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

TITLE

Natasha L. Hopkins

Abstract

1 Words

---

1 Introduction

1.1 FOXA1 Expression and ERα+ Breast Cancer

1.2 tRNAs and Gene Expression

---

2 Materials & Methods

2.1 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	FOXA1	MCF-7LP	GRCh37 (Hg19)
	SRR8393425			
	SRR8393426			
	SRR8393427	H3K27ac		
	SRR8393428			

Experiment	SRA	Factor	Tissue	Assembly
	SRR8393431 SRR8393432	None (input)		

2.1.1 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, data was visualised with "heatmap", "average", and "overlay" EaSeq tools. EaSeq[7] is avaiable at http://easeq.net.

2.2 Statistics

Statistical tests were carried out using R[8] (v 4.2.3), R Studio[9] (v 2023.03.0.386) and the tidyverse[10] package.

2.3 Motif Analysis

---

3 Results

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

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]

4 Discussion

4.1 Future

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

392 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* 2019; **116**: 26823–26834.

2 Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. *Nucleic Acids Research* 2010; **39**: D19–D21.

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* 2022; **50**: W345–W351.

4 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; **9**: 357–359.

5 Karolchik D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research* 2004; **32**: 493D–496.

6 Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research* 2015; **44**: D184–D189.

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* 2016; **23**: 349–357.

8 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, 2023. https://www.R-project.org/.

9 Posit Team. *RStudio: Integrated development environment for r*. Posit Software, PBC: Boston, MA, 2023. 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* 2019; **4**: 1686.

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* 2011; **31**: 330–350.

12 Sizer RE, Chahid N, Butterfield SP, Donze D, Bryant NJ, White RJ. TFIIIC-based chromatin insulators through eukaryotic evolution. *Gene* 2022; **835**: 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* 2011; **10**: 2779–2791.

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* 2016; **165**: 1416–1427.

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* 2022; **44**: 3001–3019.