Male *XIST* expression in cardiac pseudo-glia does not induce X chromosome inactivation

Gennady Gorin[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 204 male heart samples across 14 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 potential explanations and cross-reference the expression of *XIST* against age, sex, and pathology, finding signatures of a non-canonical inhibitory program and a possible relationship to cardiac dysfunction.

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.

Separately, a great deal of effort has been expended toward characterizing the single-cell transcriptomes of human cardiac cells in healthy and pathological tissues [15]. 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. [16], used in lieu of a control in a number of studies [17, 18]: 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, but the adipocytes and myocytes are incompatible with droplet-based single-cell protocols [16].

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 [19, 20]. 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 [15]. The original heart cell atlas reported them as neurons, but noted that a subpopulation more closely resembles (glial) Schwann cells [16]. A follow-up study by the same authors remarked that the cells fail to express canonical neuronal markers and redefined them as glia [21]. 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 [22]. Other articles have reported these cells as neurons [18, 23, 24] or glia/Schwann cells [25, 26] with little further comment. Nominally, these are considered prototypic peripheral nervous system cells [15]. However, to acknowledge the remaining uncertainty in the field, we term this cell population "pseudo-glial."

2 Results

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 [16]. 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 204 of the 221 male samples.

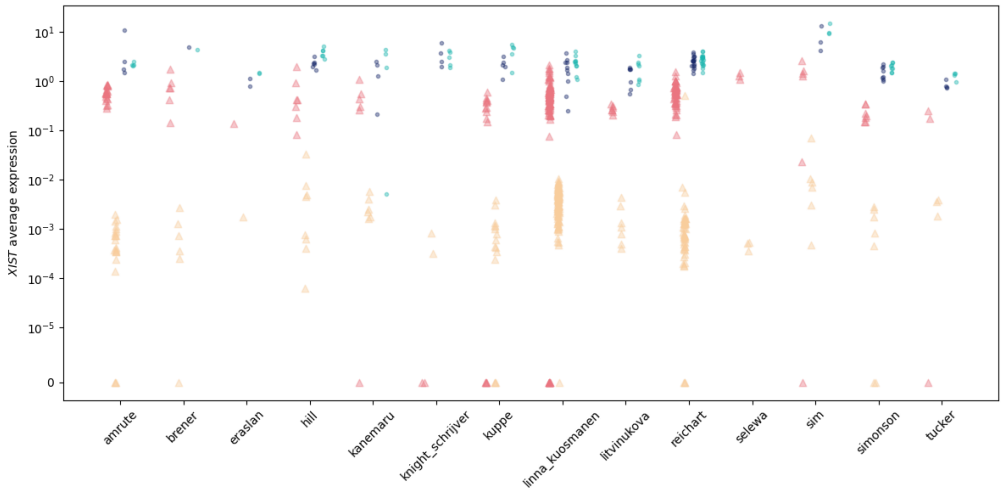

Fig 1. Male cardiac pseudo-glia show striking expression of the female-specific transcript *XIST*. Each point denotes average expression of the gene in one sample (typically one unique donor). Triangles denote samples reported as male. Circles denote samples reported as female. Red: male, pseudo-glial (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.

2.1 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. [21], reported as a Caucasian female in her early 60s, expresses negligible *XIST* outside the 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 2. Unrecorded sex chromosome abnormalities, if they occurred, would be readily detectable by tissue-wide expression of *XIST* and Y-linked genes [7, 27]. 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 [16], has a relatively high expression of *UTY* (Fig 2) 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 typical transcriptomes (high *XIST* and negligible Y-linked transcripts). In another example, the Reichart study reports that donor H40 is a male in his 60s [28]; the left ventricular

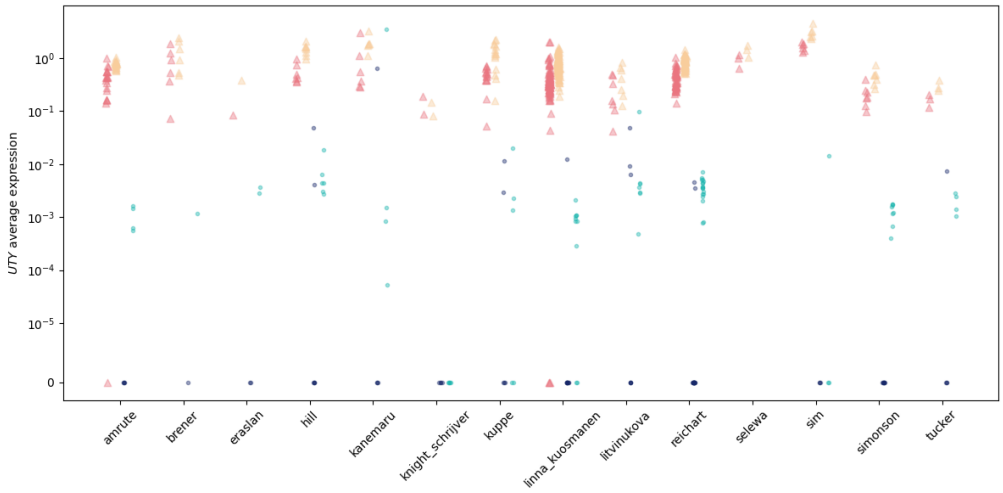

Fig 2. Average expression of *UTY* in cardiac datasets. Each point denotes one sample (typically one unique donor). Triangles denote samples reported as male. Circles denote samples reported as female. Red: male, pseudo-glial (cells annotated as Schwann cells, neurons, or glia). Yellow: male, all others. Dark blue: female, pseudo-glial. Teal: female, all others. To accommodate zero values, the region $[0, 10^{-5}]$ is linear.

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. [29], 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 contrived and 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 [30, 31]. 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 *XIST*, 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 [32] released single-nucleus alignment data generated from the interventricular septa (IVS) of two

males and two females. The reads aligned to the *XIST* region are displayed in Fig 3 (Section 4.3). The coverage in the male sample is considerably lower (129 reported UMIs, compared to 9,141 in the female sample), but the localization appears to be consistent: reads are predominantly found in exon 6 of the canonical transcript, with 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 an uncharacterized transcript with a different structure and function that happens to be localized to the same region. In sum, there does not appear to be a plausible artifactual explanation for the expression of *XIST* in male hearts.

2.2 Cross-species context

X chromosome inactivation is ubiquitous in mammals; in placental mammals, XCI is achieved through the expression of *XIST* [33, 34], although the transcript structure shows nontrivial variability [35]. 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. Curiously, 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 [36]. 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 [37]. Several other developmental datasets have been collected; however, they do not report glial or neuronal cell types [38–40].

A recent large study consisting entirely of male Wistar rats [41] reports 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 obtained part of this dataset and realigned it to the 2024 RefSeq GRCr8 assembly, which contains *Xist* (gene 100911498), finding negligible expression with 0–4 molecules per dataset, none in the relevant cell type (Section 4.1.2). In contrast, we quantified 53–204 molecules of *Tsix* per dataset.

A study of sex-specific pressure overload in cats [42] 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 [43]. 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 [44]. However, neither dataset contains *XIST* expression data, presumably due to reference limitations.

At least two single-cell pig heart datasets are available. Both report small but distinct neuronal populations, marked by *Scn7a* [45] and by *Nrxn1* [46]. However, the disclosed raw data matrices did 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 [47] and is known to be functional [47] in spite of nontrivial differences from the human gene [35], but characterization of its presence or absence in male porcine heart cell populations would require a more comprehensive transcriptome build.

It is likely that some or all of these datasets could clarify the issue upon exhaustive reanalysis, but not so readily as the immediately evident signal in human tissues.

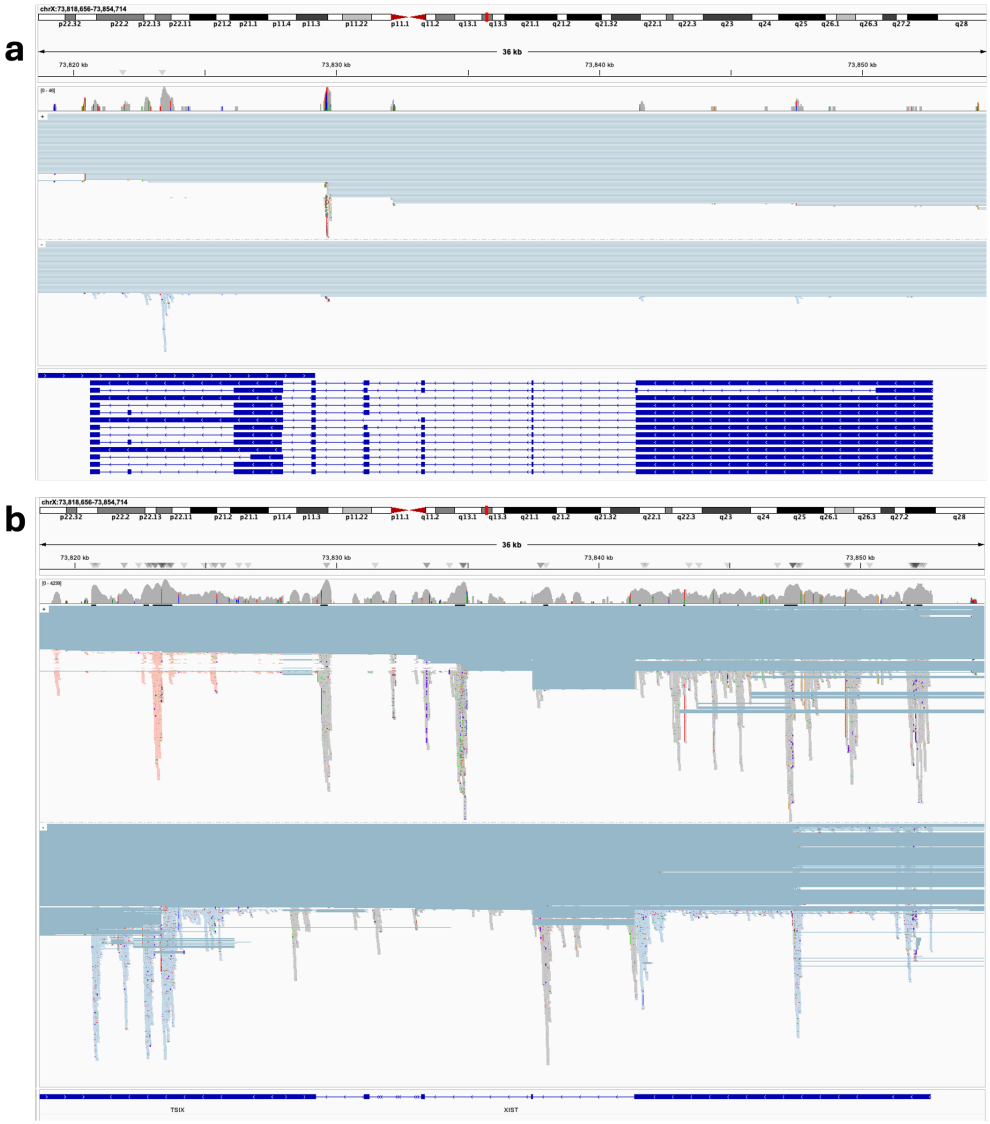

Fig 3. Read alignments of Larson and Chin interventricular septum samples collected from male (2879, **a**) and female (2867, **b**) donors. Top of each panel: X chromosome location and read coverage (log-scaled). Middle of each panel: reads localized to the *XIST* region (light blue: reads mapped to *XIST*; light red: reads mapped to *TSIX*; gray: unmapped reads; dark blue: gap regions; top: forward strand; bottom: reverse strand). Bottom of each panel: RefSeq transcripts (**a**: all transcripts; **b**: curated only).

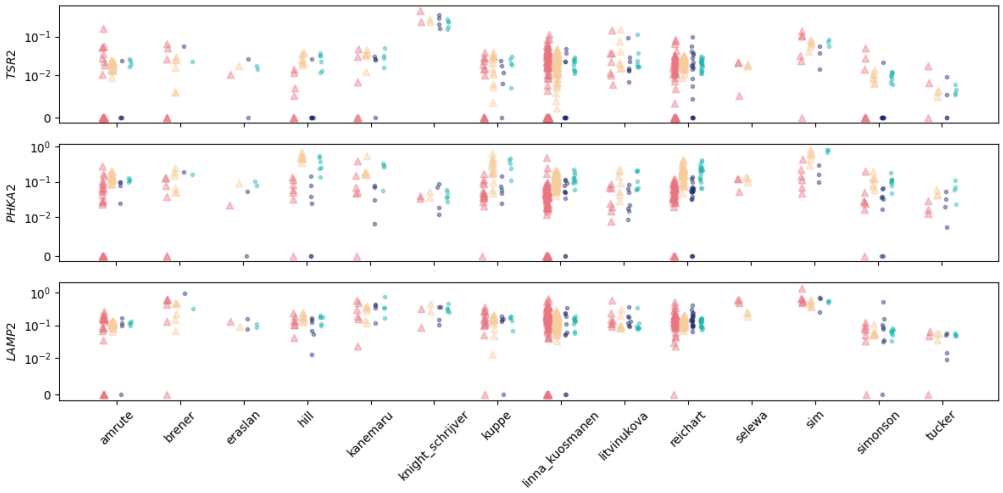

Fig 4. Average expression of example genes reported [48] as consistently silenced in XCI by Tukiainen et al. Each point denotes one sample (typically one unique donor). Triangles denote samples reported as male. Circles denote samples reported as female. Red: male, pseudo-glial (cells annotated as Schwann cells, neurons, or glia). Yellow: male, all others. Dark blue: female, pseudo-glial. Teal: female, all others. To accommodate zero values, the region $[0, 10^{-2}]$ is linear.

2.3 Biological implications

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 [48] (Fig 4). 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. A similar male neurexin- and *XIST*-expressing population is present in the Traeuble et al. reanalysis of Wirka et al. coronary artery data, originally annotated as neurons but reported as smooth muscle cells in the Traeuble resource [49, 50]. A recent single-cell atlas of arterial vasculature [51] conflates this population with lymphocytes, despite the absence of lymphocyte markers in the *PLP1*/*MPZ*/*XIST*-expressing population. A single-cell census of the eye [52] reports three male donors with *XIST* expression, predominantly in Schwann cells, but partly in basal and epithelial cells. An atlas of the human musculature [53] reports fifty *XIST*-expressing cells from male donors, restricted to non-myelinating Schwann cells. Finally, the Eraslan et al. study [26] reports relatively high expression of *XIST* in esophageal and prostate

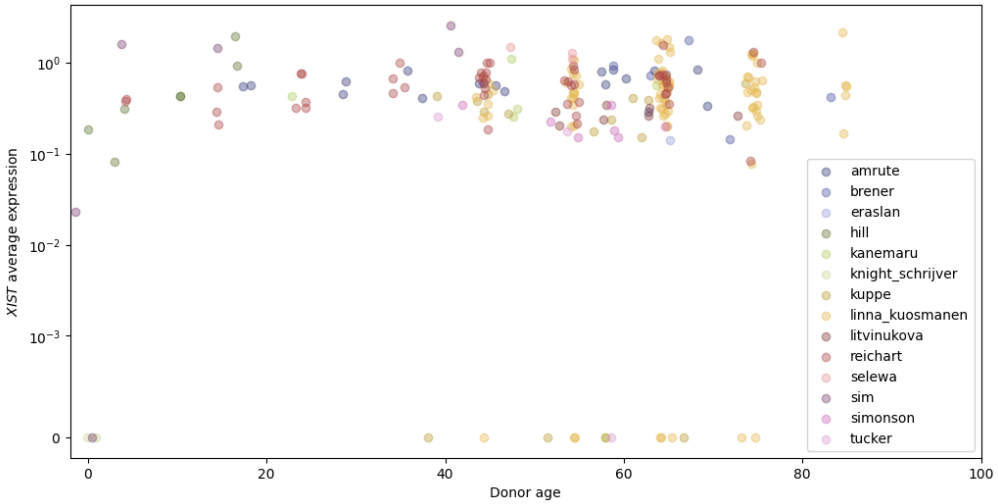

Fig 5. Average expression of *XIST* in pseudo-glial cells in male donors. Colors denote study of origin. All fetal samples are given age zero. Donors with no age information omitted. Jitter added. Data with binned ages (e.g., 20–29 years) are visualized at the bin midpoint. To accommodate zero values, the region $[0, 10^{-3}]$ is linear.

Schwann cells. Surprisingly, male donor `GTEX-1HSMQ` exhibits high expression across multiple cell types, 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. 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 [18] 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 5), 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 6, the *XIST*-positive cells exhibited a striking decrease in a diverse variety of genes in male samples, and little to no difference from the *XIST*-negative cells in female samples.

Gene Ontology (GO) analysis of the 95 genes differentially expressed in female samples reported no significant enrichments. Restricting the analysis to the 36 genes with at least modest average expression (`baseMean` > 5) yielded enrichment of laminin basement membrane constituents and integrin signaling participants (*LAMA2*, *LAMB2*), alongside miscellaneous genes associated with axon growth and regeneration (*TSPO*, *GAP43*). However, it is far from clear that even these minor differences are meaningful,

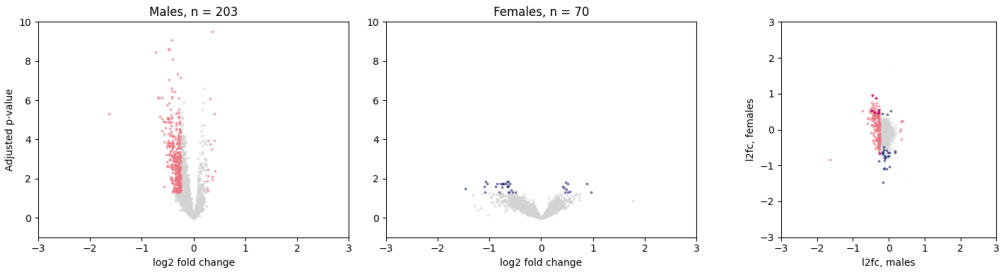

Fig 6. Results of differential expression testing between *XIST*-positive and *XIST*-negative glia-like cells. Left: male samples (red: DE genes). Middle: female samples (blue: DE genes). Right: relationship between $\log_2$ fold changes inferred for male and female samples (red: male DE genes; blue: 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.

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 169 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*, *FLNA*, 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*, 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*, *FLNA*, 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" [15], 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 challenging to ascertain the true role of the cell type or its transcriptional program based solely on bioinformatic data. A 2002 article by Occhi et al. reports a cardiac-specific form of neurexin-3 and suggests it may be involved in cell–cell connections [58]; based on this, the cell population more broadly may perform an adhesion-related function not directly related to neuronal conduction or myelination. Experimental validation would be necessary to understand the role of this cell population, although the apparent lack of homology to model organisms presents a serious barrier.

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 unclear: *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* [59, 60], or an altogether new mechanism, possibly through an uncharacterized isoform.

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. 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 several pathology datasets, and find rather equivocal results: in the two largest datasets, the expression of *XIST* appears to be higher in samples with pathology or comorbidities, but the smaller datasets are less illuminating.

A very recent article by Gadek et al. proposes that upregulation of *Plp1* in mouse brains, induced by age-related shifts in X inactivation or ectopic expression, increases performance on cognitive metrics [61]. This study is a promising example of the renewed attention to the sex determinants of health and pathology in the context of high-throughput genomics methods. However, despite superficial similarities (such as the investigation of *Plp1* and XCI), we refrain from drawing parallels to this study: the differences between mouse and human, central and peripheral systems, XX and XY cells, and exogenous and endogenous expression may be too substantial to support such analogies.

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 [62–64]. 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 [65]. The Cardiovascular Disease Knowledge Portal, which aggregates hundreds of genetic association datasets, reports one SNP (rs1474563) in the *XIST* region, associated with height [66]. The only pathogenic variant reported in dbSNP [67] is rs773396320, just upstream of the first exon of *XIST*, which leads to skewed XCI [68]. In addition, in spite of continuing efforts, variants in lncRNAs remain understudied [69]. Whether due to less study, greater exposure to natural selection [62], 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*.

Alternatively, the role of *XIST* may be elucidated by cross-referencing its expression in the pseudo-glia 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 [21, 23, 29, 70]. To our knowledge, the only large study reporting single-cell data alongside patient genotypes is the investigation of pathogenic variants in dilation cardiomyopathy by Reichart et al. [28]. 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). Therefore, characterizing the genetic correlates of male *XIST* expression in an unbiased way does not seem feasible at this time.

This analysis is deliberately 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 fairly 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 Materials and 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 [25]. 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 [17]. 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.

Eraslan. Data were obtained from CZI CELL×GENE and filtered for the heart samples, which had the tissue annotation "anterior wall of left ventricle" [26]. 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.

Hill. Single-nucleus count data were obtained from the Broad Single Cell Portal (SCP1852) and merged into an AnnData structure [75]. 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 [21]. One cell attributed to donor D11 was removed, yielding 211,060 cells, in agreement with Fig. 1. 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 [18]. 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 [23]. 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 [22]. 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 [16]. 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 [28]. 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 [29]. 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 [76]. 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) [70]. 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

Arduini. We acquired raw reads for 19 left ventricular samples from GSE280111 [41]. We generated a pseudoalignment reference from the RefSeq GRCr8 genome ("genomic" `GCF_036323735.1`), using *kallisto* | *bustools* 0.28.2 [77]. 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.

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 this assembly, the putative *Tsix* (*LOC680227*) is missing the portion of the 3' exon that covers the *Xist* promoter, previously reported by Shevchenko et al. [78] 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.

4.2 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` [79]. 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, 4, and 5 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 [80, 81]. Finally, we obtained the Wald test p-values and log fold differences for the female/male contrast and visualized them in Fig 7, providing three axis scales to accommodate the outsize effect sizes of sex chromosomes.

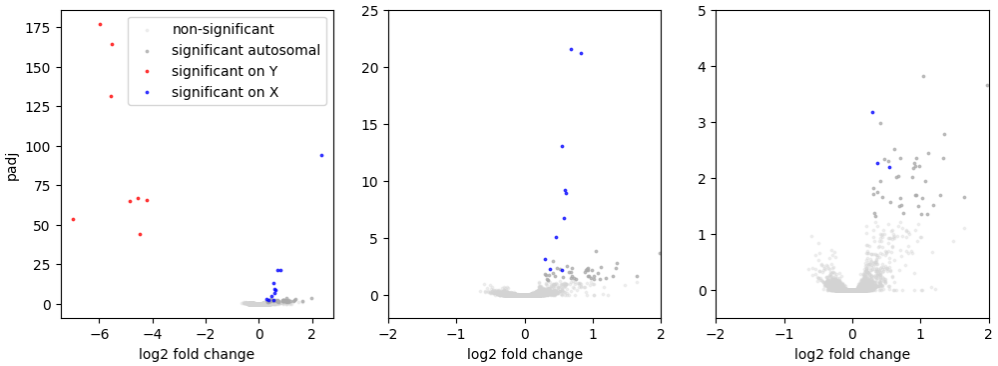

Fig 7. Results of differential expression testing between female and male samples, using sex and study as covariates. Genes with `baseMean` < 5 excluded. Differential expression is defined as $|\log_2 \mathrm{FC}| > 0.25$ and adjusted p-value < 0.05.

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. [24]. The volcano plots shown in Fig 7 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 [16, 21–23, 28]. 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 [22], 0.15% in [16], 0.1% in [21], and not readily detectable in [23, 29]). 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 2.2), 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 [82] that resemble embryonic Schwann and neural cell progenitors [83]. However, this population's markers are not expressed in adult pseudo-glia, so the generation of a model that recapitulates the adult heart

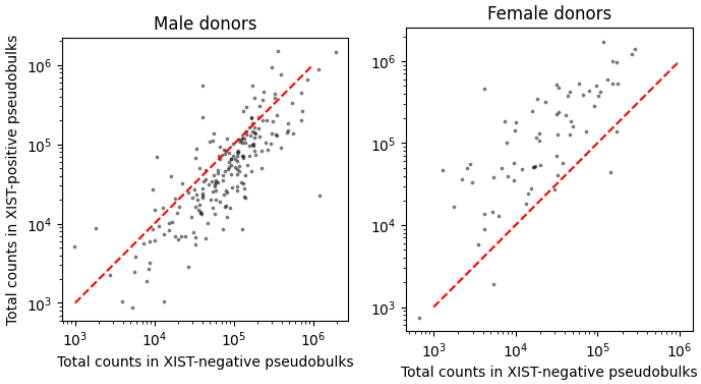

Fig 8. Total molecule counts in the pseudobulks used to evaluate differential expression between *XIST*-positive and -negative cells (points: donor pseudobulks; lines: identity).

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 [84–86], 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 204 covariates for the 203 male donors with *XIST* expression and 71 covariates for the 70 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 6 are conservative. We note this filtering is rather stringent, as it discards some 77% of the 10,391 genes intersecting between datasets. Although it is likely that some of the omitted genes are important, the small sample sizes (62% of the male pseudobulks have fewer than fifty cells; 17% 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 6 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 8, 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 8), 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 0.042. 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 0.29, with a p-value of 0.0098. In contrast, the genes with the lowest p-values in the latter analysis were *PPARG* (effect size 0.99 and $p = 2.5 \times 10^{-7}$) and *FSTL3* (effect size 0.71 and $p = 4.4 \times 10^{-6}$), both of which are known to serve roles in heart failure pathology [87–89]. 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 5), the PyDESeq2 "size-normalized" counts (Fig 9, top), or the latter with additional correction for the batch covariates (Fig 9, bottom). In contrast, the age dependence of *PPARG* and *FSTL3* is far less equivocal.

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.

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 class (dilated cardiomyopathy or DCM) and healthy controls. A PyDESeq2 Wald test yielded higher expression in the DCM class (effect size 1.0 and $p = 5.0 \times 10^{-4}$), 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

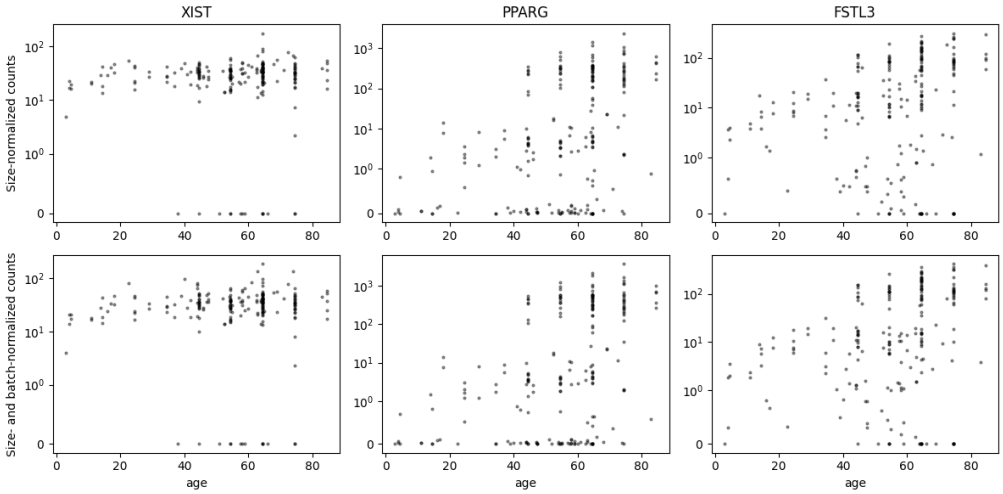

Fig 9. Expression of *XIST*, *PPARG*, and *FSTL3* in glia-like cells in male donors. Data with binned ages (e.g., 20–29 years) are visualized at the bin midpoint. Top: size-normalized pseudobulk counts. Bottom: size-normalized pseudobulk residuals, after correction for batch variables. To accommodate zero values, the region $[0, 1]$ is linear.

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 $p = 0.04$), 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 -0.93 and $p = 0.11$), with similar results under models incorporating age or the logarithm of age alongside the pathology.

Summary. Evidently, the 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 Read analysis

Read alignment files were acquired from the Sequence Read Archive depositions accompanying the GEO accession GSE161921 [32]. Specifically, we acquired the sorted BAM files generated by Cell Ranger alignments of data for male donor 2879 (GSM4923402) and female donor 2867 (GSM4923401). The alignments were filtered to retain only the reads overlapping *XIST* using the `samtools` 1.20 command `view` for the region `X:73818656-73854714`, using [90]. The resulting subsets were visualized using the Integrative Genomics Viewer [91] with reference to the GRCh38/hg38 genome.

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. [17], 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/GG_2025`.

The differential expression analyses used PyDESeq2 0.4.8 [80]. The visualization of reads used the Integrative Genomics Viewer [91]. The pseudoalignment of rat data used *kallisto* | *bustools* 0.28.2 with the GRCr8 genome.

Acknowledgments

L.G. is a co-founder and CTO of Fauna Bio. G.G. is an employee 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. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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`.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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`.

31. 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`.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. Traeuble K, Munz M, Pauli J, Sachs N, Vafadarnejad E, Carrillo-Roa T, et al.. Integrated single-cell atlas of human atherosclerotic plaques; 2024. Available from: `http://biorxiv.org/lookup/doi/10.1101/2024.09.11.612431`.

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. Occhi G, Rampazzo A, Beffagna G, Antonio Danieli G. Identification and characterization of heart-specific splicing of human neurexin 3 mRNA (NRXN3). Biochemical and Biophysical Research Communications. 2002;298(1):151–155. doi:10.1016/S0006-291X(02)02403-8.

59. 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/`.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

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. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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`.

85. 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.

86. 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.

87. 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.

88. 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.

89. 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.

90. 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.

91. 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.