An epigenomic investigation of atrial fibrillation in a matched left and right atrial human cohort

Adrian Rodriguez[1], Stephanie Frost[1,3], Alison M Thomas[2], Diego Fernandez-Aroca[1], Jishan Choudhury[1], Georgios Tikkas[1], Andrew Tinker[2], Patricia B Munroe[2], Christopher G Bell[1,2] and Diego Villar[1,#]

[1] QMUL Centre for Epigenetics, Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, United Kingdom

[2] Clinical Pharmacology & Precision Medicine, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[3] Present address: Genetic Modification Service, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, United Kingdom

[#] Correspondence to d.villarlozano@qmul.ac.uk

ABSTRACT

As the most prevalent cardiac arrythmia, atrial fibrillation is an important contributor to cardiovascular morbidity and mortality. Human population findings increasingly support its complex genetic architecture, with most genetic association signals for atrial fibrillation found in the non-coding genome. In this study, we integrated genome-wide histone modification, gene expression and DNA methylation levels in a paired left and right atrial cohort comprising permanent atrial fibrillation patients and sinus rhythm controls. First, we first identified epigenomic regions enriched in histone H3 lysine 27 acetylation (H3K27ac) across left and right atria from patients and controls, and associated them with differentially expressed genes to derive a set of dysregulated candidate loci – including *NPPB* and *SCX*. Second, by incorporating an independent replication cohort, we were able to validate gene expression and epigenomic differences for a subset of these candidate loci. Third, we profiled base-resolution DNA methylation levels and identified differentially-methylated regions (DMRs) between atrial fibrillation and sinus rhythm samples. Integration of these data with histone modification levels and gene expression allowed us to propose epigenetic mechanisms underlying transcriptomic and epigenomic changes across dysregulated loci, such as disruption of transcription factor binding by DNA methylation at the *LRRC4B* locus. The data and analyses we report constitute a systematic investigation of gene regulatory alterations across the left and right atria in permanent atrial fibrillation.

INTRODUCTION

Atrial fibrillation (AF) is the most common form of cardiac arrythmia, affecting 0.5-1% of the worldwide population [1] and increasing up to 10% in the elderly. As such, AF significantly contributes to cardiovascular morbidity and mortality, and associates with other risk factors such as male gender, obesity, diabetes or smoking [2]. Mechanistically, AF is thought to arise from a complex interplay of triggers and substrate abnormalities that initiate and maintain arrythmia in the atria [3, 4]. Triggers are commonly caused by ectopic spontaneous firing in the pulmonary veins and can be maintained by re-entrant activity or rapidly discharging focuses. Subsequent remodeling of the atria - via altered activity of ion channels and structural changes favouring fibrosis – increases re-entry and can result in permanent AF [5, 6].

The genetic architecture of AF is complex and involves both rare and common variants that increase arrythmia susceptibility [7]. Approaches such as genetic linkage have attributed familial AF cases to mutations in genes coding for atrial natriuretic peptide[8], ion channel subunits [9-12], contractile proteins [13, 14] or cardiac transcription factors [15], among others – however, these mutations are absent in most early-onset AF cases. Genome-wide association studies (GWAS) have revealed over two hundred genetic loci associated to AF susceptibility [16-19]. Most of these association signals map to common genetic variation in the non-coding genome [20], with proposed effector genes including *PITX2* [21-23], *ZFHX3* [24, 25], HAND2 [21, 26, 27] or *SYNPO2L* [28, 29]. Collectively, AF GWAS loci are thought to act via effector genes leading to cardiac structural remodeling, as observed in atrial cardiomyopathy [7, 30].

Both the genetics and epidemiology of AF suggest an important role for gene regulation and epigenetic processes in mediating arrythmia susceptibility and cardiac remodeling [20, 31], which has prompted transcriptomic and epigenomic investigations in AF patient samples. First, previous studies have compared gene expression levels in human atria for AF patients and control individuals categorised as sinus rhythm (SR) (reviewed in [32]), with some interrogating paired right and left atrial samples [33-36]. Although observed gene expression

differences across studies were often specific to patient cohorts, recent large-scale transcriptomic analyses found several putative AF susceptibility genes among those differentially expressed in the atria of AF patients [23]. Second, epigenomic studies of AF have identified genomic regions showing locus-specific variation in DNA methylation levels in AF [37-39], such as hypermethylation of the *PITX2* promoter in left atrium AF samples [40]. These analyses include epigenome-wide association studies (EWAS) of AF, albeit some have been conducted in blood rather than atrial tissue [38]. In contrast, AF epigenomic studies of other regulatory systems, such as histone modifications, have received less attention.

Here, we sought to integrate previously reported transcriptomic data in a paired left and right atrial appendage cohort [33] with epigenomic profiling of histone modifications and DNA methylation levels, obtained from a largely overlapping set of samples. By identifying candidate regions with differential epigenomic signals within the pathogenically critical tissue in AF patients and associating them with downstream gene expression, our results can inform the regulatory mechanisms underlying AF susceptibility and atrial remodeling.

RESULTS

An integrated epigenomic and transcriptomic dataset in left and right atrial appendages from atrial fibrillation patients and controls (related to Figure 1)

To investigate the regulatory landscape of human atria and its alterations in permanent AF, we conducted epigenomic profiling of histone modifications and DNA methylation in a set of paired left and right atrial appendage samples obtained from 7 atrial fibrillation (AF) patients and 5 sinus rhythm (SR) controls (Methods, Table S1, Figure S1a), and compared these data with previously obtained transcriptomes in a largely matched set of samples [33]. We first focused on genome-wide profiling of histone H3 lysine 27 acetylation (H3K27ac), a histone modification associated with active promoter and enhancer regions [41-43]. In line with previous reports [44-46], we identified an average of ~50,000 enriched regions per sample, most of which were distal to annotated transcriptional start sites (Figure S1b-f).

Principal components analysis of these data showed substantial heterogeneity in epigenomic signals across patients, with variance on the first two principal components mainly separating right atrium (RA) and left atrium (LA) samples (Figure 1A). Similarly, H3K27ac epigenomic enrichment showed higher pairwise correlations across samples from the same anatomical side, rather than disease status (Figure 1B). In agreement, examination of atrial expression markers associated to the left (*PITX2*) or right (*BMP10*) atrium showed clear differences in locus-level H3K27ac enrichment across samples from either anatomical side (Figure 1C). In contrast, cardiomyocyte loci such as *MYH6* and *NPPA* displayed similar enrichment levels in both right and left atrium samples. In sum, genome-wide H3K27ac profiling in atrial samples revealed widespread differences between anatomical sides, compared to more subtle epigenomic changes for disease status across samples.

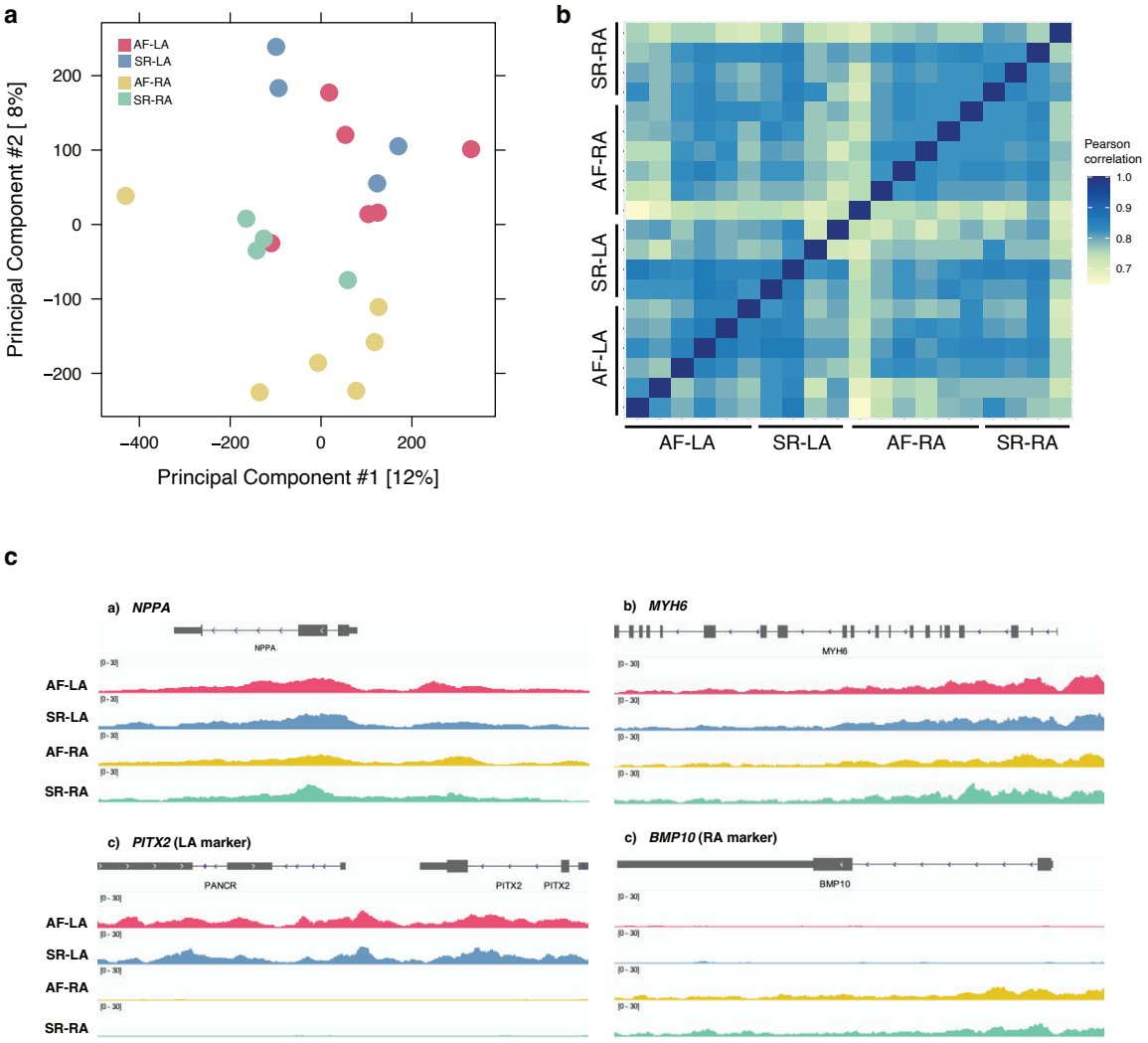

Figure 1: H3K27ac profiling across right- and left- atrial samples in a cohort of atrial fibrillation patients and sinus rhythm controls

a. H3K27ac enrichment differences across samples plotted on the first two principal components. Sample groups correspond to either left- (LA) or right- (RA) atrium samples from individuals with persistent atrial fibrillation (AF) or a normal heart rhythm (sinus rhythm, SR). Epigenomic differences more readily distinguish samples across anatomical side than disease status.

b. H3K27ac enrichment levels across samples are represented as pairwise Pearson correlations. Samples corresponding to the same anatomical side tend to show higher correlations.

c. Genome track examples of H3K27ac levels across the four sample groups for cardiac loci *MYH6* and *NPPA*, and *PITX2* and *BMP10* as left and right atrium loci, respectively. Each track represents average fold ChIP enrichment over input across samples in each group (Methods).

Definition of promoters and enhancers enriched in H3K27ac across disease and control samples (related to Figure 2)

We next aimed to identify promoters and enhancers showing an enrichment in H3K27ac levels between AF and SR samples in our data. To this end, we first used standard methods for differential analysis of ChIP-sequencing data [47, 48], comparing contrasts across anatomical side (H3K27ac differential enrichment in LA or RA) and disease status (AF or SR). Similarly to our previous results, we largely found differentially enriched regions between anatomical sides rather than disease status (TableS2), suggesting epigenomic differences across AF and SR samples may be subtle due to patient heterogeneity.

We therefore used an alternative approach to identify differential H3K27ac levels across AF and SR samples semi-quantitatively [48]. This strategy is based on (i) quantitative pairwise comparison of enrichment levels across samples to identify regions enriched in H3K27ac in each pair; and (ii) selection of enriched regions that are consistent across samples of the same type (Methods, Figure S2a-b). In this way, we defined ~1,000 regions enriched in H3K27ac levels in each of the four sample groups (AF-RA, AF-LA, SR-RA and SR-LA) (Figure 2a; Table S3), or in AF or SR samples irrespective of anatomical side (Table S3). As expected for regions defined from H3K27ac signals, most were distal to transcriptional start sites (Figure S2c).

To confirm this approach recovers significant H3K27ac enrichments, we compared normalised read coverage for enriched regions in each sample group, and verified the expected upward shift in H3K27ac levels specifically for samples in the enriched group (e.g. AF-LA samples for AF-LA enriched regions, Figure 2b). We also compared average fold enrichment levels of H3K27ac for candidate enriched regions across sample groups (Figure 2c), and observed consistently higher histone mark levels in the enriched group. In agreement with the overall

association between H3K27ac levels and gene expression [46, 49, 50], candidate loci identified in this manner include genes commonly upregulated in atrial fibrillation patients, such as *NPPB* [51]. By evaluating the intersection of enriched H3K27ac regions with gene ontology terms [52], we confirmed enriched regions in each sample group associate with cardiac biological processes (Figure S2d).

Lastly, we investigated the DNA sequences of H3K27ac enriched regions by measuring their overlap with transcription factor binding site motifs (TFBS) [53, 54], transcription factor (TF) binding locations from ChIP-sequencing catalogs [55] (Methods), and GWAS variants associated to cardiovascular traits [56]. These analyses were consistent with enriched epigenomic regions in each sample group showing an atrial regulatory signature, including sequence motifs and binding events of cardiac transcription factors (Figure S3a-b). Both GWAS variants associated to AF and other cardiovascular diseases were over-represented in H3K27ac enriched regions (Figure S3c), suggesting these regions include genetic susceptibility loci for AF and some of its co-morbidities.

In sum, by implementing a pairwise comparison approach for epigenomic profiling data in each sample group, we defined sets of candidate regions showing H3K27ac enrichment – which we next integrate with a largely matched gene expression dataset [33].

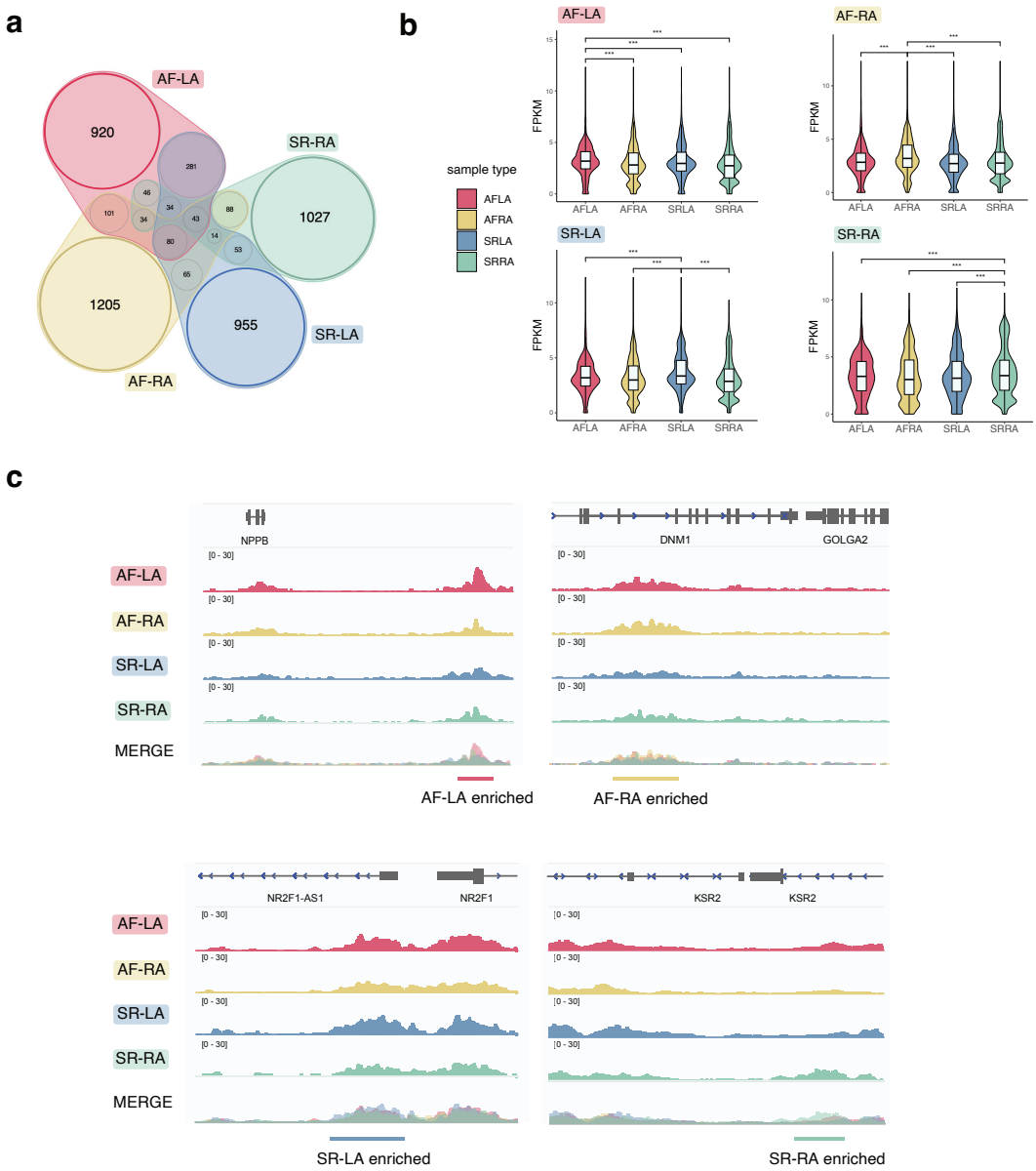

Figure 2: H3K27ac epigenomic regions enriched across sample groups

a. Venn diagram of epigenomic regions enriched in H3K27ac levels in each of the sample groups, as defined from pairwise comparisons in each category (Methods). Overlapping regions across sample groups were excluded from analysis (greyed colour shades).

b. Violin plots of H3K27ac normalised read coverage (FPKM) over epigenomic regions enriched in each sample group (coloured labels, top left of each plot). Read coverage is represented in samples corresponding to each group (x-axis in each plot). As expected, H3K27ac coverage shows an upward shift in the sample group enriched regions were identified in. p-values correspond to one-sided Wilcoxon rank-sum tests with Bonferroni correction for multiple testing. ***: $p < 0.001$

c. Genome track examples of epigenomic regions enriched in each sample group, for *NPPB* (AF-LA), *DNM1/GOLGA2* (AF-RA), *NR2F1* (SR-LA) and *KSR2* (SR-RA) loci. Tracks represent average H3K27ac fold enrichment over input across samples in each group, and the overlay of signals across the four groups (Merge).

Association of H3K27ac enrichment and differential gene expression defines candidate loci for altered gene regulation in atrial fibrillation (related to Figure 3)

We next sought to associate H3K27ac enriched regions across sample groups with downstream gene expression by leveraging previously published transcriptome data in the same sample cohort [33] (Figure S1, Table S1). First, we re-analysed this data and evaluated differential expression across anatomical side, disease status or both (Methods, Figure S4a-c, Table S4). We then associated enriched H3K27ac regions in each sample group with genes (Methods), and measured their enrichment for upregulated genes in the respective group (Fisher's exact test, Methods). Whether across AF and SR samples from each anatomical side (Figure 3a) or overall (Figure S4d), we found enriched H3K27ac regions are significantly overrepresented in the vicinity of upregulated genes for the corresponding sample group. This observation is consistent with the role of H3K27ac as an active histone modification associated with promoter and enhancer activity [42], and predictive of downstream gene expression levels [50]. Moreover, it allowed us to select candidate loci with hallmarks of gene dysregulation in atrial fibrillation patients, by focusing on the subset of loci with both enriched H3K27ac and upregulated gene expression in each sample group (Figure 3b, Figure S4e and Table S5).

As illustrative examples, candidate loci with gene regulation changes in specific sample groups included *LRRC32* (AF-LA), *SCX* (AF-RA) and *WNT5A* (SR-RA) (Figure 3c-e). First, H3K27ac enriched regions in AF-LA samples included a segment spanning the promoter and part of the *LRRC32* gene body (Figure 3c). In agreement with this candidate region

regulating proximal gene expression, we found *LRRC32* expression levels were highest in the AF-LA group, and significantly different from SR-LA samples (Figure 3c).

Second, in the *SCX* promoter region we identified an enriched H3K27ac element in AF-RA samples, suggesting increased promoter activity for *SCX* in this sample group. Accordingly, *SCX* gene expression levels were higher in AF, particularly for RA samples (Figure 3d).

Third, we identified a distal region downstream of the *WNT5A* gene (in an intronic region of the *ERC2* gene, Figure 3e) with enriched H3K27ac levels in SR-RA samples. However, examination of previously published promoter-capture HiC data in iPSC-cardiomyocytes [57] supported an interaction of this element with the *WNT5A* promoter, suggesting potentially increased enhancer activity in SR-RA. Accordingly, *WNT5A* gene expression levels were increased in this sample group (Figure 3e, bottom right boxplot). Lastly, we similarly found consistent H3K27ac enrichment and gene expression levels for candidate loci with altered gene regulation in AF or SR samples (illustrative examples in Figure S4f-g).

On the whole, by focusing on the association between epigenomic H3K27ac levels and downstream gene expression, we defined a set of candidate loci with evidence of altered gene regulation in atrial fibrillation samples – which we next sought to validate in an independent replication cohort.

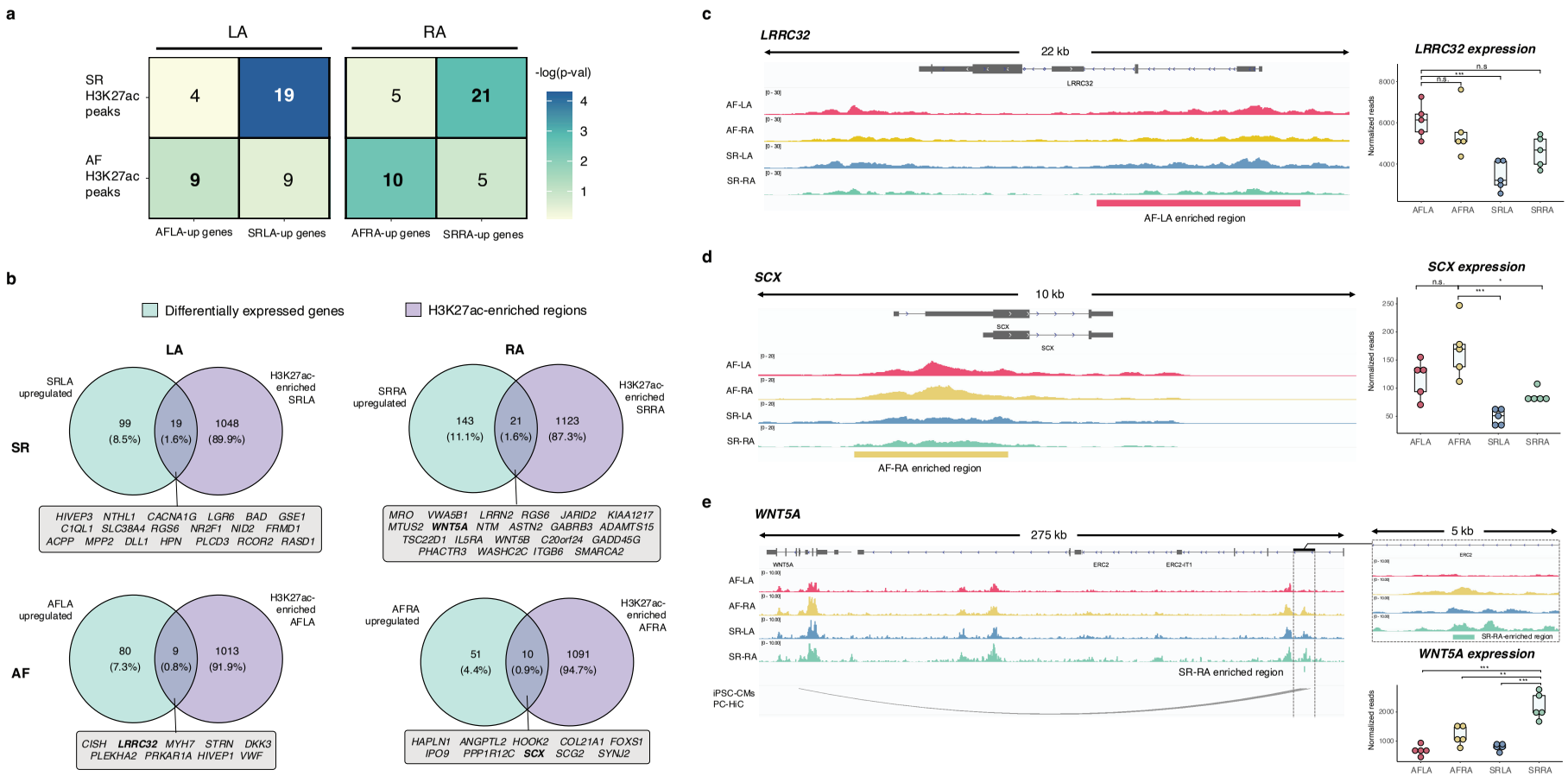
a
LA
RA
SR H3K27ac peaks
AF H3K27ac peaks
4
19
5
21
9
9
10
5
-log(p-val)
4
3
2
1
AFLA-up genes
SRLA-up genes
AFRA-up genes
SRRA-up genes
b
Differentially expressed genes
H3K27ac-enriched regions
LA
RA
SR
AF
SRLA upregulated
H3K27ac-enriched SRLA
99 (8.5%)
19 (1.6%)
1048 (89.9%)
HIVEP3 NTHL1 CACNA1G LGR6 BAD GSE1 C1QL1 SLC38A4 RGS6 NR2F1 NID2 FRMD1 ACPP MPP2 DLL1 HPN PLCD3 RCOR2 RASD1
SRRA upregulated
H3K27ac-enriched SRRA
143 (11.1%)
21 (1.6%)
1123 (87.3%)
MRO VWA5B1 LRRN2 RGS6 JARID2 KIAA1217 MTUS2 WNT5A NTM ASTN2 GABRB3 ADAMTS15 TSC22D1 IL5RA WNT5B C20orf24 GADD45G PHACTR3 WASHC2C ITGB6 SMARCA2
AFLA upregulated
H3K27ac-enriched AFLA
80 (7.3%)
9 (0.8%)
1013 (91.9%)
CISH LRRC32 MYH7 STRN DKK3 PLEKHA2 PRKAR1A HIVEP1 VWF
AFRA upregulated
H3K27ac-enriched AFRA
51 (4.4%)
10 (0.9%)
1091 (94.7%)
HAPLN1 ANGPTL2 HOOK2 COL21A1 FOXS1 IPO9 PPP1R12C SCX SCG2 SYNJ2
c
LRRC32
22 kb
AF-LA
AF-RA
SR-LA
SR-RA
AF-LA enriched region
LRRC32 expression
Normalized reads
AFLA AFRA SRLA SRRA
d
SCX
10 kb
AF-RA enriched region
SCX expression
e
WNT5A
275 kb
5 kb
SR-RA enriched region
iPSC-CMs PC-HiC
WNT5A expression

Figure 3: **Association of H3K27ac enrichment with differential gene expression across sample groups and definition of candidate loci with altered gene regulation in atrial fibrillation**

a. Association of differentially-expressed genes with enriched H3K27ac regions for each sample group. Numbers correspond to the number of gene loci with corresponding changes in gene expression (x-axis) and H3K27ac enrichment (y-axis). Cells are coloured by statistical significance (-log10 p-value, Fisher's exact test).

b. For each bolded comparison in a., Venn diagrams represent the overlap between differentially-expressed genes (cyan) and H3K27ac-enriched regions (violet). Overlapping gene loci (grey insets) correspond to candidate loci showing concordant gene expression and epigenomic changes in a particular sample group.

c. An AF-LA enriched epigenomic region at the promoter region in the *LRRC32* locus (red bar), with gene expression levels being highest in this sample group (right handside, red points). n.s.: non-significant; ***: $p < 0.001$ (DESeq2 adjusted p-values).

d. Example of an AF-RA enriched epigenomic region proximal to the *SCX* gene (yellow bar). Consistent with promoter activity, *SCX* expression levels are highest in AF-RA samples (right handside, yellow points). n.s.: non-significant; *: $p < 0.05$, ***: $p < 0.001$ (DESeq2 adjusted p-values).

e. Example of a SR-RA enriched epigenomic region downstream of the *WNT5A* gene (green bar; right handside and top-right inset). External promoter-capture HiC data in iPSC-cardiomyocytes supports a 3D interaction between this putative enhancer region and the *WNT5A* promoter. *WNT5A* expression levels across sample groups are highest in SR-RA samples (bottom-right, light green). **: $p < 0.01$, ***: $p < 0.001$ (DESeq2's adjusted p-values).

Targeted validation of candidate loci in an independent patient cohort (related to Figures 4 and 5)

The candidate loci we identified here by integration of epigenomic and transcriptomic profiles were obtained in a relatively small atrial fibrillation cohort (Table S1, Figure S1, Methods). To evaluate the robustness of these candidates relative to donor biopsies, we obtained an independent set of 3 AF and 2 SR atrial samples to use as a replication cohort. As in our discovery cohort, these samples included matched RA and LA biopsies for each donor patient (Table S1, Figure S5a). To confirm their atrial origin and correct anatomical side assignments,

we validated gene expression of cardiac, left atrium and right atrium markers in replication samples (Figure S5b).

Towards validation of candidate loci, we used locus-specific quantitative PCR to measure gene expression levels (Figure 4) and H3K27ac enrichment (Figure S5) across a subset of candidate loci (Table S6). Specifically, we focused on evaluation of gene expression and histone mark enrichments across disease status (AF or SR samples) in the smaller replication cohort. For gene expression levels (Figure 4), we found an overall good correspondence for expression differences in the discovery and replication cohorts, with most candidate genes showing the expected expression shifts in the independent sample set. We validated gene expression differences in 9 out of 12 loci tested (validation rate ~75%, p-value < 0.2 t-test), with similar rates for AF (Figure 4a, 6 out of 8 loci) or SR upregulation (Figure 4b, 3 out of 5).

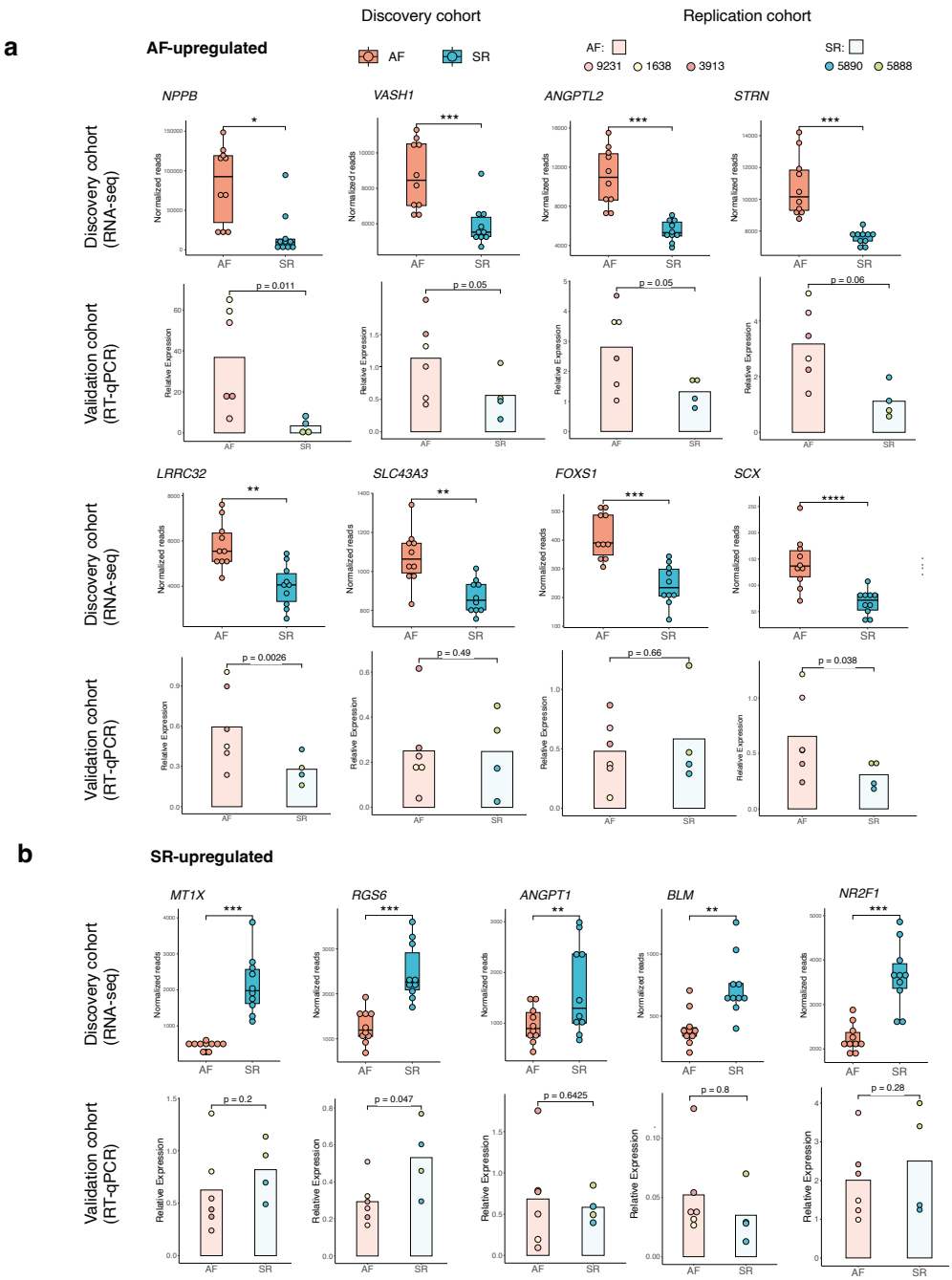

Figure 4: Targeted validation of gene expression levels at candidate loci in an independent replication cohort

Locus-specific comparison of gene expression levels in AF and SR samples for selected candidate loci in the discovery cohort (top boxplots in each category, RNA-seq), and their validation in an independent replication set (bottom barplots in each category, RT-QPCR). Gene expression differences across disease status were considered as validated in independent samples when the direction of change was maintained with a p-value < 0.2 (one-sided t-test). **a.** Gene expression levels in the discovery cohort (top boxplots) and validation

cohort (bottom barplots) across 8 AF-enriched candidate loci. **b.** As in a. for five SR-enriched candidate loci.
See also Figure S5 and Table S6.

In contrast, locus-specific H3K27ac enrichments in replication samples showed higher variability (Figure S5c-d), and a lower validation rate ~50% ($p < 0.2$, t-test). This observation is consistent with the relatively large size of H3K27ac enriched regions compared to the genomic segments we tested (Methods, Figure S5). To illustrate this difference, we examined validation results by ChIP-QPCR in comparison with H3K27ac data in the discovery cohort across three candidate loci (Figure 5). First, for an AF-enriched H3K27ac region in the *VASH1* locus, we observed slightly higher H3K27ac levels by ChIP-QPCR in independent samples (Figure 5a, P1 primer pair, $p = 0.28$), with the tested region corresponding to a genomic segment proximal to an accessible region in human atrial cardiomyocytes [58]. Second, we validated differential H3K27ac levels for an SR-enriched region in the *MT1X* locus (Figure 5b). In this case, the region tested in the replication cohort overlapped an accessible region in human atrial cardiomyocytes (CMs) (P2, $p = 0.02$). Lastly, we examined validation results for two enriched regions in the *NPPB* locus, in AF-LA and AF samples, respectively. These map to a previously defined super-enhancer in mouse heart [59], and both tested regions overlapped accessible chromatin in atrial CMs (P3 and P4, Figure 5c). For either region, we observed the highest H3K27ac levels in the expected sample group, although differences were not significant for the AF-LA region.

Overall, locus-specific evaluation of gene expression and histone mark enrichments at a subset of our candidate loci in replication samples supported the majority of candidate loci can be validated in an independent cohort, lending robustness to the putative gene regulation alterations we identified in our discovery cohort.

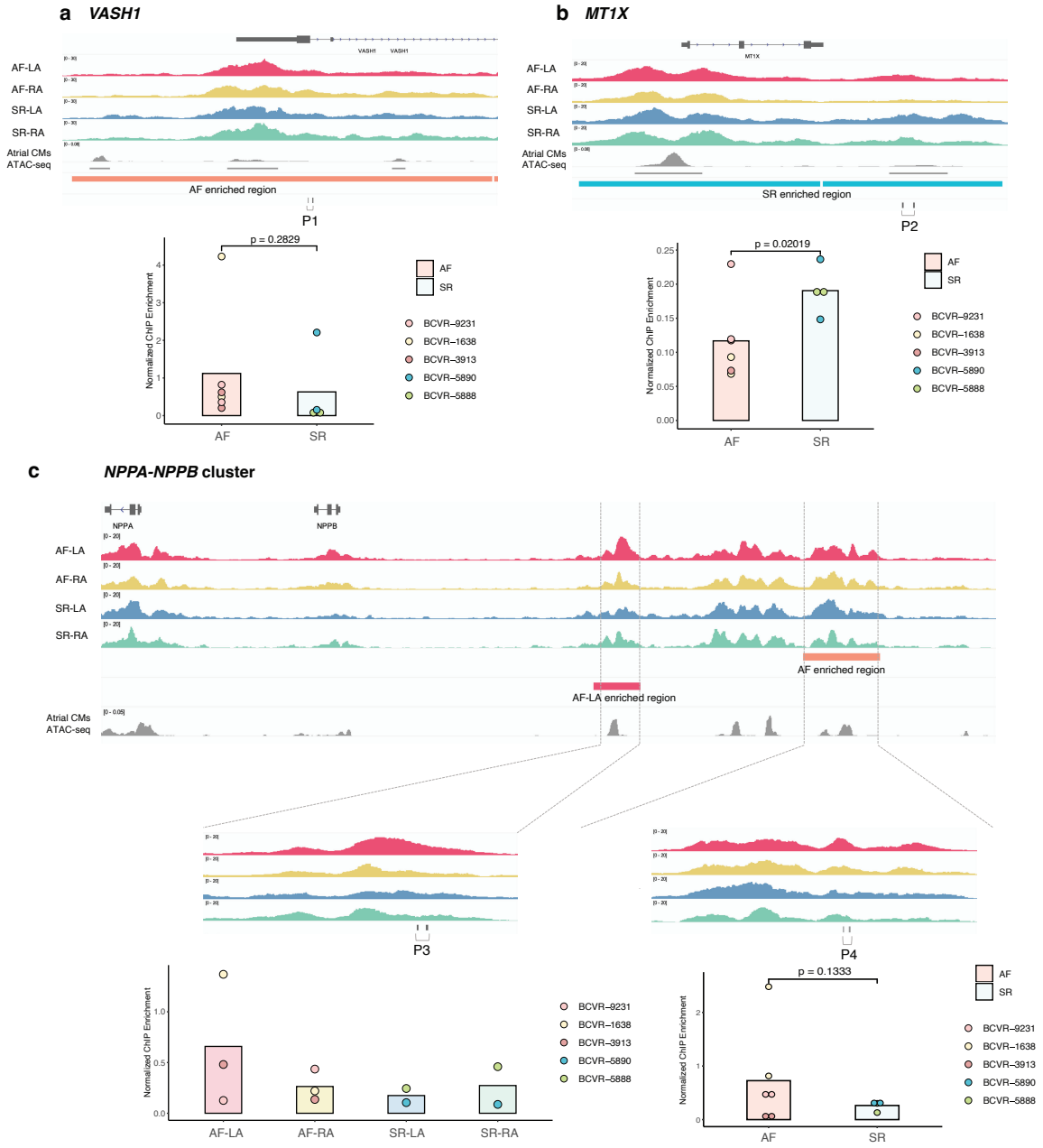

Figure 5: Targeted validation of H3K27ac enrichment at candidate loci in an independent replication cohort

a – c. Examples of ChIP-QPCR validation of H3K27ac enrichment for an AF-enriched region in the *VASH1* locus (**a**, sienna bar), a SR-enriched region in the *MTX1X* locus (**b**, cyan bar), and AF- and AF-LA enriched regions in the *NPPA/NPPB* locus (**c,** sienna and red bars, respectively). For each example, genome tracks show H3K27ac levels in each sample group (AF-RA, AF-LA, SR-RA, SR-LA), the location of validated enriched regions (coloured bars) and single-cell ATAC-seq signals corresponding to atrial cardiomyocytes in the human heart.

P1 to P4 indicate primer pairs used for validation of each region (barplot insets). Highlighted p-values correspond to one-sided t-test. For the P3 region in c, ANOVA p-value = 0.54
See also Figure S5 and Table S6.

5 ***Base-resolution DNA methylation signals at candidate loci inform mechanisms of altered gene expression and TF binding (related to Figures 6 and 7)***

To further investigate the mechanisms underlying altered gene regulation in atrial fibrillation patients for our candidate loci, we carried out genome-wide profiling of DNA methylation levels in 16 samples from our discovery cohort via the Illumina EPICv2 array (Methods, Table S1,
10 Figure S1a). First, we performed an exploratory EWAS (Figure S6a, Table S7) for differentially-methylated cytosines (DMCs), and then a differentially-methylated region (DMR) analysis (Figure 6a) to identify regions with DNA methylation differences between AF and SR samples (Methods). As expected, assessment of single CpG methylation differences was underpowered in our sample set, and most signals we identified through EWAS (Table S7)
15 did not overlap candidate loci. However, aggregating CpG methylation signals into DMRs resulted in a set of 148 regions discriminating AF and SR disease status for a majority of donors (Figure 6a-b).

To investigate DNA methylation differences across candidate loci, we next intersected DMRs with enriched H3K27ac regions (Figure 2, Figure S6b). As expected for CpG array data
20 (Methods), DMRs were relatively proximal to annotated TSSs, were often outside CpG islands and showed limited overlap with enriched H3K27ac regions across sample groups (Figure S6b-d). We thus associated DMRs with candidate regions at the locus level, and explored the association of DNA methylation differences with differential gene expression and H3K27ac enrichment (Figure S6b and Table S8).

25 DNA methylation signals at candidate loci often informed potential mechanisms of altered gene regulation in atrial fibrillation. For the *SLC43A3* locus (Figure 6c), we identified a hypomethylated DMR (lower DNA methylation levels in AF) within an enriched H3K27ac

promoter region in AF samples. Hypomethylation for this DMR associated with higher gene expression of *SLC43A3* in AF patients, as expected for increased promoter activity. In contrast, for the *FOXS1 / MYLK2* locus (Figure 6d), we identified a hypomethylated DMR in the *MYLK2* promoter region, and distally to an AF-enriched H3K27ac region with potential enhancer function. Expression levels of *MYLK2* were inconsistent with the identified DMR, and higher in SR controls compared to AF patients. However, promoter-capture HiC data [57] in this region supported an interaction between the AF-enriched H3K27ac region downstream (proximal to *FOXS1*) and the identified DMR, which may thus act as a distal enhancer of *FOXS1* expression. In agreement, *FOXS1* expression levels were found to be increased in AF patient samples. These examples illustrate how DNA methylation and H3K27ac levels can jointly inform epigenomic alterations at AF candidate loci.

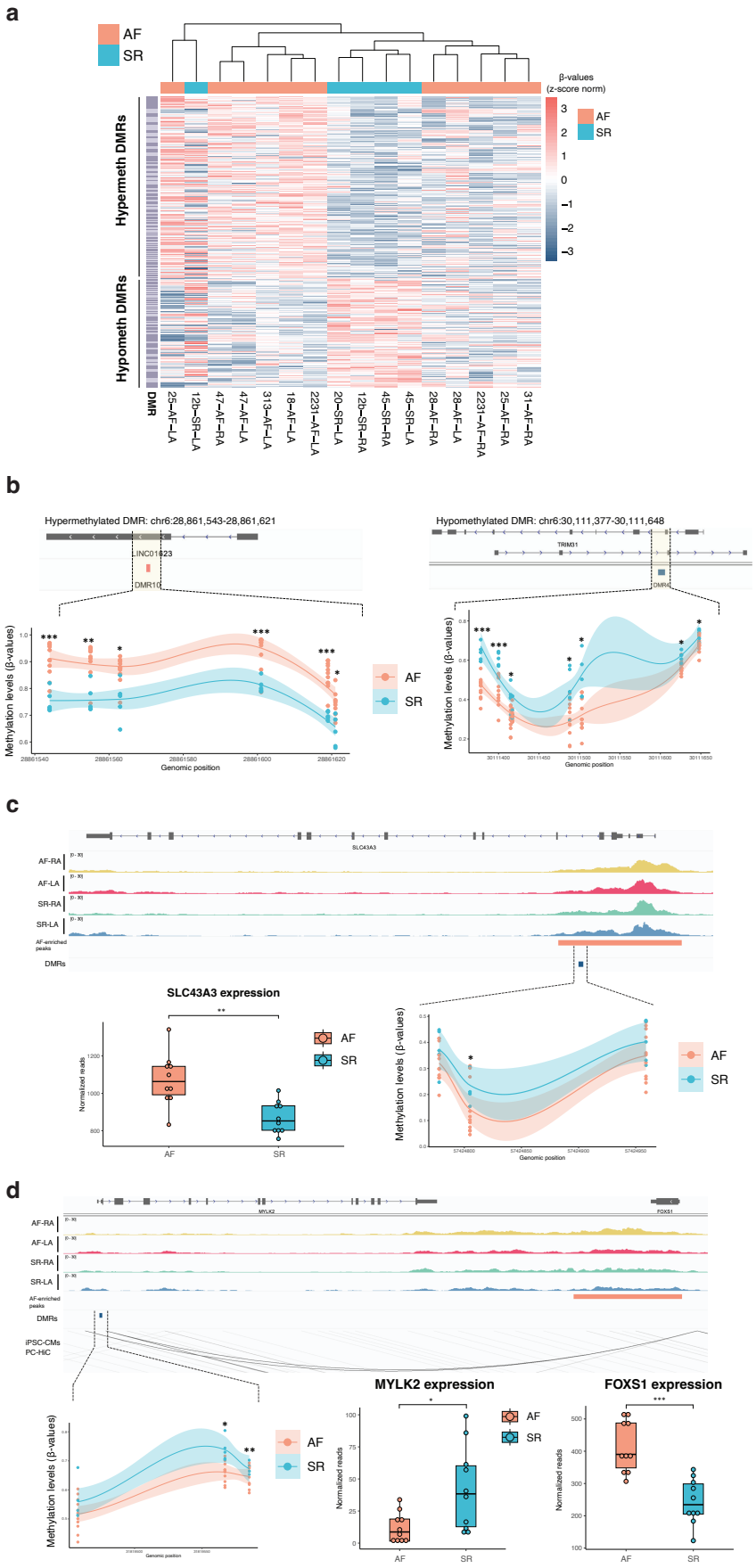
a
AF
SR
β-values
(z-score norm)
Hypermeth DMRs
Hypometh DMRs
DMR
25–AF–LA
12b–SR–LA
47–AF–RA
47–AF–LA
313–AF–LA
18–AF–LA
2231–AF–LA
20–SR–LA
12b–SR–RA
45–SR–RA
45–SR–LA
28–AF–RA
28–AF–LA
2231–AF–RA
25–AF–RA
31–AF–RA
b
Hypermethylated DMR: chr6:28,861,543-28,861,621
Hypomethylated DMR: chr6:30,111,377-30,111,648
Methylation levels (β-values)
Genomic position
AF
SR
c
SLC43A3
AF-RA
AF-LA
SR-RA
SR-LA
AF-enriched peaks
DMRs
SLC43A3 expression
Normalized reads
d
MYLK2
FOXS1
iPSC-CMs PC-HiC
MYLK2 expression
FOXS1 expression

Figure 6: **Differentially-methylated regions in AF samples inform gene dysregulation mechanisms in candidate loci**

a. Heatmap of DMRs identified as hypermethylated (y-axis, top half) or hypomethylated (y-axis, bottom half) in AF samples. Heatmap scale corresponds to Z-score normalization of beta-values. Hierarchical clustering (x-axis, top; sienna for AF and cyan for SR) shows a subset of samples associates with DNA methylation differences across DMRs.

b. Examples of DMRs hypermethylated (DMR10, left) and hypomethylated (DMR4, right) in AF patients. Shaded trendlines correspond to local regression (LOESS method) with confidence interval, and p-values to statistical significance of CpGs within a DMR.

c. A hypomethylated DMR in *SLC43A3* associates with H3K27ac enrichment and increased expression levels in AF patients. Representation as in **a** and previous Figures.

d. A hypomethylated DMR proximal to *MYLK2* acts as a potential enhancer of *FOXS1* expression and associates with H3K27ac enrichment and increased *FOXS1* expression in AF patients. Representation as in **a** and previous Figures.

FDR-corrected two-sided t-test for differentially-methylated CpGs; ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$).

DESeq2 adjusted p-values for gene expression differences: *: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$

See also Table S7.

We next sought to investigate the sequence composition of DMRs showing hyper- and hypomethylation in AF samples. To this end, we first identified TFBSs across DMRs, and observed distinct sets of TFBSs overrepresented in each group (Figure 7a, Figure S7, Methods). These include motif sequences for transcription factors previously associated to AF, such as ETV proteins [60]; and typically corresponded to transcription factors expressed at similar levels in AF and SR samples (Figure 7a). For instance, in the genomic locus spanning *LRRC4B, STY3* and *SHANK1* genes, we identified a hypermethylated DMR in which one of the CpGs overlapped an ETV-1 motif (Figure 7b). Gene expression levels of genes in this locus and promoter-capture HiC data in iPSC-cardiomyocytes [57] is consistent with this DMR acting as a distal enhancer of *LRRC4B*, a gene previously implicated in cardiac dysfunction and fibrosis [61, 62]. Integration of these data allowed us to hypothesise

hypermethylation of the ETV-1 motif may result in lower activity of this enhancer in AF patients, as suggested by reduced *LRRC4B* expression (Figure 7b, Figure S7). In contrast, we found ETV-1 expression levels were similar in both sample groups (Figure 7a).

Lastly, we also associated DMRs with GWAS variants and transcription factor binding events [55, 56], and observed enriched biological processes and cardiovascular traits consistent with atrial regulatory mechanisms (Figure S7). In agreement, we observed TFBSs corresponding to AF-associated transcription factors (TFs) [19] were often also enriched in DMRs (Figure 7c), and these TFs were expressed in atrial samples with no clear specificity towards disease status or anatomical side (Figure 7c). Instead, we found instances where differences in DNA methylation occurred within the TFBS of an AF-associated TF (Figure 7d). Within a hypomethylated DMR in the *ZNF727* promoter (Figure 7c), we identified a sequence similar to the standard HAND2 motif for which hypomethylation in AF samples could facilitate HAND2-mediated expression. In agreement, we observed higher expression levels of *ZNF727* in AF patients.

In sum, DNA methylation changes associated to AF in our sample cohort often suggested altered mechanisms of gene regulation and TF binding underlie gene expression differences, thus providing testable hypotheses for targeted investigations.

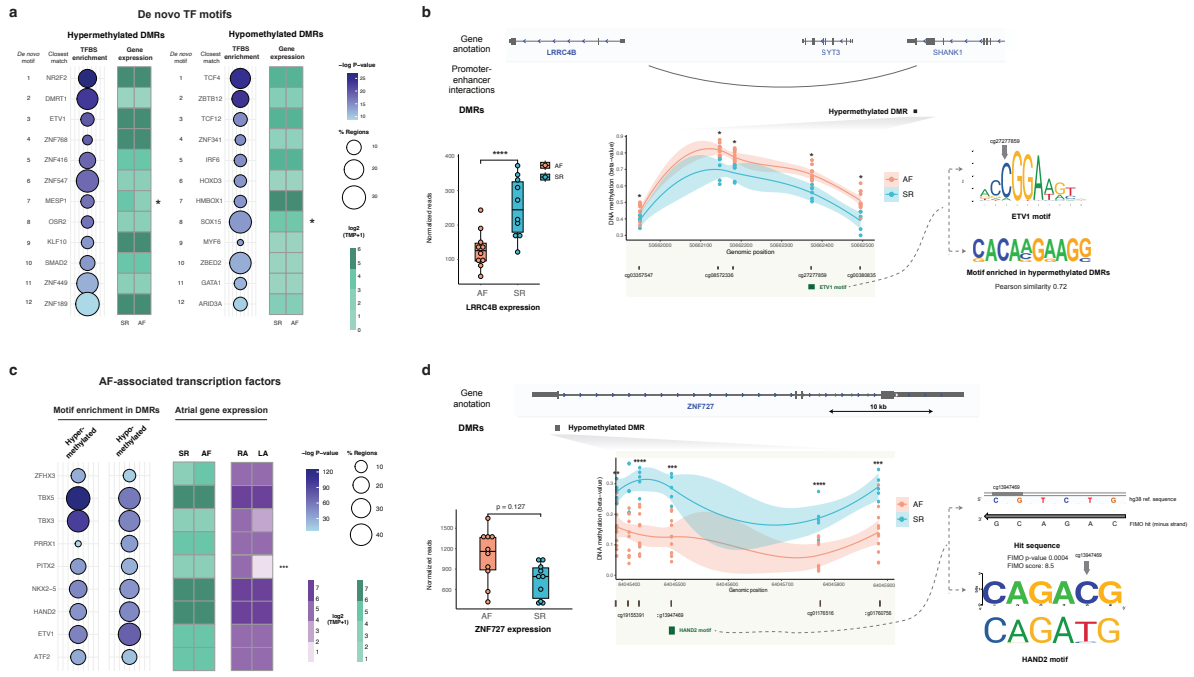

Figure 7: Transcription factor motif enrichment in AF differentially-methylated regions and examples of potential disruption of TF binding by DNA methylation

a. *De novo* transcription factor motifs enriched in hypermethylated and hypomethylated DMRs. Circle plots in each category indicate percentage of DMRs harbouring sequence motifs (circle size) and statistical significance (circle shade; -log10 p-value, hypergeometric test with Benjamini-Hochberg correction for multiple comparisons). Each enriched motif was assigned to the indicated transcription factors based on motif similarity and cardiac expression (Methods). Heatmaps indicate median expression levels (log2(TPM+1)) for each transcription factor in sinus rhythm controls (SR) or permanent AF patients (AF). Asterisks indicate differential gene expression (FDR p-value < 0.05; Methods).

b. Example of a hypermethylated DMR in the *LRRC4B* locus (DMR26), for which one of the differentially-methylated CpGs overlaps an ETV1 motif. (Top tracks) Gene and DMR annotation, and promoter-enhancer interactions in this genomic locus. (Bottom middle inset) DNA methylation differences across the DMR, shown as beta-values for AF (sienna) and SR (cyan) samples. Annotations below the plot correspond to individual CpGs and the ETV1 sequence motif (green box). *: p-value < 0.05 (Bottom right inset) The ETV1 motif is highly alignable to the *de novo* motif enriched in hypermethylated DMRs (top-right comparison of motifs logos, Pearson correlation-based similarity score = 0.72; differentially methylated CpG cg27277859 marked on the top motif by grey arrow). (Bottom left boxplot) Expression levels of *LRRC4B* expression are significantly higher in SR patients (*: p < 0.05). Expression levels of other genes in this locus are similar in SR and AF samples (Figure S7b).

c. (Circle plots, left) TFBS motif enrichment for AF-associated transcription factors in hypermethylated and hypomethylated DMRs. Representation and details as in **a.** (Heatmaps, right) Median gene expression levels of AF-associated transcription factors across SR or AF samples (shades of green); and right atrium (RA) or left atrium (LA) samples (shades of violet). Asterisks indicate differential gene expression (FDR p-value < 0.05; Methods).

d. (Top-middle inset) Example of a hypomethylated DMR in the *ZNF727* locus, which includes a differentially-methylated CpG overlapping a sequence similar to the motif for AF-associated transcription factor HAND2 (FIMO score 8.5, p-value 0.0004). (Top tracks) Gene and DMR annotation. (Bottom middle inset) DNA methylation differences across the DMR. Annotations below the plot correspond to individual CpGs and the HAND2 sequence motif (green box). P-values: **: p < 0.01; ***: p < 0.001; ****: p < 0.0001. (Bottom left logo) Sequence logo for the HAND2 motif. The hypomethylated CpG cg13947469 is highlighted by a grey arrow. (Left boxplot) Expression levels of *ZNF727* are higher in AF patients (p=0.127). Details as in **b**. See also Figure S7 and Table S7.

DISCUSSION

Human population studies increasingly support a complex genetic architecture for atrial fibrillation [3, 19], with most genetic association signals corresponding to non-coding regions with potential roles in gene regulation, such as tissue-specific enhancers [20, 63]. Accordingly, both transcriptomic and DNA methylation datasets support widespread gene regulation changes are associated with the development of persistent AF [38, 64]. Using a small set of matched left and right atrial biopsies comprising sinus rhythm controls and persistent AF patients, here we integrated epigenomic profiling of the histone modification H3K27ac and DNA methylation levels with transcriptomic data previously obtained in these samples [33]. This resource allowed us to identify epigenomic changes on histone modifications and DNA methylation in AF patients, and test their association with downstream gene expression levels. First, comparison of H3K27ac levels in each sample group with differential gene expression allowed us to define a set of candidate loci with altered gene regulation in AF, which enrich for GWAS loci associated to AF and its cardiovascular comorbidities. These candidates include both previous AF-associated loci such as *NPPB* and *ANGPTL2*, and provide

mechanistic hypotheses linking alterations of promoter and enhancer activity in AF and their potential impact on atrial gene expression. Second, we obtained genome-wide DNA methylation levels at base-resolution in a subset of these samples, and identified differentially-methylated regions in AF patients that significantly associate with candidate loci. Specific transcription factor motifs are over-represented in the sequences of these hyper- and hypomethylated regions, including those corresponding to AF-associated transcription factors ETV1 [60] and HAND2 [27]. Integration of DNA methylation changes at transcription factor binding sites with H3K72ac levels and gene expression allowed us to propose co-ordinated mechanisms underlying altered gene regulation in AF for cardiovascular loci such as *LRRC4B*. Albeit with limited sample numbers and statistical power, our results represent a systematic investigation of dysregulated gene expression across the left and right atrium in persistent AF, and add to ongoing efforts to inform the genetic and epigenetic underpinnings of human population findings in AF cohorts.

Our data revealed a number of insights into how alterations in the regulatory landscape associate with persistent AF, and delineated candidate dysregulated loci and epigenomic regions. These include known AF disease loci, such as the *NPPB* locus [65, 66]. In agreement with previous work, we observed increased *NPPB* expression in AF samples, and additionally identified epigenomic regions with increased H3K27ac levels - in AF samples overall and in AF-LA; Figures 5 and S4). In contrast, other known AF disease loci such as *PITX2* did not show significant transcriptomic and epigenomic changes in our samples, which may be due to the small numbers of patients we profiled. In this regard, previous atrial gene expression datasets did not consistently observe *PITX2* expression changes in persistent AF samples [64], which have been recently validated in larger patient cohorts [23].

Among the candidate loci we identified by integration of H3K27ac and transcriptomic profiles (Figures 3 and S4), several are involved in fibrosis-induced structural remodeling, and include loci with increased (*SCX*) or decreased (*WNT5A*) activity in AF samples. SCX enables

fibroblast to myofibroblast conversion, as evidenced in *in vivo* deficiency models with compromised cardiac matrix formation and loss of cardiac fibroblasts [67]. In this locus, the observed epigenomic and transcriptomic levels resulting from increased promoter activity in AF-RA samples from our cohort could be contributing to the activation of fibrotic processes in AF. Dysregulations of the Wnt signaling pathway are known to contribute to fibrotic processes in the atria, with decreased activity of most components, such as *WNT5A*, found in AF patient samples [68] and rapid-pacing in vivo models [69].

As part of our validation of candidate loci in an independent sample cohort, we qualitatively confirmed both H3K27ac enrichment and gene expression levels for loci such as *NPPB, VASH1* and *MT1X* (Figure 5). *VASH1* upregulation reportedly leads to reduced angiogenesis [70], and our findings support a model of endothelial dysfunction and a prothrombotic state promoting AF. *MT1X* decreased levels, on the other hand, are known to induce oxidative stress conditions [71], thus promoting both structural and electrical remodeling through alteration of ion channel activity.

Our DNA methylation data in these samples allowed us to identify potential mechanisms of gene dysregulation across candidate loci (Figure 6). *SLC43A3* is a membrane transporter protein responsible for uptake of purine nucleotide bases into the cells [72] as well as regulation of free fatty acids transport [73], and we found both DNA hypomethylation and chromatin hyperacetylation of its promoter in AF samples. Therefore, our findings suggest lipid metabolism dysregulation could contribute to AF disease through altered SLC43A3 expression. In the *MYLK2* gene locus, encoding a myosin light chain kinase 2 protein with crucial roles in cardiac muscle contraction [74], promoter hypomethylation in AF and increased H3K27ac levels at a putative distal enhancer contrasted with reduced *MYLK2* gene expression. Instead, our findings for this locus are consistent with the hypomethylated region regulating FOXS1 expression, which we found to be increased in AF samples. Because FOXS1 has been mechanistically implicated in liver fibrosis [75], it is tempting to speculate its

upregulation in atrial tissue may contribute to fibrosis in permanent AF. Nevertheless, our observation could stem from the proposed association between AF and calcific aortic valve stenosis [76], in which FOXS1 is functionally implicated [77].

Lastly, and through investigation of overrepresented transcription factor motifs in differentially-methylated regions (Figure 7 and S7), our results provide mechanistic hypotheses by associating epigenetic dysregulation in AF with potential changes in transcription factor binding. An illustrative example is the *LRRC4B* locus, where we identified a putative *LRRC4B* enhancer hypermethylated in AF samples. In this region, one of the differentially-methylated CpGs overlaps a transcription factor motif for ETV1, a known AF-associated transcription factor [60]. This observation suggests impaired ETV1 binding in this putative enhancer could mediate reduced LRRC4B expression in AF (Figure 7). LRRC4B is a lamin-associated protein with reduced protein expression in lamin A/C haploinsufficient cardiomyocytes [61]. Our results suggest reduced *LRRC4B* expression occurs via hypermethylation of a distal enhancer [78] and reduced ETV1 binding [79, 80], potentially contributing to remodeling of atrial cardiomyocytes in AF. Nevertheless, further work would be required to substantiate this hypothesis.

There are a number of limitations to our approach. First, the sample size in our study was reduced by limited availability of atrial biopsies comprising both left and right atrial appendages from each patient. This is a common limitation in epigenomic studies on human cardiac samples, and informs our choice of semi-quantitative methods to identify hyper- and hypoacetylated epigenomic regions marked by H3K27ac [48]. Nevertheless, the integration of histone modification data with DNA methylation and transcriptomic readouts across left and right atrial samples allowed us to define candidate loci with multiple indications of gene dysregulation in AF patients – some of which we could qualitatively validate in an independent set of samples.

Second, the epigenomic profiling experiments we report targeted H3K27ac as a histone modification consistently associated with the activity of mammalian promoters and enhancers [42, 81], and DNA methylation levels as a second epigenetic modification with documented and widespread changes during AF development [38, 39]. Albeit H3K27ac is known to be dysregulated across a range of cardiovascular diseases [45, 46, 82], epigenomic regions defined by other histone modifications, such as those associated to non-canonical enhancers [83], are also likely to be modified in permanent AF. Similarly, additional epigenetic mechanisms known to be altered in AF patients, such as microRNA or long non-coding RNA expression, were not investigated in our study. Thus, our data represents a practical compromise between sample availability and epigenomic completeness. Nevertheless, denser sampling of AF patients and sinus rhythm controls would impact performance of our approach by improving sensitivity and reducing false discovery rate.

Third, our analyses used bulk readouts on cardiac tissue samples from left and right atrium, thus measuring an average of signals arising from cardiac cell types in each sample. Despite some of our identified loci containing genes with documented roles in cardiomyocytes, such as *LRRC4B*, multiple lines of evidence suggest our results are likely to inform mechanism of gene dysregulation in additional cell types. First, several of our candidate loci include genes expressed in cardiac fibroblasts and known to regulate fibrosis, such as *SCX* [84]. Second, epigenomic regions with changes in H3K27ac and DNA methylation both associate with open chromatin regions across multiple cell types in the human heart (Table S10). Although outside the scope of our work, computational deconvolution approaches [85, 86] could be applied to the datasets we report here to resolve epigenomic signals in specific cardiac cell types. The significant rise in cost notwithstanding, future studies may extend our approach by leveraging single-cell technologies for transcriptomic and epigenomic profiling [87, 88].

In summary, this study presents an integrated investigation of epigenomic and transcriptomic changes associated with persistent atrial fibrillation in the right and left atrium – thereby

addressing the potential impact of dysregulated gene expression in AF susceptibility and progression. By systematically identifying epigenomic regions dysregulated in AF and their association with gene expression, we show the value of functional genomics analyses to identify candidate loci dysregulated in AF patients, and connect these changes in gene regulation with mechanistic hypotheses at the molecular level.

MATERIALS AND METHODS

Details of samples used in the study

Left and right adult human atrial appendages were collected from patients undergoing open-heart surgery (bypass grafting and/or atrial/mitral valve repair) at Barts Health Centre, Barts Health National Health Service (NHS) Trust. Ethical approval was obtained from the East of England-Cambridge Central Research Committee (14/EE/0007). Patients classified as SR who developed post-operative AF were excluded from the study. Samples were cryopreserved in RNAlater and stored at -80°C.

A discovery cohort of 20 samples was used for the generation of ChIP-seq and EPIC array DNA methylation data. These samples comprised 12 individuals, 7 classified as persistent AF and 5 as sinus rhythm (SR). Paired left and right atrial samples were obtained in 8 of the 12 individuals. A further 10 paired left and right atrial samples were obtained from an independent cohort of 5 individuals (3 AF and 2 SR) and used for validation of epigenome and transcriptome-wide findings in the discovery cohort through locus-level quantification of gene expression or histone modification enrichment by quantitative PCR.

Details regarding patients, samples and data generated from each sample can be found in Table S1.

Chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq)

ChIP and high-throughput sequencing of the discovery cohort were performed across four different batches. The OSAT R package was used to avoid sample allocation biases and ensure even distribution of biological groups across independent batches [89]. An additional batch of ChIP was performed for the 10 samples in the validation cohort for downstream qPCR experiments. All ChIP experiments were performed as described in existing protocols with minor modifications [90], using commercial antibodies for H3K27ac (ab4729, Abcam) and H3K4me3 (05-1339, Sigma-Aldrich). The latter antibody was only used for a subset of samples (as detailed in our GEO dataset, see Data Availability section).

Computational analysis of ChIP-seq data

<u>Basic alignment and peak calling</u>: Aligned bam files were obtained with bwa 0.7.17 and the Ensembl v114 hg38 human genome assembly. We used macs2 [91] to call peaks for each ChIP-Seq replicate, using default parameters and "--keep-dup all" to retain duplicate reads. Before peak-calling, multi-mapping reads were removed and read-depth adjusted to 20 million uniquely mapped reads (or all available reads for low-depth libraries).

<u>Definition of regulatory regions across each sample type</u>: we first constructed sets of reproducible peaks for each of the four sample groups by calling peaks with Genrich

(https://github.com/jsh58/Genrich) on all ChIP and input libraries within each group. We then used the chip-greylist tool to identify peaks reproducibly called from input samples alone ("greylisted"), which we excluded from analysis.

Definition of enriched H3K27ac regions across sample groups from pairwise replicate comparisons

H3K27ac-enriched peaks identified with MACS2 within individual samples were used to define a consensus set of enriched regions in each of the four biological categories (AF-LA, AF-RA, SR-LA and SR-RA) or the two overall disease groups (AF and SR). To do this, regions with differential signal were identified for pairwise comparisons across all individual samples in the dataset (for n = 20, 190 possible comparisons) using MACS2 'bdgdiff' function with a likelihood ratio cutoff of 100 (c = 2), 300bp of differential region minimum size (l = 300) and 100 bp maximum gap between peaks to be merged into a single one (g = 100). For each pairwise comparison, two peaksets were obtained (corresponding to peaks enriched in either sample pair, 190x2 = 380 enriched peaksets). These peaksets were then used to build a consensus set of enriched regions for each category (excluding within-category pairwise comparisons). Enriched peaks within each category with at least 1bp overlap were merged into a single consensus region. To achieve similar numbers of enriched regions across sample groups, the number of sets required to overlap for a region to be included in the consensus set was adjusted relative to the number of replicates within each category. To exclude enriched regions resulting from merging and concatenation, at least 30% of the length of enriched peaks was required to reference peaks collectively called for all replicates in each category (obtained as detailed above with Genrich). Lastly, enriched regions overlapping multiple categories were removed, retaining uniquely enriched regions for further analyses.

To obtain normalised coverage values, we used the "cov" function in BAMscale [92] to calculate FPKM read coverage in H3K27ac ChIP and input samples for enriched regions across sample groups, as described previously [44]. H3K27ac coverage was then normalised by input coverage and quantile normalisation, and represented in logarithmic scale.

RNA-sequencing analyses of differential gene expression

Raw RNA-seq data generated for a former study [33] in largely overlapping samples (Table S1) was obtained from the National Centre for Biotechnology Information's Gene Expression Omnibus (GEO) (GSE128188) and re-analysed to identify DEGs between sample groups.

Reads were aligned to the hg38 reference human genome using STAR with default parameters [93]. A count matrix of all reads was created with featureCounts [94] and differential expression analysis was performed using DESeq2 (v1.42.1) [95]. After incorporating Ensembl v114 bioMart human gene annotations [96], protein-coding genes with

at least 5 read counts across all samples were selected for downstream analysis (n = 18,370 genes). Read counts were normalised by library size factors, and p-values were calculated using a negative bimodal distribution and corrected for multiple testing using Benjamini and Hochberg method.

Integration of DEGs with ChIP-seq H3K27ac enrichment data was investigated by exploring the overlap of DEGs and genes in proximity to H3K27ac-enriched regions using the default gene-to-peak association rule in GREAT (v4.0.4) [97]. Significance was tested with Fisher's Exact Test, with all genes used in differential expression analysis as background.

Experimental validation of selected candidate regions in an independent patient cohort

Validation of gene expression and H3K27ac enrichment levels across candidate loci in an independent patient cohort was performed with qPCR technology. For RT-qPCR validation, gene-specific primers were designed with NCBI Primer Designing tool (Table S9). Total RNA was isolated from 30-80 mg of atrial tissue. Complete disruption and homogenisation of tissue was achieved with 3 cycles of 15 seconds at 4.5 m/s in a FastPrep-24 bead beating and lysis system (MP Biomedicals). The RNeasy Plus Universal kit (Qiagen) and TRIzol reagent (Ambion) were used for RNA extraction. RNA integrity was assessed with the RNA 6000 Nano Kit for Bioanalyzer 2100 (Agilent). 1000 ng of RNA were reversed transcribed into complementary DNA (cDNA) using the High-Capacity cDNA Reverse Transcription Kit (Fisher Scientific), and RT-qPCR reactions were performed using 5 ng of template cDNA. Ct values for each gene of interest were normalised to those of *TPB* reference gene using the ΔCt method [98].

H3K27ac ChIP of atrial samples in the validation cohort was performed as above. qPCR primers were designed to target selected enriched regions identified in the discovery cohort's ChIP-seq data, with the NCBI Primer Designing tool (Table S9). ChIP and input DNA samples were diluted 1:5 and 1:10 respectively for qPCR reactions. ChIP enrichments were normalised to those of a region in the *GAPDH* reference locus over input samples ($2^{\wedge} - (Ct_{Target} - Ct_{GAPDH})_{ChIP} - (Ct_{Target} - Ct_{GAPDH})_{Input}$).

Universal KAPA SYBR reagents (KAPA Biosystems) were used for all qPCR reactions on a Roche LightCycler480 system, with the following parameters: 95°C for three minutes (pre-incubation), 40 cycles at 95°C for 10 seconds, 60°C for 20 seconds and 72°C for 1 second. Generation of melt curves after 40 cycles was carried out to control for amplification of non-specific products.

Experimental assessment of DNA methylation levels with EPIC array technology

Genomic DNA was isolated from 16 atrial samples (AF-LA, n=6; AF-RA, n=5; SR-LA, n=4; SR-RA, n=4) using the DNeasy Blood & Tissue Kit (Qiagen). Bisulfite treatment and

subsequent amplification and hybridisation to DNA methylation profiling microarray was performed by the QMUL Genome Centre. The Infinium MethylationEPIC v2.0 kit was used, providing coverage of approximately 930K CpGs, including RefSeq annotated genes, CpG islands and enhancer regions. Even and random distribution of biological groups across two chips of 8 samples each was allocated using the OSAT R Bioconductor package [89].

Computational analysis of DNA methylation data

We quantified methylation levels for a total of 930,698 CpG sites using the R Bioconductor package *meffil* (v1.3.6) [99]. Briefly, functional normalisation was applied to array intensity data (".idat" files) to correct for between-array technical variation. Low-quality and sex chromosome probes were removed. Methylation status was reported for all measured CpG sites as β-values ranging from 0 to 1 representing the fraction of DNAm signal compared to the total locus signal intensity.

An EWAS model was constructed to test for association of DNA methylation CpG sites with disease status, incorporating surrogate variables to correct for confounders. Details about the genomic locations of all CpG sites were loaded from the Illumina methylation EPIC array v2.0 annotation package (IlluminaHumanMethylationEPICv2anno.20a1.hg38). Differentially methylated regions (DMRs) between AF and SR samples were called with the *dmrff* package (https://github.com/perishky/dmrff) by combining the EWAS summary statistics from CpG sites located close together and showing strong associations. DMRs were defined as regions containing more than one CpG with at most 500 bp between consecutive sites and effect estimates with the same direction. Significant DMRs were called with p-values < 0.05, with Bonferroni multiple-testing correction.

Analyses of enriched transcription factor motifs, gene ontology terms and binding regions

For *de novo* and known motif enrichment analysis, the findMotifsGenome.pl script (-size given) from HOMER v5.1 [54] was used on H3K27ac-enriched sets of interest and hyper- or hypomethylated DMRs, using the sets of reference peaks from each category and all candidate regions with aggregated CpG methylation signal respectively as background. In addition, significance for TF binding sites from ChIP-seq experimental data generated from the ReMap catalogue was assessed using the shuffling procedure implemented in the ReMapEnrich R package [55]. A subset of ChIP-seq TF binding experiments from the ReMap catalogue was selected from biologically relevant samples (i.e., human cardiac tissue or *in vitro* heart models).

For enrichment of AF-associated TFs on the DMRs, these were selected from recent GWAS [19] and binding motifs were collected from JASPAR [100]. FIMO online tool (Find Individual Motif Occurrences) from the MEME suite (v.5.5.7) (match p-value < 0.001) was used to annotate motifs over DMRs and the background set of regions [101]. A hypergeometric test was used to calculate statistical significance of motif occurrences across hypermethylated and hypomethylated regions (Bonferroni correction was applied for multiple testing).

For gene ontology enrichment analyses, genes were assigned to enriched regions based on the default association rule in GREAT v4.0.4 [102], and evaluated for ontology term enrichments against reference peaks from each category using the enrichGO function from clusterProfiler v4.15.1 [103].

Enrichment of GWAS signals and cell-type specific chromatin accessibility

Genetic lead variants associated to cardiovascular traits at genome-wide significance (p-value < 10^{-8}) were obtained from the NHGRI-EBI GWAS catalog [56] (version e113, 2025), for all heart-related traits (search term “heart” across associated traits in the catalog), and four major cardiovascular diseases and traits (search terms “atrial fibrillation, “QT interval”, “coronary artery disease” and “cardiomyopathy”). Each set of lead variants was annotated with proxy variants in strong LD (R^2 > 0.8) using SNIPA [104]. The enrichment of GWAS SNPs in heart promoters and enhancers across epigenomically-conserved, primate-specific and human only categories calculated with a hypergeometric test.

Signatures of gene regulation across heart cell types and cardiomyocytes were obtained from previously published datasets: RPKM chromatin accessibility levels across cardiac cell types [58], and significant promoter-enhancer interactions in human iPSCs and iPSC-derived cardiomyocytes were obtained from [57].

AUTHOR CONTRIBUTIONS

AR, SF, DFA, JC, GT and DV designed and performed experiments. AR, CGB and DV analysed data. AMT, AT and PBM provided essential reagents and technical advice. DV secured funding and coordinated the study. DV and AR wrote the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

We thank the QMUL Genome Centre for assistance with DNA methylation analyses on EPIC v2 arrays and members of the QMUL Centre for Epigenetics, Prof. Miguel Manzanares (CBMSO, Madrid, Spain) and Dr. Ines Cebola (Imperial College London, UK) for helpful criticism and discussions.

DECLARATIONS

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

Research involving human participants was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained in collaboration with the National Institutes for Health Research (NIHR) Barts Biomedical Research Centre Cardiovascular Bio-Registry from the East of England-Cambridge Central Research Committee (14/EE/0007). Patients at Barts Heart Centre, Barts Health National Health Service (NHS) Trust gave informed consent to participate in the study.

FUNDING

This work has been funded by the British Heart Foundation (Fellowship FS/18/39/33684 to D.V.), Barts Charity (large grant MGU0501 and seed grant G-002635 to D.V.). AR has been supported by a PhD stipend from Barts Charity (MGU0501). AT and PBM acknowledge support from the NIHR Biomedical Research Centre at Barts for provision of atrial appendage samples used in this study.

AVAILABILITY OF DATA AND MATERIALS

The ChIP-sequencing and DNA methylation data reported here has been deposited to GEO with accession numbers GSE227793 and GSE291249, respectively.
Analysis code is available in Github:
https://github.com/adrian-rodr/Rodriguez_2025_AFepigenome.git

REFERENCES

1. Linz D, Gawalko M, Betz K, Hendriks JM, Lip GYH, Vinter N, Guo Y, Johnsen S: **Atrial fibrillation: epidemiology, screening and digital health**. *Lancet Reg Health Eur* 2024, **37**:100786.
2. Lippi G, Sanchis-Gomar F, Cervellin G: **Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge**. *Int J Stroke* 2021, **16**(2):217-221.
3. Vinciguerra M, Dobrev D, Nattel S: **Atrial fibrillation: pathophysiology, genetic and epigenetic mechanisms**. *Lancet Reg Health Eur* 2024, **37**:100785.
4. Staerk L, Sherer JA, Ko D, Benjamin EJ, Helm RH: **Atrial Fibrillation: Epidemiology, Pathophysiology, and Clinical Outcomes**. *Circ Res* 2017, **120**(9):1501-1517.
5. Voigt N, Heijman J, Wang Q, Chiang DY, Li N, Karck M, Wehrens XHT, Nattel S, Dobrev D: **Cellular and molecular mechanisms of atrial arrhythmogenesis in patients with paroxysmal atrial fibrillation**. *Circulation* 2014, **129**(2):145-156.
6. Nattel S, Heijman J, Zhou L, Dobrev D: **Molecular Basis of Atrial Fibrillation Pathophysiology and Therapy: A Translational Perspective**. *Circ Res* 2020, **127**(1):51-72.
7. Andersen JH, Andreasen L, Olesen MS: **Atrial fibrillation-a complex polygenetic disease**. *Eur J Hum Genet* 2021, **29**(7):1051-1060.
8. Hodgson-Zingman DM, Karst ML, Zingman LV, Heublein DM, Darbar D, Herron KJ, Ballew JD, de Andrade M, Burnett JC, Jr., Olson TM: **Atrial natriuretic peptide frameshift mutation in familial atrial fibrillation**. *N Engl J Med* 2008, **359**(2):158-165.
9. Chen YH, Xu SJ, Bendahhou S, Wang XL, Wang Y, Xu WY, Jin HW, Sun H, Su XY, Zhuang QN *et al*: **KCNQ1 gain-of-function mutation in familial atrial fibrillation**. *Science* 2003, **299**(5604):251-254.
10. Sinner MF, Pfeufer A, Akyol M, Beckmann BM, Hinterseer M, Wacker A, Perz S, Sauter W, Illig T, Nabauer M *et al*: **The non-synonymous coding IKr-channel variant KCNH2-K897T is associated with atrial fibrillation: results from a systematic candidate gene-based analysis of KCNH2 (HERG)**. *Eur Heart J* 2008, **29**(7):907-914.
11. Savio-Galimberti E, Weeke P, Muhammad R, Blair M, Ansari S, Short L, Atack TC, Kor K, Vanoye CG, Olesen MS *et al*: **SCN10A/Nav1.8 modulation of peak and late sodium currents in patients with early onset atrial fibrillation**. *Cardiovasc Res* 2014, **104**(2):355-363.
12. Macri V, Mahida SN, Zhang ML, Sinner MF, Dolmatova EV, Tucker NR, McLellan M, Shea MA, Milan DJ, Lunetta KL *et al*: **A novel trafficking-defective HCN4 mutation is associated with early-onset atrial fibrillation**. *Heart Rhythm* 2014, **11**(6):1055-1062.
13. Ahlberg G, Refsgaard L, Lundegaard PR, Andreasen L, Ranthe MF, Linscheid N, Nielsen JB, Melbye M, Haunso S, Sajadieh A *et al*: **Rare truncating variants in the sarcomeric protein titin associate with familial and early-onset atrial fibrillation**. *Nat Commun* 2018, **9**(1):4316.
14. Orr N, Arnaout R, Gula LJ, Spears DA, Leong-Sit P, Li Q, Tarhuni W, Reischauer S, Chauhan VS, Borkovich M *et al*: **A mutation in the atrial-specific myosin light chain gene (MYL4) causes familial atrial fibrillation**. *Nat Commun* 2016, **7**:11303.
15. Postma AV, van de Meerakker JB, Mathijssen IB, Barnett P, Christoffels VM, Ilgun A, Lam J, Wilde AA, Lekanne Deprez RH, Moorman AF: **A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation**. *Circ Res* 2008, **102**(11):1433-1442.
16. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, Almgren P, Alonso A, Anderson CD, Aragam KG *et al*: **Multi-ethnic genome-wide association study for atrial fibrillation**. *Nat Genet* 2018, **50**(9):1225-1233.
17. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, Herron TJ, McCarthy S, Schmidt EM, Sveinbjornsson G *et al*: **Biobank-driven genomic discovery yields new insight into atrial fibrillation biology**. *Nat Genet* 2018, **50**(9):1234-1239.
18. Miyazawa K, Ito K, Ito M, Zou Z, Kubota M, Nomura S, Matsunaga H, Koyama S, Ieki H, Akiyama M *et al*: **Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction**. *Nat Genet* 2023, **55**(2):187-197.
19. Roselli C, Surakka I, Olesen MS, Sveinbjornsson G, Marston NA, Choi SH, Holm H, Chaffin M, Gudbjartsson D, Hill MC *et al*: **Meta-analysis of genome-wide associations and polygenic risk prediction for atrial fibrillation in more than 180,000 cases**. *Nat Genet* 2025, **57**(3):539-547.

20. van Ouwerkerk AF, Hall AW, Kadow ZA, Lazarevic S, Reyat JS, Tucker NR, Nadadur RD, Bosada FM, Bianchi V, Ellinor PT *et al*: **Epigenetic and Transcriptional Networks Underlying Atrial Fibrillation**. *Circ Res* 2020, **127**(1):34-50.

21. van Ouwerkerk AF, Bosada FM, van Duijvenboden K, Hill MC, Montefiori LE, Scholman KT, Liu J, de Vries AAF, Boukens BJ, Ellinor PT *et al*: **Identification of atrial fibrillation associated genes and functional non-coding variants**. *Nat Commun* 2019, **10**(1):4755.

22. Schulz C, Lemoine MD, Mearini G, Koivumaki J, Sani J, Schwedhelm E, Kirchhof P, Ghalawinji A, Stoll M, Hansen A *et al*: **PITX2 Knockout Induces Key Findings of Electrical Remodeling as Seen in Persistent Atrial Fibrillation**. *Circ Arrhythm Electrophysiol* 2023, **16**(3):e011602.

23. Reyat JS, Sommerfeld LC, O'Reilly M, Roth Cardoso V, Thiemann E, Khan AO, O'Shea C, Harder S, Muller C, Barlow J *et al*: **PITX2 deficiency leads to atrial mitochondrial dysfunction**. *Cardiovasc Res* 2024, **120**(15):1907-1923.

24. Kao YH, Hsu JC, Chen YC, Lin YK, Lkhagva B, Chen SA, Chen YJ: **ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes**. *Int J Cardiol* 2016, **210**:85-92.

25. Jameson HS, Hanley A, Hill MC, Xiao L, Ye J, Bapat A, Ronzier E, Hall AW, Hucker WJ, Clauss S *et al*: **Loss of the Atrial Fibrillation-Related Gene, Zfhx3, Results in Atrial Dilation and Arrhythmias**. *Circ Res* 2023, **133**(4):313-329.

26. Wass SY, Offerman EJ, Sun H, Hsu J, Rennison JH, Cantlay CC, McHale ML, Gillinov AM, Moravec C, Smith JD *et al*: **Novel functional atrial fibrillation risk genes and pathways identified from coexpression analyses in human left atria**. *Heart Rhythm* 2023, **20**(9):1219-1226.

27. George RM, Guo S, Firulli BA, Rubart M, Firulli AB: **Neonatal Deletion of Hand1 and Hand2 within Murine Cardiac Conduction System Reveals a Novel Role for HAND2 in Rhythm Homeostasis**. *J Cardiovasc Dev Dis* 2022, **9**(7).

28. Clausen AG, Vad OB, Andersen JH, Olesen MS: **Loss-of-Function Variants in the SYNPO2L Gene Are Associated With Atrial Fibrillation**. *Front Cardiovasc Med* 2021, **8**:650667.

29. Wass SY, Sun H, Tchou G, Liu N, Van Wagoner DR, Chung MK, Barnard J, Smith JD: **Transcriptomic Insights into the Atrial Fibrillation Susceptibility Locus near the MYOZ1 and SYNPO2L Genes**. *Int J Mol Sci* 2024, **25**(19).

30. Goette A, Kalman JM, Aguinaga L, Akar J, Cabrera JA, Chen SA, Chugh SS, Corradi D, D'Avila A, Dobrev D *et al*: **EHRA/HRS/APHRS/SOLAECE expert consensus on Atrial cardiomyopathies: Definition, characterisation, and clinical implication**. *J Arrhythm* 2016, **32**(4):247-278.

31. Lozano-Velasco E, Franco D, Aranega A, Daimi H: **Genetics and Epigenetics of Atrial Fibrillation**. *Int J Mol Sci* 2020, **21**(16).

32. Huiskes FG, Creemers EE, Brundel B: **Dissecting the Molecular Mechanisms Driving Electropathology in Atrial Fibrillation: Deployment of RNA Sequencing and Transcriptomic Analyses**. *Cells* 2023, **12**(18).

33. Thomas AM, Cabrera CP, Finlay M, Lall K, Nobles M, Schilling RJ, Wood K, Mein CA, Barnes MR, Munroe PB *et al*: **Differentially expressed genes for atrial fibrillation identified by RNA sequencing from paired human left and right atrial appendages**. *Physiol Genomics* 2019, **51**(8):323-332.

34. Hsu J, Hanna P, Van Wagoner DR, Barnard J, Serre D, Chung MK, Smith JD: **Whole genome expression differences in human left and right atria ascertained by RNA sequencing**. *Circ Cardiovasc Genet* 2012, **5**(3):327-335.

35. Lipovsky CE, Jimenez J, Guo Q, Li G, Yin T, Hicks SC, Bhatnagar S, Takahashi K, Zhang DM, Brumback BD *et al*: **Chamber-specific transcriptional responses in atrial fibrillation**. *JCI Insight* 2020, **5**(18).

36. Cubukcuoglu Deniz G: **Transcriptomic Landscape of Human Left and Right Atria Associated with Atrial Fibrillation**. *Anatol J Cardiol* 2025, **29**(5):242-248.

37. Han MR, Jeong JH, Kim YG, Yang HH, Seo CO, Kim Y, Lee HS, Shim J, Kim YH, Choi JI: **Epigenetic regulation on left atrial function and disease recurrence after catheter ablation in atrial fibrillation**. *Clin Epigenetics* 2024, **16**(1):183.

38. Lin H, Yin X, Xie Z, Lunetta KL, Lubitz SA, Larson MG, Ko D, Magnani JW, Mendelson MM, Liu C *et al*: **Methylome-wide Association Study of Atrial Fibrillation in Framingham Heart Study**. *Sci Rep* 2017, **7**:40377.

39. Zhao G, Zhou J, Gao J, Liu Y, Gu S, Zhang X, Su P: **Genome-wide DNA methylation analysis in permanent atrial fibrillation**. *Mol Med Rep* 2017, **16**(4):5505-5514.

40. Donate Puertas R, Meugnier E, Romestaing C, Rey C, Morel E, Lachuer J, Gadot N, Scridon A, Julien C, Tronc F *et al*: **Atrial fibrillation is associated with hypermethylation in human left atrium, and treatment with decitabine reduces atrial tachyarrhythmias in spontaneously hypertensive rats**. *Transl Res* 2017, **184**:57-67 e55.

41. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans**. *Nature* 2011, **470**(7333):279-283.

42. Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R *et al*: **Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution**. *Science* 2014, **346**(6212):1007-1012.

43. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ *et al*: **Enhancer evolution across 20 mammalian species**. *Cell* 2015, **160**(3):554-566.

44. Parey E, Fernandez-Aroca D, Frost S, Uribarren A, Park TJ, Zottl M, St John Smith E, Berthelot C, Villar D: **Phylogenetic modeling of enhancer shifts in African mole-rats reveals regulatory changes associated with tissue-specific traits**. *Genome Res* 2023, **33**(9):1513-1526.

45. Tan WLW, Anene-Nzelu CG, Wong E, Lee CJM, Tan HS, Tang SJ, Perrin A, Wu KX, Zheng W, Ashburn RJ *et al*: **Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype**. *Circ Res* 2020, **127**(6):761-777.

46. Pei J, Harakalova M, Treibel TA, Lumbers RT, Boukens BJ, Efimov IR, van Dinter JT, Gonzalez A, Lopez B, El Azzouzi H *et al*: **H3K27ac acetylome signatures reveal the epigenomic reorganization in remodeled non-failing human hearts**. *Clin Epigenetics* 2020, **12**(1):106.

47. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR *et al*: **Differential oestrogen receptor binding is associated with clinical outcome in breast cancer**. *Nature* 2012, **481**(7381):389-393.

48. Eder T, Grebien F: **Comprehensive assessment of differential ChIP-seq tools guides optimal algorithm selection**. *Genome Biol* 2022, **23**(1):119.

49. Berthelot C, Villar D, Horvath JE, Odom DT, Flicek P: **Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression**. *Nat Ecol Evol* 2018, **2**(1):152-163.

50. Hou L, Xiong X, Park Y, Boix C, James B, Sun N, He L, Patel A, Zhang Z, Molinie B *et al*: **Multitissue H3K27ac profiling of GTEx samples links epigenomic variation to disease**. *Nat Genet* 2023, **55**(10):1665-1676.

51. Geelhoed B, Borschel CS, Niiranen T, Palosaari T, Havulinna AS, Fouodo CJK, Scheinhardt MO, Blankenberg S, Jousilahti P, Kuulasmaa K *et al*: **Assessment of causality of natriuretic peptides and atrial fibrillation and heart failure: a Mendelian randomization study in the FINRISK cohort**. *Europace* 2020, **22**(10):1463-1469.

52. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.

53. Worsley Hunt R, Mathelier A, Del Peso L, Wasserman WW: **Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment**. *BMC Genomics* 2014, **15**(1):472.

54. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities**. *Mol Cell* 2010, **38**(4):576-589.

55. Hammal F, de Langen P, Bergon A, Lopez F, Ballester B: **ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments**. *Nucleic Acids Res* 2022, **50**(D1):D316-D325.

56. Cerezo M, Sollis E, Ji Y, Lewis E, Abid A, Bircan KO, Hall P, Hayhurst J, John S, Mosaku A *et al*: **The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity**. *Nucleic Acids Res* 2025, **53**(D1):D998-D1005.

57. Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, Bozek G, Moskowitz IP, McNally EM, Nobrega MA: **A promoter interaction map for cardiovascular disease genetics**. *Elife* 2018, **7**.

58. Hocker JD, Poirion OB, Zhu F, Buchanan J, Zhang K, Chiou J, Wang TM, Zhang Q, Hou X, Li YE *et al*: **Cardiac cell type-specific gene regulatory programs and disease risk association**. *Sci Adv* 2021, **7**(20).

59. Man JCK, van Duijvenboden K, Krijger PHL, Hooijkaas IB, van der Made I, de Gier-de Vries C, Wakker V, Creemers EE, de Laat W, Boukens BJ *et al*: **Genetic Dissection of a Super Enhancer Controlling the Nppa-Nppb Cluster in the Heart**. *Circ Res* 2021, **128**(1):115-129.

60. Rommel C, Rosner S, Lother A, Barg M, Schwaderer M, Gilsbach R, Bomicke T, Schnick T, Mayer S, Doll S *et al*: **The Transcription Factor ETV1 Induces Atrial Remodeling and Arrhythmia**. *Circ Res* 2018, **123**(5):550-563.

61. Bertero A, Fields PA, Smith AST, Leonard A, Beussman K, Sniadecki NJ, Kim DH, Tse HF, Pabon L, Shendure J *et al*: **Chromatin compartment dynamics in a haploinsufficient model of cardiac laminopathy**. *J Cell Biol* 2019, **218**(9):2919-2944.

62. Ghosh AK, Kalousdian AA, Shang M, Lux E, Eren M, Keating A, Wilsbacher LD, Vaughan DE: **Cardiomyocyte PAI-1 influences the cardiac transcriptome and limits the extent of cardiac fibrosis in response to left ventricular pressure overload**. *Cell Signal* 2023, **104**:110555.

63. Villar D, Frost S, Deloukas P, Tinker A: **The contribution of non-coding regulatory elements to cardiovascular disease**. *Open Biol* 2020, **10**(7):200088.

64. Gomez-Ochoa SA, Mohn M, Malz MV, Ottenheijm R, Lanzer JD, Wiedmann F, Kraft M, Muka T, Schmidt C, Freichel M *et al*: **The transcriptional landscape of atrial fibrillation: A systematic review and meta-analysis**. *PLoS One* 2025, **20**(5):e0323534.

65. Cubukcuoglu Deniz G, Durdu S, Dogan Y, Erdemli E, Ozdag H, Akar AR: **Molecular Signatures of Human Chronic Atrial Fibrillation in Primary Mitral Regurgitation**. *Cardiovasc Ther* 2021, **2021**:5516185.

66. Jonmundsson T, Steindorsdottir AE, Austin TR, Frick EA, Axelsson GT, Launer L, Psaty BM, Loureiro J, Orth AP, Aspelund T *et al*: **A proteomic analysis of atrial fibrillation in a prospective longitudinal cohort (AGES-Reykjavik study)**. *Europace* 2023, **25**(11).

67. Bagchi RA, Roche P, Aroutiounova N, Espira L, Abrenica B, Schweitzer R, Czubryt MP: **The transcription factor scleraxis is a critical regulator of cardiac fibroblast phenotype**. *BMC Biol* 2016, **14**:21.

68. Wolke C, Antileo E, Lendeckel U: **WNT signaling in atrial fibrillation**. *Exp Biol Med (Maywood)* 2021, **246**(9):1112-1120.

69. Chilukoti RK, Giese A, Malenke W, Homuth G, Bukowska A, Goette A, Felix SB, Kanaan J, Wollert HG, Evert K *et al*: **Atrial fibrillation and rapid acute pacing regulate adipocyte/adipositas-related gene expression in the atria**. *Int J Cardiol* 2015, **187**:604-613.

70. Chen CY, Salomon AK, Caporizzo MA, Curry S, Kelly NA, Bedi K, Bogush AI, Kramer E, Schlossarek S, Janiak P *et al*: **Depletion of Vasohibin 1 Speeds Contraction and Relaxation in Failing Human Cardiomyocytes**. *Circ Res* 2020, **127**(2):e14-e27.

71. Ding Y, Fang J, Chen M, Xu Y, Liu N, Fang S, Xiang W, Chen R, Wu C, Yu H: **MT1X is an oncogene and indicates prognosis in ccRCC**. *Biosci Rep* 2022, **42**(10).

72. Furukawa J, Inoue K, Maeda J, Yasujima T, Ohta K, Kanai Y, Takada T, Matsuo H, Yuasa H: **Functional identification of SLC43A3 as an equilibrative nucleobase transporter involved in purine salvage in mammals**. *Sci Rep* 2015, **5**:15057.

73. Hasbargen KB, Shen WJ, Zhang Y, Hou X, Wang W, Shuo Q, Bernlohr DA, Azhar S, Kraemer FB: **Slc43a3 is a regulator of free fatty acid flux**. *J Lipid Res* 2020, **61**(5):734-745.

74. Ran Q, Li A, Tan Y, Zhang Y, Zhang Y, Chen H: **Action and therapeutic targets of myosin light chain kinase, an important cardiovascular signaling mechanism**. *Pharmacol Res* 2024, **206**:107276.

75. Bates EA, Kipp ZA, Lee WH, Martinez GJ, Weaver L, Becker KN, Pauss SN, Creeden JF, Anspach GB, Helsley RN *et al*: **FOXS1 is increased in liver fibrosis and regulates TGFbeta responsiveness and proliferation pathways in human hepatic stellate cells**. *J Biol Chem* 2024, **300**(3):105691.

76. Chai C, Li S, Chen L, Song X: **Causal association of calcific aortic valve stenosis and atrial fibrillation: a Mendelian randomization study**. *Sci Rep* 2023, **13**(1):20284.

77. Jiang C, Yao D, Shen Q, Tian R, Fan L, Zheng Q, Qian X, Liu Z, Huang Y, Dong N: **Oxidized LDL-induced FOXS1 mediates cholesterol transport dysfunction and inflammasome activation to drive aortic valve calcification**. *Cardiovasc Res* 2025.

78. Schubeler D: **Function and information content of DNA methylation**. *Nature* 2015, **517**(7534):321-326.

79. Cooper CD, Newman JA, Aitkenhead H, Allerston CK, Gileadi O: **Structures of the Ets Protein DNA-binding Domains of Transcription Factors Etv1, Etv4, Etv5, and Fev: DETERMINANTS OF**

DNA BINDING AND REDOX REGULATION BY DISULFIDE BOND FORMATION. *J Biol Chem* 2015, **290**(22):13692-13709.

80. Stephens DC, Poon GM: **Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains**. *Nucleic Acids Res* 2016, **44**(18):8671-8681.

81. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V *et al*: **Rapid and pervasive changes in genome-wide enhancer usage during mammalian development**. *Cell* 2013, **155**(7):1521-1531.

82. Spurrell CH, Barozzi I, Kosicki M, Mannion BJ, Blow MJ, Fukuda-Yuzawa Y, Slaven N, Afzal SY, Akiyama JA, Afzal V *et al*: **Genome-wide fetalization of enhancer architecture in heart disease**. *Cell Rep* 2022, **40**(12):111400.

83. Pradeepa MM, Grimes GR, Kumar Y, Olley G, Taylor GC, Schneider R, Bickmore WA: **Histone H3 globular domain acetylation identifies a new class of enhancers**. *Nat Genet* 2016, **48**(6):681-686.

84. Chattopadhyaya S, Nagalingam RS, Ledingham DA, Moffatt TL, Al-Hattab DS, Narhan P, Stecy MT, O'Hara KA, Czubryt MP: **Regulation of Cardiac Fibroblast GLS1 Expression by Scleraxis**. *Cells* 2022, **11**(9).

85. Teschendorff AE, Zhu T, Breeze CE, Beck S: **EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data**. *Genome Biol* 2020, **21**(1):221.

86. Li H, Sharma A, Luo K, Qin ZS, Sun X, Liu H: **DeconPeaker, a Deconvolution Model to Identify Cell Types Based on Chromatin Accessibility in ATAC-Seq Data of Mixture Samples**. *Front Genet* 2020, **11**:392.

87. Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, Hannani MT, Tanevski J, Wunnemann F, Nagai JS *et al*: **Spatial multi-omic map of human myocardial infarction**. *Nature* 2022, **608**(7924):766-777.

88. Kanemaru K, Cranley J, Muraro D, Miranda AMA, Ho SY, Wilbrey-Clark A, Patrick Pett J, Polanski K, Richardson L, Litvinukova M *et al*: **Spatially resolved multiomics of human cardiac niches**. *Nature* 2023, **619**(7971):801-810.

89. Yan L, Ma C, Wang D, Hu Q, Qin M, Conroy JM, Sucheston LE, Ambrosone CB, Johnson CS, Wang J *et al*: **OSAT: a tool for sample-to-batch allocations in genomics experiments**. *BMC Genomics* 2012, **13**:689.

90. Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT: **ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions**. *Methods* 2009, **48**(3):240-248.

91. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al*: **Model-based analysis of ChIP-Seq (MACS)**. *Genome Biol* 2008, **9**(9):R137.

92. Pongor LS, Gross JM, Vera Alvarez R, Murai J, Jang SM, Zhang H, Redon C, Fu H, Huang SY, Thakur B *et al*: **BAMscale: quantification of next-generation sequencing peaks and generation of scaled coverage tracks**. *Epigenetics Chromatin* 2020, **13**(1):21.

93. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics* 2013, **29**(1):15-21.

94. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.

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

96. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W: **BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis**. *Bioinformatics* 2005, **21**(16):3439-3440.

97. McLellan MA, Skelly DA, Dona MSI, Squiers GT, Farrugia GE, Gaynor TL, Cohen CD, Pandey R, Diep H, Vinh A *et al*: **High-Resolution Transcriptomic Profiling of the Heart During Chronic Stress Reveals Cellular Drivers of Cardiac Fibrosis and Hypertrophy**. *Circulation* 2020, **142**(15):1448-1463.

98. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative C(T) method**. *Nat Protoc* 2008, **3**(6):1101-1108.

99. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M: **Meffil: efficient normalization and analysis of very large DNA methylation datasets**. *Bioinformatics* 2018, **34**(23):3983-3989.

100. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, Lemma RB, Lucas J, Cheneby J, Baranasic D *et al*: **JASPAR 2024: 20th anniversary of the open-**

access database of transcription factor binding profiles. *Nucleic Acids Res* 2024, **52**(D1):D174-D182.

101. Grant CE, Bailey TL, Noble WS: **FIMO: scanning for occurrences of a given motif**. *Bioinformatics* 2011, **27**(7):1017-1018.

102. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions**. *Nat Biotechnol* 2010, **28**(5):495-501.

103. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L *et al*: **clusterProfiler 4.0: A universal enrichment tool for interpreting omics data**. *Innovation (Camb)* 2021, **2**(3):100141.

104. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmuller G: **SNiPA: an interactive, genetic variant-centered annotation browser**. *Bioinformatics* 2015, **31**(8):1334-1336.