, E. P. Henske, J. Blenis, B. D. Levy, D. Kwiatkowski, and J. J. Yu. Estradiol and mTORC2 cooperate to enhance prostaglandin biosynthesis and tumorigenesis in TSC2-deficient LAM cells. *J Exp Med*, 211(1):15–28, Jan 2014.

[16] P. Dalle Pezze, A. G. Sonntag, A. Thien, M. T. Prentzell, M. del, S. Fischer, E. Neumann-Haefelin, T. B. Huber, R. Baumeister, D. P. Shanley, and K. Thedieck. A dynamic network model of mTOR signaling reveals TSC-independent mTORC2 regulation. *Sci Signal*, 5(217):ra25, Mar 2012.

[17] M. Canpolat, H. Gumus, S. Kumandas, A. Coskun, and H. Per. The use of rapamycin in patients with tuberous sclerosis complex: Long-term results. *Epilepsy Behav*, 88:357–364, Nov 2018.

[18] Y. Tang, D. J. Kwiatkowski, and E. P. Henske. Midkine expression by stem-like tumor cells drives persistence to mTOR inhibition and an immune-suppressive microenvironment. *Nat Commun*, 13(1):5018, Aug 2022.

[19] A. M. Taveira-DaSilva and J. Moss. Optimizing treatments for lymphangioleiomyomatosis. *Expert Rev Respir Med*, 6(3):267–276, Jun 2012.

[20] L. Lee, P. Sudentas, and S. L. Dabora. Combination of a rapamycin analog (CCI-779) and interferon-gamma is more effective than single agents in treating a mouse model of tuberous sclerosis complex. *Genes Chromosomes Cancer*, 45(10):933–944, Oct 2006.

[21] M. Karbowniczek, T. Cash, M. Cheung, G. P. Robertson, A. Astrinidis, and E. P. Henske. Regulation of B-Raf kinase activity by tuberin and Rheb is mammalian target of rapamycin (mTOR)-independent. *J Biol Chem*, 279(29):29930–29937, Jul 2004.

[22] M. Karbowniczek, G. P. Robertson, and E. P. Henske. Rheb inhibits C-raf activity and B-raf/C-raf heterodimerization. *J Biol Chem*, 281(35):25447–25456, Sep 2006.

[23] Y. Liu, K. Marks, G. S. Cowley, J. Carretero, Q. Liu, T. J. Nieland, C. Xu, T. J. Cohoon, P. Gao, Y. Zhang, Z. Chen, A. B. Altabef, J. H. Tchaicha, X. Wang, S. Choe, E. M. Driggers, J. Zhang, S. T. Bailey, N. E. Sharpless, D. N. Hayes, N. M. Patel, P. A. Janne, N. Bardeesy, J. A. Engelman, B. D. Manning, R. J. Shaw, J. M. Asara, R. Scully, A. Kimmelman, L. A. Byers, D. L. Gibbons, I. I. Wistuba, J. V. Heymach, D. J. Kwiatkowski, W. Y. Kim, A. L. Kung, N. S. Gray, D. E. Root, L. C. Cantley, and K. K. Wong. Metabolic and functional genomic studies identify deoxythymidylate kinase as a target in LKB1-mutant lung cancer. *Cancer Discov*, 3(8):870–879, Aug 2013.

[24] A. Delprato, B. Bonheur, M. P. o, P. Rosay, L. Lu, R. W. Williams, and W. E. Crusio. Systems genetic analysis of hippocampal neuroanatomy and spatial learning in mice. *Genes Brain Behav*, 14(8):591–606, Nov 2015.

[25] A. Delprato, M. P. o, B. Bonheur, J. A. Bubier, L. Lu, R. W. Williams, E. J. Chesler, and W. E. Crusio. QTL and systems genetics analysis of mouse grooming and behavioral responses to novelty in an open field. *Genes Brain Behav*, 16(8):790–799, Nov 2017.

[26] A. Delprato, B. Bonheur, M. P. o, A. Murillo, E. Dhawan, L. Lu, R. W. Williams, and W. E. Crusio. A quantitative trait locus on chromosome 1 modulates intermale aggression in mice. *Genes Brain Behav*, 17(7):e12469, Sep 2018.

[27] N. Lozovaya, S. Gataullina, T. Tsintsadze, V. Tsintsadze, E. Pallesi-Pocachard, M. Minlebaev, N. A. Goriounova, E. Buhler, F. Watrin, S. Shityakov, A. J. Becker, A. Bordey, M. Milh, D. Scavarda, C. Bulteau, G. Dorfmuller, O. Delalande, A. Represa, C. Cardoso, O. Dulac, Y. Ben-Ari, and N. Burnashev. Selective suppression of excessive GluN2C expression rescues early epilepsy in a tuberous sclerosis murine model. *Nat Commun*, 5:4563, Aug 2014.

[28] K. D. Huynh and J. T. Lee. X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. *Nat Rev Genet*, 6(5):410–418, May 2005.

[29] T. Goto and M. Monk. Regulation of X-chromosome inactivation in development in mice and humans. *Microbiol Mol Biol Rev*, 62(2):362–378, Jun 1998.

[30] M. A. Bogue, V. M. Philip, D. O. Walton, S. C. Grubb, M. H. Dunn, G. Kolishovski, J. Emerson, G. Mukherjee, T. Stearns, H. He, V. Sinha, B. Kadakkuzha, G. Kunde-Ramamoorthy, and E. J. Chesler. Mouse Phenome Database: a data repository and analysis suite for curated primary mouse phenotype data. *Nucleic Acids Res*, 48(D1):D716–D723, Jan 2020.

[31] B. Han and E. Eskin. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet*, 8(3):e1002555, 2012.

[32] R. T. Coffey, Y. Shi, M. J. Long, M. T. Marr, and L. Hedstrom. Ubiquilin-mediated Small Molecule Inhibition of Mammalian Target of Rapamycin Complex 1 (mTORC1) Signaling. *J Biol Chem*, 291(10):5221–5233, Mar 2016.

[33] M. rk, G. Lin, Z. Zuo, D. Mao, E. Watson, A. G. Mikos, and H. J. Bellen. Ubiquilins regulate autophagic flux through mTOR signalling and lysosomal acidification. *Nat Cell Biol*, 21(3):384–396, Mar 2019.

[34] R. S. Yeung, H. Gu, M. Lee, and T. A. Dundon. Genetic identification of a locus, Mot1, that affects renal tumor size in the rat. *Genomics*, 78(3):108–112, Dec 2001.

[35] Y. Kikuchi, A. Sudo, H. Mitani, and O. Hino. Presence of a modifier gene(s) affecting early renal carcinogenesis in the Tsc2 mutant (Eker) rat model. *Int J Oncol*, 24(1):75–80, Jan 2004.

[36] H. Onda, A. Lueck, P. W. Marks, H. B. Warren, and D. J. Kwiatkowski. Tsc2(+/-) mice develop tumors in multiple sites that express gelsolin and are influenced by genetic background. *J Clin Invest*, 104(6):687–695, Sep 1999.

[37] C. Wilson, S. Idziaszczyk, L. Parry, C. Guy, D. F. Griffiths, E. Lazda, R. A. Bayne, A. J. Smith, J. R. Sampson, and J. P. Cheadle. A mouse model of tuberous sclerosis 1 showing background specific early post-natal mortality and metastatic renal cell carcinoma. *Hum Mol Genet*, 14(13):1839–1850, Jul 2005.

[38] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population. *Mamm Genome*, 23(9-10):713–718, Oct 2012.

[39] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.

[40] M. Civelek and A. J. Lusis. Systems genetics approaches to understand complex traits. *Nat Rev Genet*, 15(1):34–48, Jan 2014.

[41] I. Gerdes Gyuricza, J. M. Chick, G. R. Keele, A. G. Deighan, S. C. Munger, R. Korstanje, S. P. Gygi, and G. A. Churchill. Genome-wide transcript and protein analysis highlights the role of protein homeostasis in the aging mouse heart. *Genome Res*, 32(5):838–852, May 2022.

[42] Y. Takemon, J. M. Chick, I. Gerdes Gyuricza, D. A. Skelly, O. Devuyst, S. P. Gygi, G. A. Churchill, and R. Korstanje. Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. *Elife*, 10, Mar 2021.

[43] T. Gu, D. M. Gatti, A. Srivastava, E. M. Snyder, N. Raghupathy, P. Simecek, K. L. Svenson, I. Dotu, J. H. Chuang, M. P. Keller, A. D. Attie, R. E. Braun, and G. A. Churchill. Genetic Architectures of Quantitative Variation in RNA Editing Pathways. *Genetics*, 202(2):787–798, Feb 2016.

[44] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, 534(7608):500–505, Jun 2016.

[45] A. R. Coffey, T. L. Smallwood, J. Albright, K. Hua, M. Kanke, D. Pomp, B. J. Bennett, and P. Sethupa-

thy. Systems genetics identifies a co-regulated module of liver microRNAs associated with plasma LDL cholesterol in murine diet-induced dyslipidemia. *Physiol Genomics*, 49(11):618–629, Nov 2017.

[46] T. Melia and D. J. Waxman. Sex-Biased lncRNAs Inversely Correlate With Sex-Opposite Gene Coexpression Networks in Diversity Outbred Mouse Liver. *Endocrinology*, 160(5):989–1007, May 2019.

[47] M. N. Huda, M. VerHague, J. Albright, T. Smallwood, T. A. Bell, E. Que, D. R. Miller, B. Roshanravan, H. Allayee, F. P. Manuel de Villena, and B. J. Bennett. Dissecting the Genetic Architecture of Cystatin C in Diversity Outbred Mice. *G3 (Bethesda)*, 10(7):2529–2541, Jul 2020.

[48] E. Que, K. L. James, A. R. Coffey, T. L. Smallwood, J. Albright, M. N. Huda, D. Pomp, P. Sethupathy, and B. J. Bennett. Genetic Architecture Modulates Diet-Induced Hepatic mRNA and miRNA Expression Profiles in Diversity Outbred Mice. *Genetics*, 216(1):241–259, Sep 2020.

[49] D. A. Skelly, A. Czechanski, C. Byers, S. Aydin, C. Spruce, C. Olivier, K. Choi, D. M. Gatti, N. Raghupathy, G. R. Keele, A. Stanton, M. Vincent, S. Dion, I. Greenstein, M. Pankratz, D. K. Porter, W. Martin, C. O'Connor, W. Qin, A. H. Harrill, T. Choi, G. A. Churchill, S. C. Munger, C. L. Baker, and L. G. Reinholdt. Mapping the Effects of Genetic Variation on Chromatin State and Gene Expression Reveals Loci That Control Ground State Pluripotency. *Cell Stem Cell*, 27(3):459–469, Sep 2020.

[50] A. R. Ouellette, S. M. Neuner, L. Dumitrescu, L. C. Anderson, D. M. Gatti, E. R. Mahoney, J. A. Bubier, G. Churchill, L. Peters, M. J. Huentelman, J. H. Herskowitz, H. S. Yang, A. N. Smith, C. Reitz, B. W. Kunkle, C. C. White, P. L. De Jager, J. A. Schneider, D. A. Bennett, N. T. Seyfried, E. J. Chesler, N. Hadad, T. J. Hohman, and C. C. Kaczorowski. Cross-Species Analyses Identify Dlgap2 as a Regulator of Age-Related Cognitive Decline and Alzheimer's Dementia. *Cell Rep*, 32(9):108091, Sep 2020.

[51] J. A. Bubier, H. He, V. M. Philip, T. Roy, C. M. Hernandez, R. Bernat, K. D. Donohue, B. F. O'Hara, and E. J. Chesler. Genetic variation regulates opioid-induced respiratory depression in mice. *Sci Rep*, 10(1):14970, Sep 2020.

[52] T. Melia and D. J. Waxman. Genetic factors contributing to extensive variability of sex-specific hepatic gene expression in Diversity Outbred mice. *PLoS One*, 15(12):e0242665, 2020.

[53] J. R. Bagley, E. J. Chesler, V. M. Philip, and J. D. Jentsch. Heritability of ethanol consumption and pharmacokinetics in a genetically diverse panel of collaborative cross mouse strains and their inbred founders. *Alcohol Clin Exp Res*, 45(4):697–708, Apr 2021.

[54] Y. Takemon, V. Wright, B. Davenport, D. M. Gatti, S. M. Sheehan, K. Letson, H. S. Savage, R. Lennon, and R. Korstanje. Uncovering Modifier Genes of X-Linked Alport Syndrome Using a Novel Multiparent Mouse Model. *J Am Soc Nephrol*, 32(8):1961–1973, Aug 2021.

[55] R. L. Gould, S. W. Craig, S. McClatchy, G. A. Churchill, and R. Pazdro. Genetic mapping of renal glutathione suggests a novel regulatory locus on the murine X chromosome and overlap with hepatic glutathione regulation. *Free Radic Biol Med*, 174:28–39, Oct 2021.

[56] D. Koyuncu, M. K. K. Niazi, T. Tavolara, C. Abeijon, M. L. Ginese, Y. Liao, C. Mark, A. Specht, A. C. Gower, B. I. Restrepo, D. M. Gatti, I. Kramnik, M. Gurcan, B. Yener, and G. Beamer. CXCL1: A new diagnostic biomarker for human tuberculosis discovered using Diversity Outbred mice. *PLoS Pathog*, 17(8):e1009773, Aug 2021.

[57] B. T. Keenan, R. J. Galante, J. Lian, L. Zhang, X. Guo, O. J. Veatch, E. J. Chesler, W. T. O'Brien, K. L. Svenson, G. A. Churchill, and A. I. Pack. The dihydropyrimidine dehydrogenase gene contributes to heritable differences in sleep in mice. *Curr Biol*, 31(23):5238–5248, Dec 2021.

[58] E. G. Williams, N. Pfister, S. Roy, C. Statzer, J. Haverty, J. Ingels, C. Bohl, M. Hasan, J. uklina, P. hlmann,

N. Zamboni, L. Lu, C. Y. Ewald, R. W. Williams, and R. Aebersold. Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Syst*, 13(1):43–57, Jan 2022.

[59] C. C. Parker, V. M. Philip, D. M. Gatti, S. Kasparek, A. M. Kreuzman, L. Kuffler, B. Mansky, S. Masneuf, K. Sharif, E. Sluys, D. Taterra, W. M. Taylor, M. Thomas, O. Polesskaya, A. A. Palmer, A. Holmes, and E. J. Chesler. Genome-wide association mapping of ethanol sensitivity in the Diversity Outbred mouse population. *Alcohol Clin Exp Res*, 46(6):941–960, Jun 2022.

[60] J. B. Hackett, J. E. Glassbrook, M. C. iz, M. Bross, A. Fielder, G. Dyson, N. Movahhedin, J. McCasland, C. McCarthy-Leo, and H. M. Gibson. A diversity outbred F1 mouse model identifies host-intrinsic genetic regulators of response to immune checkpoint inhibitors. *Oncoimmunology*, 11(1):2064958, 2022.

[61] A. E. Al-Shaer, A. Pal, Q. Shi, M. S. Carson, J. Regan, M. Behee, N. Buddenbaum, C. Drawdy, T. Davis, R. Virk, and S. R. Shaikh. Modeling human heterogeneity of obesity with diversity outbred mice reveals a fat mass-dependent therapeutic window for resolvin E1. *FASEB J*, 36(6):e22354, Jun 2022.

[62] J. G. Xenakis, C. Douillet, T. A. Bell, P. Hock, J. Farrington, T. Liu, C. E. Y. Murphy, A. Saraswatula, G. D. Shaw, G. Nativio, Q. Shi, A. Venkatratnam, F. Zou, R. C. Fry, M. blo, and F. Pardo-Manuel de Villena. An interaction of inorganic arsenic exposure with body weight and composition on type 2 diabetes indicators in Diversity Outbred mice. *Mamm Genome*, 33(4):575–589, Dec 2022.

[63] M. M. Moran, F. C. Ko, L. D. Mesner, G. M. Calabrese, B. M. Al-Barghouthi, C. R. Farber, and D. R. Sumner. Intramembranous bone regeneration in diversity outbred mice is heritable. *Bone*, 164:116524, Nov 2022.

[64] G. R. Keele, T. Zhang, D. T. Pham, M. Vincent, T. A. Bell, P. Hock, G. D. Shaw, J. A. Paulo, S. C. Munger, F. P. de Villena, M. T. Ferris, S. P. Gygi, and G. A. Churchill. Regulation of protein abundance in genetically diverse mouse populations. *Cell Genom*, 1(1), Oct 2021.

[65] B. K. Cornes, C. Paisie, E. Swanzey, P. D. Fields, A. Schile, K. Brackett, L. G. Reinholdt, and A. Srivastava. Protein coding variation in the J:ARC and J:DO outbred laboratory mouse stocks provides a molecular basis for distinct research applications. *G3 (Bethesda)*, 13(4), Apr 2023.

[66] J. R. Bagley, L. S. Bailey, L. H. Gagnon, H. He, V. M. Philip, L. G. Reinholdt, L. M. Tarantino, E. J. Chesler, and J. D. Jentsch. Behavioral phenotypes revealed during reversal learning are linked with novel genetic loci in diversity outbred mice. *Addict Neurosci*, 4, Dec 2022.

[67] Y. Gupta, A. L. Ernst, A. Vorobyev, F. Beltsiou, D. Zillikens, K. Bieber, S. Sanna-Cherchi, A. M. Christiano, C. D. Sadik, R. J. Ludwig, and T. Sezin. Impact of diet and host genetics on the murine intestinal mycobiome. *Nat Commun*, 14(1):834, Feb 2023.

[68] J. B. Jacob, K. C. Wei, G. Bepler, J. D. Reyes, A. Cani, L. Polin, K. White, S. Kim, N. Viola, J. McGrath, A. Guastella, C. Yin, Q. S. Mi, B. L. Kidder, K. U. Wagner, S. Ratner, V. Phillips, J. Xiu, P. Parajuli, and W. Z. Wei. Identification of actionable targets for breast cancer intervention using a diversity outbred mouse model. *iScience*, 26(4):106320, Apr 2023.

[69] T. R. Price, D. S. Stapleton, K. L. Schueler, M. K. Norris, B. W. Parks, B. S. Yandell, G. A. Churchill, W. L. Holland, M. P. Keller, and A. D. Attie. ) as an enzyme that metabolizes phosphatidylcholine and cardiolipin. *bioRxiv*, Mar 2023.

[70] J. E. Glassbrook, J. B. Hackett, M. C. iz, M. Bross, G. Dyson, N. Movahhedin, A. Ullrich, and H. M. Gibson. Host genetic background regulates the capacity for anti-tumor antibody-dependent phagocytosis. *bioRxiv*, May 2023.

[71] S. Aydin, D. T. Pham, T. Zhang, G. R. Keele, D. A. Skelly, J. A. Paulo, M. Pankratz, T. Choi, S. P. Gygi,

L. G. Reinholdt, C. L. Baker, G. A. Churchill, and S. C. Munger. Genetic dissection of the pluripotent proteome through multi-omics data integration. *Cell Genom*, 3(4):100283, Apr 2023.

[72] V. M. Philip, H. He, M. C. Saul, P. E. Dickson, J. A. Bubier, and E. J. Chesler. Gene expression genetics of the striatum of Diversity Outbred mice. *bioRxiv*, May 2023.

[73] K. Chella Krishnan, E. J. El Hachem, M. P. Keller, S. G. Patel, L. Carroll, A. D. Vegas, I. Gerdes Gyuricza, C. Light, Y. Cao, C. Pan, K. E. Kaczor-Urbanowicz, V. Shravah, D. Anum, M. Pellegrini, C. F. Lee, M. M. Seldin, N. A. Rosenthal, G. A. Churchill, A. D. Attie, B. Parker, D. E. James, and A. J. Lusis. Genetic architecture of heart mitochondrial proteome influencing cardiac hypertrophy. *Elife*, 12, Jun 2023.

[74] M. Kim, M. N. Huda, L. W. Evans, E. Que, E. R. Gertz, N. Maeda-Smithies, and B. J. Bennett. Integrative analysis of hepatic transcriptional profiles reveals genetic regulation of atherosclerosis in hyperlipidemic Diversity Outbred-F1 mice. *Sci Rep*, 13(1):9475, Jun 2023.

[75] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, S. Sen, B. S. Yandell, and G. A. Churchill. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.

[76] A. L. Tyler, W. Lu, J. J. Hendrick, V. M. Philip, and G. W. Carter. CAPE: an R package for combined analysis of pleiotropy and epistasis. *PLoS Comput Biol*, 9(10):e1003270, Oct 2013.

[77] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.

[78] A. L. Tyler, B. Ji, D. M. Gatti, S. C. Munger, G. A. Churchill, K. L. Svenson, and G. W. Carter. Epistatic Networks Jointly Influence Phenotypes Related to Metabolic Disease and Gene Expression in Diversity Outbred Mice. *Genetics*, 206(2):621–639, Jun 2017.

[79] A. L. Tyler, C. Spruce, R. Kursawe, A. Haber, R. L. Ball, W. A. Pitman, A. D. Fine, N. Raghupathy, M. Walker, V. M. Philip, C. L. Baker, J. M. Mahoney, G. A. Churchill, J. J. Trowbridge, M. L. Stitzel, K. Paigen, P. M. Petkov, and G. W. Carter. Variation in histone configurations correlates with gene expression across nine inbred strains of mice. *Genome Res*, May 2023.

[80] J. M. Winter, D. E. Gildea, J. P. Andreas, D. M. Gatti, K. A. Williams, M. Lee, Y. Hu, S. Zhang, J. C. Mullikin, T. G. Wolfsberg, S. K. McDonnell, Z. C. Fogarty, M. C. Larson, A. J. French, D. J. Schaid, S. N. Thibodeau, G. A. Churchill, and N. P. Crawford. Mapping Complex Traits in a Diversity Outbred F1 Mouse Population Identifies Germline Modifiers of Metastasis in Human Prostate Cancer. *Cell Syst*, 4(1):31–45, Jan 2017.

[81] W. Z. Wei, H. M. Gibson, J. B. Jacob, J. A. Frelinger, J. A. Berzofsky, H. Maeng, G. Dyson, J. D. Reyes, S. Pilon-Thomas, S. Ratner, and K. C. Wei. Diversity Outbred Mice Reveal the Quantitative Trait Locus and Regulatory Cells of HER2 Immunity. *J Immunol*, 205(6):1554–1563, Sep 2020.

[82] A. L. Tyler, B. El Kassaby, G. Kolishovski, J. Emerson, A. E. Wells, J. M. Mahoney, and G. W. Carter. Effects of kinship correction on inflation of genetic interaction statistics in commonly used mouse populations. *G3 (Bethesda)*, 11(7), Jul 2021.

[83] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*, 21(9):1281–1289, Sep 2018.

[84] A. B. Wiltschko, M. J. Johnson, G. Iurilli, R. E. Peterson, J. M. Katon, S. L. Pashkovski, V. E. Abraira, R. P. Adams, and S. R. Datta. Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135, Dec 2015.

[85] T. D. Pereira, D. E. Aldarondo, L. Willmore, M. Kislin, S. S. Wang, M. Murthy, and J. W. Shaevitz. Fast animal pose estimation using deep neural networks. *Nat Methods*, 16(1):117–125, Jan 2019.

[86] J. M. Graving, D. Chae, H. Naik, L. Li, B. Koger, B. R. Costelloe, and I. D. Couzin. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife*, 8, Oct 2019.

[87] J. E. Markowitz, W. F. Gillis, C. C. Beron, S. Q. Neufeld, K. Robertson, N. D. Bhagat, R. E. Peterson, E. Peterson, M. Hyun, S. W. Linderman, B. L. Sabatini, and S. R. Datta. The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell*, 174(1):44–58, Jun 2018.

[88] L. E. Hession, G. S. Sabnis, G. A. Churchill, and V. Kumar. A machine-vision-based frailty index for mice. *Nat Aging*, 2(8):756–766, Aug 2022.

[89] B. Geuther, M. Chen, R. J. Galante, O. Han, J. Lian, J. George, A. I. Pack, and V. Kumar. High-throughput visual assessment of sleep stages in mice using machine learning. *Sleep*, 45(2), Feb 2022.

[90] K. Sheppard, J. Gardin, G. S. Sabnis, A. Peer, M. Darrell, S. Deats, B. Geuther, C. M. Lutz, and V. Kumar. Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation. *Cell Rep*, 38(2):110231, Jan 2022.

[91] B. Q. Geuther, A. Peer, H. He, G. Sabnis, V. M. Philip, and V. Kumar. Action detection using a neural network elucidates the genetics of mouse grooming behavior. *Elife*, 10, Mar 2021.

[92] T. Gschwind, A. Zeine, I. Raikov, J. E. Markowitz, W. F. Gillis, S. Felong, L. L. Isom, S. R. Datta, and I. Soltesz. Hidden behavioral fingerprints in epilepsy. *Neuron*, 111(9):1440–1452, May 2023.

[93] N. Cheng, Y. Ren, J. Zhou, Y. Zhang, D. Wang, X. Zhang, B. Chen, F. Liu, J. Lv, Q. Cao, S. Chen, H. Du, D. Hui, Z. Weng, Q. Liang, B. Su, L. Tang, L. Han, J. Chen, and C. Shao. Deep Learning-Based Classification of Hepatocellular Nodular Lesions on Whole-Slide Histopathologic Images. *Gastroenterology*, 162(7):1948–1961, Jun 2022.

[94] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*, 25(8):1301–1309, Aug 2019.

[95] G. Backer, S. Eddy, S. M. Sheehan, Y. Takemon, A. Reznichenko, H. S. Savage, M. Kretzler, and R. Korstanje. FAR2 is associated with kidney disease in mice and humans. *Physiol Genomics*, 50(8):543–552, Aug 2018.

[96] S. Sheehan, S. Mawe, R. E. Cianciolo, R. Korstanje, and J. M. Mahoney. Detection and Classification of Novel Renal Histologic Phenotypes Using Deep Neural Networks. *Am J Pathol*, 189(9):1786–1796, Sep 2019.

[97] J. A. King and D. M. Stamilio. Maternal and fetal tuberous sclerosis complicating pregnancy: a case report and overview of the literature. *Am J Perinatol*, 22(2):103–108, Feb 2005.

[98] N. Sharma, S. Sharma, J. L. Thiek, S. S. Ahanthem, A. Kalita, and D. Lynser. Maternal and Fetal Tuberous Sclerosis: Do We Know Enough as an Obstetrician? *J Reprod Infertil*, 18(2):257–260, 2017.

[99] N. Gupta, N. Singh, S. Sarangi, S. Dalmia, and S. Mittal. Fetal cardiac rhabdomyoma with maternal tuberous sclerosis complicating pregnancy. *Arch Gynecol Obstet*, 278(2):169–170, Aug 2008.

[100] B. Sanderson, P.D. Fields, and Lloyd M.W. JAX NGS Operations Nextflow DSL2 Pipelines, 2023.

[101] B. Li and C. N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12:323, Aug 2011.

[102] Broad Institute. Picard tools, 2023.

[103] P. Ewels, M. Magnusson, S. Lundin, and M. ller. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, Oct 2016.

[104] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12):550, 2014.

[105] C. Soneson and M. Delorenzi. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14:91, Mar 2013.

[106] J. S. Sigmon, M. W. Blanchard, R. S. Baric, T. A. Bell, J. Brennan, G. A. Brockmann, A. W. Burks, J. M. Calabrese, K. M. Caron, R. E. Cheney, D. Ciavatta, F. Conlon, D. B. Darr, J. Faber, C. Franklin, T. R. Gershon, L. Gralinski, B. Gu, C. H. Gaines, R. S. Hagan, E. G. Heimsath, M. T. Heise, P. Hock, F. Ideraabdullah, J. C. Jennette, T. Kafri, A. Kashfeen, M. Kulis, V. Kumar, C. Linnertz, A. Livraghi-Butrico, K. C. K. Lloyd, C. Lutz, R. M. Lynch, T. Magnuson, G. K. Matsushima, R. McMullan, D. R. Miller, K. L. Mohlke, S. S. Moy, C. E. Y. Murphy, M. Najarian, L. O'Brien, A. A. Palmer, B. D. Philpot, S. H. Randell, L. Reinholdt, Y. Ren, S. Rockwood, A. R. Rogala, A. Saraswatula, C. M. Sassetti, J. C. Schisler, S. A. Schoenrock, G. D. Shaw, J. R. Shorter, C. M. Smith, C. L. St Pierre, L. M. Tarantino, D. W. Threadgill, W. Valdar, B. J. Vilen, K. Wardwell, J. K. Whitmire, L. Williams, M. J. Zylka, M. T. Ferris, L. McMillan, and F. P. Manuel de Villena. Content and Performance of the MiniMUGA Genotyping Array: A New Tool To Improve Rigor and Reproducibility in Mouse Research. *Genetics*, 216(4):905–930, Dec 2020.

[107] The Jackson Laboratory. Haploqa, 2022.

[108] D. M. Gatti, K. L. Svenson, A. Shabalin, L. Y. Wu, W. Valdar, P. Simecek, N. Goodwin, R. Cheng, D. Pomp, A. Palmer, E. J. Chesler, K. W. Broman, and G. A. Churchill. Quantitative trait locus mapping methods for diversity outbred mice. *G3 (Bethesda)*, 4(9):1623–1633, Sep 2014.

[109] R. W. Doerge and G. A. Churchill. Permutation tests for multiple loci affecting a quantitative character. *Genetics*, 142(1):285–294, Jan 1996.

[110] H. M. Kang, N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723, Mar 2008.

[111] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent, S. Allen, A. T. Broman, R. Bacher, C. Kendziorski, K. W. Broman, B. S. Yandell, G. A. Churchill, and A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.

[112] A. L. Tyler, L. R. Donahue, G. A. Churchill, and G. W. Carter. Weak Epistasis Generally Stabilizes Phenotypes in a Mouse Intercross. *PLoS Genet*, 12(2):e1005805, Feb 2016.

[113] A. Tyler, J. M. Mahoney, and G. W. Carter. Genetic Interactions Affect Lung Function in Patients with Systemic Sclerosis. *G3 (Bethesda)*, 10(1):151–163, Jan 2020.

[114] V. M. Philip, A. L. Tyler, and G. W. Carter. Dissection of complex gene expression using the combined analysis of pleiotropy and epistasis. *Pac Symp Biocomput*, pages 200–211, 2014.

[115] A. L. Tyler, T. C. McGarr, B. J. Beyer, W. N. Frankel, and G. W. Carter. A genetic interaction network model of a complex neurological disease. *Genes Brain Behav*, 13(8):831–840, Nov 2014.

[116] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186, Jun 2017.

[117] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell*, 177(4):1022–1034, May 2019.

[118] O. Y. n, C. Crainiceanu, E. L. Ogburn, B. S. Caffo, T. D. Wager, and M. A. Lindquist. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19(2):121–136, Apr 2018.

[119] E. C. Neto, M. P. Keller, A. D. Attie, and B. S. Yandell. CAUSAL GRAPHICAL MODELS IN SYSTEMS GENETICS: A UNIFIED FRAMEWORK FOR JOINT INFERENCE OF CAUSAL NETWORK AND GENETIC ARCHITECTURE FOR CORRELATED PHENOTYPES. *Ann Appl Stat*, 4(1):320–339, Mar 2010.

[120] R. H. Blair, D. J. Kliebenstein, and G. A. Churchill. What can causal networks tell us about metabolic pathways? *PLoS Comput Biol*, 8(4):e1002458, 2012.

[121] Kenneth A. Bollen. *Structural equations with latent variables*, volume 210. John Wiley & Sons, 1989.

[122] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–15550, Oct 2005.

[123] J. Millstein and D. Volfson. Computationally efficient permutation-based confidence interval estimation for tail-area FDR. *Front Genet*, 4:179, 2013.

[124] J. L. Brabec, M. K. Lara, A. L. Tyler, and J. M. Mahoney. System-Level Analysis of Alzheimer's Disease Prioritizes Candidate Genes for Neurodegeneration. *Front Genet*, 12:625246, 2021.

[125] A. L. Tyler, A. Raza, D. N. Krementsov, L. K. Case, R. Huang, R. Z. Ma, E. P. Blankenhorn, C. Teuscher, and J. M. Mahoney. Network-Based Functional Prediction Augments Genetic Association To Predict Candidate Genes for Histamine Hypersensitivity in Mice. *G3 (Bethesda)*, 9(12):4223–4233, Dec 2019.

[126] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[127] Muhammad Ali Farooq, Asma Khatoon, Viktor Varkarakis, and Peter Corcoran. Advanced deep learning methodologies for skin cancer classification in prodromal stages. *arXiv preprint arXiv:2003.06356*, 2020.

[128] Pramit Dutta, Khaleda Akhter Sathi, and Md Saiful Islam. Multi-classification of brain tumor images using transfer learning based deep neural network. In *International Conference on Artificial Intelligence for Smart Community: AISC 2020, 17–18 December, Universiti Teknologi Petronas, Malaysia*, pages 927–933. Springer, 2022.

[129] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak, and H. J. W. L. Aerts. Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin Cancer Res*, 25(11):3266–3275, Jun 2019.

[130] J. Yu, Y. Deng, T. Liu, J. Zhou, X. Jia, T. Xiao, S. Zhou, J. Li, Y. Guo, Y. Wang, J. Zhou, and C. Chang. Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat Commun*, 11(1):4807, Sep 2020.

[131] T. Eggermann, S. Spengler, B. Denecke, K. Zerres, and C. J. Mache. Multi-exon deletion in the XDH gene as a cause of classical xanthinuria. *Clin Nephrol*, 79(1):78–80, Jan 2013.

[132] X. Y. Liu, F. R. Zhang, J. Y. Shang, Y. Y. Liu, X. F. Lv, J. N. Yuan, T. T. Zhang, K. Li, X. C. Lin, X. Liu, Q. Lei, X. D. Fu, J. G. Zhou, and S. J. Liang. Renal inhibition of miR-181a ameliorates 5-fluorouracil-induced mesangial cell apoptosis and nephrotoxicity. *Cell Death Dis*, 9(6):610, May 2018.

[133] X. Wang, J. Park, K. Susztak, N. R. Zhang, and M. Li. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun*, 10(1):380, Jan 2019.

[134] Koby Crammer and Gal Chechik. A needle in a haystack: local one-class optimization. In *Proceedings of the twenty-first international conference on Machine learning*, page 26, 2004.