TITLE

by Natasha Louise Hopkins

Master of Biology (Honours), Molecular Cell Biology
University of York, UK

Project Director
Prof. Robert J White

Examination Date
17 April, 2023

Word Count
Abstract:
Main:

Contents

Introduction **1**

FOXA1 Expression and ERα+ Breast Cancer 1

tRNAs and Gene Expression 1

Materials & Methods **1**

Acquisition of Public ChIP-seq Datasets 1

EaSeq for Chip-seq Peak Quantification 2

Motif Analysis 2

Statistics 2

Results **3**

Binding of FOXA1 and H3k27ac to tRNA Genes 3

Figure 2 5

Why? 5

What? 5

Suggests 5

Figure 3 6

Figure ? 8

Figure 4 8

What? 8

Figure 5 10

Discussion **11**

Conclusion **11**

References **11**

TITLE

Natasha L. Hopkins

Abstract

1 Words

Introduction

FOXA1 Expression and ERα+ Breast Cancer

tRNAs and Gene Expression

Materials & Methods

Acquisition of Public ChIP-seq Datasets

ChIP-seq was performed on genetically modified MCF7L cells (*insertion, using a lentiviral cDNA delivery system to express Dox-inducible FOXA1*)[1]. Datasets were deposited into the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA)[2] under accession no. PRJNA512997 (Table 1). Using "Genetic Manipulation Tools" within the Galaxy[3] environment (v 23.0.rc1), SRAs were converted to FastQ files. FastQ files were then aligned to the human genome assembly GRCh37 (hg19) using Bowtie2 (v 2.5.0)[4] to output BAM files.

Table 1. Publicly available ChIP-seq SRA files aquired from the NCBI SRA database (accession no. PRJNA512997).

Experiment	SRA	Factor	Tissue	Assembly
PRJNA512997	SRR8393424 SRR8393425	FOXA1	MCF-7LP	GRCh37 (Hg19)

Experiment	SRA	Factor	Tissue	Assembly
	SRR8393426			
	SRR8393427	H3K27ac		
	SRR8393428			
	SRR8393431	None (input)		
	SRR8393432			

EaSeq for Chip-seq Peak Quantification

BAM files were uploaded into EaSeq (v1.111) as "Datasets" using the standard settings for Chip-seq data. GRCh37 (hg19) tRNA sequences (n = 606) were downloaded as a "Geneset" from the UCSC Table Browser[5], (available at https://genome.ucsc.edu). High-confidence tRNAs (n = 416) identified in the GtRNAdb[6] were extracted as a "Regionset".

Signal peak intensities surrounding tRNAs were quantified using the EaSeq "quantify" tool. Here the default settings "Normalize to reads per million" and "Normalize counts to DNA-fragments" were left checked. The default setting "Normalise to a signal of 1000 bp" was unchecked. The window size was offset ±500bp from the start of each tRNA gene. Outputs are referred to as "Q-values".

To quantify upstream and downstream signals, the "quantify" tool was used with adjusted window sizes. The upstream region was defined as 500 bp preceding and the first nucleotide of tRNA loci. Thus, the start position was offset to 0 bp, and the end position was offset to -500 bp. The downstream region constitutes the 500 bp region beginning with the first nucleotide of tRNA gene body. The start position was offset to 1 bp, and the end position was offset to 500 bp.

Following quantification, tRNA binding events were arranged in ascending order -DOX Q-value and visualised as heatmaps. Data was also visualised with "average", and "overlay" EaSeq tools.

EaSeq[7] is avaiable at http://easeq.net.

Motif Analysis

Multiple EM for Motif Elicitation ChIP (MEME) Suite

Statistics

Statistical tests and graphs were generated with R[8] (v 4.2.3), R Studio[9] (v 2023.03.0.386) and the tidyverse[10] package.

Results

Binding of FOXA1 and H3k27ac to tRNA Genes

To investigate the impact of FOXA1 on tRNA enhancers in ER+ MCF-7 cells, public ChIP-Seq datasets from Fu et al. (2019)[1] were interrogated. In this paper, a doxycycline (Dox) inducible OE system was used to achieve FOXA1 OE akin to tamoxifen-resistant (TamR) MCF-7 cells[1]. Quantification of 416 high-confidence tRNAs revealed that FOXA1 interacts with 100% of tRNAs and H3k27ac with 99.5%. of tRNA genes, relative to ±500 bp flanking regions (Figure 1A). Whereas input reads generated minimal peak enrichment (Supplementary Figure X). The number of genes bound was unchanged by OE of FOXA1.

Mapped reads of FOXA1 and H3K27ac binding were visualised as heatmaps and ordered by increasing -DOX Q-value. This revealed a concentration of FOXA1 and H3k27ac at approximately half of tRNAs, relative to ±10 kb flanking regions. Upon FOXA1 OE, FOXA1 binding increased at a small proportion of tRNAs genes and H3K27ac binding decreases at approximately half of tRNA genes (Figure 1A). This was confirmed by average signal intensity plots of FOXA1 and H3K27ac binding (Figure 1B).

Upon FOXA1 OE, the mean Q-values for FOXA1 binding significantly decreased 0.16-fold and insignificantly decrease by 0.62-fold for H3K27ac (Mann-Whitney: $p < 0.005$) (Supplementary Figures X). However, there is a significant difference in binding between -DOX and +DOX tRNAs for both FOXA1 and H3K27ac (Wilcoxon Signed Rank Test : $p < 0.0001$) (Figure 1C). Changes in FOXA1 and H3K27ac binding has been visualised as a parameter heatmap in Figure 1D. Together, these results suggest that FOXA1 overexpression alters the binding landscape of FOXA1 and H3K27ac at tRNAs.

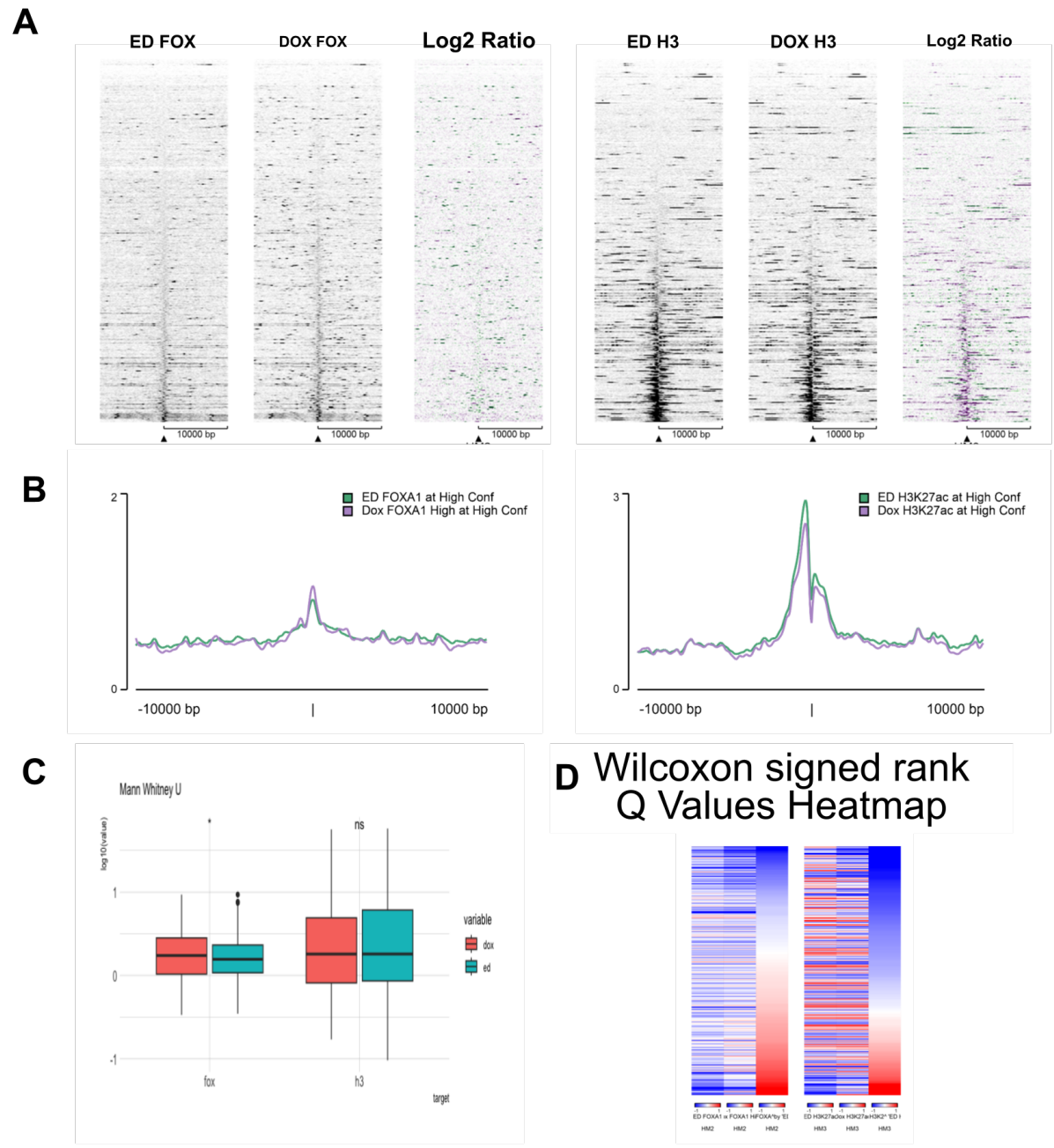


Figure 1. (B) Ratiometric heatmaps of the log2 ratio between the binding of FOXA1 or H3k27ac with endogenous FOXA1 expression vs. the binding of FOXA1 or H3k27ac with FOXA1 OE.

Figure 2

Why?

FOXA1 OE alters binding at tRNAs (fig1).

What?

Using Q-values, tRNAs that are differently enriched upon FOXA1 OE were categorised as 'GAIN' or 'LOSS'. This discovered substantially more tRNAs with increased (GAIN) than decreased (LOSS) FOXA1 (96 vs. 23) (Figure 2A). However, for H3K27ac, the number of tRNAs with an increase (GAIN) was comparable to those with a decrease (LOSS) (54 vs. 52)(Figure 2B).

Of the tRNAs which gained FOXA1, 14% also gained H3K27ac. and 13.5% lost H3K27ac. These co-binding events represent approximately 25% of all tRNAs that each gain H3K27ac or lose H3k27ac (Figure 2C).

Examples of these tRNAs are shown in Figure 2D.

Suggests

FOXA1 alone is insufficient in increasing tRNA activity.

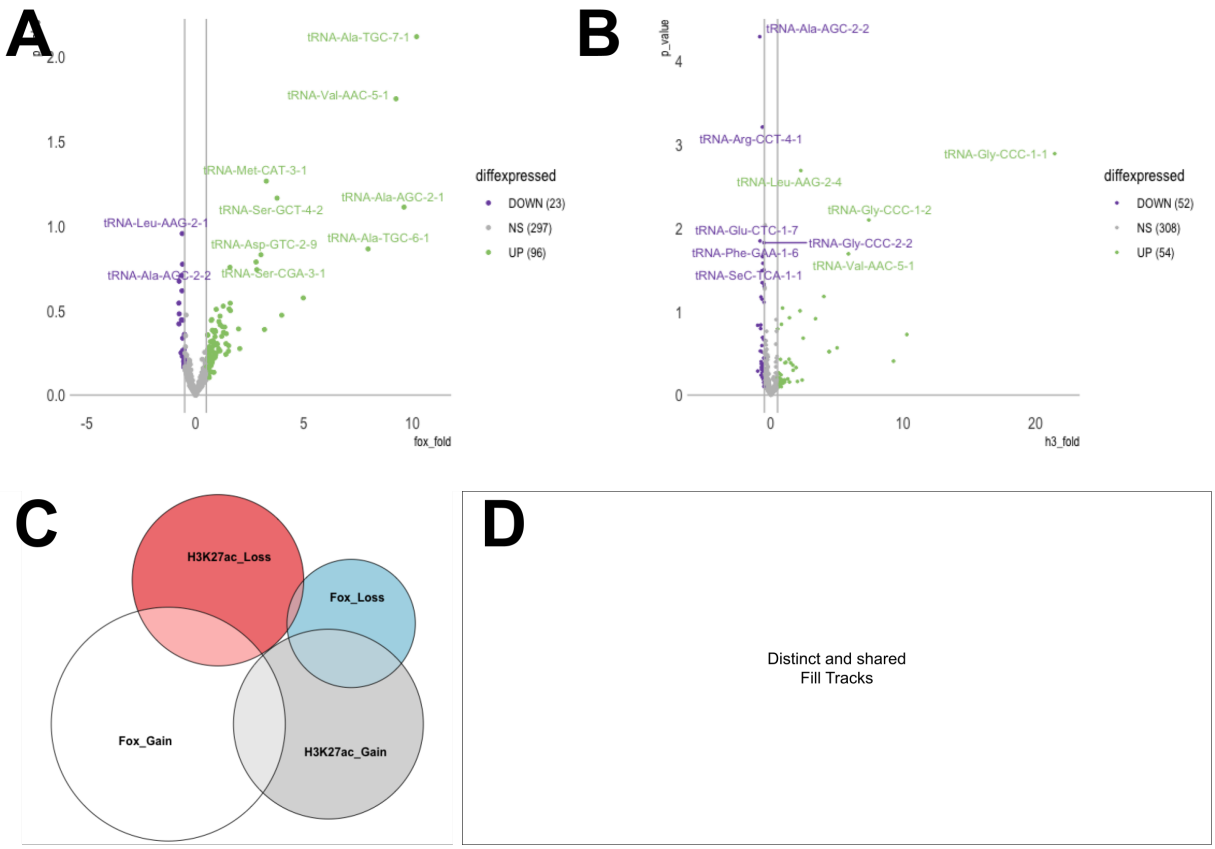


Figure 2. (B) Ratiometric heatmaps of the log2 ratio between the binding of FOXA1 or H3k27ac with endogenous FOXA1 expression vs. the binding of FOXA1 or H3k27ac with FOXA1 OE.

Figure 3

How many genes are 'activated' by FOXA1?

What?

- tRNAs were classified as 'active' if H3K27ac Q-values exceeded the median -DOX value (Q > 1.81).
- How many genes are 'active' in -DOX vs +DOX
  - Number is similar
  - Are +DOX genes the same genes? Compare -DOX and +DOX tRNA lists
  - Parameter heatmap of GAIN/LOSS genes?

Suggests?

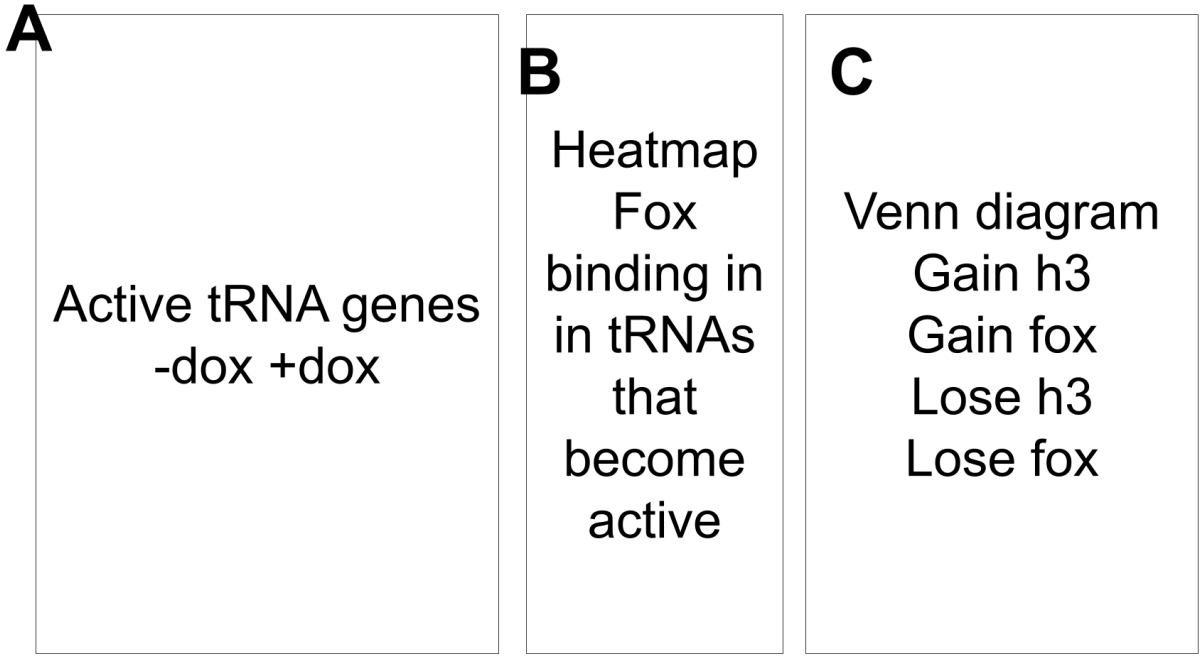

A
Active tRNA genes
-dox +dox
B
Heatmap
Fox
binding in
in tRNAs
that
become
active
C
Venn diagram
Gain h3
Gain fox
Lose h3
Lose fox

Figure 3. .

Figure ?

- Relative position (not very interesting)
- Binding at isotopes? Certain AA more up than others?

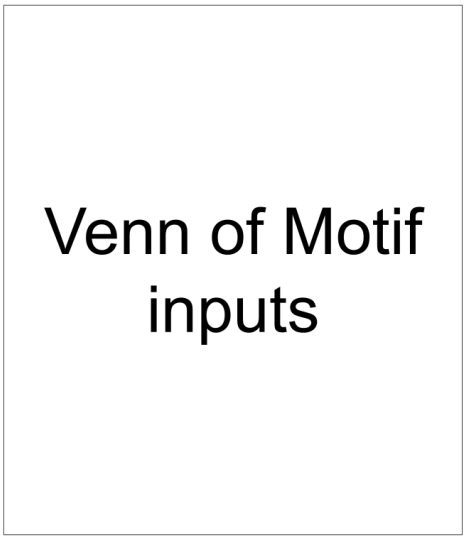
Figure 4

Why?

What?

MEME CentriMo to identify de novo motifs that are enriched at tRNAs which gain both FOXA1 and H3k27ac

- relative to other tRNAS (gain/lose, lose/lose, lose/gain)
- FOXA1 not enriched
- Look at ERE, AP-1, others?
- Top 3 motifs (fisher E values) - IRF7, ERR3, NR4A1
- A and B box motifs as a control? How?
  - All downstream
  - tRNAs where ‘matching sequences’ all best matches = code for valine
  - Branched amino acids associated with lower BC risk


Venn of Motif inputs


De novo Identified Motifs

Suggests?

Figure 5

- Localisation of FOXA1 at individual tRNA genes in MCF-7 cells

Why?

- tRNAs implicated in cancer
- Are they upregulated?
- Look at gain function/gain h3/fox
- Motif ontology
-

What?

Suggests?

Table 2. .

Group	Function
ALOXE	Insulator Function[11,12]
Ebersole	Insulator Function[12,13]
HES7	
Per1	
TMEM107	Insulator Function[11,12]
Arg-CCG	Implicated in Cancer Progression[14]
Glu-TTC	Implicated in Cancer Progression[14]
iMET	Proliferation of Breast Cancer
Met	iMet Control
SeC	Involved in REDOX[15]

Discussion

- FOXA1 alone not efficient to increase activity
  - p300
- FOXA1 moves nucleosomes to make other TF accessible?
- Loses fox = weak binding?
- Dynamic vs stable marks
- ATAC-seq

Conclusion

1009 Words

References

1. Fu X, Pereira R, De Angelis C, Veeraraghavan J, Nanda S, Qin L, et al. FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. Proceedings of the National Academy of Sciences [Internet]. 2019 Dec 11;116(52):26823–34. Available from: http://dx.doi.org/10.1073/pnas.1911584116
2. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. Nucleic Acids Research [Internet]. 2010 Nov 9;39(Database):D19–21. Available from: http://dx.doi.org/10.1093/nar/gkq1019
3. Afgan E, Nekrutenko A, Grüning BA, Blankenberg D, Goecks J, Schatz MC, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Research [Internet]. 2022 Apr 21;50(W1):W345–51. Available from: http://dx.doi.org/10.1093/nar/gkac247
4. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods [Internet]. 2012 Mar 4;9(4):357–9. Available from: http://dx.doi.org/10.1038/nmeth.1923
5. Karolchik D. The UCSC Table Browser data retrieval tool. Nucleic Acids Research [Internet]. 2004 Jan 1;32(90001):493D–496. Available from: http://dx.doi.org/10.1093/nar/gkh103

6. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Research [Internet]. 2015 Dec 15;44(D1):D184–9. Available from: http://dx.doi.org/10.1093/nar/gkv1309

7. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nature Structural & Molecular Biology [Internet]. 2016 Feb 29;23(4):349–57. Available from: http://dx.doi.org/10.1038/nsmb.3180

8. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2023. Available from: https://www.R-project.org/

9. Posit Team. RStudio: Integrated development environment for r [Internet]. Boston, MA: Posit Software, PBC; 2023. Available from: http://www.posit.co/

10. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. Journal of Open Source Software [Internet]. 2019 Nov 21;4(43):1686. Available from: http://dx.doi.org/10.21105/joss.01686

11. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, et al. Human tRNA genes function as chromatin insulators. The EMBO Journal [Internet]. 2011 Nov 15;31(2):330–50. Available from: http://dx.doi.org/10.1038/emboj.2011.406

12. Sizer RE, Chahid N, Butterfield SP, Donze D, Bryant NJ, White RJ. TFIIIC-based chromatin insulators through eukaryotic evolution. Gene [Internet]. 2022 Aug;835:146533. Available from: http://dx.doi.org/10.1016/j.gene.2022.146533

13. Ebersole T, Kim J-H, Samoshkin A, Kouprina N, Pavlicek A, White RJ, et al. tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. Cell Cycle [Internet]. 2011 Aug 15;10(16):2779–91. Available from: http://dx.doi.org/10.4161/cc.10.16.17092

14. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell [Internet]. 2016 Jun;165(6):1416–27. Available from: http://dx.doi.org/10.1016/j.cell.2016.05.046

15. Sangha AK, Kantidakis T. The Aminoacyl-tRNA Synthetase and tRNA Expression Levels Are Deregulated in Cancer and Correlate Independently with Patient Survival. Current Issues in Molecular Biology [Internet]. 2022 Jul 2;44(7):3001–19. Available from: http://dx.doi.org/10.3390/cimb44070207