Regulatory effects of the Uty/Ddx3y locus on neighboring chromosome Y genes and autosomal mRNA transcripts in adult mouse non-reproductive cells

Christian F. Deschepper

In addition to sperm-related genes, the male-specific chromosome Y (chrY) contains a class of ubiquitously expressed and evolutionary conserved dosage-sensitive regulator genes that include the neighboring Uty, Ddx3y and (in mice) Eif2s3y genes. However, no study to date has investigated the functional impact of targeted mutations of any of these genes within adult non-reproductive somatic cells. We thus compared adult male mice carrying a gene trap within their Uty gene (UtyGT) to their wild-type (WT) isogenic controls, and performed deep sequencing of RNA and genome-wide profiling of chromatin features in extracts from either cardiac tissue, cardiomyocyte-specific nuclei or purified cardiomyocytes. The apparent impact of UtyGT on gene transcription concentrated mostly on chrY genes surrounding the locus of insertion, i.e. Uty, Ddx3y, long non-coding RNAs (lncRNAs) contained within their introns and Eif2s3y, in addition to possible effects on the autosomal Malat1 lncRNA. Notwithstanding, UtyGT also caused coordinate changes in the abundance of hundreds of mRNA transcripts related to coherent cell functions, including RNA processing and translation. The results altogether indicated that tightly co-regulated chrY genes had nonetheless more widespread effects on the autosomal transcriptome in adult somatic cells, most likely due to mechanisms other than just transcriptional regulation of corresponding protein-coding genes.

Although the contributions of genes on the male sex chrY have long believed to be restricted to their effects on reproductive functions in sex organs, increasing evidence indicates that their impact may in fact extend to somatic cells as well. Thus, several types of genetic rodent models have shown that chrY and/or some of its genetic variants affect a range of phenotypes within the cardiovascular, immune, metabolic or neurologic systems. In humans, one specific haplotype of the male-specific portion of chrY (MSY) associates with increased coronary artery disease risk, along with altered gene expression within macrophages, while mosaic loss of chrY in blood cells (a relatively frequent and age-related event) associates with a host of disease manifestations, including shorter survival, increased cancer risk, cardiovascular events, Alzheimer disease and age-related macular degeneration.

In addition of the male sex-determination Sry gene, MSY carries only a handful of protein-coding genes, which fall for the most part in two classes. The first one corresponds to genes expressed only in male reproductive tissues and/or specializing in sperm-related functions. Genes within the second class share a particular set of properties, as they: (1) are expressed ubiquitously in all tissues; (2) are active genes referred to as “X-degenerates” by virtue of the fact that they have on chromosome X (chrX) counterparts that escape chrX-inactivation in females; (3) exert their effects in a dose-related fashion that depends on expression of copies from both sex chromosomes; and (4) belong to an evolutionary conserved core of genes found in MSY from all placental vertebrates. Although there are some species-specific variation concerning the exact nature of genes found in that shared core group, some of them (including Uty, Ddx3y and Kdm5d) are common to all investigated species.
In addition to the above properties, the so-called “X-degenerate genes” on MSY have functional annotations that relate mostly to some of the most basic cell functions (such as chromatin modification, ubiquitination, translation, splicing and chromatin modification). To date, there is evidence (albeit limited) that these MSY genes show at least some level functional redundancy with their homologs on chrX. For instance, DDX3X has been shown to be functionally interchangeable with DDX3Y in cultured mammalian cells\(^\text{13,14}\). Likewise, studies using mice with inactivating gene traps of either Utx or Uty demonstrated that either the presence of Uty or overexpression of Utx mutants devoid of any lysine demethylase activity rescued both the developmental effects of Utx during development\(^\text{15}\) or the preleukemic phenotypes induced by Utx deficiency within the hematopoietic lineage\(^\text{16}\), showing that both genes shared high levels of redundancy for some functions that are exerted independently of the lysine demethylase activity known to be carried by one domain of Utx. However, there is still no study to date exploring whether targeted mutations of MSY genes do have an impact on somatic non-reproductive cells in adult mice. To address this particular question, we used male mice carrying an inactivating gene trap that introduces within the fourth intron of Uty a splicing acceptor site followed by coding sequences carrying a stop codon. These animals (Uty\(^\text{GT}\) in short) were previously shown to be fully viable\(^\text{15}\), but not used for studying phenotypes other than those occurring during the developmental stage. In particular, we performed the following experiments on material extracted from tissues or cells from either adult UtX\(^\text{GT}\) mice or their isogenic wild-type (WT) controls: (1) 3′-end deep sequencing of RNA (RNA-Seq) of ribosomal RNA-depleted total RNA extracted from cardiac left ventricles; (2) full-length RNA-Seq of ribosomal RNA-depleted total RNA extracted from cardiomyocytes isolated and purified from adult whole hearts; (3) chromatin immunoprecipitation and sequencing (ChIP-Seq) assays of Histone H3 acetyl lysine-27 (H3K27ac) residues (i.e. marks of active enhancers) in extracts from adult whole cardiac left ventricles; (4) assays for transposase-accessible chromatin using sequencing (ATAC-Seq) on cardiomyocyte-specific nuclei extracted and purified from cardiac left ventricles by cytofluorometry, as well as full-length RNA-Seq on total RNA extracted from the same material (mRNA-Seq); and (5) Western blot analyses for total Ddx3 Immunoreactivity on extracts from either adult whole hearts or primary cultures of adult lung fibroblasts. Some of the results from these experiments were verified and complemented by RT-qPCR quantification of particular mRNA transcripts in either cardiac or non-cardiac tissues. Comparisons of the results of the above experiments revealed that, despite significant effects on hundreds of mRNA transcripts in cardiac cells, the genomic regulatory effects Uty\(^\text{GT}\) were much more restricted, as they concentrated mostly on genomic regions on MSY. The data suggested that while the transcriptional effects of the Uty locus were restricted to a handful of regulators, the latter could nonetheless further regulate the transcriptomic profile of cells, presumably via post-transcriptional or post-translational mechanisms.

### Results

#### Regulatory effects of Uty\(^\text{GT}\) on surrounding MSY genes and genomic sequences.

Combined analysis of the results of RNA-Seq, ChIP-Seq and ATAC-Seq demonstrated that the greatest impact of Uty\(^\text{GT}\) was observed on either the abundance of transcripts from surrounding genes or chromatin marks in corresponding genomic regulatory regions. RNA-Seq revealed that Ddx3y and Eif2s3y were among the top most affected genes, either in terms of statistical significance or absolute fold change (Table 1). Ddx3y was the gene whose expression was (after Uty itself) affected most significantly, with its abundance in either hearts or cardiomyocytes from Uty\(^\text{GT}\) being about ~15% of that in their WT counterparts (Table 1). Eif2s3y was among the 60 most affected genes but its expression was (contrary to Ddx3y) increased in hearts or cardiomyocytes from Uty\(^\text{GT}\) mice. RT-qPCR experiments were performed to further validate those results. Since Uty, Ddx3y and Eif2s3y are all ubiquitously expressed MSY genes, non-cardiac tissues (including skeletal muscle, spleen, liver and/or primary lung fibroblasts) were included in the analysis. Regardless of the type of tissue, Ddx3y showed (similarly as detected by RNA-Seq) a greater than five-fold downregulation in samples from Uty\(^\text{GT}\) mice (Fig. 1a). Likewise, Eif2s3y was upregulated in all tested tissues, the abundance of its mRNA transcripts in Uty\(^\text{GT}\) tissues being about 150% of that found in their WT counterparts (Fig. 1b). In additional experiments, we verified whether Uty\(^\text{GT}\) had any effect on expression of the chrX homologs of the above genes. Expression of Utx was unaffected and Ddx3x showed a

|       | Left ventricles | Dispersed cardiocytes |
|-------|----------------|-----------------------|
|       | Uty Ddx3y Eif2s3y | Uty Ddx3y Eif2s3y |
| KO FPKM | 3.2 ± 1.6 | 126.3 ± 49.9 | 16.4 ± 3 |
| WT FPKM | 126.3 ± 49.9 | 741 ± 90.3 | 53 ± 233.3 |
| KO/WT | 5.4% | 17.1% | 32.3% |
| FDR | 1.2 E−40 | 6.6 E−37 | 4.7 E−06 |
| Rank FDR | 1 | 2 | 1 |
| Rank abs. FC | 1 | 2 | 1 |

Table 1. Effects of Uty\(^\text{GT}\) on abundance of transcripts from MSY genes (KO/WT) and significance of the effects in either left ventricles or dispersed isolated adult cardiomyocytes. KO FPKM: FPKM values obtained in Uty\(^\text{GT}\) mice (mean ± SD, n = 7); WT FPKM: FPKM values obtained in WT mice (mean ± SD, n = 7); KO/WT: ratios of the FPKM values obtained for Uty\(^\text{GT}\) mice vs. that of WT mice; FDR: false discovery rate; Rank FDR: relative rank for FDR values obtained for all tested genes; Rank abs. FC: relative rank for absolute fold values obtained for all tested genes.
Regulatory relationships between MSY genes. In primary lung fibroblasts from Uty\textsuperscript{GT} mice, Ddx3y is down-regulated to the same extent as in heart or other non-cardiac tissues (Fig. 1a). To test whether such low
Ddx3y levels could be rescued by re-initiation of expression of the protein-coding part of Uty, we thus infected such cells with viral particles carrying a lentiviral expression vector containing CMV-driven sequences for both Uty and the reporter eGFP gene, and isolated Uty-overexpressing cells by eGFP cytofluorometry. We found that Uty was clearly overexpressed in eGFP-positive cells, but that it was not associated with any changes in the low levels of Ddx3y expression (Fig. 4).

To test whether MSY genes could display high levels of co-expression constitutively, we used tools provided by the COEXPRESdb database to analyze data obtained on multiple tissues with a total of total 2,236 Affymetrix arrays across 154 experiments17. Among all 21,036 tested gene transcripts tested, Uty was found to interact most strongly with the protein coding Ddx3y, Eif2s3y and Uba1y genes, as well as with two other gene products: (1) Gm39552, a three-exon IncRNA gene located within intron 25 of Uty; and (2) and C030026M15Rik, which corresponds to expressed sequences from Uty intron 14 (Supplemental Table S1). No autosomal gene showed co-expression scores approaching those found for MSY genes.

Since steady-state mRNA levels in cell cytoplasm are influenced by a variety of post-transcriptional mechanisms, it has been proposed that deep-sequencing of nuclear RNA (nucRNA-Seq) provides data that are more closely related to true transcriptional output18,19. We thus performed nucRNA-Seq on part of the nuclear material because their coordinates were not included in the standard ENSEMBL mrc GRCm38 gene build annotation file. We thus performed a novel nucRNA-Seq analysis that used a custom annotated file where the positions of potential changes in their expression levels could not have been detected by the nucRNA-Seq as it was performed, transcripts could indeed be observed in samples from WT cardiomyocyte nuclei. However, unlike Gm4017, potential changes in their expression levels could not have been detected by the nucRNA-Seq as it was performed, because their coordinates were not included in the standard ENSEMBL mrc GRCm38 gene build annotation file. We thus performed a novel nucRNA-Seq analysis that used a custom annotated file where the positions of Gm39552 and C030026M15Rik were manually added to the GRCm38 gene build, in addition to that of AK153586, a previously detected expressed RNA transcript whose sequence matched genomic sequences within intron 17 (ENSMUSG00000101059). This intronic localization was reminiscent to that observed for Gm39552 and C030026M15Rik (i.e. the two transcripts reported as co-expressed with Uty in the COEXPRESdb database). Visual inspection of the nucRNA-Seq sequencing tracks showed that these two transcripts could indeed be observed in samples from WT cardiomyocyte nuclei. However, unlike Gm4017, potential changes in their expression levels could not have been detected by the nucRNA-Seq as it was performed, because their coordinates were not included in the standard ENSEMBL mrc GRCm38 gene build annotation file. We thus performed a novel nucRNA-Seq analysis that used a custom annotated file where the positions of Gm39552 and C030026M15Rik were manually added to the GRCm38 gene build, in addition to that of AK153586, a previously detected expressed RNA transcript whose sequence matched genomic sequences within Uty intron 4 (Supplementary Fig. S3). When using these additional parameters, all transcripts emerged as significantly decreased (FDR < 0.05) in samples from UtyGT mice (Table 2). The presence of these transcripts in material from WT cardiomyocyte nuclei was further supported by the fact that FPKM values obtained in corresponding genomic regions were significantly higher (P < 0.05) than values found in remaining intron sequences from the genes they were inbedded in. In contrast, none of these transcripts were detected by the RNA-Seq analysis of cardiomyocyte extracts, where RNA originates predominantly from the cytoplasm.

Effects of UtyGT on autosomal chromatin. Outside of MSY, the H3K27ac ChIP-Seq analysis detected only two autosomal regions with significant differences. The first one comprised three peaks on chr17 (with matching positions of ENCODE tracks for H3K27Ac, p300, Pol2 and DNase hypersensitivity) whose intensity was lower in UtyGT hearts than in their WT counterparts (Supplementary Fig. S3). However, neither RNA-Seq results nor additional RT-qPCR experiments performed on heart extracts (results not shown) showed any evidence of differential expression for any of the genes in the proximity of these peaks (Dynlt1b, Tmem181a, Tmem181b, and the IncRNA gene Gm29719). The second peak was found in a gene-poor intergenic region on chr19, with no matching peak in published ENCODE tracks.

With ATAC-Seq, differently accessible regions in autosomal chromatin were detected chr19 (in the intergenic regions between the long non-coding RNA (IncRNA) gene Malat1 and Neat1 (FDR = 9.6E–19) (Fig. 5).
This signal was matched by increased expression of Malat1 in UtyGT samples, as found in heart samples by RNA-Seq, and further confirmed by RT-qPCR, both for heart as well as non-cardiac tissues (including skeletal muscle, liver and primary fibroblasts) (Fig. 6). We detected no change in expression of the Malat1-neighboring Neat1 gene, either by RNA-Seq or additional RT-qPCR experiments (results not shown).

Increased chromatin accessibility (FDR $1.0 \times 10^{-31}$) was also found on chr5 in the region of the first intron of En2. However, we found no evidence indicating that the corresponding transcript was present in our samples, a finding in keeping with ENCODE transcriptome data showing that expression of this gene is mostly restricted to the developing central nervous system. In addition to En2 and Malat1, some levels of differential accessibility were detected for about 100 other autosomal regions (Supplemental Table S2). However, the level of significance of other differentially accessibility regions dropped sharply compared to the top five most significantly affected regions, and there was no accompanying evidence for differential regulation of nearby genes.

**Impact of UtyGT on the adult heart transcriptome.** The RNA-Seq experiments revealed that UtyGT was accompanied: (2) in whole heart left ventricles, by down- and upregulation of 671 and 716 genes (FDR < 0.05), with absolute fold changes averaging 0.52 and 0.68 fold, respectively; and (2) in adult cardiocytes, by down-

---

**Figure 3.** Graphic representation of results from the H3K27ac ChIP-Seq and ATAC-Seq analyses in MSY regions. Results represent those obtained within the genomic regions of the Uty locus and its two flanking genes. RefSeq gene coordinates are shown on top. Peaks for either H3K27ac ChIP-Seq or ATAC-Seq are shown in either the top or bottom framed areas, respectively. The areas framed in green indicate highlight peaks detected in WT, but not in UtyGT mice. All detected peaks align with those published by ENCODE (in red) and obtained using extracts from adult mouse hearts for either H3K27ac, p300 or Pol2 ChIP-Seq or Dnase hypersensitivity assays.
and upregulation of 669 and 476 genes (FDR < 0.05), with absolute fold changes averaging 0.49 and 0.63 fold, respectively. A total of 151 genes were common to both datasets. To test whether the transcriptomic changes could concern particular cellular functions, we looked for GO ontology terms showing significant enrichment (FDR < 0.01) among the UtyGT-affected genes (FDR < 0.05) with > twofold enrichment, then regrouped the terms meeting these criteria within broader categories that could encompass them. Among genes down-regulated in UtyGT mice, highly enriched GO terms could be consolidated with 2–3 overarching categories, accounting for 59% and 70% of GO terms enriched in differentially expressed genes in left ventricles and cardiomyocytes, respectively (Table 3) (Supplemental Table S6). Likewise, among genes up-regulated in UtyGT mice, enriched GO terms could be consolidated with 3–5 overarching categories, accounting for 45% and 50% of all enriched GO terms in left ventricles and cardiomyocytes, respectively (Table 4) (Supplemental Table S6). Both in ventricular extracts and isolated cardiomyocytes, genes up-regulated in UtyGT mice showed high and significant enrichment for ribosome-related GO terms.

**Discussion**

Despite mounting correlative evidence that MSY genes have effects outside of reproductive tissues, studies using gene targeting experiments to demonstrate in a direct fashion their impact within adult somatic cells and/or their mechanisms of action are still lacking. The current study demonstrates that a gene trap within the Uty gene does have demonstrable effects on adult somatic cell functions, but also that MSY genes may have some idiosyncratic properties. The most prominent effect of the UtyGT gene trap is that it is accompanied by a profound and ubiquitous downregulation of its neighboring Ddx3y gene and, by a smaller extent, by up-regulation of its other flanking gene (Eif2s3y). The dual effects of UtyGT on expression of both Uty and Ddx3y were accompanied by direct regulatory genomic changes involving a disappearance of both H3K27Ac and ATAC-Seq peaks in the proximal enhancer regions of these two genes. In contrast, the protein-coding product of the Uty gene did not appear to have direct regulatory actions on Ddx3y expression, because its re-expression in Uty GT primary cells did not alleviate the downregulation of Ddx3y. One possible explanation for these observations (for which there is actually some support within the literature) could be that non-coding sequences within the Uty/Ddx3y locus

---

**Table 2.** Top most differentially abundant RNA transcripts detected by the nucRNA-Seq analysis with custom annotation parameters. The FPKM columns contain the values found in either UtyGT or WT mice (mean ± SD, n = 3).

| Ensembl gene ID | Gene name | mRNA | Gene_biotype | Localization | KO FPKM | WT FPKM | KO/WT | P-value | P-adj (FDR) |
|-----------------|-----------|------|--------------|--------------|---------|---------|-------|---------|------------|
| ENSMUSG00000068457 | Uty | Protein_coding | MSY | 18.3 ± 3.7 | 1,146.2 ± 72.6 | 0.02 | 2.02E−135 | 6.03E−131 |
| AK153586 | AK153586 | Expressed seq | Uty intron 4 | 0.0 | 158.9 ± 23.2 | 0.00 | 1.08E−15 | 1.23E−11 |
| Gm39552 | IncRNA | Uty intron 25 | 0.0 | 160.8 ± 36.4 | 0.00 | 1.24E−15 | 1.23E−11 |
| Cj00026M15Rik | AK049876 | IncRNA | Uty intron 14 | 0.0 | 65.5 ± 7.3 | 0.00 | 1.54E−11 | 1.14E−07 |
| ENSMUSG00000069045 | Ddx3y | Protein_coding | MSY | 191.9 ± 91.1 | 1,229.1 ± 97.5 | 0.16 | 1.97E−09 | 1.17E−05 |
| ENSMUSG000000101059 | Gm4017 | Processed_pseudogene | Ddx3y intron 17 | 4.3 ± 3.7 | 33.1 ± 4.7 | 0.13 | 7.31E−05 | 3.63E−01 |
| ENSMUSG00000068745 | Mybphl | Protein_coding | Chr3 | 18.7 ± 12.1 | 72.5 ± 29.1 | 0.26 | 1.65E−04 | 6.71E−01 |
| ENSMUSG00000089049 | Eif2s3y | Protein_coding | MSY | 471.9 ± 93.1 | 732.6 ± 54.7 | 0.64 | 1.92E−04 | 6.71E−01 |
| ENSMUSG0000056673 | Kdm5d | Protein_coding | MSY | 780.9 ± 127.6 | 1,135 ± 27.3 | 0.69 | 2.03E−04 | 6.71E−01 |
actually play a role in the co-regulation of these two genes. Indeed, comparisons of chrY sequences from five mammalian species have revealed that, despite highly divergent gene order among all of them, the \( Uty/Ddx3y \) gene pair is part of a microsyntenic region that has been preserved over 340 million years of independently sampled evolutionary history\(^{20} \). Other phylogenomic studies have shown previously that: (1) such conserved ancient microsyntenic pairs arise when conservation of \( cis \)-regulatory sequences was required for co-regulation of the genes, and (2) such sequences were enriched within non-coding regions of the locus, and often resided within particularly large introns of the genes within the conserved pair\(^{21} \). In fact, our nucRNA-Seq analysis revealed that four transcripts corresponding to genomic sequences within three introns of \( Uty \) and one of \( Ddx3y \) could indeed be detected in nuclei from WT mice, with some of them having been previously been detected in transcriptomic studies performed across a wide range of tissues and whose results have been compiled in publically available databases. In the present study, we also observed that abundance of all these non-coding transcripts was greatly
decreased in nuclei from Uty<sup>GT</sup> mice cardiomyocytes. It thus appeared that the disappearance of H3K27ac and ATAC-Seq peak in the proximal enhancer regions of Uty and Ddx3y was accompanied by downregulation of not only the protein-coding parts of these genes, but also of transcripts originating from sequences within their introns.

In a more general fashion, the co-expression databases showed that high levels of co-expression were a defining feature of MSY genes in general. As this occurred not only in mice, but in humans as well, it may constitute an evolutionary conserved characteristic of MSY genes in vertebrates. Of note, such patterns of co-expression between neighboring genes are rarely seen within vertebrate genomes, and do almost never occur at such high levels. Nonetheless, such characteristics may have emerged within the male chrY because evolutionary forces that are different from those exerted on autosomes have resulted in different types of regulatory mechanism. One other evolutionary conserved feature appeared to be that, among the most highly co-expressed MSY genes, a high proportion of them [four out of eight of genes (50%) in mice, and 15 out of 23 (65%) in humans] were in fact non-coding. Interestingly, a recent study analyzing transcripts in brains from human and chimpanzee embryos discovered six novel non-annotated lncRNAs originating from chrY. Such RNA transcripts may thus constitute an evolutionary conserved characteristic of MSY genes in general. As this occurred not only in mice, but in humans as well, it may constitute a class of agents that mediate some MSY effects or, alternatively, participate to the high levels of co-expression among MSY genes. Indeed, while three-dimensional looping events might be invoked in mice (where all of the so-called "X-degenerate MSY genes" are contained within a relatively small 1 MB genomic region) as possible mechanisms underlying high levels of co-expression among MSY genes, such mechanisms are less likely to occur in humans, where the loci of MSY genes show much greater dispersion across much larger regions. Moreover, we found that the non-coding RNAs co-expressed with Uty were detected in samples from cardiomyocyte nuclei but not in those extracted from whole cardiomyocytes, indicating that these transcripts correspond in fact nuclear lncRNAs. The latter are known to participate (among other biological processes) to chromatin organization and accessibility, in part via recruitment of chromatin-modulating proteins to nearby chromatin regions. In particular, given that Uty resides within the same Uty intron that harbors AK153586, decreased production of the latter could constitute one initial event triggering other chromatin rearrangements within the whole locus.

Although DDX3Y has been shown to be functionally interchangeable with DDX3X in cultured mammalian cells, some controversy remains concerning the production of the cognate protein outside of reproductive tissues. Indeed, some have argued that the production of the DDX3Y protein was (in contrast to the ubiquitous expression of the gene) restricted to the male germline in cultured mammalian cells. In contrast, other studies have detected DDX3Y peptides in either neural progenitor or hematopoietic stem cells. Our Western blot analysis are in line with the latter studies, providing evidence that Ddx3y can indeed be translated into a protein in at least some mouse somatic cells. Moreover, the downregulation of Ddx3y caused by Uty<sup>GT</sup> is accompanied by an equivalent reduction in overall Ddx3 immunoreactivity. The importance of this finding is many-fold. For instance, it has been shown that hematopoietic loss of DDX3X results is lethal in developing females, but viable in males; however, without

### Table 3. Common themes encompassing GO terms showing both high (> twofold) and significant (FDR < 0.01) enrichment among genes downregulated in Uty<sup>GT</sup> mice. More details about each specific GO term and statistical values for their enrichment are present in Supplemental Table S6.

| # of GO terms | % of GO terms | Common theme | Mean FE | # of GO terms | % of GO terms | Common theme | Mean FE |
|---------------|---------------|--------------|---------|---------------|---------------|--------------|---------|
| 56            | 100           | Total        |         | 261           | 100           | Total        |         |
| 17            | 30            | Post-translational fate of proteins | 2.6     | 107           | 41            | Organo-/morpho-genesis | 4.6     |
| 16            | 29            | Organization of intracellular membrane structures | 4.3     | 41            | 16            | Responses to EC stimuli | 4.8     |
|               |               |              | 35       | 13            | Cell migration/proliferation/contacts | 4.1     |

### Table 4. Common themes encompassing GO terms showing both high (> twofold) and significant (FDR < 0.01) enrichment among genes upregulated in Uty<sup>GT</sup> mice. More details about each specific GO term and statistical values for their enrichment are present in Supplemental Table S6.

| # of GO terms | % of GO terms | Common theme | Mean FE | # of GO terms | % of GO terms | Common theme | Mean FE |
|---------------|---------------|--------------|---------|---------------|---------------|--------------|---------|
| 111           | 100           | Total        |         | 224           | 100           | Total        |         |
| 20            | 18            | Ribosomal activity | 6.7     | 42            | 18            | Mitochondrial activity | 7.5     |
| 18            | 16            | RNA processing | 3.3     | 36            | 16            | Muscle contractile apparatus | 5.9     |
| 12            | 11            | Chromosome maintenance | 3       | 14            | 6             | Protein catabolism | 4.8     |
|               |               |              | 13       | 6             | Ribosome organization | 7.3     |
|               |               |              | 13       | 6             | Cytoskeleton |         |
knowing whether production of the DDX3Y protein can occur in somatic cells, it was not possible to determine whether this compensatory effect of MSY could be due to DDX3Y itself. Likewise, since the actions of DDX3X are dosage-related, confirmation of the production of DDX3Y protein supports its possible contribution to dosage-related effects. Of note, due to differences in their promoters, the DDX3Y and DDX3X genes have been shown to be independently regulated33, indicating that male organisms may regulate overall DDX3X immunoreactivity in part via mechanisms that are different from those in females.

When the results of H3K27ac ChIP-Seq, ATAC-Seq, RNA-Seq and nucRNA-Seq are all considered collectively, the genes contained within the Uty/Ddx3y locus were the only ones for which evidence for transcriptional regulation via direct regulatory genomic effects was entirely consistent across experiments. For Malat1, increased chromatin accessibility was matched by increased abundance of mRNA abundance in tissue extracts, but not in nuclear extracts. For Eif2s3y, the results were more difficult to interpret, as decreased chromatin accessibility in its proximal promoter region was matched by a trend towards decreased abundance in nuclear extracts (Table 2), which contrasted with a significant (FDR < 0.01) increase in the abundance of its transcript in tissues (see discussion below). Other indications of effects of UtyGT on either H3K27ac abundance or accessibility in autosomal chromatin were not matched by evidence for changes in the abundance of the transcripts of genes in corresponding regions. While the direct transcriptional effects of UtyGT thus appeared to be relatively restricted, the gene trap nonetheless had substantial effects on the transcriptome of cardiac cells, as hundreds of mRNA transcripts showed differential abundance. The magnitudes of the average fold-changes (ranging from 0.49 to 0.62) were relatively modest, but within the range of those reported by others that have studied sex biases in gene expression32,34. Moreover, there was functional coherence in the transcript abundance changes, as GO term analyses revealed that the majority of terms showing enrichment in differentially regulated genes showed great overlap within a few well-defined functional categories. Interestingly, others have also reported recently that a knockdown of UTY in human macrophages affects the abundance of many mRNA transcripts, and functional coherence was shown by the fact that the affected mRNA transcripts showed statistically significantly enrichment within 59 particular pathways6. Altogether, these findings indicate that changes in expression of MSY genes can indeed affect particular cell functions, these effects being mediated via coordinated changes in the abundance of mRNA transcripts within functionally coherent pathways.

In the present work, the functions of genes downregulated in left ventricles of UtyGT mice corresponded to post-translational protein processes (including protein localization, oligomerization and phosphorylation) and maintenance of intracellular membranes (including organelles, mitochondria and sarcolemma). For genes upregulated in left ventricles of UtyGT mice, these functions corresponded to ribosomal activity (including ribosomal organization and translation), processing and metabolism of RNA, and chromosome maintenance. In one additional analysis, we compared the genes whose transcript abundance was affected by the UtyGT in mouse cardiac left ventricles to a group of 470 transcripts whose abundance was previously found to be affected (at the FDR < 0.05 level) by UTY knockdown in human macrophages (J. Eales and M. Tomaszewski, personal communication). We found overlap for a total of 31 genes, with the latter showing significant enrichment (FDR < 0.05) for GO ontology terms related to regulation of histone modification, histone H3K9/H3K4 methylation, protein modification and cellular responses to fatty acids (Supplementary Table S9). Overall, these findings remain compatible with the proposed concept that the MSY genes with functional counterparts on chrX are regulators of some of the most basic cellular homeostatic maintenance mechanisms12. In the current study, the genes belonging to categories related to ribosomal activity and RNA processing might partly explain the somewhat contradictory results obtained for Eif2s3y. Indeed, the RNA-Seq data for left ventricular extracts indicated that, in addition to increased Eif2s3y, the transcripts of three other eukaryotic initiation factors (Eif4a3, Eif4e3, and Eif5) were significantly increased, whereas those of three others (Eif4g1, Eif4f3, and Eif3h) were significantly decreased (FDR < 0.01). Accordingly, UtyGT might affect the stability of several eukaryotic initiation factor transcripts, and affect that of Eif2s3y in a manner that is contrary to its effects on transcription of the gene.

Interestingly, at least two of the genes for which we observed evidence of increased transcription are known regulators of post-transcriptional mechanisms. On one hand, the DDX3 proteins are RNA helicases genes involved in all aspects of RNA metabolism, including mRNA nuclear export, translation and decay, and ribosome biogenesis33,35. In particular, they affect protein translation by a variety of mechanisms that include interactions with translation initiation factors36,37, assembly of functional ribosomes38 and N6-methyladenosine modifications39,40. On the other hand, Malat1 may regulate: (1) mRNA transcript stability via its ability to behave as a miRNA-binding competitive endogenous RNA; and (2) regulate the activity of proteins and/or assist in their cellular localization via its direct protein-binding capabilities41. All these functions are related to the GO categories showing highest levels of enrichment among UtyGT-regulated gene transcripts (i.e. to ribosomal activity and RNA processing). The above two genes may even work in complementary fashion, as DDX3X regulates the abundance of N6-methyladenosine residues in Malat1, which in turn may modulate its biological activity42,43. Altogether, the current study shows that MSY genes may have demonstrable effects on regulators of adult somatic cell functions. However, such effects might result from MSY behaving as an entire integrated regulatory unit rather than from individually regulated MSY genes, and might involve mechanisms other than just transcriptional regulation of autosomal protein-coding genes. These findings could constitute a stepping stone to better understand in which fashion chrY-related genetic differences may affect health-related outcomes.

Material and methods
Mouse strains and cells. All procedures involving mice were conducted following approval by the animal ethic committee of the Institut de Recherches Cliniques de Montréal (IRCM) and in agreement with the guidelines of the Canadian Council for Animal Care. Uty+/+X;Ddx3y+/+ (UtyGT in short) ES cells (derived from the original parental E14Tg2a.4 ES cell line from 129P2/OlaHsd mice) were obtained from the Mutant Mouse
Resource & Research Centers (MMRRC) and injected into C57BL/6J host blastocysts for chimera generation. Male chimeras were crossed to female C57BL/6J to assess germine transmission, and resulting males harboring the Y<sup>CM</sup>-<sup>WT</sup> were mated for more than ten generations to female C57BL/6J mice. Isogenic control WT mice were generated by backcrossing 129P2/OlaHsd male mice to female C57BL/6J mice for the same number of generations. Primary fibroblasts were obtained from lungs of adult mice as described previously. When cells reached confluence 6 days after initial plating, they were split 1:4 for passage 1 (P1) and allowed to reach confluence again. For experiments, all fibroblasts were passage 2. Mouse adult cardiocytes were obtained by digesting adult mouse hearts by retrograde perfusion of a solution of collagenase type 2 (Worthington, Lakewood, NJ) again. For experiments, all fibroblasts were passage 2. Mouse adult cardiocytes were obtained by digesting adult mouse hearts by retrograde perfusion of a solution of collagenase type 2 (Worthington, Lakewood, NJ) in calcium-free solutions, followed by gravity sedimentation within 5% bovine serum albumin (BSA) solutions supplemented in a step-wise fashion with increasing CaCl<sub>2</sub> concentrations, as described. After isolation, the cells were plated on laminin-coated dishes, and used 4 h after plating for RNA isolation.

**Isolation and flow-cytometry sorting of cardiomyocyte nuclei.** Cardiomyocyte-specific nuclei were isolated from frozen left ventricular cardiac sample as described, with modifications to optimize maintenance of RNA integrity. In short, left ventricle tissue (priorly powdered under liquid nitrogen) was homogenized with a Polytron in a 40 mM Tris/10 mM citrate buffer (to inhibit ribonucleases) containing 5 mM KCl, 12% polyethylene glycol 8,000 (to allow for isolation of nuclei in the absence of ribonuclease-activating ions), 1 mM DTT, 1 mM PMSE, protease inhibitors, 1.5 mM spermine and 20 mM sodium propionate (for chromatin integrity). The extracts were processed with a Dounce homogenizer and passed sequentially through 98 µm and 38 µm cell selectors. After spinning, the nuclei were resuspended in 1 ml of the same buffer supplemented with 3 µl of an Alexa fluor 642-coupled antibody raised against the pericentriolar 1 protein (PCM1) (SCBT, Dallas, TX), i.e. an antigen that accumulates specifically around nuclei from cardiomyocytes. A sub-aliquot of that material was used for counterstaining with Hoechst bis-benzidine. After 30 min incubation at 4 °C, the pellets were washed and resuspended in 1 ml of Tris/acetate buffered saline supplemented with 2% ribonuclease-free BSA, sorted out by cytfluorometry, and collected in tubes containing 200U of the Ribolock ribonuclease inhibitor (Thermofisher). The number of nuclei isolated in this manner were 50,000 and 150,000 for ATAC-Seq and RNA-Seq, respectively. Parallel analysis of the Hoechst-counterstained sub-aliquots showed that ~95% of PCM1 positive particles were Hoechst-positive nuclei, and comprised in average about 30% of all nuclei in the preparation.

**Sequencing-based procedures.** Total RNA was isolated from tissues and cells using RNeasy micro kits with in-column DNase treatments (Qiagen, Toronto, Canada). For cardiac left ventricular samples (n = 7 for each mouse strain), preparation of 3’ end RNA sequencing libraries were prepared (using the Lexogen QuantSeq 3’ mRNA-Seq Library Prep Kit FWD library) and sequenced by the UCLA Neuroscience Genomics Core (UNGRC). Read-quality confirmation and alignment on the mouse GRCh38/mm10 genome were performed using Lexogen’s QuantSeq Bluebee analysis platform. For RNA from isolated adult cardiomyocytes (n = 3 for each mouse strain), ribodepletion was performed using the Fast Select mouse RNA removal kit (Qiagen), the RNA-seq libraries were prepared according to manufacturer’s instructions using the Illumina TrueSeq Stranded mRNA Kit and 50 bp single reads sequencing was performed with the Illumina HiSeq 4,000 Sequencer. For the smaller amounts of RNA obtained from cardiomyocyte nuclei (n = 3 for each mouse strain), RNA-seq library preparation was performed using the SMARTer Stranded Total RNA-Seq Kit v2—Pico (Takara Bio, Mountain View, CA) prior to 100 bp paired-end sequencing. For the last two sequencing analyses, after assessing the read quality using FastQc 0.11.8, alignment on the mouse GRCh38/mm10 genome was performed using STAR 2.5.1b. For all RNA-seq data, feature quantification was extracted with featureCounts v1.6.0 on the GRCh38 v98 Ensembl annotation. We also quantified additional transcript in the Uty region (AK142197, AK153586, Gm39552 and C030026M15s) as were not included in the Ensembl reference annotation. Differential expression analysis was performed with DESeq2 1.22.2 R package. For the mncRNA-Seq analysis, additional steps were taken to exclude artificial contamination with genomic DNA. In particular, we identified a total of two megabases of regions where the absence of transcription was assessed based on the absence of repetitive signal overlapping criteria: (1) lack of data showing expression of corresponding transcripts in hearts (on the basis of mouse ENCODE transcriptome data), in contrast to other non-cardiac tissues where the pattern of tissue expression was restricted to just a few tissue types; and (2) lack of data for chromatin accessibility in the same regions, on the basis of ENCODE DNase hypersensitivity tracks obtained in mouse adult hearts. UCSC exons and introns of these regions were extracted with GenomicFeatures R package and quantified as FPKM values with featureCounts. Absence of genomic contamination was confirmed by verifying that FPKM values were either zero or very low (Supplemental Table S6).

ChIP-Seq was performed on two replicate samples from each mouse strain as we described previously, using the ChIP grade ab4729 anti-H3K27 antibody from Abcam (Cambridge, MA). For the analysis, read quality was confirmed using FastQC 0.11.8 before alignment using Bowtie2 2.2.6 on the mouse GRCh37/mm9 genome (that latter build was chosen in order to allow for comparisons with other ChIP tracks published by ENCODE). Peak calling and differential enrichment were performed using the MACS2 2.0.10 “callpeaks” and “bdgdiff” functions, respectively. For ATAC-Seq, preparations containing 50,000 nuclei from either WT or Uty<sup>CM</sup> hearts were performed on three different days, and processed as described previously. For the analysis, sequences from the adapters (NexteraPE-PE) were removed with Trimomatic 0.36 after read quality control. Alignment was performed as for ChIP-Seq post-processing to remove PCR duplicates (Picard tool 2.4.1) and reads mapping to mitochondrial DNA (samtools 1.8). In order to represent the real Tn5 transposase binding sites of 9 bp, the coordinates of the reads were shifted by 4 bp for the plus strand and by 5 bp for the minus strand, using DeepTools 3.0.1. The former was also used to remove ENCODE's blacklisted regions (signal artefact regions) and convert
Bam files to BEDPE format. MACS2 was used to identify significant peaks using q-value < 0.01. DiffBind 2.10.0 R package was used to generate the count matrix of Tn5 insertion site numbers for each consensus peak (peaks that were present in at least two samples). Differentially accessibility regions (DARs) were identified between conditions using edgeR R package 3.24.3, including its generalized linear model functionalities to perform a multi-factor analysis to control for batch effects.

Data deposition. Results from all the sequencing-based experiments were deposited in the Gene Expression Omnibus database as SuperSeries GSE147652.

Molecular biology. Full-length mUtx cDNA was obtained by PCR amplification of reverse-transcribed mouse heart RNA, then subcloned into the pLRES-V5 vector (gift of N. Seidah). After full sequence verification, the whole cDNA was subcloned into the lentiviral CMV GFP destination vector 736-1 (Addgene #19732), and further used along with packaging vectors to generate lentiviral particles (control lentiviral particles were generated in the same manner using empty lentiviral vectors). For infection, confluent P1 mouse fibroblasts were passaged at a concentration of 50,000 cells/2 cm² dishes. Twenty four hours after replating, the cells were exposed for 6 h to lentiviral particles (M.O.I = 3) in growth medium supplemented with 6 μg/ml dextran sulfate, then washed and grown further overnight. Cells were then cryopreserved for cytofluorometric separation of GFP-positive (i.e. cells with lentiviral integration) and GFP-negative cells. A total of 40–60,000 cells were then used for each sample to extract total RNA using the RNeasy micro RNA kit.

RT-qPCR was performed on cDNA obtained from total RNA using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher), using the primers detailed in Supplementary Table S5. Results were expressed as relative expression ratios of the 2(− ΔΔCt) values of the gene of interest vs. of that of the Rps16 normalizing gene.

Western blotting. After separation of gels and transfer to PVDF membranes, the latter were incubated with the following primary antibodies: monoclonal anti DDX3C C4, raised against the 114N-terminal portion of human DDX3x (Santa Cruz Biotechnology); polyclonal anti DDX3, raised against recombinant human DDX3x (Genetex, Irvine, CA); polyclonal anti GAPDH (Proteintech, Rosemont, IL). After treatment with horseradish peroxidase coupled second antibodies, signal was revealed by chemiluminescence, then visualized and analyzed using the Image lab software (Bio-Rad Laboratories, Saint-Laurent, QC, Canada).

Received: 10 April 2020; Accepted: 13 August 2020
Published online: 10 September 2020

References
1. Prokop, J. W. & Deschepper, C. F. Chromosome Y genetic variants: Impact in animal models and on human disease. Physiol. Genom. 47, 525–537 (2015).
2. Case, L. K. & Teuscher, C. Y genetic variation and phenotypic diversity in health and disease. Biol. Sex Differ. https://doi.org/10.1186/s13293-015-0024-z (2015).
3. Kremenetsov, D. N. et al. Genetic variation in chromosome Y regulates susceptibility to influenza A virus infection. PNAS 114, 3491–3496 (2017).
4. Prokop, J. W. et al. The phenotypic impact of the male-specific region of chromosome-Y in inbred mating: the role of genetic variants and gene duplications in multiple inbred rat strains. Biol. Sex Diff. 7, 10 (2016).
5. ALSiraj, Y. et al. XX sex chromosome complement promotes atherosclerosis in mice. Nat. Commun. 10, 1–13 (2019).
6. Eales, J. M. et al. Human Y chromosome exerts pleiotropic effects on susceptibility to atherosclerosis. Arterioscler. Thromb. Vasc. Biol. 39, 2386–2401 (2019).
7. Fonseca, I. A. et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat. Genet. 46, 624–628 (2014).
8. Duman, J. et al. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. Am. J. Hum. Genet. 98, 1208–1219 (2016).
9. Grassmann, F. et al. Y chromosome mosaicism is associated with age-related macular degeneration. Eur. J. Hum. Genet. 27, 36–41 (2019).
10. Hattat, S. et al. Loss of Y chromosome in blood is associated with major cardiovascular events during follow-up in men after carotid endarterectomy. Circ. Cardiovasc. Genet. https://doi.org/10.1161/CIRCGENETICS.116.001544 (2017).
11. Skalsky, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837 (2003).
12. Bellott, D. W. et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508, 494–499 (2014).
13. Sekiguchi, T., Iida, H., Fukumura, J. & Nishimoto, T. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. Exp. Cell Res. 300, 213–222 (2004).
14. Wang, T. et al. Identification and characterization of essential genes in the human genome. Science 350, 1096–1101 (2015).
15. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. PLoS Genet. 8, e1002964 (2012).
16. Gozdecka, M. et al. UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemiaogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat. Genet. 50, 883–894 (2018).
17. Obayashi, T. & Kinoshita, K. COXPRESdb: A database to compare gene coexpression in seven model animals. Nucleic Acids Res. 39, D1016–1022 (2011).
18. Mitchell, J. A. et al. Nuclear RNA sequencing of the mouse erythroid cell transcriptome. PLoS ONE 7, e49274 (2012).
19. Preissl, S. et al. Deciphering the epigenetic code of cardiac myocyte transcriptionnovelty and significance. Circ. Res. 117, 413–423 (2015).
20. Li, G. et al. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. 23, 1486–1495 (2013).
The work was partly funded by Grant MOP-93583 from the Canadian Institutes for Health Research (CIHR). I am deeply grateful for the help of the IRCM Bioinformatics, Molecular Biology, Cytofluorometry and Animal Model Laboratories, and acknowledge in particular the contributions of Virginie Calderon, Caroline Grou, Odile Neyret, Myriam Rondeau, Eric Massicotte, Julie Lord and Manon Laprise. I also thank Sylvie Picard for technical help, Mariana Bego and Mélanie Laporte for help with generation of lentiviral particles, and Pierre Grou for technical help. I also thank Sylvie Picard for technical help, Mariana Bego and Mélanie Laporte for help with generation of lentiviral particles, and Pierre Grou for technical help. I also thank Sylvie Picard for technical help, Mariana Bego and Mélanie Laporte for help with generation of lentiviral particles, and Pierre Grou for technical help.

Acknowledgements

The work was partly funded by Grant MOP-93583 from the Canadian Institutes for Health Research (CIHR). I am deeply grateful for the help of the IRCM Bioinformatics, Molecular Biology, Cytofluorometry and Animal Model Laboratories, and acknowledge in particular the contributions of Virginie Calderon, Caroline Grou, Odile Neyret, Myriam Rondeau, Eric Massicotte, Julie Lord and Manon Laprise. I also thank Sylvie Picard for technical help, Mariana Bego and Mélanie Laporte for help with generation of lentiviral particles, and Pierre Bensidoun for immunocytochemistry and microscopy. I thank J. Eales and M. Tomaszewski for sharing data pertaining to one of their previous publication.

Author contributions

C.F.D. was sole author responsible for the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-71447-3.

Correspondence and requests for materials should be addressed to C.F.D.

Reprints and permissions information is available at www.nature.com/reprints.
