

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:



UNIVERSITY
of York

Contents

Introduction	1
FOXA1 Expression and ER α + 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	4
Why?	4
.	6
Figure 3	7
Figure ?	9
Figure 4	9
What?	9
Figure 5	11
Discussion	12
Conclusion	12
References	12

TITLE

Natasha L. Hopkins

Abstract

1 Words

Introduction

FOXA1 Expression and ER α 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 ± 500 bp 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 available 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].

FOXA1 and H3K27ac peaks of 416 high-confidence tRNAs were quantified relative to the ± 500 bp flanking regions. 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). Input reads generated minimal peak enrichment (Supplementary Figure X).

Peaks were classified as binding events if Q-values exceeded input values. FOXA1 interacted with 329 (79.1%) of tRNA genes and H3K27ac with 293 (70.4%) of tRNA genes. FOXA1 co-binds with 89.4% of H3K27ac sites. Upon FOXA1 OE, FOXA1 binding decreased to 319 tDNAs (76.7%). H3K27ac decreased to 266 (63.9%) (Figure 1C). Here, FOXA1 co-binds represents 92.5% of H3K27ac sites. Co-binding events represent 62% (262) of high confidence tDNAs, decreasing to 59.1% (246) when FOXA1 is overexpressed; these subsets differ by 28 genes.

Upon FOXA1 OE, mean Q-values significantly increased 1.18-fold for FOXA1 binding ($p < 0.0001$), and significantly decreased 0.86-fold for H3K27ac ($p < 0.01$) (Supplementary Figures X). However, FOXA1 OE leads to a significant difference in FOXA1 and H3K27ac binding between individual tDNAs ($p < 0.0001$) (Figure 1D).

Together, these results support the notion that FOXA1 overexpression alters the binding landscape of FOXA1 and H3K27ac at tRNAs.

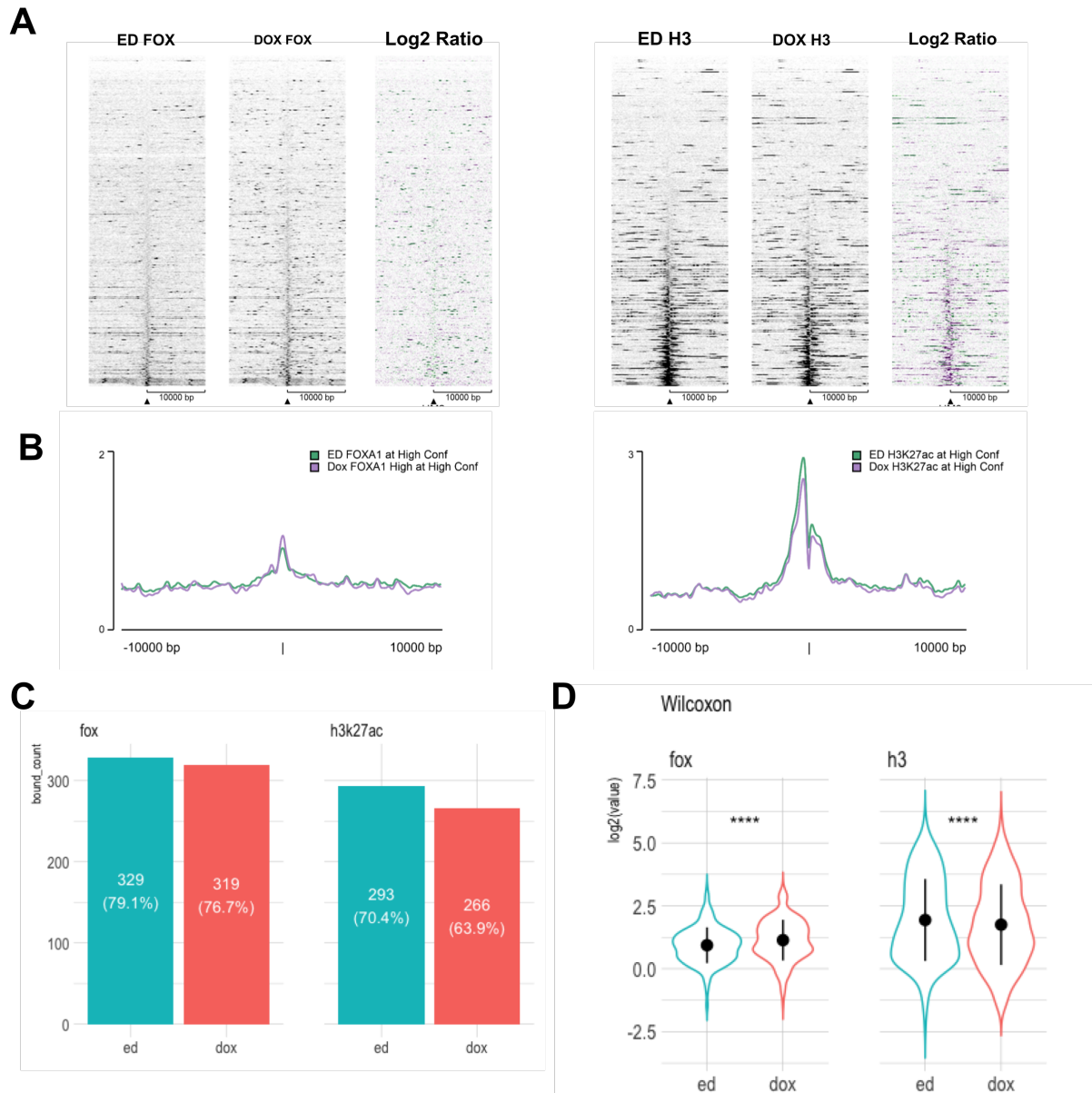


Figure 1. (B) Ratiometric heatmaps of the log2 ratio between the binding of FOXA1 or H3k27ac with endogenous FOXA1 expression vs. the binding of FOXA1 or H3k27ac with FOXA1 OE.

Figure 2

Why?

How does FOXA1 alter H3K27ac binding?

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 (92 vs. 21) (Figure 2A). However, for H3K27ac, the number of tRNAs with an increase (GAIN)

was comparable to those with a decrease (LOSS) (41 vs. 44) (Figure 2B). Of the tDNAs which GAIN H3k27ac, 51% (21) also gain (GAIN) FOXA1; none lose (LOSS) FOXA1. Of the tDNAs which lose (LOSS) H3k27ac, 22.7% (10) also lose FOXA1, with just 1 tDNA gaining FOXA1 (Figure 2C). FOXA1-H3K27ac colocalisation events represent 63% (262) of -Dox tDNAs, and 59.1% (246) of +Dox tDNAs. Examples of these tDNAs are shown in Figure 2D.

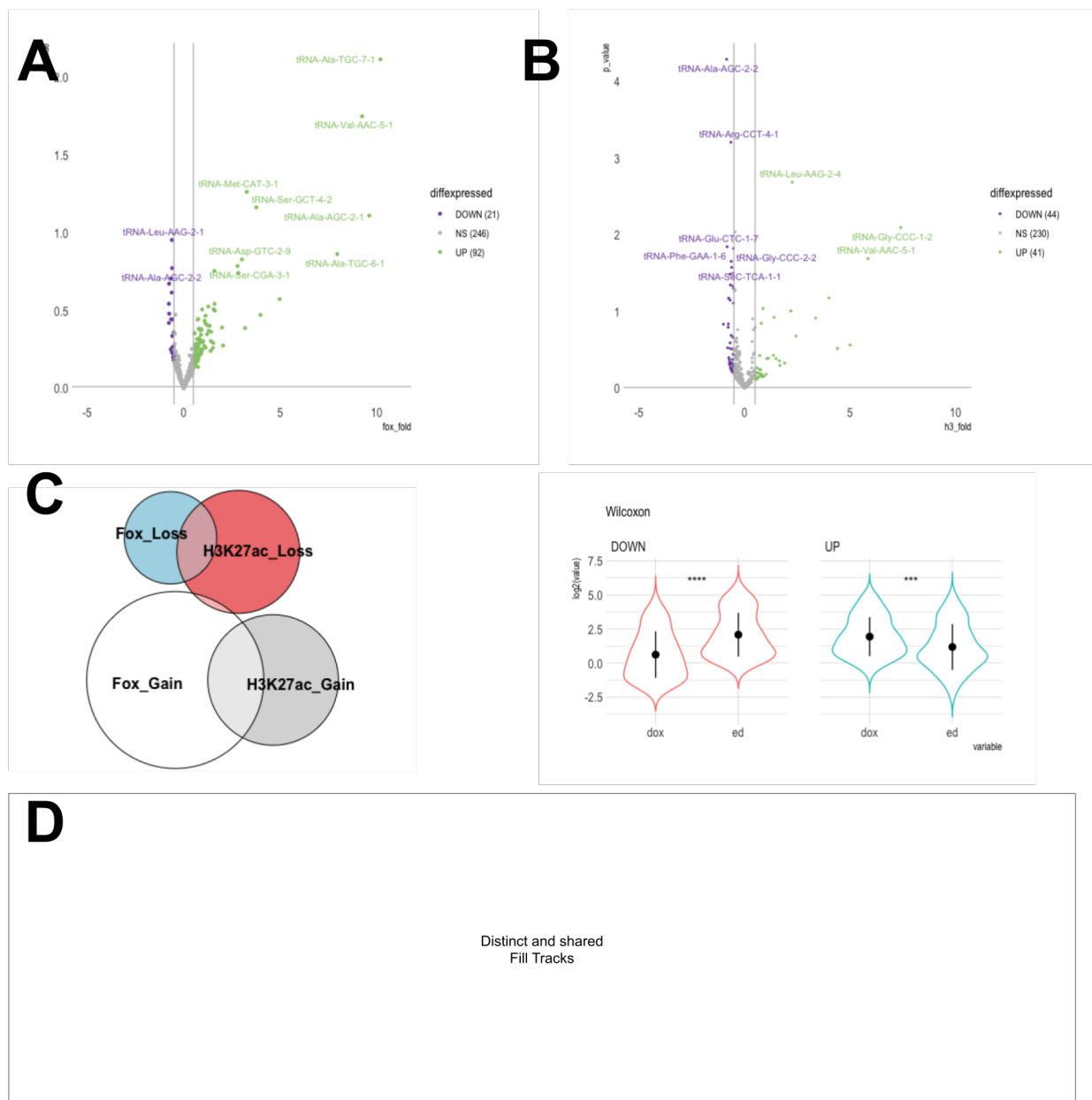


Figure 2. (B) Ratiometric heatmaps of the log₂ ratio between the binding of FOXA1 or H3k27ac with endogenous FOXA1 expression vs. the binding of FOXA1 or H3k27ac with FOXA1 OE.

Figure 3

How many genes are 'activated' by FOXA1?

What?

- 50% tDNAs active/inactive
- tRNAs were classified as 'active' if H3K27ac Q-values exceeded the median -DOX value (Q > 1.808).
- 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?
- where fox/h3 gainers, how many gain activity? how many were already ac

FOXA1 alone is insufficient in increasing tRNA activity.

Suggests?

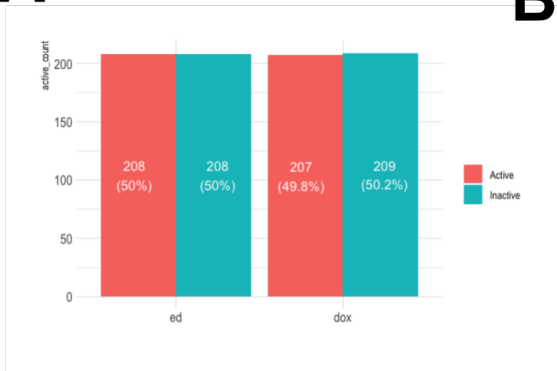
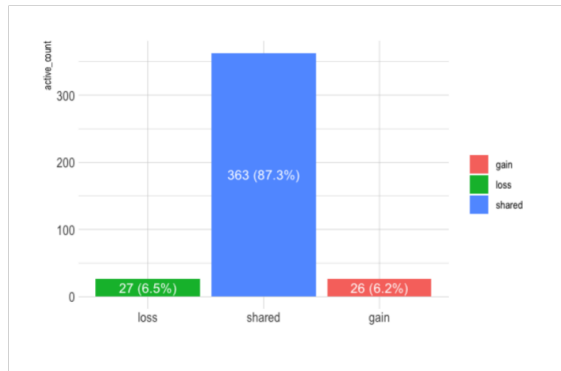
A**B**

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

1085 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>