Unexpected *XIST* Expression in Male Hearts Associates with Disease

Gennady Gorin[1], Natalie DeForest[1], Linda Goodman[1*]

1 Fauna Bio, Emeryville, CA

* `linda[at]faunabio.com`

Abstract

XIST is a long non-coding RNA that mediates the process of X chromosome inactivation in females, and has not been previously observed in males outside of developmental, pathological, and ectopic contexts. We report the unprecedented endogenous expression of *XIST* in 226 male heart samples across 15 single-cell RNA sequencing studies. The male expression of *XIST* is specific to a rare cell type typically annotated as neurons or glia. We investigate a variety of explanations and cross-reference the expression of *XIST* against age, sex, and pathology, finding signatures of a non-canonical inhibitory program, potentially mediated by a truncated transcript. Interestingly, the highest-powered datasets exhibit *XIST* upregulation in heart disease and comorbidity donors relative to healthier controls. Furthermore, in humans we identify a male-specific association of the *XIST* genomic locus with myocardial infarction ($p = 1.3 \times 10^{-4}$). Taken together, our findings suggest that the pathology of cardiomyopathy may involve uncharacterized, therapeutically actionable non-coding RNA pathways that operate through the cardiac nervous system, and underscore the importance of genomic investigations of the X chromosome.

1 Introduction

X-chromosome inactivation (XCI) establishes monoallelic expression of X-linked genes in female cells by condensing one of the X chromosomes. Although the details of the process are relatively complex and differ by species and tissue, several elementary points are generally considered uncontroversial:

- In humans, XCI is **ubiquitous** in and **specific** to females (XX) [1].
- XCI is mediated and continuously enforced by *XIST*, a long non-coding RNA (lncRNA) on the X chromosome [1].
- XCI leads to dramatic changes in the topology and accessibility of one of the copies of the X chromosome [2].
- Gene expression in the resulting Barr body (X_i) ranges from near-total silencing to a slight decrease relative to the active X chromosome (X_a) [3, 4], depending on the gene.
- There are rare exceptions to the ubiquity in females: in some restricted immune cell contexts, cells escape XCI [5, 6], leading to overexpression of X-linked genes.

Some rare exceptions to the specificity to females are more contentious. XCI and *XIST* expression transiently occur in the male (XY) germline [7–9]. However, *XIST* has also been reported in certain tumors, more tentatively in conditions with supernumerary sex chromosomes (e.g., with XXY or XXXY karyotypes) [7, 10], and recently claimed to be quantifiable by qPCR in male lymphocytes in a conference abstract, albeit without any follow-up reports [11]. *In vitro*, it has been observed in differentiating stem cells [12]. Ectopic induction in XY stem cells results in somewhat weaker XCI [13]. The expression of a truncated, non-silencing variant in male mice triggers autoimmunity to a *Xist* ribonucleoprotein complex [14]. In other words, there is some precedent for such exceptions, but they are well-understood to be atypical in somatic, differentiated tissues, and generally require ectopic expression.

The genomics and health implications of *XIST* have been almost exclusively studied in the context of XCI. Outside the recently reported non-canonical connections to autoimmunity [14], *XIST* and XCI escapees have been tentatively implicated in affective disorders [15] and aging [16]. Conversely, the induction of XCI has been proposed as a promising therapeutic approach to silence lethal trisomies [17]. However, the X chromosome and non-coding RNA are rarely included in genome-wide association studies [18–21], and remain an untapped reservoir of variation related to disease.

Separately, a great deal of effort has been expended toward characterizing the single-cell transcriptomes of human cardiac cells in healthy and pathological tissues [22]. At this time, there are several dozen such published datasets. Despite the variety of study designs and goals, the cell type compositions tend to be consistent with the "canonical" 2020 Human Cell Atlas dataset by Litviňuková et al. [23], used in lieu of a control in a number of studies [24–26]: cardiac tissue contains cardiomyocytes, vascular cells, fibroblasts, immune cells, adipocytes, and neurons. As in other tissues, all of these cell types can be sampled by single-nucleus sequencing, although adipocytes and myocytes are generally (but not universally [27]) considered morphologically incompatible with droplet-based single-cell protocols [23].

As the cardiac nervous system controls signaling and conduction, cardiac neurons have been investigated as potentially impactful and accessible targets for disease intervention long before single-cell sequencing. However, the architecture, role, and pathology of the heart's neural population remains poorly characterized [28–30]. Analogously, in the single-cell atlases, the cardiac neurons are slightly more contentious than other cell types. The relevant cell population strongly expresses neurexins (*NRXN1*/*NRXN3*), proteolipid protein 1 (*PLP1*), myelin protein zero (*MPZ*), and a number of other nervous system markers [22]. The original heart cell atlas reported them as neurons, but noted that a subpopulation more closely resembles (glial) Schwann cells [23]. A follow-up study by the same authors remarked that the cells fail to express canonical neuronal markers and redefined them as glia [31]. This study additionally used staining to validate the presence of this distinct PLP1-expressing cell population and characterize its association with P cells. Similarly, spatial transcriptomics was performed to characterize the expression of the Schwann cell marker *MPZ* in a more recent study [32]. Other articles have reported these cells as neurons [25, 33, 34] or glia/Schwann cells [35, 36] with little further comment. Nominally, these are considered "prototypic" peripheral nervous system cells [22]. However, to acknowledge the remaining uncertainty in the field, we term this cell population "pseudo-glial."

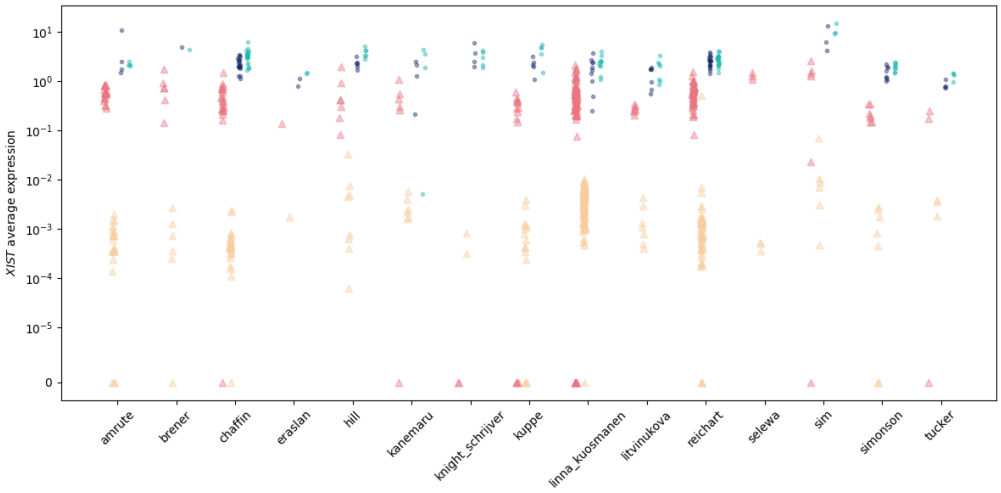

Fig 1. Male cardiac pseudo-glia show striking expression of the female-specific transcript *XIST*. Each point denotes average expression of the gene in one sample (typically one unique donor). Triangles denote samples reported as male. Circles denote samples reported as female. Red: male, pseudo-glial (cells annotated as Schwann cells, neurons, or glia). Yellow: male, all others. Dark blue: female, pseudo-glial. Teal: female, all others. To accommodate zero values, the region $[0, 10^{-5}]$ is linear.

2 Results

2.1 Male pseudo-glia express *XIST*

Somewhat surprisingly, these pseudo-glia demonstrate high and specific expression of the *XIST* transcript in males (Fig 1). Although some variability is evident, the usual difference between male and female expression of *XIST* is 2–3 orders of magnitude: female samples show strong *XIST* expression in over half of all cells, almost certainly an underestimate due to experimental limitations; on the other hand, the expression in male samples is, although nonzero, vanishingly sparse and well within error bounds generally explainable by contamination, doublets, or barcoding errors. For instance, *XIST* expressors make up 63% of cells in female and 0.3% in male samples from the Litviňuková atlas [23]. Yet in the pseudo-glia, male and female cells have comparable amounts of *XIST*, with the latter showing slightly but not drastically higher expression. In fact, 67% of the rare *XIST*-positive male cells in the atlas are annotated as neural, and another 16% is assigned to an "unknown" population that strongly expresses the Schwann cell markers *PLP1* and *MPZ*. Overall, we observe average expression values that exceed 10^{-2} in 226 of the 248 male samples across fifteen studies of the human heart.

2.2 Exogenous or artifactual explanations are inadequate

As expression of *XIST* in male somatic tissues is unusual, it is necessary to rule out potential exogenous sources. These include donor metadata errors, sample metadata errors, background contamination, doublets, and alignment errors.

Donor metadata errors refer to a tissue donor's sex being reported incorrectly. This explanation cannot account for cell type-specific enrichment of *XIST*, as the transcript would be expressed at high levels elsewhere, as in other female samples. In addition, this type of error is fairly easy to spot: for instance, the donor AV3 in Kanemaru et al. [31], reported as a Caucasian female in her early 60s, expresses negligible *XIST* outside the

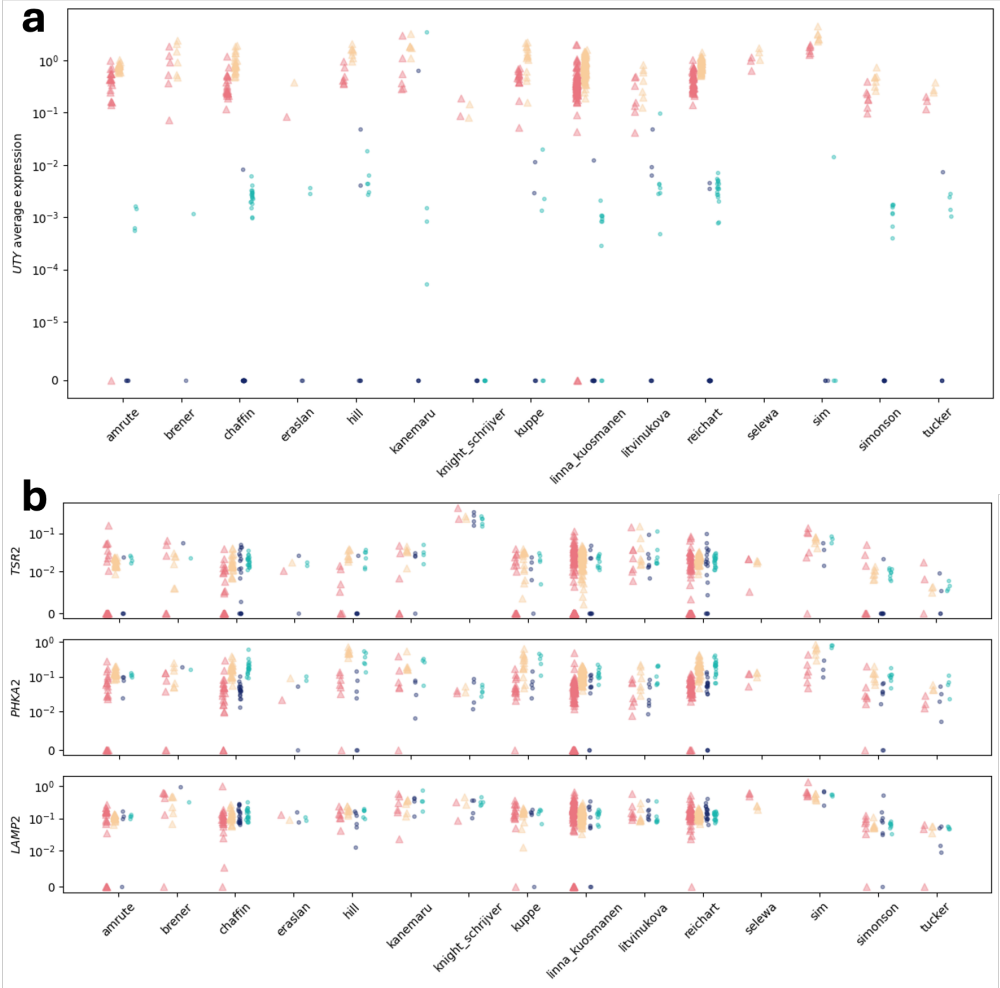

Fig 2. Expression of sex-associated genes in the heart. Conventions as in Fig 1.
a. Average expression of *UTY* in cardiac datasets (linear region $[0, 10^{-5}]$).
b. Average expression of example genes reported [39] as consistently silenced in XCI by Tukiainen et al. (linear region $[0, 10^{-2}]$).

pseudo-glia, but shows substantial expression of Y chromosome transcripts *UTY* (65% of cells), *DDX3Y* (43%), and *ZFY* (32%). This sample is readily evident in Fig 1, and displays the atypical high expression of *UTY* in Fig 2a. Unrecorded sex chromosome abnormalities, if they occurred, would be readily detectable by tissue-wide expression of *XIST* and Y-linked genes [7, 37, 38]. It appears most reasonable to conclude that the hundreds of donors under consideration, all of whom are annotated as male, express Y-linked genes, and do not express *XIST* outside the pseudo-glia, are indeed male (XY).

Sample metadata errors refer to the relationship between a sequencing sample and a donor being reported incorrectly. A single sequencing run may contain several multiplexed donor samples, or tissues from single donor may have been sequenced separately, e.g. split by cardiac chamber, and the appropriate mapping may have been lost. Again, this explanation cannot account for the enrichment of *XIST* in a rare cell type and nowhere else. In addition, this type of error is also fairly easy to spot. For instance, in the Litviňuková atlas, donor H5, reported as a Caucasian female in her early 50s [23], has a relatively high expression of *UTY* (Fig 2a) exclusive to the right

ventricular sample `H0015_RV`, which also has negligible *XIST*; the other five samples assigned to this donor (LV, RA, LA, apex, and septum) have more typically female transcriptomes (high *XIST* and negligible Y-linked transcripts). In another example, the Reichart study reports that donor H40 is a male in his 60s [40]; the left ventricular sample `BO_H40_LVW_premrna`, one of four for this donor, shows a female-like transcriptome, where the others – two LV and one septal – show entirely typical male expression of *XIST* and Y-linked genes. Thus, metadata errors do not appear to be responsible for this effect.

Background contamination and doublets (e.g., due to incorporation of exogenous *XIST* from female samples processed in parallel) are implausible explanations, for a variety of reasons: the robustness of the expression across many studies, the presence of the same expression pattern in the male-only data by Selewa et al. [41], and the lack of a plausible mechanism for background transcript enrichment in a specific cell type. In other words, there is no clear source for exogenous *XIST*, nor a clear reason why it should specifically associate with pseudo-glia regardless of protocol.

Alignment errors are somewhat challenging to rule out. It is conceivable that a rare, cell type-specific transcript, generated from a different gene and unattested until now, shares enough structure to be mistakenly quantified as *XIST*. Such an explanation, appealing to genomic "dark matter," is somewhat implausible if only because it would predict the total amount of "*XIST*" should be *higher* in female glia-like samples, rather than equal to or slightly lower than in other female cells, as actually observed in sequencing data. We do, however, note that observed data may be consistent with an uncharacterized *XIST*-like gene whose expression is cell type- and male-specific. Incidentally, overlap with *known* transcript sequences would artificially deflate "*XIST*" quantification. This is observed, e.g., in the Commons Cell Atlas reanalysis of the Litviňuková data (accession ERX4319109, sample `H0026_RV`, lane 3), which reports zero *XIST* expression [42, 43]. This loss of signal appears to be a consequence of an unstranded pseudoalignment strategy, which typically maps more reads, but cannot uniquely assign, and must discard, reads that can originate from either gene in a sense–antisense pair (i.e., exon 6 of the canonical transcript *XIST*-204, which overlaps *TSIX*).

The identification of the precise transcript identity requires investigation of the raw read alignments. None of the datasets reanalyzed in Fig 1 are accompanied by such alignments. However, a small unrelated study by Larson and Chin [44] released single-nucleus alignment data generated from the interventricular septa (IVS) of two males and two females. The reads aligned to genes in the *XIST*/*TSIX* region are displayed in Fig 7a (Section 4.3). The coverage in the male sample is considerably lower, but the localization appears to be consistent: reads are predominantly found in exon 6 of the canonical transcript, with far sparser expression in exon 1. Although these short-read data do not provide unequivocal evidence that the *XIST* transcript observed in human heart pseudo-glia is canonical, they do suggest that the count data are not likely to be attributable to, e.g., a poorly characterized unrelated transcript that happens to be localized to the same region, such as the pseudogene *LNX2BP*. These results agree with our reanalysis of a recent long-read single-cell heart dataset [45], which shows that pseudo-glial *XIST* read coverage is exclusive to exon 6 of the canonical transcript (Fig 7b, Section 4.4), biased toward the 3' end and internal poly(A)-rich regions. The read coverage distribution is consistent with the relevant species being a truncated transcript that covers exon 6 of *XIST*-204, but we cannot rule out a more complex structure without full-length sequencing. In sum, there does not appear to be a plausible artifactual explanation for the expression of *XIST* in male hearts.

2.3 Cross-species context and biological implications

X chromosome inactivation is ubiquitous in mammals; in placental mammals, XCI is achieved through the expression of *XIST* [46, 47], although the transcript structure shows nontrivial variability [48]. Therefore, non-human mammals provide a natural point of comparison, as well as a potential choice of model organisms for follow-up *in vivo* investigations. However, the presence of *XIST* in non-human male pseudo-glia is ambiguous. As we discuss in Section 4.1.2, the absence of the lncRNA in genome references poses a barrier. Our preliminary investigation suggests that, in males, *Xist* is not expressed in the adult rat, mouse, pig, or cynomolgus monkey hearts, but shows some non-specific expression in the neonatal mouse heart. It is likely that other public datasets could shed light on the issue upon exhaustive reanalysis, but not so readily as the immediately evident signal in human tissues.

Despite the usual identification of *XIST* expression with XCI, there does not appear to be any evidence for canonical XCI in this cell population. Genes such as *TSR2*, *PHKA2*, and *LAMP2*, reported to be robustly inactivated in female tissues, appear to be no less expressed in male pseudo-glial cells than in others [39] (Fig 2b). Were the XCI process effective, one would expect their expression in XY cells to be far lower than in XX cells. However, this does not appear to occur; in fact, the non-pseudoautosomal (i.e., susceptible to XCI) X-linked gene *PLP1* is one of the strongest markers for this cell type. The inability to induce XCI is, to some extent, consistent with the results reported by Gayen et al. in an ectopic expression scenario [13]. This study's results suggest that the inactivation mechanism relies on a ploidy counting mechanism mediated by escape products, and inactivation only occurs when the cell has two X chromosomes. However, despite a weaker magnitude, qualitative signatures of "canonical" silencing, such as *XIST* coating and silencing of X-linked genes, were observed in XY cells in that study. As the biological context is drastically different and the characteristics of *XIST* coating in the heart are unknown, we cannot establish whether the mechanisms proposed by Gayen et al. are relevant here.

The male expression of *XIST* does not appear to be restricted to these datasets, nor the heart. Similar male neurexin- and *XIST*-expressing populations are present in arterial atlases [49–51], albeit reported as smooth muscle cells or lymphocytes [49, 51]. Single-cell surveys of the human eye [52], musculature [53], prostate, and esophagus [36] report male donor *XIST* expression in rare Schwann cell populations, with some caveats (Section 4.1.1). In sum, the expression of *XIST* in male glia-like cells may not be tissue-specific; however, further investigation is hampered by the rarity of the cell type and the availability of data.

It is interesting to note that the fetal "neuronal" cell type reported by Knight-Schrijver [25] does not exhibit *XIST* expression in the two fetal male samples. This is consistent with the unusually low expression in the two fetal male samples reported by Sim et al. [54] (SAMN15889155 and SAMN15889156, visible in Fig 1). Aside from this dearth of expression in developing heart tissues, the expression of *XIST* is remarkably stable, varying across approximately an order of magnitude across the entire human lifespan. This consistency is more striking yet due to the small sample sizes and the lack of adjustments for technology, batch effects, pathology, or discrepancies in reference annotations (Fig 6a), all of which should be expected to introduce variability. After accounting for age and source study as covariates, the age effects on expression, if any, appear to be minimal (Section 4.2.4).

To investigate the transcriptome-wide associations of *XIST* expression while accounting for the many technical and biological covariates, we performed differential expression (DE) testing between *XIST*-positive and *XIST*-negative cells annotated as neurons or glia, encoding the donor ID as a model covariate (Section 4.2.3). As shown in Fig 3, the *XIST*-positive cells exhibited a striking decrease in a diverse variety of

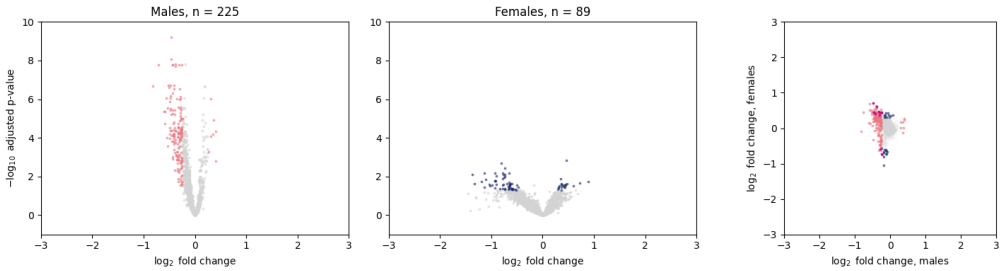

Fig 3. Results of differential expression testing between *XIST*-positive and *XIST*-negative pseudo-glial cells. Left: male samples (red: DE genes). Middle: female samples (blue: DE genes). Right: relationship between $\log_2$ fold changes inferred for male and female samples (red: male DE genes; blue: female DE genes; pink: genes DE in both comparisons). Differential expression is defined as $|\log_2 \mathrm{FC}| > 0.25$ and adjusted p-value < 0.05. *XIST* excluded from visualization. Visualization restricted to genes with `baseMean` > 5.

genes in male samples, and little to no difference from the *XIST*-negative cells in female samples.

Gene Ontology (GO) analysis of the 225 genes differentially expressed in female samples reported few significant enrichments, primarily the cellular and mitochondrial ribosomal proteins. Restricting the analysis to the 75 genes with at least modest average expression (`baseMean` > 5) yielded analogous enrichment of cellular ribosomal proteins and electron transport chain components, with modest statistical significance. It is far from clear that even these minor differences are meaningful, as the distinction between *XIST*-positive and -negative cells in female samples is likely to be predominantly governed by incidental loss of molecules in library construction and sequencing, rather than true lack of *XIST*, which is understood to be ubiquitous in XX cells. This is supported by the lack of detectable signatures of, e.g., escape from XCI in supposedly *XIST*-negative cells. In contrast, XCI escape *is* readily detectable through analogous statistical testing between male and female donors (Section 4.2.2).

GO analysis of 167 DE genes in male samples (likewise restricted to `baseMean` > 5) reported a variety of enriched biological processes, including genes associated with myofiber assembly and myogenesis (various components and forms of titin, fibulin, myomesin, fibronectin, actinin, and actin), with action potential regulation (channels and receptors *SLC8A1*, *CACNB2*, *CACNA1C*, and *RYR2*, alongside *FGF12*), with neuron migration regulation (*PHACTR1*, *UNC5C*, *ERBB4*, and *NRG3*), among others.

These genes were predominantly downregulated; the few exceptions with higher expression in *XIST*-positive cells included several relatively sparsely observed genes, including the potassium and choline transporters *KCNMA1* and *SLC5A7*, the glutamate receptor *GRID2*, the immunoglobulin *OPCML*, the obscure sulfotransferase *CHST9* (tentatively associated with heart conditions in genomics studies [55–57]), the neuronal development mediator *ALK*, the receptor modulators *LYPD6* and *OPCML*, and the poorly characterized gene *FAM135B*. None of these genes are sex-linked. Three X-linked genes, all outside the pseudoautosomal regions, were downregulated in *XIST*-positive cells: *TMSB4X*, *ARHGAP6*, and *DIAPH2*. All are involved in actin binding.

The pervasive depletion of canonical muscle transcripts in *XIST*-positive cells is suggestive, although some caution is warranted: these gene products are ubiquitous in, e.g., the cardiomyocytes, and a nontrivial fraction of the expression is likely exogenous due to contamination from the cell lysate. However, it is not clear how such a mechanism might specifically contaminate *XIST*-negative cells in male samples. We discuss these potential explanations in Section 4.2.3.

3 Conclusion

The potential biological impact of this finding is unclear. It is possible that the male *XIST* expression is incidental or explicable by artifacts not considered here, and the cells may be "glia as usual." However, the fundamental divergence between these cells and supposedly universal patterns of gene expression regulation elsewhere in the human body leads us to urge caution in treating them as "prototypic" [22], as the functional differences may be far more profound. At the very least, the expression patterns reported here are striking counterexamples to the conventional wisdom that *XIST* is sufficient for XCI and that its expression is strictly pathological or developmental in males.

It is tempting to speculate that the suppression of the myogenesis program in the *XIST*-positive population is effected by a direct mechanism analogous to XCI. It is also likely mistaken: the suppression is genome-wide and seemingly specific, rather than localized to a single chromosome. It appears more plausible that *XIST* is involved upstream to a set of regulators that induce the usual myogenesis program, which are suppressed as the result of its action on some narrow, and unknown, set of mediators. The ultimate physiologic purpose is not immediately clear, although the context of the niche may provide a clue: the extracellular environment of the heart is likely to be awash in endocrine and paracrine signals that induce myogenesis; these signals may well infiltrate non-muscular cells and induce this transcriptional program; some baseline level of its expression may be generally tolerable, but interfere with the function of this particular cell type; *XIST*-mediated inhibition happens to be the mechanism of choice for suppressing it. However, the causal and molecular details are unknown: *XIST* expression may mediate the effect, or merely be an incidental consequence of establishing the pseudo-glial transcription program. Even if *XIST* plays an active role, the mechanism may involve the regulation of X-linked genes in *cis*, a non-canonical pathway in *trans* [58, 59], or an altogether new mechanism, possibly through an uncharacterized isoform. Overall, it is challenging to ascertain the role of the cell type or its transcriptional program based solely on bioinformatic data; experimental validation would be necessary, although the apparent lack of homology to model organisms presents a serious barrier. Nevertheless, a recent study [60] independently identified this expression trend, proposed that the relevant *XIST* species is a truncated transcript generated from an alternative promoter, and characterized the non-XCI effects of its overexpression in HEK-293T cells, suggesting the relevant mechanisms may be identifiable *in vitro*.

It is additionally tempting to connect this atypical cardiac expression of a sex-linked gene in males to the increased incidence of heart failure in males [61]. Based on the relatively scant evidence, the expression of *XIST* and the suppression of the reported program may plausibly be protective, deleterious, or both depending on the broader context of the organ and the organism. In Section 4.2.5, we examine this question, and find that the expression of *XIST* appears to be higher in samples with pathology or comorbidities in three relatively large datasets, but the smaller datasets are less illuminating.

In principle, this scenario should be an ideal application of genomics: the gene is specific to a single cell type, so any genome-wide association study (GWAS) results connecting variability in male *XIST* to cardiac phenotypes should reflect the functional role of *XIST* in the pseudo-glia. Unfortunately, sex chromosomes are generally omitted from GWAS [18–20]. For instance, a recent GWAS of the determinants of dilated cardiomyopathy in tens of thousands of patients and over a million controls analyzes and reports some fifteen million variants, all on the autosomes [62]. The Cardiovascular Disease Knowledge Portal, which aggregates hundreds of genetic association datasets, reports one SNP (rs1474563) in the *XIST* region, associated with height [63]. The only

pathogenic variant explicitly reported in dbSNP [64] is rs773396320, just upstream of the first exon of *XIST*, which leads to skewed XCI [65]. In addition, in spite of continuing efforts, variants in lncRNAs remain understudied [21]. Whether due to less study, greater exposure to natural selection [18], or selection biases (for instance, if *XIST* loss of function in males leads to embryonic mortality), the data are sparse and do not illuminate the cardiac role of *XIST*.

To enhance sensitivity to male-specific effects, we interrogated the *XIST* genomic locus in a sex-stratified association study for myocardial infarction [66]. In males, we identified a significant association of the *XIST* locus with myocardial infarction ($p = 1.3 \times 10^{-4}$, n =160,316 cases / 6,672 controls), defined by the variant rs187334110 for which *XIST* is the nearest gene. This association of *XIST* with myocardial infarction is not seen in females ($p = 0.3$, n =192,586 cases / 1,567 controls), suggesting that variation at this locus may contribute to heart disease in a sex-dependent manner. This connection is intriguing, and suggests that the dearth of large-scale X-chromosomal studies represents a lacuna that may reveal ties to sex-specific pathology under further investigation.

Alternatively, the role of *XIST* may be elucidated by cross-referencing its pseudo-glial expression against the male donors' genotypes. For instance, a particular polymorphism may be associated with higher *XIST* expression in one of studies considered here, and with greater risk of cardiac conditions in the context of a larger GWAS. Several of these studies summarize the representation of known heart condition-associated variants in expression or chromatin accessibility data, albeit on the level of peaks or counts and without donor genotyping [31, 33, 41, 67]. To our knowledge, the only large study reporting single-cell data alongside patient genotypes is the investigation of pathogenic variants in dilated cardiomyopathy by Reichart et al. [40]. In this study, variant identification was predominantly performed through 46- or 174-gene Illumina TruSight sequencing panels. Nine male donors were genotyped through whole exome or whole genome sequencing; however, neither raw nor intermediate genotyping data were released alongside the controlled-access transcriptomic data (EGAS00001006374). Many studies associate bulk RNA expression with variants [68, 69], but drawing conclusions about cell type-specific expression from bulk data remains problematic [70]. Therefore, characterizing the genetic correlates of male *XIST* expression in an unbiased way does not seem feasible at this time.

This analysis is necessarily cursory. Ideally, such analysis should be performed on harmonized data, quantified using the same reference and conventions. For instance, each of the investigated datasets quantified between 24,250 and 39,464 features; 10,391 overlapped between all studies, and only those were used for DE gene identification. It is plausible that biologically interesting effects are masked by the incidental omission of one or another gene due to choice of reference. In addition, processing and cell type assignment are highly subjective and manual, and seemingly arbitrary choices of quality control thresholds or hyperparameters typically lead to drastic differences [71]. The pseudo-glial population investigated here is distinctive and the detection of *XIST* is robust, but takes existing annotations at face value, and may be more or less reproducible in a comprehensive, consistent reanalysis. Such reanalysis would make this study more systematic, but would necessarily be a manual task, as any tools automating the process would have been trained and tested on the very datasets under scrutiny. Naturally, the reanalysis of raw data is impractical — in most cases impossible without institutional approval due to patient privacy concerns — and would encounter the same issues related to ambiguous or contradictory metadata reported here. It is also immaterial to the key finding, which is clear even on a superficial and qualitative investigation of publicly available data.

Finally, this unusual finding in plain sight, revealed by lncRNA expression, echoes,

and is directly inspired, by recent discovery of severe, previously uncharacterized artifacts in single-cell data, wherein nuclear transcripts are absent from subpopulations of certain cell types, perhaps suggesting quality control issues [72, 73]. We suggest that the biological result, as well as severe metadata errors in large studies, should urge stronger scrutiny of results, as well as a healthy amount of skepticism regarding approaches for automating analyses by summarizing the existing literature, or by tacitly accepting the results of potentially incomplete analyses as "ground truth."

4 Methods

4.1 Dataset curation and analyses

4.1.1 Human data acquisition

Amrute. The dataset released by Amrute et al. was acquired from GEO at GSE226314 [35]. The Seurat RDS file was converted to AnnData using the `SeuratDisk` R package. The `orig.ident` field was used as donor ID and `cell.type` field was used as cell type. All non-pathological donor samples were omitted, as they had been collected as part of an earlier study [74].

Sexes and ages were annotated manually based on metadata in Supplementary Table 1. Donors 388 and 378, listed as nonresponders in the dataset, were not listed in metadata, nor disclosed on the GEO entries. Sexes were inferred to be male based on expression of Y-chromosomal genes. Donors 370 and 371, listed as female and male nonresponders in the metadata, were not present in the dataset.

Pre- and post-LVAD sample IDs were retained in their original form, somewhat arbitrarily treating the time points as though they had been collected from distinct donors. This is a simplification, targeted at avoiding the elision of nontrivial remodeling effects reported by Amrute et al. It is possible that a more sophisticated statistical design would obviate the need for this simplification.

Brener. The count data for the dataset released by Brener et al. was acquired from GEO at GSE185457 [24]. These data do not have cell type annotations. The annotations were obtained from a version of the dataset previously available at the Broad Single Cell Portal (SCP1464), since removed from the database. Sexes and ages were annotated manually, based on metadata in Supplementary Table 5; donor `15_144548-RV` was female and all others were male. Only COVID-19 samples, collected for this study, were retained. The `batches` field was used as donor ID and `leiden` field was used as cell type.

Chaffin. The AnnData object released by Chaffin et al. was acquired from the Broad Single Cell Portal (SCP1303) [75]. The `donor_id` field was used as donor ID and `cell_type_leiden0.6` field was used as cell type. Ages were parsed using the `age` field.

Eraslan. Data were obtained from CZI CELL×GENE and filtered for the heart samples, which had the tissue annotation "anterior wall of left ventricle" [36]. Ages were annotated manually, based on metadata in Table S1; these data used different bins than the released AnnData file (suggesting, e.g., that the age of donor `GTEX-1ICG6` was 70, as it is listed in the 70–79 bin in one source and the 61–70 bin in the other). The `donor_id` field was used as donor ID and `cell_type` field was used as cell type.

Somewhat surprisingly, male donor `GTEX-1HSMQ` exhibited high *XIST* expression outside the heart Schwann cells, albeit only in the lung tissue sample. This does not

appear to be a sample metadata artifact, as *UTY* is also ubiquitous throughout. This expression pattern is somewhat challenging to account for.

Hill. Single-nucleus count data were obtained from the Broad Single Cell Portal (SCP1852) and merged into an AnnData structure [76]. The `donor_id` field was used as donor ID and `MainCellType` field was used as cell type. Ages were parsed using the `age` field, rounding down to the year. Donor `13_198`, which had two distinctly named samples corresponding to left and right ventricular samples, was kept split. UK1 and UK2 samples were mapped to FC3CB and 3B62D based on manual cross-referencing of ages.

Kanemaru. Data were obtained from CZI CELL×GENE and subset to primary data [31]. One cell attributed to donor D11 was removed, yielding 211,060 cells, in agreement with Fig. 1 of the original publication. Ages were annotated manually, based on metadata in Supplementary Table 1; these values largely agreed with those reported in the released AnnData structure, with the exception of donor AH2, reported as 45–50 years old in the table but 40–45 in the data file. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type.

Knight-Schrijver. Data were obtained from CZI CELL×GENE and subset to primary data [25]. As all primary data were fetal, ages were set to 0. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type.

Kuppe. Data were obtained from CZI CELL×GENE [33]. Ages were parsed from the `development_stage` field. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type.

Linna–Kuosmanen. Data were obtained from CZI CELL×GENE [32]. The two datasets (Periheart and Carebank) were concatenated. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type. The number of donors was rather larger than that reported in the original article and its metadata. We did not filter the dataset further. Ages were parsed from the `development_stage` field.

Litviňuková. Data were obtained from CZI CELL×GENE [23]. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type. Ages were annotated manually, based on metadata in Supplementary Table 1; the bins reported in the latter were finer than those disclosed in the dataset.

Reichart. Data were obtained from CZI CELL×GENE and subset to primary data [40]. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type. Ages were parsed from the `development_stage` field using the metadata provided in Supplementary Table 1.

Selewa. Data were obtained from CZI CELL×GENE and subset to primary data [41]. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type.

Sim. Data were obtained from CZI CELL×GENE and subset to primary data [54]. The `donor_id` field was used as donor ID and `cell_type` field was used as cell type. Ages were parsed from the `development_stage` field. All fetal samples' ages were set to 0.

Simonson. Data were obtained from the Human Cell Atlas Data Explorer [77]. The `donor_id` field was used as donor ID and `cell_type_leiden0.5` field was used as cell type.

Tucker. Data were obtained from the Broad Single Cell Portal (SCP498) [67]. The `biological.individual` field was used as donor ID and `Cluster` field was used as cell type. Ages and sexes were manually annotated based on Table 1.

4.1.2 Non-human

There is no evidence for *Xist* expression in pseudo-glia outside the humans, and relatively little data to draw upon to confirm or disconfirm its presence, despite the publication of several nonhuman heart sequencing studies. In general, analogous cell populations are robustly identifiable, but *Xist* is either not observed or not included in the genome reference. In this section, we summarize potential comparable data sources.

A 2018 study by Skelly et al. reports Schwann cells in mouse heart and uses *Xist* and Y-linked gene expression to effectively demultiplex pooled male and female samples, reporting Schwann cell differences between sexes but not remarking on any detection of, e.g., cells simultaneously positive for *Xist* and *Uty*, making their presence dubious [78]. The 10X scRNA-seq dataset generated as part of the Tabula Muris Senis project reports 58 cardiac neurons from male subjects, with no *Xist* expression, but this sample size may be too small to be conclusive or representative [79]. Several other developmental datasets have been collected; however, they do not report glial or neuronal cell types [80–82].

A study of sex-specific pressure overload in cats [83] reported a "neuronal/conduction" cluster with prominent expression of *NRXN1* but not *PLP1* or *NRXN3*. However, this dataset did not contain *XIST* in the reference, although the presence of *XIST* in cats is well-known and often serves as the first pedagogical introduction to X inactivation, using the calico pattern as a case study [84]. Similarly, the PANDORA atlas of non-model species reports small *NRXN1*/*PLP1*-expressing clusters in the cat and tiger hearts, annotated as cardiomyocytes and fibroblasts respectively [85]. However, neither dataset contains *XIST* expression data, presumably due to reference limitations.

Mouse: Mesquita, Li, Patrick, and Ranjbarvaziri via scBaseCount. A recent initiative, scBaseCount, attempts to facilitate the consistent analysis of transcriptomic data by applying a common quantification procedure to a large cross-section of public 10X single-cell datasets and hosting the results [86]. Without aiming to be exhaustive, we acquired a diversity of adult mouse heart datasets: two samples from Mesquita et al. [87], five from Li et al. [88], six from Patrick et al. [89], and four from Ranjbarvaziri et al. [90]. All mice were male, as confirmed by the expression of Y-chromosomal markers. However, all had relatively low expression of *Xist* (6–33 UMIs per sample). Clustering the *scVI* embedding of the data yields some 2,000 cells substantially enriched in the pseudo-glial markers [91, 92], but 0–1 UMIs of *Xist* per sample, suggesting that the population is present, but does not express the gene. Therefore, there does not appear to be any convincing evidence suggesting that the expression of *Xist* occurs in adult mouse hearts.

Mouse: Shen. A small study characterizing the transcriptomic landscape of the neonatal mouse heart, using four donors from the first six days after birth, reported a very small *Plp1*-containing neuronal population [27, 93]. We acquired the count matrices, generated using Cell Ranger 5.0.0, from GSE232466. The study did not

determine the pups' sex and did not disclose the cell type assignments. However, inspection of the count matrices suggested that the pups collected on days 1, 2, and 4 were male, as evidenced by relatively high (> 200 UMIs) expression of the Y-chromosomal genes *Uty*, *Ddx3y*, and *Kdm5d*; these genes showed negligible expression (< 5 UMIs) in the pup collected on day 6.

Interestingly, we observed remarkably high *Xist* expression ($> 40,000$ UMIs) in all samples, with the highest expression ($\approx 150,000$ UMIs) in the putative female sample. These quantifications translate to average male expression of 3.5–6.5 UMIs per barcode. Clustering the *scVI* embedding of the data yields 127 cells substantially enriched in the pseudo-glial markers [91, 92]. The presence of *Xist* was ubiquitous, and not restricted to the neuronal subpopulation, which had a similar average *Xist* expression (0.6–6.5 UMIs per barcode). In sum, although the male neonatal mouse heart expresses *Xist* to a surprisingly large degree, it does not appear to do so in a cell type-specific fashion.

For the sake of completeness, we note that *Xist* was depleted in one cluster, which expressed canonical muscle marker genes (*Myl2*, *Myl3*, *Tnnt2*), but did not express *Malat1*, a ubiquitous nuclear transcript. The lack of nuclear content — noted by the original authors — strongly suggests that this cluster is artifactual, and represents the *technical* failure to lyse the cardiomyocyte nucleus rather than the detection of a novel physiological cell type, previously reported in other tissues [72, 73].

Rat: Arduini. A recent large study consisting entirely of male Wistar rats [94] reported analogous glia-like cell types, but was based on a bespoke version of the Ensembl *R. norvegicus* `Rnor_6.0.96` genome without *Xist*. We realigned the raw data to quantify *Xist* expression.

We acquired raw reads for 19 left ventricular samples from GSE280111 [94]. We generated a pseudoalignment reference from the RefSeq GRCr8 genome ("genomic" `GCF_036323735.1`), using *kallisto* | *bustools* 0.28.2 [95]. We ran `kb ref` in the `nac` configuration, using the `d-list` option to add potential artifactual sequences — the 10X template-switch oligo, its reverse complement, and the $(A)_{30}$ oligomer — to the D-list for masking. We pseudoaligned the left ventricular samples using `kb count` in the `nac` configuration and inspected the unfiltered total count matrices to characterize the *Xist* content of the sequencing libraries. The datasets had negligible expression, with 0–4 molecules of *Xist* per dataset, none in the relevant cell type.

For comparison, we also quantified the amount of *Tsix*. This gene is not explicitly annotated in the reference. However, *LOC134484082* is a reasonable candidate: this transcript is antisense to *Xist*, and its longest exon almost completely covers exon 1 of *Xist*. Interestingly, this transcript is absent in the mRatBN7.2/rn7 assembly, which annotates *Xist* as *LOC100911498* or transcript `NR_132635.1`. In that assembly, the putative *Tsix* (*LOC680227*) is missing the portion of the 3' exon that covers the *Xist* promoter, previously reported by Shevchenko et al. [96] in their study of rodent *Xist* conservation.

Despite being based on the Rnor6 assembly, the Arduini et al. genome (acquired from GEO at GSE280111) did not contain *LOC100911498* or *LOC134484082*, but did contain putative *Tsix* (*LOC680227*), annotated as a protein-coding gene. We acquired count data from the Broad Single Cell Portal (SCP2828). The released dataset contained modest amounts of *Tsix* (9–44 UMIs per LV sample). This is somewhat lower than the *kallisto* | *bustools* quantification (53–204 UMIs), likely due to a combination of alignment strategy differences, filtering, and truncation of an exon.

Cynomolgus: Qu. A single-cell atlas of the cynomolgus monkey (*Macaca fascicularis*) released heart transcriptomic data. Although XCI has been studied in *M. fascicularis* [97, 98], the standard reference genomes do not contain *XIST*. The human

XIST is fairly similar to the *Macaca mulatta* long non-coding RNA *LOC106995245* (cover 30%, identity 92.3% relative to human transcript variant 1), suggesting it as a reasonable candidate for the *M. mulatta XIST* [99]. We used Liftoff [100] to map the RefSeq version of the *M. mulatta* genome `Mmul_10`/rheMac10 to the Ensembl *M. fascicularis* assembly `Macaca_fascicularis_6.0`, localizing the putative gene to X:67,721,448–67,753,629 in the latter genome. After extracting and harmonizing [101] the *LOC106995245* reference, we concatenated it to `Macaca_fascicularis_6.0` and generated a pseudoalignment reference using *kallisto* | *bustools* 0.28.2 [95], using `kb ref` in the `nac` configuration.

We pseudoaligned the male heart, male liver, and female liver samples released by Qu et al. using `kb count` in the `nac` configuration and inspected the unfiltered total count matrices to characterize the *LOC106995245* content of the sequencing libraries [102]. The female and male liver samples had 11,104 and 187 UMIs of *LOC106995245*, respectively, supporting the hypothesis that *LOC106995245* is the sex-specific gene *XIST*. The heart sample had a single UMI of *LOC106995245*.

The Qu et al. atlas does not explicitly report the detection of pseudo-glia, and it is conceivable that this rare cell type could simply be absent in this particular dataset. However, clustering the *scVI* embedding of `bustools`-filtered data yields some 300 cells substantially enriched in the pseudo-glial markers, but without any detectable *LOC106995245* [91, 92], suggesting that the population is present but does not express *XIST*.

Pig: Miao, Mendelson, Wang. At least four single-cell pig heart studies are available [103–106]. Three report small but distinct neuronal populations, marked by *Scn7a* [103, 106] and by *Nrxn1* [104]. In addition, a recent cross-tissue atlas sequenced heart samples from a mixture of three male pigs [107], albeit without explicitly reporting pseudo-glia (Figure S2 of [107]).

These studies' disclosed raw data matrices do not contain either *Xist* or *LOC102165344*, the gene's identifier in the 2011 Sscrofa10.2 assembly. Neither identifier is listed in the standard 2017 genome Sscrofa11.1. Porcine *Xist* has been annotated as locus KC149530 [108] and is known to be functional [108] in spite of nontrivial differences from the human gene [48], but characterization of its presence or absence in male porcine heart cell populations requires more comprehensive transcriptome build.

We used Liftoff [100] to map the RefSeq version of the *S. scrofa* genome Sscrofa10.2 to the Ensembl assembly Sscrofa11.1, localizing the putative gene to X:59,277,353–59,310,014 in the latter genome. After extracting and harmonizing [101] the *LOC102165344* reference, we concatenated it to Sscrofa11.1 and generated a pseudoalignment reference using *kallisto* | *bustools* 0.28.2 [95], using `kb ref` in the `nac` configuration.

Pig sexes were not reported in the Nakada and Wei studies [103, 105], leading us to reanalyze the remaining three studies. The Miao and Mendelson data were generated using the standard 10X Genomics 3' v3 technology. The single-nucleus data in the Wang atlas were generated using an MGI DNBelab C4 instrument, which uses two 10-base barcodes selected from a common whitelist and 10-base UMIs. The read structure and whitelists are disclosed by the manufacturer (`DNBelabC4_scRNA_readStructure.json` in [109]); we manually generated the corresponding 20-base cell barcode whitelist for *kallisto* | *bustools* . From the annotation, we inferred that the read direction was forward. This inference is consistent with trial pseudoalignments (e.g., SRR17661076 produces 11 million UMIs in the forward configuration but 6 million UMIs in the reverse configuration).

We pseudoaligned all samples using `kb count` in the `nac` configuration, setting the *kallisto* | *bustools* parameters appropriately. We analyzed the 10X and DNBelab

datasets separately to avoid embedding artifacts stemming from technology differences.

Clustering the *scVI* embedding of `bustools`-filtered data from the Miao and Mendelson studies yields 3,587 cells substantially enriched in the pseudo-glial markers, but with 0–3 molecules of *LOC102165344* per sample [104, 106], suggesting that the population is present but does not express *XIST*.

Clustering the *scVI* embedding of `bustools`-filtered data from the Wang study did not yield an identifiable pseudo-glial population (consistently with the original article), and the expression of the cell type's markers was sparse. This may be an artifact of the isolation procedure, an incidental consequence of the low cell count ($\approx$ 3000 cells passing QC in the authors' analysis), or, somewhat less plausibly, a consequence of the sequencing technology. The entire dataset contained $\approx$ 100 UMIs of *LOC102165344* across five samples, with no clear cell type enrichment.

None of the three studies had female samples that could serve as comparable positive controls for *Xist* expression. However, the Wang et al. atlas reports four samples generated from a single female pig (lung, blood, and two adipose samples), albeit using a single-cell protocol and the 10X chemistry (as opposed to single-nucleus and DNBelab used for the heart). Reanalyses of these data confirmed the substantial expression of *LOC102165344* in the lung and blood samples. However, the adipose samples had low expression of putative *Xist*, alongside high expression of the Y-linked genes *ZFY*, *KDM5D*, and *EIF1AY*, suggesting they were incorrectly labeled and originated from a distinct male pig.

4.2 Human data analysis and statistics

We used the cell type annotations and donor IDs, defined as in Section 4.1.1, to generate count sum pseudobulks using the Python package `decoupler` [110]. The parameters were as lenient as possible, with molecule and cell count thresholds set to zero.

Identification of cells in the pseudo-glial population was performed by string matching to cell type annotations. We extracted all cells whose cell type names included `Glia`, `glia`, `Neuron`, `neuron`, `Neural`, `neural`, or `Schwann`. This strategy takes the clusters at face value, and omits, e.g., plausibly glia-like cells assigned to other cell types or annotated as unknown.

XIST-positive pseudo-glia were defined naïvely, by identifying all cells belonging to the relevant clusters, binning them into zero- and nonzero-*XIST* categories based on expression, and pseudobulking based on donor ID and *XIST* presence. The presence or absence of a gene may be different between "processed" (e.g., `anndata.X`) and "raw" expression matrices (e.g., `anndata.raw.X`), which occurs when processing involves imputation (in the Eraslan, Selewa, and Sim datasets). We used the sparsity pattern of the raw data to split the datasets.

4.2.1 Expression comparisons

Figures 1, 2, and 6a were generated by computing sample means: extracting all relevant pseudobulks (glial or non-glial) for a donor, adding the gene counts, then dividing them by the sum of the corresponding `psbulk_n_cells` entries, i.e., the total number of cells.

4.2.2 Differential expression between male and female donors

To characterize the sex-dependent expression differences within the pseudo-glial cells, we aggregated all cells in these populations into coarser sum pseudobulks. Next, we fit a PyDESeq2 model using the donor sex and the source study as categorical covariates [111, 112]. Finally, we obtained the Wald test p-values and log fold differences

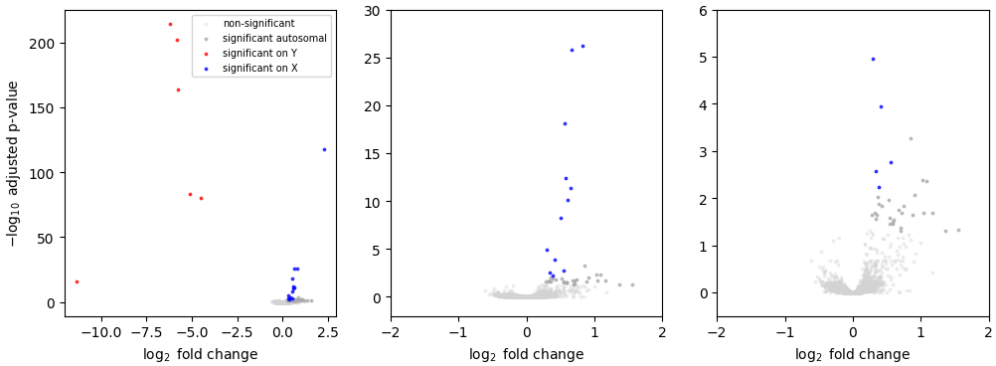

Fig 4. Results of differential expression testing between female and male samples, using sex and study as covariates. Genes with `baseMean` < 5 excluded from visualization. Differential expression is defined as $|\log_2 \mathrm{FC}| > 0.25$ and adjusted p-value < 0.05.

for the female/male contrast and visualized them in Fig 4, providing three axis scales to accommodate the outsize effect sizes of sex chromosomes.

Evidently, female samples have little to no expression of Y-chromosomal genes. However, they also have higher expression of certain X-chromosomal genes – most prominently by far, *XIST*, but also the XCI mediators *JPX* and *ZFX*. In addition, the expression of X-chromosomal genes *KDM6A*, *NLGN4X*, *AP1S2*, *EIF1AX*, and *CA5B* is 50–80% higher in the female samples, in the correct range for XCI escape (i.e., higher than in male, but not quite doubled).

It is curious to note that the sex analysis is analogous to a cross-study investigation very recently released by Gao et al. [34]. The volcano plots shown in Fig 4 partially reproduce Supplemental Figure 15a of that preprint. In addition, the results of this differential expression test immediately imply the key finding of the current study: the expression of *XIST* in the pseudo-glia (the "neuronal" population) is approximately fourfold lower in males than in females, about half an order of magnitude (Fig 1). This is in stark contrast to other cell types, where the male expression of *XIST* is far more sparse ($\log_2$ fold change below -5, and frequently even lower).

For the sake of completeness, we note that the analysis of epicardial cells (*ITLN1*/*MSLN*) in Supplemental Figure 9a of Gao et al. reports sex-specific enrichment of *XIST* quantitatively similar to the pseudo-glia, with $\log_2$ fold change ≈ -2. However, the cells expressing these markers (typically annotated as "mesothelial") do not appear to exhibit systematic male *XIST* expression [23, 31–33, 40]. Therefore, the reported enrichment of *XIST* in male epicardial cells may be an artifact of the drastic differences in this cell type's representation between studies (7% of cells in [32], 0.15% in [23], 0.1% in [31], and not readily detectable in [33, 41]). This imbalance may lead to suboptimal clustering or to inaccurate estimation of pseudobulk "size factors," either of which can lead to miscalibration of differential expression testing.

4.2.3 Differential expression between *XIST*-positive and -negative cells

Ideally, the study of *XIST* in the pseudo-glial population should fully characterize the impact of its presence or absence on the molecular makeup and phenotype of the cell population, its niche, the tissue, and the organism. Thus far, there is no evidence that this expression pattern occurs outside humans (Section 4.1.2), precluding *in vivo* study. It is possible that part of the context may be recapitulated with an appropriately constructed tissue model. A recent study reported that a heart organoid model produces a small population of conductance cells [113] that resemble embryonic

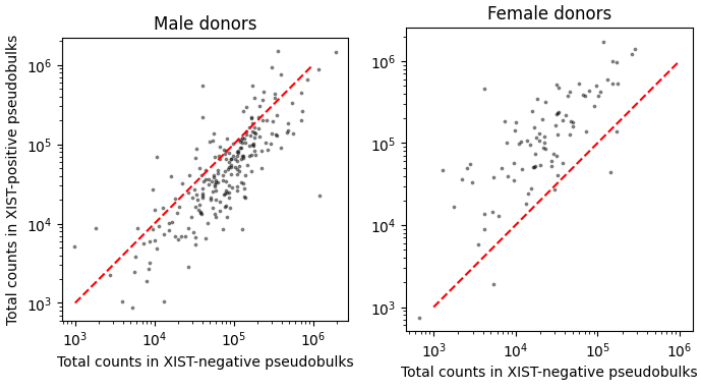

Fig 5. Total molecule counts in the pseudobulks used to evaluate differential expression between *XIST*-positive and -negative cells (points: donor pseudobulks; lines: identity).

Schwann and neural cell progenitors [114]. However, this population's markers are not expressed in adult pseudo-glia, so the generation of a model that recapitulates the adult heart remains an outstanding problem.

In the absence of an experimental system that can be perturbed (e.g., by knocking out *XIST*), it is possible to approximate regulatory relationships between genes using endogenous patterns of co-expression. This approach falls under the umbrella of gene regulatory network (GRN) inference. Although this is a thriving field with numerous tools, the problem of GRN inference is far from solved [115–117], and applying this class of algorithms to a rare cell type scattered across hundreds of donor samples collected under different conditions is unlikely to be informative.

In lieu of attempting to discover genome-wide networks, we restrict ourselves to describing the immediate impacts of *XIST* presence or absence. This classification is necessarily imperfect. Due to contamination, cells with no *XIST* may be reported as having *XIST*. Conversely, due to the limitations of cDNA library construction and sequencing, some molecules may not be captured. The background contamination – if comparable to that in the heart overall (tentatively, the yellow triangles in Fig 1) – is unlikely to be more than a minuscule fraction of the total. The issue of false negatives is more challenging to bypass. However, the misassignment of *XIST*-positive cells as *XIST*-negative attenuates, rather than increases differences between the categories, so this approach is conservative.

Bearing these limitations in mind, we investigated the differential expression of genes between the *XIST*-positive and -negative cells, applying PyDESeq2 to pseudobulks. As we are interested in isolating the effects of intra-individual variation, we encoded the donor ID as a categorical covariate alongside *XIST* status. This results in a relatively complex problem, with 226 covariates for the 225 male donors with *XIST* expression and 90 covariates for the 89 corresponding female donors. We fit the model using the intersection of genes identified in all datasets. We used a PyDESeq2 Wald test to compute p-values, then filtered the results further for genes with `baseMean` > 5. As the p-values are adjusted based on the entire model, the values shown in Fig 3 are conservative. We note this filtering is rather stringent, as it discards some 78% of the 10,391 genes intersecting between datasets. Although it is likely that some of the omitted genes are important, the small sample sizes (61% of the male pseudobulks have fewer than fifty cells; 16% have fewer than ten) produce an considerable risk of false positives for low-expression genes.

Some reasonable artifactual explanations for the differential expression patterns observed in Fig 3 involve the diffusion of abundant myogenic transcripts into droplets

containing pseudo-glia. In other words, the enrichment of background genes may take place due to insufficiently accounted-for "size" effects, which may occur through technical or combination technical and biological mechanisms. In the first scenario, *XIST*-negative populations incidentally have fewer overall sequencing primers, and capture fewer endogenous and exogenous transcripts. As shown in Fig 5, this intuition holds for female samples but does not hold for male samples, which tend to have somewhat *less* RNA.

In the second scenario, *XIST*-positive populations systematically have lower overall endogenous genome-wide expression (presumably due to *XIST*-mediated repression, incidentally consistent with Fig 5), and relatively more of the cDNA library is sourced from the contaminating transcripts. In both scenarios, one would expect *XIST*-positive populations to have *higher* rather than lower expression of ubiquitous muscle genes, the opposite of what is observed. Although it is challenging to make definitive conclusions based on summary data collected using heterogeneous procedures, the comparison of paired samples is rather more suggestive of endogenous suppression than exogenous contamination.

4.2.4 *XIST* relationship to age

To characterize the age-dependent expression variability within the pseudo-glial cells, we followed the procedure in Section 4.2.2 to construct coarse pseudobulks, subset the data to non-fetal males with age annotations, then fit a PyDESeq2 model using the source study and the age as covariates. As elsewhere, when only coarse bins were available, we used the bin midpoint as an imperfect proxy for the true age.

First, we used a model which directly used the age as a continuous covariate. The fit yielded an effect size of 0.007, with a p-value of 0.05. This effect is small and dubiously significant. Next, we used a model which used the logarithm of the age as the covariate. This analysis yielded an effect size of 0.27, with a p-value of 0.01. In contrast, the genes with the lowest p-values in the latter analysis, filtering for those with `baseMean` > 10, were *PPARG* (effect size 0.89 and $p_{\text{adj}} = 5.9 \times 10^{-3}$) and *FSTL3* (effect size 0.71 and $p_{\text{adj}} = 7.8 \times 10^{-3}$), both of which are known to serve roles in heart failure pathology [118–120]. We note that the targeted inspection of *XIST* uses the nominal p, whereas the untargeted inspection differential expression tables uses the p-value adjusted for multiple comparisons. Although it may be the case that *XIST* slightly increases in adolescence, there does not appear to be strong evidence for a continuing increase in expression during adulthood, whether considering the sample mean (Fig 6a), the PyDESeq2 "size-normalized" counts (Fig 6b, top), or the latter with additional correction for the batch covariates (Fig 6b, bottom). In contrast, the age dependence of *PPARG* and *FSTL3* is somewhat less equivocal, if slight. However, these results are somewhat challenging to interpret without appropriately accounting for the uncertainty represented by binned age metadata.

4.2.5 *XIST* relationship to pathology

Given the abundance of pathological samples in several publicly available studies, as well as the lack of evidence regarding the protective or pathological roles of *XIST*, we selected the largest datasets for exploratory analysis of the relationship between *XIST* expression and heart failure phenotypes.

Chaffin. The dataset released by Chaffin et al. had data for healthy controls as well as hypertrophic and dilated cardiomyopathy (HCM and DCM) patients. We separately compared *XIST* expression between the pathology classes and healthy controls. A PyDESeq2 Wald test yielded higher expression in the HCM/control comparison (effect

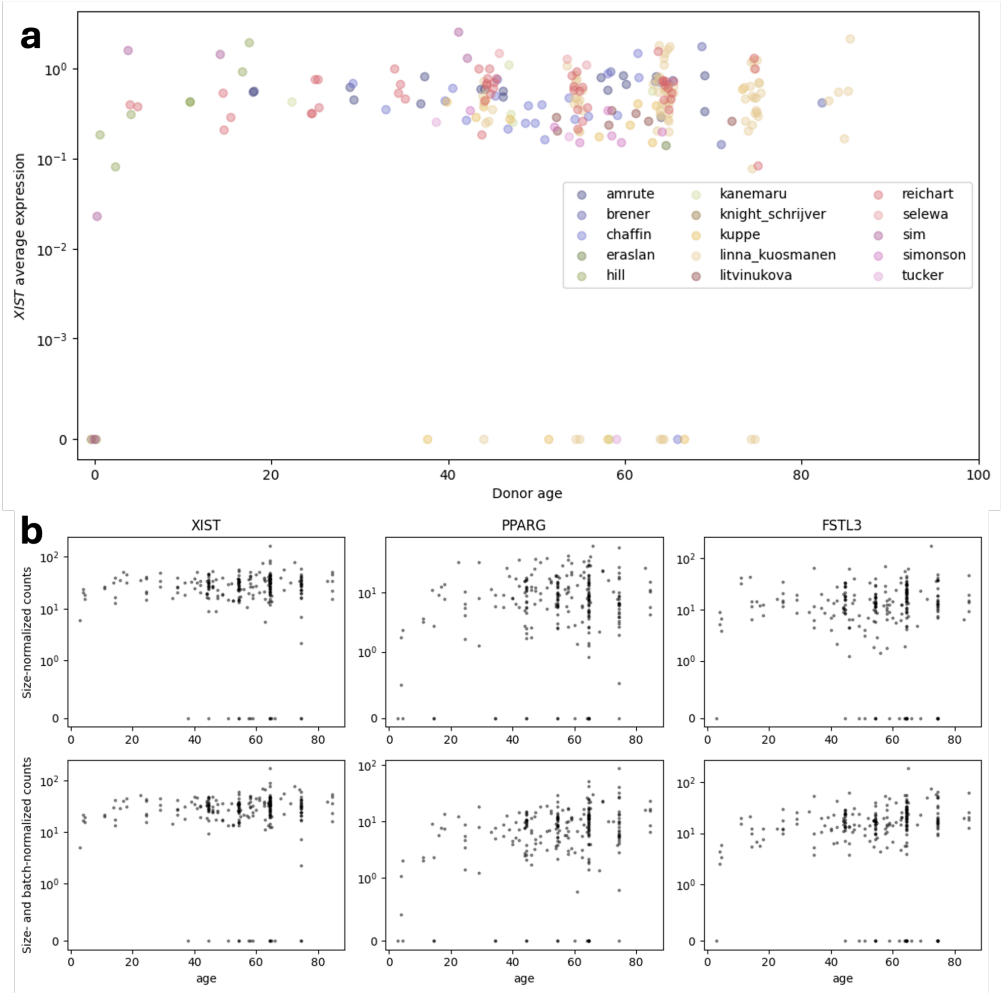

Fig 6. Age dependence of cardiac gene expression.
a. Average expression of *XIST* in pseudo-glia in male donors across ages. Colors denote study of origin. All fetal samples are given age zero. Donors with no age information omitted. Jitter added. Data with binned ages (e.g., 20–29 years) are visualized at the bin midpoint. To accommodate zero values, the region $[0, 10^{-3}]$ is linear.
b. Expression of *XIST*, *PPARG*, and *FSTL3* in pseudo-glia in male donors. Data with binned ages (e.g., 20–29 years) are visualized at the bin midpoint. Top: size-normalized pseudobulk counts. Bottom: size-normalized pseudobulk residuals, after correction for batch variables, inferred from a batch/log-age fit. To accommodate zero values, the region $[0, 1]$ is linear.

size 0.88 and $p = 1.3 \times 10^{-3}$) and in the DCM/control comparison (effect size 0.98 and $p = 1.3 \times 10^{-3}$), with similar results under models incorporating age or the logarithm of age alongside the pathology.

Reichart. The dataset released by Reichart et al. had data for healthy controls as well as several classes of cardiomyopathy. We compared *XIST* expression between the largest classes (DCM and arrhythmogenic right ventricular cardiomyopathy, ARVCM) and healthy controls. A PyDESeq2 Wald test yielded higher expression in the DCM/control comparison (effect size 1.0 and $p = 5.9 \times 10^{-4}$) and in the ARVCM/control comparison (effect size 1.3 and $p = 9.9 \times 10^{-4}$) with similar results under models incorporating age or the logarithm of age alongside the pathology.

Linna–Kuosamanen. The dataset released by Linna–Kuosamanen et al. had patients with a variety of conditions; all patients had some form of valve disease. In addition, the released AnnData files annotated numerous donors as having atrial fibrillation (AF), in conflict with the metadata: for instance, of the fourteen non-control Periheart donors, zero are reported as having AF in the article, but four are reported as having AF in the dataset. As the original publication treated valvular disease patients as "controls," we binned the samples accordingly, comparing patients with comorbidities to valve disease-only patients. A PyDESeq2 Wald test yielded slightly higher expression in the comorbidity class (effect size 0.27 and $p = 0.035$), with similar results under models incorporating age or the logarithm of age alongside the pathology.

Kuppe. The dataset released by Kuppe et al. compared twelve myocardial infarction (MCI) patients to three controls. A PyDESeq2 Wald test yielded slightly, but non-significantly lower expression in the MCI class (effect size -0.55 and $p = 0.55$), with similar results under models incorporating age or the logarithm of age alongside the pathology.

Simonson. The dataset released by Simonson et al. compared three ischemic cardiomyopathy (ICM) patients to four controls. A PyDESeq2 Wald test yielded slightly, but non-significantly lower expression in the ICM class (effect size -0.95 and $p = 0.18$), with similar results under models incorporating age or the logarithm of age alongside the pathology.

Summary. Evidently, the Chaffin, Reichart, and Linna–Kuosamanen datasets suggest stronger *XIST* expression in more pathological samples, whereas the others are not informative. As suggested in Section 3, reanalysis would be necessary to draw quantitative conclusions. For instance, plausible pseudo-glia may have been removed in the authors' QC pipelines, or incidentally assigned to other cell types, biasing the expression profiles, and a reanalysis may recover them. However, it is unlikely that even such follow-up would be conclusive: subtle details may differ by age, sex, chamber, pathology, its severity, and perhaps subpopulation. Given these caveats, as well as the rarity of the cell type, it appears likely that the pathological role, if any, can only be determined with thorough characterization of its function in healthy tissues, followed by the extensive and targeted collection of new data.

4.3 Short read analysis

Read alignment files were acquired from the Sequence Read Archive depositions accompanying the GEO accession GSE161921 [44]. Specifically, we acquired the sorted BAM files generated by Cell Ranger alignments of data for male donors 2879 and 2880,

as well as female donors 2867 and 2884. The alignments were filtered to retain only the reads overlapping *XIST* using the `samtools` 1.20 command `view` for the region `X:73818656-73854714`, then filtered further to include only reads with the `GN` tag [121]. This procedure retains all reads that would normally be quantified, but omits, e.g., uncharacterized transcripts. The resulting subsets were visualized using the Integrative Genomics Viewer [122] with reference to the GRCh38/hg38 genome.

4.4 Long read analysis

A recent study by Pan et al. [45] constructed 10X cDNA libraries from isolated heart nuclei, then sequenced them using long-read nanopore (ONT PromethION) sequencing, aiming to characterize the isoform diversity of the healthy and pathological human heart. We cross-referenced the findings from the 10X/Illumina datasets from other studies against the Pan et al. data, but did not analyze them jointly.

Gene quantification results were obtained from GEO accession GSE288222. Sample- and cell-level metadata were obtained from Tables S1 and S8 of the original publication, with 28,534 out of 59,488 barcodes having cell type annotations.

The pseudo-glial population was only reported in 4/6 female samples and 5/6 male samples. Three male samples had no *XIST* expression in this cell type, and one had negligible expression at a level of ≈ 0.01 UMIs per barcode (as opposed to 0.9–3.1 UMIs per barcode in female samples). We hypothesized that this result is a consequence of the alignment strategy: as the reads reported in Section 4.3 are predominantly localized to the 3' exon of *XIST*, an unstranded alignment would be unable to disambiguate *XIST* reads from antisense *TSIX* reads. The authors do not explicitly report the alignment setting, which is chosen automatically as part of the *Minimap2* pipeline [123], but one corroborating line of evidence is the amount of *TTN-AS1*, a transcript antisense to titin. Previous heart atlases (e.g., [23, 31]) report considerable amounts of *TTN-AS1* in cardiomyocytes, but the counts are relatively low in the Pan et al. data. However, it is challenging to judge the discrepancies: *TTN-AS1* reported in short-read atlases may itself be a *TTN*-derived antisense artifact that the nanopore protocol happens to avoid. In addition, the discrepancies may result from the omission or misassignment of cells in the rare population due to pre-processing choices.

We obtained six male datasets and one female dataset (ND245) from the Sequence Read Archive and split the nanopore reads into technical/biological read pairs using the *scNanoGPS* utility Scanner [124] per the authors' procedure. This produced a set of reads in the reverse configuration, starting with a poly(T) tail (as opposed to forward-configuration reads conventionally produced by the 10X protocol).

At this juncture, we diverged from the Pan et al. procedure. We generated a *kallisto* | *bustools* index from the human genome (10X Genomics, 2024-A version) using $k = 63$, then used the long-read `kb count` pipeline [125] in the `nac` configuration, assuming reverse strandedness, to pseudoalign the reads.

Clustering the *scVI* embedding of knee plot-filtered data yields 364 cells substantially enriched in the pseudo-glial markers (10–136 per male sample). Pseudo-glia show average male *XIST* expression between 0.06 and 0.5 UMIs per barcode (1–29 UMIs per pseudobulk), whereas other cells (excluding empty drops) show expression below 0.008 UMIs per barcode, consistently with Figure 1. As expected, the female sample shows high *XIST* across the cellular landscape (≈ 1.4 UMIs per barcode in pseudo-glia, 1.9 otherwise). Curiously, *TTN-AS1* is present in relatively high amounts (tens of thousands of UMIs, vs. dozens of UMIs in disclosed data), suggesting that the antisense titin cDNA are either derived from physiological molecules or pre-fragmentation artifacts.

In the interest of being exhaustive, we tested the differential expression of *XIST* between 4 heart failure and 2 non-diseased samples, as in Section 4.2.5. As may be

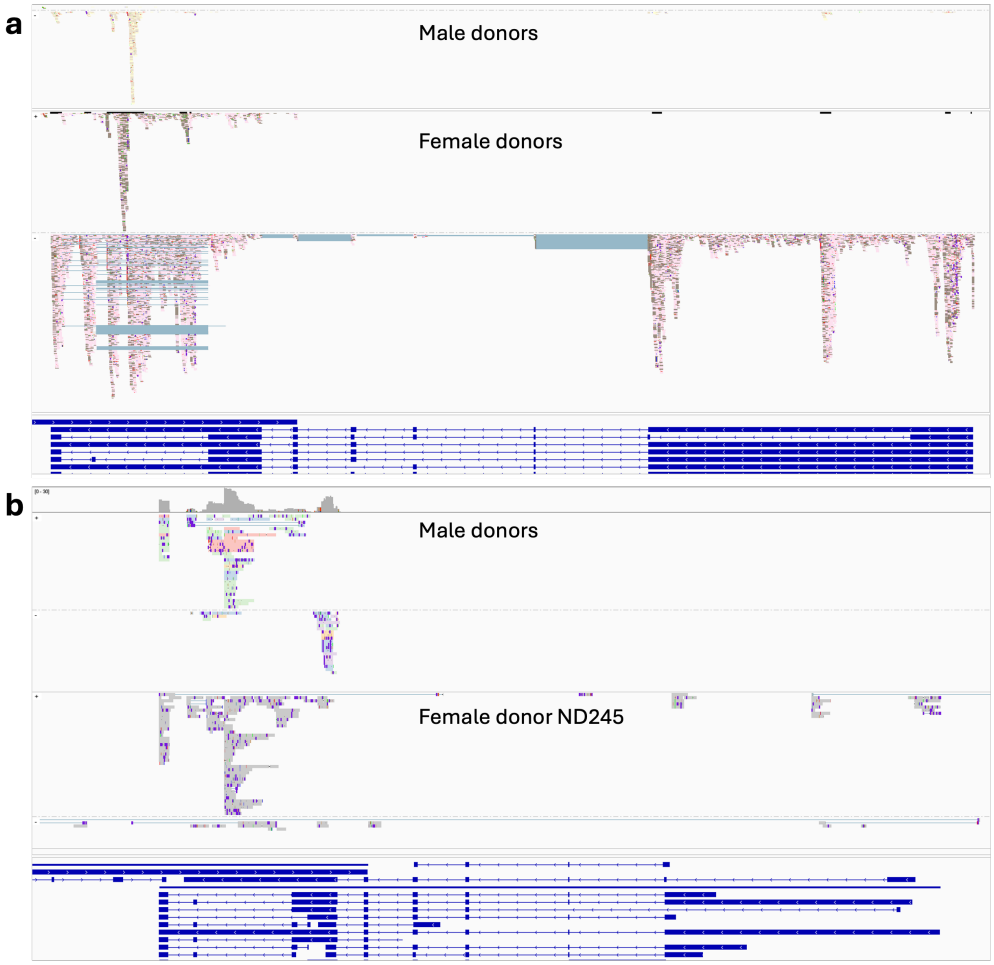

Fig 7. Read coverage and distribution of *XIST* transcripts.
a. Read alignments of Larson and Chin [44] septum samples. Top panel: male donors (color-coding: donor identity; dark blue lines: splice junctions; reads are split by strand, with positive at the top). Middle panel: female donors (conventions as in top panel). Bottom panel: Representative RefSeq transcripts in the region.
b. Read alignments of Pan et al. [45], restricted to pseudo-glial samples (conventions as in **a**, with the negative strand at the top of each subpanel).

expected from the small sample size, the comparison was inconclusive (effect size -0.41 and $p = 0.64$).

In principle, long-read sequencing can be used to elucidate transcript identities better than short-read sequencing. To check the transcript structures, we identified barcode/UMI pairs for reads mapping to *XIST* and *TSIX*, extracted the corresponding entries from the biological read files, trimmed the poly(T) sequences, and remapped them to a section of the genome immediately surrounding this gene pair using *Minimap2* [123].

Data availability

No new experimental data were generated for this study. All datasets investigated here are available at, and were obtained from, CZI CELL×GENE, Gene Expression Omnibus, the Broad Single Cell Portal, the Human Cell Atlas Data Explorer, or the Sequence Read Archive. The sole exception is the set of annotated data originally released on the Single Cell Portal by Brener et al. [24], which was obtained prior to its removal from the database. The dataset released on GEO contains data for the same cells and genes, and information regarding donor identities and various quality control metrics, but does not contain cluster identities; to our knowledge, there are no publicly available resources that continue to host this dataset. The details of the data acquisition and curation process are given in Section 4.1.

The Python notebooks used to perform the analyses in this study are hosted at `https://github.com/Fauna-Bio/GDG_2025`. The processed data (differential expression tables, *XIST*-aligning reads, count matrices, pseudobulks used for statistical tests, and genome reference entries) are hosted on Zenodo [126]. The intermediate data (split long reads, pseudoalignments) are available on request.

The differential expression analyses used PyDESeq2 0.4.8 [111]. The visualization of reads used the Integrative Genomics Viewer [122]. The pseudoalignment of rat data used *kallisto* | *bustools* 0.28.2 with the GRCr8 genome. The pseudoalignment of cynomolgus data used *kallisto* | *bustools* 0.28.2 with the `Macaca_fascicularis_6.0` genome, augmented with the putative *M. mulatta XIST*. The pseudoalignment of pig data used *kallisto* | *bustools* 0.28.2 with the Sscrofa11.1 genome, augmented with the putative Sscrofa10.2 *XIST*. The pseudoalignment of long-read heart data used *kallisto* | *bustools* 0.29.5 with the 2024 version of the `GRCh38` genome released by 10X Genomics. The alignment of putative *XIST* long reads used *Minimap2*.

Declaration of interests

L.G. is a co-founder and CTO of Fauna Bio. G.G. and N.D. are employees of Fauna Bio.

Acknowledgments

L.G. is a co-founder and CTO of Fauna Bio. G.G. and N.D. are employees of Fauna Bio. The authors would like to acknowledge Tara Chari, Sina Booeshaghi, Prashant Bhat, Phil McNamara, Evelyn Tran, Alan Cohen, Natalie DeForest, and Alex Burr for their assistance in the preparation of this manuscript. The visualizations in this manuscript use the `LaCroixColoR` palette by Dave Armitage and Johannes Bjork.

References

1. Augui S, Nora EP, Heard E. Regulation of X-chromosome inactivation by the X-inactivation centre. Nature Reviews Genetics. 2011;12(6):429–442. doi:10.1038/nrg2987.

2. Dixon-McDougall T, Brown C. The making of a Barr body: the mosaic of factors that eXIST on the mammalian inactive X chromosome. Biochemistry and Cell Biology. 2016;94(1):56–70. doi:10.1139/bcb-2015-0016.

3. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 2005;434(7031):400–404. doi:10.1038/nature03479.

4. Berletch JB, Yang F, Xu J, Carrel L, Disteche CM. Genes that escape from X inactivation. Human Genetics. 2011;130(2):237–245. doi:10.1007/s00439-011-1011-z.

5. Souyris M, Cenac C, Azar P, Daviaud D, Canivet A, Grunenwald S, et al. *TLR7* escapes X chromosome inactivation in immune cells. Science Immunology. 2018;3(19):eaap8855. doi:10.1126/sciimmunol.aap8855.

6. Jacobson EC, Pandya-Jones A, Plath K. A lifelong duty: how Xist maintains the inactive X chromosome. Current Opinion in Genetics & Development. 2022;75:101927. doi:10.1016/j.gde.2022.101927.

7. Sadagopan A, Nasim IT, Li J, Achom M, Zhang CZ, Viswanathan SR. Somatic XIST activation and features of X chromosome inactivation in male human cancers. Cell Systems. 2022;13(11):932–944.e5. doi:10.1016/j.cels.2022.10.002.

8. McCarrey JR, Watson C, Atencio J, Ostermeier GC, Marahrens Y, Jaenisch R, et al. X-chromosome inactivation during spermatogenesis is regulated by an Xist/Tsix-independent mechanism in the mouse. genesis. 2002;34(4):257–266. doi:10.1002/gene.10163.

9. Yan W, McCarrey JR. Sex Chromosome Inactivation in the Male. Epigenetics : official journal of the DNA Methylation Society. 2009;4(7):452–456.

10. Weakley SM, Wang H, Yao Q, Chen C. Expression and Function of a Large Non-coding RNA Gene XIST in Human Cancer. World Journal of Surgery. 2011;35(8):1751–1756. doi:10.1007/s00268-010-0951-0.

11. Meester I, Frausto GAP, Solis-Soto JM. *XIST* expression in male leukocytes. The Journal of Immunology. 2023;210(Supplement_1):167.37–167.37. doi:10.4049/jimmunol.210.Supp.167.37.

12. Sousa EJ, Stuart HT, Bates LE, Ghorbani M, Nichols J, Dietmann S, et al. Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. Cell Stem Cell. 2018;22(6):919–928.e6. doi:10.1016/j.stem.2018.05.001.

13. Gayen S, Maclary E, Hinten M, Kalantry S. Sex-specific silencing of X-linked genes by Xist RNA. Proceedings of the National Academy of Sciences. 2016;113(3). doi:10.1073/pnas.1515971113.

14. Dou DR, Zhao Y, Belk JA, Zhao Y, Casey KM, Chen DC, et al. Xist ribonucleoproteins promote female sex-biased autoimmunity. Cell. 2024;187(3):733–749.e16. doi:10.1016/j.cell.2023.12.037.

15. Ji B, Higa KK, Kelsoe JR, Zhou X. Over-expression of XIST, the Master Gene for X Chromosome Inactivation, in Females With Major Affective Disorders. EBioMedicine. 2015;2(8):909–918. doi:10.1016/j.ebiom.2015.06.012.

16. Gadek M, Shaw CK, Abdulai-Saiku S, Saloner R, Marino F, Wang D, et al. Aging activates escape of the silent X chromosome in the female mouse hippocampus. Science Advances. 2025;11(10):eads8169. doi:10.1126/sciadv.ads8169.

17. Gupta K, Czerminski JT, Lawrence JB. Trisomy silencing by XIST: translational prospects and challenges. Human Genetics. 2024;143(7):843–855. doi:10.1007/s00439-024-02651-8.

18. Gorlov IP, Amos CI. Why does the X chromosome lag behind autosomes in GWAS findings? PLOS Genetics. 2023;19(2):e1010472. doi:10.1371/journal.pgen.1010472.

19. Sun L, Wang Z, Lu T, Manolio TA, Paterson AD. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? American Journal of Human Genetics. 2023;110(6):903–912. doi:10.1016/j.ajhg.2023.04.009.

20. Bellavance J, Wang L, Gagliano Taliun SA. Eight quick tips for including chromosome X in genome-wide association studies. PLOS Computational Biology. 2024;20(6):e1012160. doi:10.1371/journal.pcbi.1012160.

21. Yang Y, Wang D, Miao YR, Wu X, Luo H, Cao W, et al. lncRNASNP v3: an updated database for functional variants in long non-coding RNAs. Nucleic Acids Research. 2023;51(D1):D192–D198. doi:10.1093/nar/gkac981.

22. Miranda AMA, Janbandhu V, Maatz H, Kanemaru K, Cranley J, Teichmann SA, et al. Single-cell transcriptomics for the assessment of cardiac disease. Nature Reviews Cardiology. 2023;20(5):289–308. doi:10.1038/s41569-022-00805-7.

23. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, et al. Cells of the adult human heart. Nature. 2020;588(7838):466–472. doi:10.1038/s41586-020-2797-4.

24. Brener MI, Hulke ML, Fukuma N, Golob S, Zilinyi RS, Zhou Z, et al. Clinico-histopathologic and single-nuclei RNA-sequencing insights into cardiac injury and microthrombi in critical COVID-19. JCI Insight. 2022;7(2):e154633. doi:10.1172/jci.insight.154633.

25. Knight-Schrijver VR, Davaapil H, Bayraktar S, Ross ADB, Kanemaru K, Cranley J, et al. A single-cell comparison of adult and fetal human epicardium defines the age-associated changes in epicardial activity. Nature Cardiovascular Research. 2022;1(12):1215–1229. doi:10.1038/s44161-022-00183-w.

26. Zanders L, Glaser SF, Ruz Jurado M, Brandt M, John D, Nicin L, et al.. Deciphering human heart failure with preserved ejection fraction (HFpEF) at single cell resolution; 2025. Available from: `http://biorxiv.org/content/early/2025/04/03/2025.04.02.646923.abstract`.

27. Shen J, Ma L, Hu J, Li Y. Single-Cell Atlas of Neonatal Mouse Hearts Reveals an Unexpected Cardiomyocyte. Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease. 2023;12(23):e028287. doi:10.1161/JAHA.122.028287.

28. Wake E, Brack K. Characterization of the intrinsic cardiac nervous system. Autonomic Neuroscience. 2016;199:3–16. doi:10.1016/j.autneu.2016.08.006.

29. Herring N, Paterson DJ. The Heart's Little Brain: Shedding New Light and CLARITY on the "Black Box". Circulation Research. 2021;128(9):1297–1299. doi:10.1161/CIRCRESAHA.121.319148.

30. Scherschel K, Wolf H, Ajijola OA, Shivkumar K, Lindner D, Gomez-Sanchez JA, et al. Glia of the heart's nervous system. Nature Reviews Neuroscience. 2025;doi:10.1038/s41583-025-00974-7.

31. Kanemaru K, Cranley J, Muraro D, Miranda AMA, Ho SY, Wilbrey-Clark A, et al. Spatially resolved multiomics of human cardiac niches. Nature. 2023;619(7971):801–810. doi:10.1038/s41586-023-06311-1.

32. Linna-Kuosmanen S, Schmauch E, Galani K, Ojanen J, Boix CA, Örd T, et al. Transcriptomic and spatial dissection of human ex vivo right atrial tissue reveals proinflammatory microvascular changes in ischemic heart disease. Cell Reports Medicine. 2024;5(5):101556. doi:10.1016/j.xcrm.2024.101556.

33. Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, et al. Spatial multi-omic map of human myocardial infarction. Nature. 2022;608(7924):766–777. doi:10.1038/s41586-022-05060-x.

34. Gao W, Hu P, Qiu Q, Kang X, Bedi K, Sasaki K, et al.. Integrated single-cell and spatial multiomic analysis reveals widespread reactivation of developmental programs in diseased human hearts; 2025. Available from: `http://biorxiv.org/content/early/2025/04/07/2025.04.01.646460.abstract`.

35. Amrute JM, Lai L, Ma P, Koenig AL, Kamimoto K, Bredemeyer A, et al. Defining cardiac functional recovery in end-stage heart failure at single-cell resolution. Nature Cardiovascular Research. 2023;2(4):399–416. doi:10.1038/s44161-023-00260-8.

36. Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Science. 2022;376(6594):eabl4290. doi:10.1126/science.abl4290.

37. Mahyari E, Guo J, Lima AC, Lewinsohn DP, Stendahl AM, Vigh-Conrad KA, et al. Comparative single-cell analysis of biopsies clarifies pathogenic mechanisms in Klinefelter syndrome. American Journal of Human Genetics. 2021;108(10):1924–1945. doi:10.1016/j.ajhg.2021.09.001.

38. Karczewski KJ, Solomonson M, Chao KR, Goodrich JK, Tiao G, Lu W, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. Cell Genomics. 2022;2(9):100168. doi:10.1016/j.xgen.2022.100168.

39. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, et al. Landscape of X chromosome inactivation across human tissues. Nature. 2017;550(7675):244–248. doi:10.1038/nature24265.

40. Reichart D, Lindberg EL, Maatz H, Miranda AMA, Viveiros A, Shvetsov N, et al. Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. Science. 2022;377(6606):eabo1984. doi:10.1126/science.abo1984.

41. Selewa A, Luo K, Wasney M, Smith L, Sun X, Tang C, et al. Single-cell genomics improves the discovery of risk variants and genes of atrial fibrillation. Nature Communications. 2023;14(1):4999. doi:10.1038/s41467-023-40505-5.

42. Booeshaghi AS, Galvez-Merchán Á, Pachter L. Algorithms for a Commons Cell Atlas; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.03.23.586413`.

43. Galvez-Merchán Á, Booeshaghi AS, Pachter L. A human commons cell atlas reveals cell type specificity for OAS1 isoforms; 2024. Available from: `https://www.biorxiv.org/content/10.1101/2024.03.23.586412v1`.

44. Larson A, Chin MT. A method for cryopreservation and single nucleus RNA-sequencing of normal adult human interventricular septum heart tissue reveals cellular diversity and function. BMC Medical Genomics. 2021;14:161. doi:10.1186/s12920-021-01011-z.

45. Pan T, Lu L, Youker K, Shiau CK, Wang M, Lin HY, et al. Single-Cell Splicing Isoform Atlas of the Adult Human Heart and Heart Failure. Circulation. 2025; p. CIRCULATIONAHA.125.074959. doi:10.1161/CIRCULATIONAHA.125.074959.

46. Payer B, Lee JT. X Chromosome Dosage Compensation: How Mammals Keep the Balance. Annual Review of Genetics. 2008;42(1):733–772. doi:10.1146/annurev.genet.42.110807.091711.

47. Chaumeil J, Waters PD, Koina E, Gilbert C, Robinson TJ, Marshall Graves JA. Evolution from XIST-Independent to XIST-Controlled X-Chromosome Inactivation: Epigenetic Modifications in Distantly Related Mammals. PLoS ONE. 2011;6(4):e19040. doi:10.1371/journal.pone.0019040.

48. Liang M, Zhang L, Lai L, Li Z. Unraveling the role of Xist in X chromosome inactivation: insights from rabbit model and deletion analysis of exons and repeat A. Cellular and Molecular Life Sciences. 2024;81(1):156. doi:10.1007/s00018-024-05151-0.

49. Traeuble K, Munz M, Pauli J, Sachs N, Vafadarnejad E, Carrillo-Roa T, et al. Integrated single-cell atlas of human atherosclerotic plaques. Nature Communications. 2025;16(1):8255. doi:10.1038/s41467-025-63202-x.

50. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nature Medicine. 2019;25(8):1280–1289. doi:10.1038/s41591-019-0512-5.

51. Zhao Q, Pedroza A, Sharma D, Gu W, Dalal A, Weldy C, et al.. A cell and transcriptome atlas of the human arterial vasculature; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.09.10.612293`.

52. Van Zyl T, Yan W, McAdams AM, Monavarfeshani A, Hageman GS, Sanes JR. Cell atlas of the human ocular anterior segment: Tissue-specific and shared cell types. Proceedings of the National Academy of Sciences. 2022;119(29):e2200914119. doi:10.1073/pnas.2200914119.

53. Kedlian VR, Wang Y, Liu T, Chen X, Bolt L, Tudor C, et al. Human skeletal muscle aging atlas. Nature Aging. 2024;doi:10.1038/s43587-024-00613-3.

54. Sim CB, Phipson B, Ziemann M, Rafehi H, Mills RJ, Watt KI, et al. Sex-Specific Control of Human Heart Maturation by the Progesterone Receptor. Circulation. 2021;143(16):1614–1628. doi:10.1161/CIRCULATIONAHA.120.051921.

55. Wang T, Sun Y, Zhao Y, Huang J, Huang Y. Identification of hub genes in heart failure by integrated bioinformatics analysis and machine learning. Frontiers in Cardiovascular Medicine. 2024;10:1332287. doi:10.3389/fcvm.2023.1332287.

56. Peng H, Lv B, Du J, Huang Y, Cui Q, Cui C, et al. Identification of Metabolism-Related Hub Genes in Heart Failure via Comprehensive Transcriptome Analysis. Genes. 2025;16(3):305. doi:10.3390/genes16030305.

57. Lenhart KC, Robasky K, Jones W, Halladay J, Ammerman A, Patterson C, et al. Genome-Wide Association Analysis Identified Candidate Variants Associated With Responsiveness to a Blood Pressure Reduction Intervention in Rural North Carolina. The FASEB Journal. 2016;30(S1):1177.15–1177.15.

58. Yao S, Jeon Y, Kesner B, Lee JT. Xist RNA binds select autosomal genes and depends on Repeat B to regulate their expression; 2024. Available from: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11291044/`.

59. Dror I, Chitiashvili T, Tan SYX, Cano CT, Sahakyan A, Markaki Y, et al. XIST directly regulates X-linked and autosomal genes in naïve human pluripotent cells. Cell. 2024;187(1):110–129.e31. doi:10.1016/j.cell.2023.11.033.

60. O'Leary K, Li MY, Jackson K, Shi L, Ezhkova E, Morrow BE, et al. Transcription and potential functions of a novel XIST isoform in male peripheral glia. Genome Research. 2025; p. gr.280832.125. doi:10.1101/gr.280832.125.

61. Mehta PA, Cowie MR. Gender and heart failure: a population perspective. Heart. 2006;92(Suppl 3):iii14–iii18. doi:10.1136/hrt.2005.070342.

62. Zheng SL, Henry A, Cannie D, Lee M, Miller D, McGurk KA, et al. Genome-wide association analysis provides insights into the molecular etiology of dilated cardiomyopathy. Nature Genetics. 2024;56(12):2646–2658. doi:10.1038/s41588-024-01952-y.

63. Costanzo MC, Roselli C, Brandes M, Duby M, Hoang Q, Jang D, et al. Cardiovascular Disease Knowledge Portal: A Community Resource for Cardiovascular Disease Research. Circulation: Genomic and Precision Medicine. 2023;16(6). doi:10.1161/CIRCGEN.123.004181.

64. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research. 2001;29(1):308–311.

65. Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, et al. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nature Genetics. 1997;17(3):353–356. doi:10.1038/ng1197-353.

66. Neale BM, Churchhouse C, Walters RK, Howrigan DP, Bryant S. UK Biobank GWAS: Phenotype 20002-1075; 2018. Available from: `http://www.nealelab.is/uk-biobank/`.

67. Tucker NR, Chaffin M, Fleming SJ, Hall AW, Parsons VA, Bedi KC, et al. Transcriptional and Cellular Diversity of the Human Heart. Circulation. 2020;142(5):466–482. doi:10.1161/CIRCULATIONAHA.119.045401.

68. Enzan N, Miyazawa K, Koyama S, Kurosawa R, Ieki H, Yoshida H, et al. Genome-wide analysis of heart failure yields insights into disease heterogeneity and enables prognostic prediction in the Japanese population. Nature Communications. 2025;16(1):9680. doi:10.1038/s41467-025-64659-6.

69. Murray CS, Yang C, Chen SN, Graw S, Karimpour-Fard A, Cleveland J, et al.. Integration of dilated cardiomyopathy genomics with transcriptomics from the human heart implicates regulatory molecular mechanisms; 2026. Available from: `https://www.medrxiv.org/content/10.64898/2026.01.12.26343934v1`.

70. Hu M, Zhang MJ, Chikina M. DALE-Eval: A comprehensive cell type-specific expression deconvolution benchmark for transcriptomics data; 2025. Available from: `https://www.biorxiv.org/content/10.1101/2025.07.31.667984v1`.

71. Rich JM, Moses L, Einarsson PH, Jackson K, Luebbert L, Booeshaghi AS, et al.. The impact of package selection and versioning on single-cell RNA-seq analysis; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.04.04.588111`.

72. Montserrat-Ayuso T, Esteve-Codina A. High content of nuclei-free low-quality cells in reference single-cell atlases: a call for more stringent quality control using nuclear fraction. BMC Genomics. 2024;25(1):1124. doi:10.1186/s12864-024-11015-5.

73. Clarke ZA, Bader GD. *MALAT1* expression indicates cell quality in single-cell RNA sequencing data; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.07.14.603469`.

74. Koenig AL, Shchukina I, Amrute J, Andhey PS, Zaitsev K, Lai L, et al. Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. Nature Cardiovascular Research. 2022;1(3):263–280. doi:10.1038/s44161-022-00028-6.

75. Chaffin M, Papangeli I, Simonson B, Akkad AD, Hill MC, Arduini A, et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. Nature. 2022;608(7921):174–180. doi:10.1038/s41586-022-04817-8.

76. Hill MC, Kadow ZA, Long H, Morikawa Y, Martin TJ, Birks EJ, et al. Integrated multi-omic characterization of congenital heart disease. Nature. 2022;608(7921):181–191. doi:10.1038/s41586-022-04989-3.

77. Simonson B, Chaffin M, Hill MC, Atwa O, Guedira Y, Bhasin H, et al. Single-nucleus RNA sequencing in ischemic cardiomyopathy reveals common transcriptional profile underlying end-stage heart failure. Cell Reports. 2023;42(2):112086. doi:10.1016/j.celrep.2023.112086.

78. Skelly DA, Squiers GT, McLellan MA, Bolisetty MT, Robson P, Rosenthal NA, et al. Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Reports. 2018;22(3):600–610. doi:10.1016/j.celrep.2017.12.072.

79. Consortium TM. A Single Cell Transcriptomic Atlas Characterizes Aging Tissues in the Mouse. Nature. 2020;583(7817):590–595. doi:10.1038/s41586-020-2496-1.

80. Feng W, Bais A, He H, Rios C, Jiang S, Xu J, et al. Single-cell transcriptomic analysis identifies murine heart molecular features at embryonic and neonatal stages. Nature Communications. 2022;13(1):7960. doi:10.1038/s41467-022-35691-7.

81. Ren H, Zhou X, Yang J, Kou K, Chen T, Pu Z, et al. Single-cell RNA sequencing of murine hearts for studying the development of the cardiac conduction system. Scientific Data. 2023;10(1):577. doi:10.1038/s41597-023-02333-6.

82. Li G, Tian L, Goodyer W, Kort EJ, Buikema JW, Xu A, et al. Single cell expression analysis reveals anatomical and cell cycle-dependent transcriptional shifts during heart development. Development. 2019;146(12):dev173476. doi:10.1242/dev.173476.

83. Eaton DM, Berretta RM, Lynch JE, Travers JG, Pfeiffer RD, Hulke ML, et al. Sex-specific responses to slow progressive pressure overload in a large animal model of HFpEF. American Journal of Physiology-Heart and Circulatory Physiology. 2022;323(4):H797–H817. doi:10.1152/ajpheart.00374.2022.

84. Osgood MP. X-Chromosome Inactivation: The Case of the Calico Cat. American Journal of Pharmaceutical Education. 1994;58(2):204–205. doi:10.1016/S0002-9459(24)02621-4.

85. Chen D, Sun J, Zhu J, Ding X, Lan T, Wang X, et al. Single cell atlas for 11 non-model mammals, reptiles and birds. Nature Communications. 2021;12(1):7083. doi:10.1038/s41467-021-27162-2.

86. Youngblut ND, Carpenter C, Prashar J, Ricci-Tam C, Ilango R, Teyssier N, et al.. scBaseCount: an AI agent-curated, uniformly processed, and continually expanding single cell data repository; 2025. Available from: `https://www.biorxiv.org/content/10.1101/2025.02.27.640494v2`.

87. Mesquita T, Miguel-dos Santos R, Liu W, Fournier M, Rogers RG, Alfaro J, et al. Upregulated FoxO1 promotes arrhythmogenesis in mice with heart failure and preserved ejection fraction. Nature Communications. 2025;16(1):1184. doi:10.1038/s41467-025-56186-1.

88. Li G, Zhao H, Cheng Z, Liu J, Li G, Guo Y. Single-cell transcriptomic profiling of heart reveals ANGPTL4 linking fibroblasts and angiogenesis in heart failure with preserved ejection fraction. Journal of Advanced Research. 2025;68:215–230. doi:10.1016/j.jare.2024.02.006.

89. Patrick R, Janbandhu V, Tallapragada V, Tan SSM, McKinna EE, Contreras O, et al. Integration mapping of cardiac fibroblast single-cell transcriptomes elucidates cellular principles of fibrosis in diverse pathologies. Science Advances. 2024;10(25):eadk8501. doi:10.1126/sciadv.adk8501.

90. Ranjbarvaziri S, Zeng A, Wu I, Greer-Short A, Farshidfar F, Budan A, et al. Targeting HDAC6 to treat heart failure with preserved ejection fraction in mice. Nature Communications. 2024;15:1352. doi:10.1038/s41467-024-45440-7.

91. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nature Methods. 2018;15(12):1053–1058. doi:10.1038/s41592-018-0229-2.

92. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports. 2019;9(1):5233. doi:10.1038/s41598-019-41695-z.

93. Cao L, Chang R, Wang X, Shen J, Yang Z, Ma L, et al. Integrative Single-Cell and Spatial Transcriptomics Reveal Functional and Spatial Heterogeneity of Atrial and Ventricular Cardiomyocytes in the Heart. Molecular Biotechnology. 2025;doi:10.1007/s12033-025-01443-3.

94. Arduini A, Fleming SJ, Xiao L, Hall AW, Akkad AD, Chaffin MD, et al. Transcriptional profile of the rat cardiovascular system at single-cell resolution. Cell Reports. 2025;44(1):115091. doi:10.1016/j.celrep.2024.115091.

95. Sullivan DK, Min KHJ, Hjörleifsson KE, Luebbert L, Holley G, Moses L, et al. kallisto, bustools and kb-python for quantifying bulk, single-cell and single-nucleus RNA-seq. Nature Protocols. 2025;20(3):587–607. doi:10.1038/s41596-024-01057-0.

96. Shevchenko AI, Malakhova AA, Elisaphenko EA, Mazurok NA, Nesterova TB, Brockdorff N, et al. Variability of Sequence Surrounding the Xist Gene in Rodents Suggests Taxon-Specific Regulation of X Chromosome Inactivation. PLOS ONE. 2011;6(8):e22771. doi:10.1371/journal.pone.0022771.

97. Okamoto I, Nakamura T, Sasaki K, Yabuta Y, Iwatani C, Tsuchiya H, et al. The X chromosome dosage compensation program during the development of cynomolgus monkeys. Science. 2021;374(6570):eabd8887. doi:10.1126/science.abd8887.

98. Reinius B, Saetre P, Leonard JA, Blekhman R, Merino-Martinez R, Gilad Y, et al. An Evolutionarily Conserved Sexual Signature in the Primate Brain. PLoS Genetics. 2008;4(6):e1000100. doi:10.1371/journal.pgen.1000100.

99. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

100. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37(12):1639–1643. doi:10.1093/bioinformatics/btaa1016.

101. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Research. 2020;9:304. doi:10.12688/f1000research.23297.1.

102. Qu J, Yang F, Zhu T, Wang Y, Fang W, Ding Y, et al. A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys. Nature Communications. 2022;13:4069. doi:10.1038/s41467-022-31770-x.

103. Nakada Y, Zhou Y, Gong W, Zhang EY, Skie E, Nguyen T, et al. Single Nucleus Transcriptomics: Apical Resection in Newborn Pigs Extends the Time Window of Cardiomyocyte Proliferation and Myocardial Regeneration. Circulation. 2022;145(23):1744–1747. doi:10.1161/CIRCULATIONAHA.121.056995.

104. Miao J, Zhang K, Yang Y, Xu S, Du J, Wu T, et al. Single-nucleus transcriptomics reveal cardiac cell type-specific diversification in metabolic disease transgenic pigs. iScience. 2024;27(6):110015. doi:10.1016/j.isci.2024.110015.

105. Wei Y, Walcott G, Nguyen T, Geng X, Guragain B, Zhang H, et al. Follistatin From hiPSC-Cardiomyocytes Promotes Myocyte Proliferation in Pigs With Postinfarction LV Remodeling. Circulation Research. 2025;136(2):161–176. doi:10.1161/CIRCRESAHA.124.325562.

106. Mendelson JB, Sternbach JD, Kim M, Moon RA, Blake JC, Raveendran RM, et al. Single-Nucleus RNAseq Defines Cellular and Molecular Landscape of Dysfunctional Right Atrium in Pulmonary Artery Banded Swine. Circulation Research. 2025; p. CIRCRESAHA.125.326892. doi:10.1161/CIRCRESAHA.125.326892.

107. Wang F, Ding P, Liang X, Ding X, Brandt CB, Sjöstedt E, et al. Endothelial cell heterogeneity and microglia regulons revealed by a pig cell landscape at single-cell level. Nature Communications. 2022;13(1):3620. doi:10.1038/s41467-022-31388-z.

108. Zeng F, Huang Z, Yuan Y, Shi J, Cai G, Liu D, et al. Effects of RNAi-mediated knockdown of Xist on the developmental efficiency of cloned male porcine embryos. The Journal of Reproduction and Development. 2016;62(6):591–597. doi:10.1262/jrd.2016-095.

109. Tech M. DNBelab C Series Single-Cell RNA Analysis Software; 2021. Available from: `https://github.com/MGI-tech-bioinformatics/DNBelab_C_Series_scRNA-analysis-software/tree/master`.

110. Badia-i Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances. 2022;2(1):vbac016. doi:10.1093/bioadv/vbac016.

111. Muzellec B, Teleńczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. Bioinformatics. 2023;39(9):btad547. doi:10.1093/bioinformatics/btad547.

112. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550. doi:10.1186/s13059-014-0550-8.

113. Volmert B, Kiselev A, Juhong A, Wang F, Riggs A, Kostina A, et al. A patterned human primitive heart organoid model generated by pluripotent stem cell self-organization. Nature Communications. 2023;14(1):8245. doi:10.1038/s41467-023-43999-1.

114. Asp M, Giacomello S, Larsson L, Wu C, Fürth D, Qian X, et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell. 2019;179(7):1647–1660.e19. doi:10.1016/j.cell.2019.11.025.

115. Badia-i Mompel P, Casals-Franch R, Wessels L, Müller-Dott S, Trimbour R, Yang Y, et al.. Comparison and evaluation of methods to infer gene regulatory networks from multimodal single-cell data; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.12.20.629764`.

116. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nature Methods. 2020;17(2):147–154. doi:10.1038/s41592-019-0690-6.

117. Kernfeld E, Keener R, Cahan P, Battle A. Transcriptome data are insufficient to control false discoveries in regulatory network inference. Cell Systems. 2024;15(8):709–724.e13. doi:10.1016/j.cels.2024.07.006.

118. Ke ZP, Tao WQ, Zhao G, Cheng K. Role of PPAR-related genes in chronic heart failure: evidence from large populations. BMC Cardiovascular Disorders. 2023;23(1):552. doi:10.1186/s12872-023-03554-8.

119. Panse KD, Felkin LE, López-Olañeta MM, Gómez-Salinero J, Villalba M, Muñoz L, et al. Follistatin-Like 3 Mediates Paracrine Fibroblast Activation by Cardiomyocytes. Journal of Cardiovascular Translational Research. 2012;5(6):814–826. doi:10.1007/s12265-012-9400-9.

120. Montaigne D, Butruille L, Staels B. PPAR control of metabolism and cardiovascular functions. Nature Reviews Cardiology. 2021;18(12):809–823. doi:10.1038/s41569-021-00569-6.

121. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

122. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 2013;14(2):178–192. doi:10.1093/bib/bbs017.

123. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

124. Shiau CK, Lu L, Kieser R, Fukumura K, Pan T, Lin HY, et al. High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. Nature Communications. 2023;14(1):4124. doi:10.1038/s41467-023-39813-7.

125. Loving RK, Sullivan DK, Booeshagi AS, Reese F, Rebboah E, Sakr J, et al.. Long-read sequencing transcriptome quantification with lr-kallisto; 2025. Available from: `https://www.biorxiv.org/content/10.1101/2024.07.19.604364v4`.

126. Gorin G. GDG_2025 data supplement; 2026. Available from: `https://zenodo.org/records/18329880`.