Study on FSHD2 Myotube Nucleus Disease Based on Mononuclear RNA-seq Identification

Zihan Liao[1],* and Virgia Wang[2]

[1] KCS Chengdu Dipont, Sichuan, China
[2] Shanghai Jiao Tong University, Shanghai, China; Email: virgiathekite@gmail.com (V.W.)
*Correspondence: 2924337170@qq.com (Z.L.)

Abstract—**Facioscapulohumeral Muscular Dystrophy (FSHD) is a genetic disease that leads to progressively wasting of facial, shoulder, and upper arm musculature. We performed a PCA analysis on the bulk RNA-seq datasets to observe the variances in gene expression profiles that could distinguish the impacts of genotypes and myoblast differentiation. We found out that the variance of differentiation (51.9%) was much higher than the percentage of variance of genotype difference (5.9%). Corresponding to this, we also found that only a small portion of differentially expressed genes were driven by the FSHD effect only, and the vast rest were driven by differentiation effect. These results further confirmed the difficulties in detecting FSHD induced or DUX4 downstream genes from bulk RNA-seq datasets and highlighted our downstream analysis by using a single-nucleus RNA-seq dataset to dissect the pathology in disease-associated populations.**

I. INTRODUCTION

Facioscapulohumeral Muscular Dystrophy (FSHD) is a genetic disease that leads to progressively wasting of facial, shoulder, and upper arm musculature, with the earliest symptoms often appearing around the age of 20, and the prevalence of FSHD is about 1/8000 [1]. There are two types of FSHD, FSHD1 and FSHD2, with FSHD1 accounting for nearly 95% of all cases. FSHD1 and FSHD2 are caused by the misregulation of DUX4 expression, which is mediated by different genetic changes [2]. DUX4 is silenced in normal muscle cells, and its gene body is covered by the D4Z4 repeats with the length between 11 to 100 units, which cover the polyA tail and prevent transcription. In addition, DUX4 can also be silenced through the interaction with SMCHD1, which stabilizes CpG methylation over the gene body [3]. DUX4 is expressed only in the embryonic stage and it will be silenced afterward [4]. However, abnormal expression of DUX4 can cause muscle cell differentiation defects and cytotoxicity in myotubes [5]. Patients with FSHD1 have a length of D4Z4 repeats less than 10 units, which causes the DUX4 gene body to be less covered and

the poly-A tail gets exposed to be transcribed [6]. In FSHD2 patients, mutations in SMCHD1 lead to less condensed methylation status around the gene body and DUX4 gets access to the activated transcription status [2]. These two mechanisms will both lead to the transcription of DUX4, and DUX4 protein will be further translated, causing the death of muscle cells. Although DUX4 has been identified as the disease-causing gene, it is hard to directly detect its expression from patient cells since DUX4 is only expressed in 1/200 (0.5%) of myotube nuclei [7]. In addition, DUX4 is a human-specific gene, which means we are not able to study FSHD with other animal models. Symptoms of FSHD are irreversible and there's no treatment for it. Also, as DUX4 is expressed at an extremely low level, most patients are diagnosed after they have severe symptoms.

Here we focused on finding an alternative biomarker to replace DUX4 and detect FSHD at an early stage. We hypothesized that DUX4 downstream genes, meaning not only protein-coding genes but long non-coding RNA (lncRNAs) genes, have higher expression frequency and expression level than DUX4, which can be used as candidates for disease detection. LncRNAs are known to be important regulators in normal muscle development [LncRNAs are known to be important regulators in normal muscle development]. Therefore, it is reasonable to hypothesize that they make contributions to regulating the disease progression as well. In order to identify these alternative biomarkers to replace DUX4 and detect FSHD, we first used bulk RNA-seq to find out which DUX4 downstream genes changed according to the FSHD pathology. Then, we used single-cell RNA-seq to find out how these genes express in FSHD-associated populations, and in our case, we compared between 'FSHD Lo' population, which has lower expression of DUX4 downstream genes and is in day 3 differentiation and the 'FSHD Hi' population, which has higher expression of these genes and is in day 5 differentiation. We found that lncRNAs have significant enrichment in DUX4 downstream genes during the progression of FSHD pathology. Compared with DUX4 downstream protein-coding genes, these lncRNAs not only share the comparable fold change between healthy and disease patients but also have higher specificity in disease-associated nucleus populations and more enrichment in disease-associated pathways.

II. LITERATURE REVIEW

A. Data Collection

GSE143493 [8] was used for both bulk RNA-sequencing analysis and single-nucleus RNA-seq analysis.

B. Data Preprocessing

Bulk RNA-sequencing reads were mapped to hg38, and gene expression data was normalized to TPM (Transcripts per million mapped reads). Single-nucleus RNA-seq reads were also aligned to hg38 and cell clusters have been identified by using the published results GSE143493 [8].

C. Principal Component Analysis (PCA)

Bulk RNA-seq expression data (TPM) was quantile normalized by using limma package [9] and log transformed. Then the PCA analysis was performed on the matrix by using prcomp function in R with "center=TRUE, scale=FALSE".

D. Time-Course Differential Analysis

The differential gene expression analysis between control and FSHD2 patients across the differentiation time-course was performed by maSigPro [10] with the parameters "rsq=0.8, T.fit alpha=0.05". maSigpro identified expression modules, in which genes with similar expression profiles were clustered together. To identify those highly correlated genes, we used CEMiTool [11] to get the co-expressed genes for each expression module identified by maSigPro.

E. Gene Ontology Analysis

Gene ontology analysis was done by using Metascape [12] with input species. Enriched terms were selected with BH-corrected p-value < 0.05 based on hypergeometric test.

F. lncRNA Functional Annotation

Guilty by association analysis was performed to associate lncRNAs with their nearby protein-coding genes. Then the nearest protein-coding genes were input for gene ontology analysis, which would be representative of the functional annotation of lncRNAs.

G. Single-Nucleus RNA-seq Analysis

A volcano plot was performed by using ggplot2 [13] to show the differential genes between 'FSHD Hi' and 'FSHD Lo' single nucleus populations. Heatmap was performed by using heatmap2 [14] to show the average expression of DUX4 downstream genes, protein-coding, and lncRNAs, in 'FSHD Hi' and 'FSHD Lo' single nucleus populations. The proportion of nuclei that expressed each DUX4 downstream gene was summarized by using a boxplot, done by ggplot2 [13], to observe the expression frequency between genotypes in 'FSHD Hi' and 'FSHD Lo' single nucleus populations.

2.8 Co-expression of DUX4 downstream genes in disease-associated nucleus populations

DUX4 downstream genes were coexpressed in the same nucleus. These coexpression relationships included protein-protein genes, protein-lncRNA genes, and lncRNA-lncRNA genes co-expressed together. We observed these co-expression patterns and compared them between the 'FSHD Hi' and 'FSHD Lo' single nucleus populations. In Fig. 8, the x-axis represents the number of lncRNA or protein-coding genes and the y-axis represents the number of cells that do the co-expression. In Fig. 9, the x-axis represents the number of lncRNA and the y-axis represents the number of core set protein-coding genes that do the co-expression.

III. RESULTS

A. Enrichment of lncRNAs in DUX4 Downstream Genes

GSE143493 [8] was used for both bulk RNA-sequencing analysis and single-nucleus RNA-seq analysis

In this study, we sought to comprehensively understand the molecular mechanisms underlying Facioscapulohumeral Muscular Dystrophy (FSHD), a complex neuromuscular disorder. Specifically, we aimed to explore the enrichment of long non-coding RNAs (lncRNAs) in the downstream genes regulated by DUX4, the identified disease-associated gene, shedding light on potential alternative biomarkers and contributing to early FSHD detection.
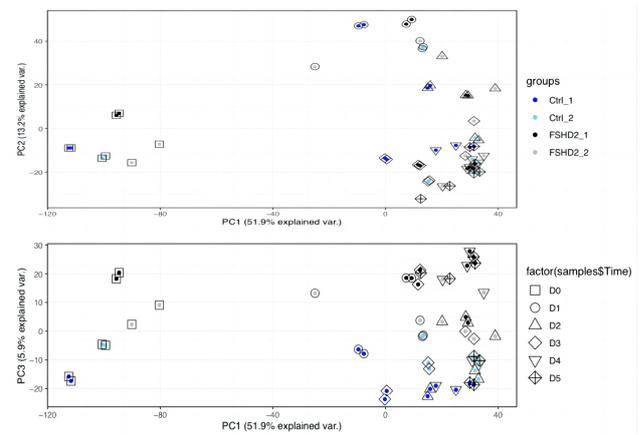

Figure 1. Principal component analysis on samples with different genotypes and across differentiation time-course. Genotype is denoted by color; time-course is denoted by shape.

In order to identify genes that were activated during FSHD pathology by DUX4, we performed differential analysis on the published bulk RNA-seq data [8]. We first performed the Principal Component Analysis (PCA) to observe variances among samples, which were contributed from either muscle differentiation time-course or FSHD-related pathologies. We found that PC1 took about 51.9% of the total variances and it represented the effects of muscle differentiation time-course for both control and FSHD patients (Fig. 1, top panel). We also found that PC3, taking about 5.9% of total variances, clearly separated samples based on their genotype differences (Fig. 1, bottom panel). Our results revealed that the samples had distinct expression profiles based on their genotypes or differentiation status; however, differentiation time-course contributed significantly more

variances to the sample differences compared with FSHD pathologies, indicating DUX4 associated or downstream genes might take only a small set in the differentially expressed genes.

Then, we performed a time-course differential gene expression analysis to identify genes changing according to the muscle differentiation and FSHD pathological progression. We identified 5511 differentially expressed genes by using maSigPro [10] and k-means clustering was further performed on them to group genes with similar expression profiles into six clusters. For example, 1260 genes in cluster 1 were down-regulated along with the differentiation while 1549 genes in cluster 5 were up-regulated, both of which reflected the effects of muscle cell differentiation. Genes in these two clusters were significantly involved in metabolism and muscle development (Fig. 2). For another example, 276 genes in cluster 6 were up-regulated only in FSHD patients but kept constant in control samples, reflecting the effects of FSHD pathologies. These genes were significantly associated with negative regulation of cell differentiation and cell destruction development (Fig. 2). Note that these 276 FSHD-associated genes only accounted for about 5% of total differentially expressed genes (276/5511 = 0.05), corresponding to the limited contributions of FSHD pathology in sample variances (Fig. 1). Although genes were grouped together based on their similar expression profiles, their relationships associated with each other were still unknown. We conducted co-expression analysis on each gene cluster and found that 57 protein-coding genes were highly co-expressed in cluster 6, driven by the progression of FSHD pathology. We treated them as the core set of DUX4 downstream protein-coding genes, and further investigated their relationships with lncRNAs at the single-nucleus level.

We then looked at the enrichment of lncRNAs in these clusters driven by either the differentiation effect or the FSHD effect. We found that the proportion of lncRNA in cluster 5 and cluster 6, in which genes were up-regulated along with differentiation, is much higher than the one in cluster 1, in which genes were down-regulated along with differentiation (Table I). In addition, lncRNA enrichment was higher in cluster 6 than in cluster 5 (Table I), indicating that lncRNAs were highly enriched in FSHD downstream genes and this enrichment became more significant with increasing gene expression during muscle differentiation.

TABLE I. COMPARISON OF LNCRNAS ENRICHMENT IN GENE EXPRESSION MODULES AFFECTED BY MUSCLE DIFFERENTIAL AND FSHD PATHOLOGY.

Cluster	Effects	Protein-coding	Antisense	lncRNA	Ratio (%)
1	Differentiation	1224	13	14	2.21
5	Differentiation	1432	39	56	6.63
6	FSHD	254	3	14	6.69

B. Advantages of lncRNAs in FSHD-Associated Cell Population

In this section, we explore the advantages of utilizing long non-coding RNAs (lncRNAs) as potential biomarkers in FSHD-associated cell populations. While protein-coding genes have long been the focus of genetic studies, recent research has highlighted the regulatory roles of lncRNAs in various biological processes, including disease progression. Our investigation into the lncRNA landscape within FSHD-associated cell populations sheds light on their potential as robust diagnostic indicators.

We have identified a core set of protein-coding genes that were up-regulated with the progression of FSHD pathology and found that lncRNAs were significantly enriched in the DUX4 downstream genes. In order to understand the performance of these lncRNAs and their advantages as biomarkers compared to protein-coding genes, we continued focusing on the previously identified FSHD-associated cell populations [8] and observing the expression dynamics of lncRNAs in them. FSHD-associated cells were classified into 'FSHD-Hi' and 'FSHD-Lo' populations based on the expression of the core set DUX4 downstream protein-coding genes and stage of muscle differentiation. Specifically, 'FSHD-Hi' cells had higher expression levels and frequency of the DUX4 downstream genes and they were prone to the late stage of muscle differentiation compared with 'FSHD-Lo' cells [8].

Firstly, in order to understand whether DUX4 downstream genes identified in time-course differentiation (Figs. 1 and 2) were also differentially expressed between 'FSHD-Hi' and 'FSHD-Lo' cells, we performed an intersected analysis on the differential genes between bulk RNA-seq and single-nucleus RNA-seq data (Fig. 3). We detected more differentially expressed genes in the single-nucleus data, and this was according to the disease-associated populations
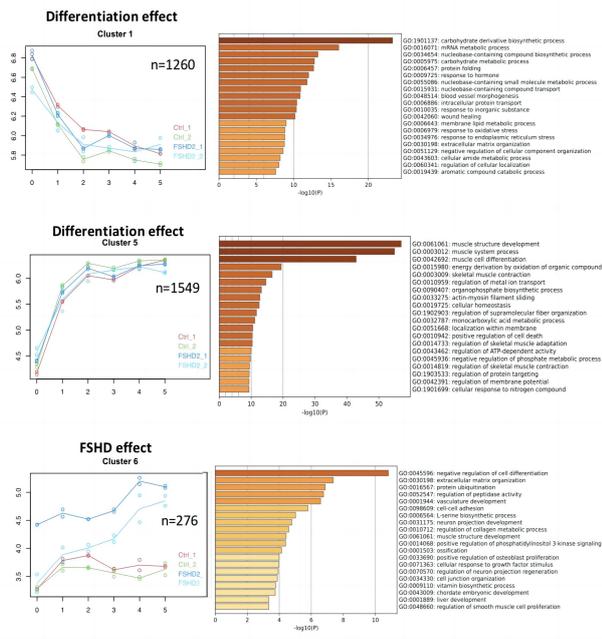

Figure 2. Time-course differential gene expression analysis between control and FSHD samples. Genotype is denoted by color, and gene ontology terms are listed.

comparison versus FSHD time-course data comparison which mainly focused on the difference between healthy and FSHD patients. However, we noticed that those 57 core set DUX4 downstream genes were not only up-regulated in FSHD pathology (cluster 6 in Fig. 2) but also had higher expression in 'FSHD-Hi' than 'FSHD-Lo' cells, indicating the essential role of these co-expression protein-coding genes in mediating DUX4 signals, and distinguish FSHD patients from the healthy. In addition, these genes could also have continuous co-expression patterns since they differed between early and late disease progression. Interestingly, we observed that the intersection of lncRNAs between bulk and single-nucleus RNA-seq data was much lower compared with protein-coding genes, indicating that lncRNAs were more specific in terms of the purpose of detection. For example, the lncRNAs to distinguish FSHD patients from the healthy should be a different set compared with the lncRNAs to distinguish between early and late FSHD patients.
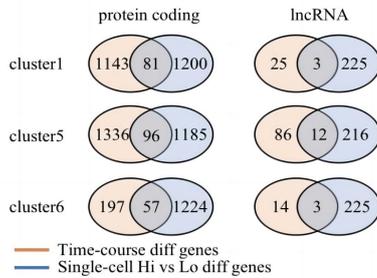

Figure 3. Venn diagram of the number of protein-coding genes and lncRNAs that are essential in time-course differentiation and single-cell Hi vs Lo differentiation in FSHD affected gene population and time-course related gene populations.

Then we looked at the specificity of lncRNAs and protein-coding genes in 'FSHD-Hi', 'FSHD-Lo' and control cells by comparing their expression at the single-nucleus level.

We found that almost all of the lncRNAs were highly expressed only on day 5 and in FSHD2 patients (Fig. 4, left panel). Nevertheless, about half of the protein-coding genes were expressed in day 3 FSHD patients or day 5 controls and the difference between FSHD2 and controls was not significant (Fig. 4, right panel). Our results indicate that the specificity of lncRNA is higher than protein-coding genes in FSHD-associated cells with late-stage pathologies.
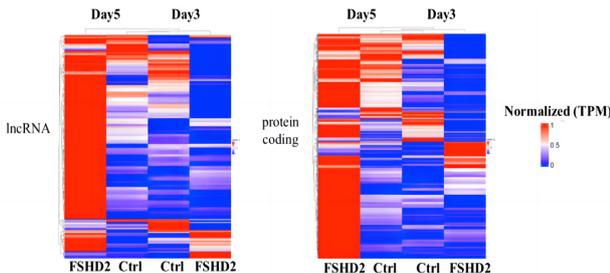

Figure 4. Heatmap of the expression level of the protein coding genes and lncRNA in FSHD Hi control and FSHD samples, and FSHD Lo control and FSHD samples.

Next, we observed the fold change and p-value of lncRNAs and protein-coding genes in 'FSHD-Hi' and 'FSHD-Lo' single nucleus comparison in order to find out whether the fold change or significant level of lncRNAs are comparable to protein-coding genes. We found that both the fold change and p-value were comparable between lncRNA and protein-coding genes (Fig. 5). However, over 95% (228/239) of differentially expressed lncRNAs were up-regulated in 'FSHD-Hi' cells while about 92% (1281/1381) of those differentially expressed protein-coding genes were up-regulated in 'FSHD Hi' cells (Fig. 5). These results indicate that the function of lncRNAs and protein-coding genes as the biomarker were comparable.
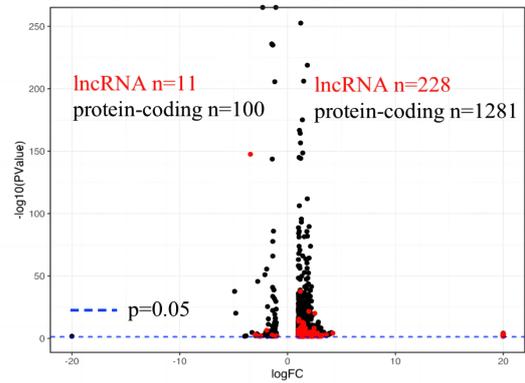

Figure 5. Volcano plot of the significant level and fold change of lncRNAs and protein-coding genes that are essential in FSHD Hi vs Lo differentiation.

At last, we observed the biological pathways associated with these differential lncRNAs and protein-coding genes between 'FSHD-Hi' and 'FSHD-Lo' cells. For lncRNAs, we hypothesized that the function of lncRNAs has the same function as their nearby protein-coding genes as they may be transcribed by those protein-coding genes. By finding the nearby protein-coding genes of those lncRNAs (guilty by association analysis), we found that lncRNAs were significantly enriched in regulating muscle differentiation and negative regulation of the metabolic processes (Fig. 6, upper panel). However, gene ontology on protein-coding genes showed that they were related to basic cell cycle and cell development processes (Fig. 6, bottom panel). Based on these we conclude that the functions of lncRNAs are more specific to muscle differentiation compared to protein-coding genes and they have a strong relationship with the progression of the disease as their several GO terms were associated with muscle development, cell adhesion, and immune response, suggesting the multifaceted impact of these co-expression networks on various aspects of FSHD pathology.
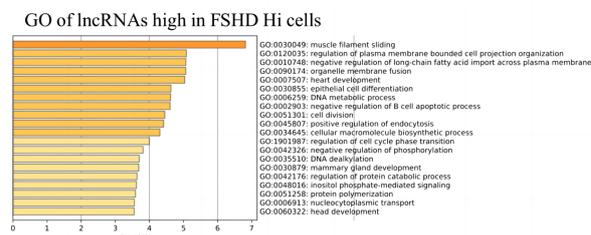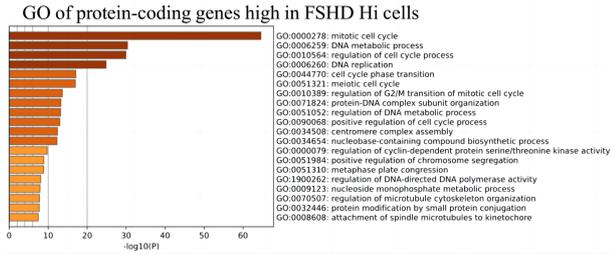
Figure 6. Genotype is denoted by color, and gene ontology terms of the highly expressed protein-coding genes in FSHD Hi and lncRNAs' nearby protein-coding genes are listed.

C. Co-expression of lncRNAs and Protein-Coding Genes in Disease-Associated Cells

In this section, we delve into the intricate landscape of gene expression within disease-associated cell populations. By employing advanced techniques, we sought to decipher the co-expression patterns of long non-coding RNAs (lncRNAs) and protein-coding genes, shedding light on their potential roles in the pathogenesis of FSHD. This investigation holds significant promise, as lncRNAs have emerged as pivotal players in gene regulation and disease progression.

To delve deeper into the functional implications of gene expression alterations, we conducted co-expression analyses on both FSHD-induced protein-coding genes and long non-coding RNA (lncRNA) genes. These analyses provided a comprehensive view of how genes within the same functional pathways or networks might respond to FSHD-associated perturbations.

Previous studies have shown that FSHD-induced protein-coding genes had low expression frequency in tissue, in which each gene was only expressed in a small proportion of cells/nuclei.

Furthermore, they rarely coexpressed within the same nucleus and they tended to express sparsely in different nuclei within the same myotube cell [15].

In order to find out whether the expression patterns of FSHD-induced lncRNAs are comparable with the ones of protein-coding genes, we first analyzed single-nucleus RNA-seq data and observed the expression frequency of those DUX4 downstream candidates in "FSHD-Hi" and "FSHD-Lo" populations (Fig. 7). As expected, the proportion of nuclei expressed in each of those 57 core set DUX4 downstream protein-coding genes was significantly higher in "FSHD-Hi" (Day5 FSHD2) compared with "FSHD-Lo" (Day3 FSHD2) population ($p < 2.2e{-}16$, t-test). In addition, this high expression frequency was only shown in FSHD genotype and disappeared in Day5 controls ($p = 5.181e{-}09$, t-test) (Fig. 7; left panel). We then observed the similar expression patterns for each of the FSHD-induced lncRNAs, which also showed the highest expression frequency in "FSHD-Hi" population compared with "FSHD-Lo" or controls (Fig. 7; right panel). Although a higher proportion of "FSHD-Hi" nuclei expressed DUX4 downstream genes, they only expressed in around 1% of nuclei by looking at each of them no matter if the candidate was protein-coding or lncRNA genes (Fig. 7).
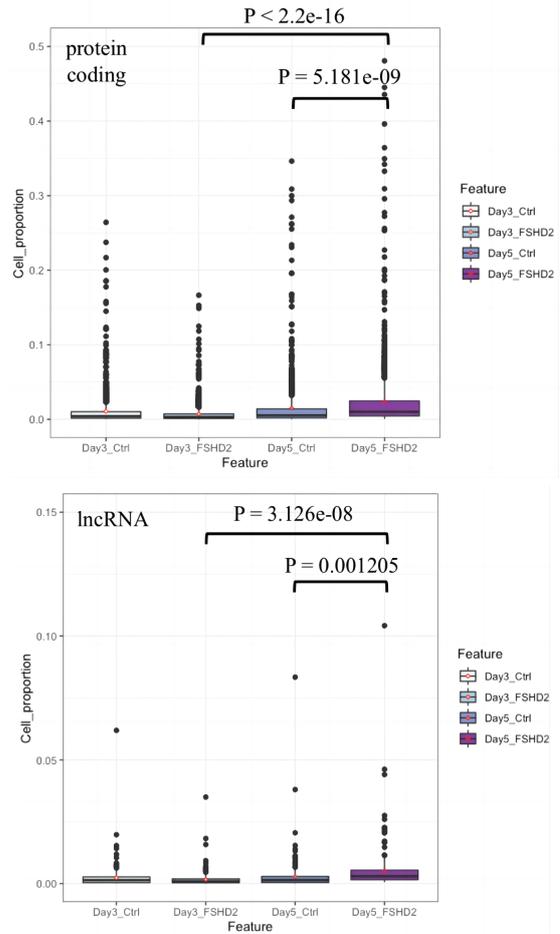

Figure 7. Boxplot of the frequency of cell expression of FSHD induced protein coding genes and lncRNAs in FSHD Hi control and FSHD samples, and FSHD Lo control and FSHD samples.

We next looked at the co-expression patterns of these FSHD-induced protein-coding and lncRNA genes, and tried to understand whether they were comparable. We also looked at the co-expression between protein-coding and lncRNAs genes in "FSHD-Hi" and "FSHD-Lo" populations and tried to understand whether "FSHD-Hi" had more co-expression frequency than "FSHD-Lo" population. In order to find out whether the co-expression of lncRNAs can function as well as the 57 core set protein-coding genes, we used a gene co-expression analysis to find out the number of genes related to FSHD co-expressed with how many lncRNA or protein-coding genes respectively (Fig. 8). By observing the relationship between the number of nuclei and the number of co-expressed lncRNA (Fig. 8; left panel) and core set protein-coding genes (Fig. 8; right panel), we found that the co-expression patterns of lncRNA and protein-coding genes were similar. Except for "FSHD-Hi" (FSHD in day5), most nuclei only expressed one lncRNA or protein-coding genes once. The co-expression levels of "FSHD-Hi" (FSHD on day5) were much higher compared with nuclei in "FSHD-Lo" (FSHD on day3) or in controls. We concluded that both protein-coding and lncRNA genes shared a low co-expression level, but "FSHD-Hi" (FSHD on day5) nuclei had a higher co-

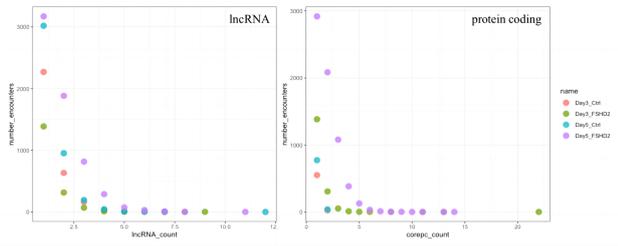expression frequency compared with the ones of other genotypes.

Figure 8. Co-expression analysis between the number of lncRNA, FSHD induced protein-coding genes and the number of nuclei in FSHD Hi control and FSHD samples, and FSHD Lo control and FSHD samples.

We then looked at the co-expression between these 57 core set protein-coding and lncRNA genes and compared them between "FSHD-Hi" and "FSHD-Lo" populations. We observed the relationship between the number of expressed core set protein-coding genes and the number of expressed lncRNAs, and we found that the level of co-expression increased from "FSHD-Lo" (FSHD in day3) to "FSHD-Hi" (FSHD in day5). Specifically, most core set protein-coding genes only co-expressed with one lncRNA in "FSHD-Lo" (Fig. 9; left panel) while more than one lncRNA co-expressed with more than one core set protein-coding gene in "FSHD-Hi" (Fig. 9; right panel). We concluded that FSHD-induced lncRNAs had similar expression frequency and co-expression patterns as the ones of core set protein-coding genes. In addition, "FSHD-Hi" nuclei had significantly higher co-expression levels between DUX4 downstream lncRNA and protein-coding genes.
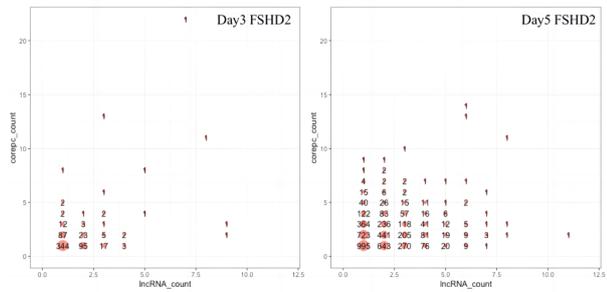

Figure 9. Co-expression analysis between FSHD induced protein coding genes and lncRNAs in FSHD samples in FSHD Hi and FSHD Lo.

IV. DISCUSSION

We performed a PCA analysis on the bulk RNA-seq datasets to observe the variances in gene expression profiles that could distinguish the impacts of genotypes and myoblast differentiation. We found out that the variance of differentiation (51.9%) was much higher than the percentage of variance of genotype difference (5.9%). Corresponding to this, we also found that only a small portion of differentially expressed genes was driven by FSHD effect only and the vast rest was driven by differentiation effect. These results further confirmed the difficulties in detecting FSHD induced or DUX4 downstream genes from bulk RNA-seq datasets and

highlighted our downstream analysis by using a single-nucleus RNA-seq dataset to dissect the pathology in disease-associated populations.

In addition, we successfully detected FSHD-induced lncRNAs, which could be further developed into candidates for replacing DUX4 genes in disease diagnosis. In time-course analysis, the expression level of these lncRNAs was significantly increased along with the myoblast differentiation and they were poorly enriched in gene clusters that had decreased expression along with the differentiation. Importantly, they were mostly enriched in the FSHD-specific cluster, in which genes were not only expressed increasingly with differentiation but also only in FSHD genotypes. Further in the single-nucleus analysis, compared with the FSHD-induced protein-coding genes, these lncRNA have some advantages, including higher specificity in FSHD-associated cell populations and more association with muscle development and immune response. Given these solid results, we look forward to performing in vitro / in vivo experiments by knocking out these lncRNA candidates and observing their effects on myoblast differentiation or dissection of disease populations in FSHD.

We also performed a co-expression analysis on both FSHD-induced protein-coding and lncRNA genes and compared their expression patterns between "FSHD-Hi" and "FSHD-Lo" populations. Although the levels of co-expression or expression frequency of these genes were significantly higher in nuclei with more severe FSHD pathology, they were rarely expressed within the same nucleus or in the same cell. This gives the challenges when selecting the panel for genes for cell selection since they may pick up different subsets of cells with different disease statuses. Future studies should focus on understanding how these FSHD-induced genes work collaboratively to mediate FSHD pathology. Furthermore, experiments should be performed to understand which markers were expressed within the same myotube cell even in different nuclei, as the proteins or lncRNA molecules may be transported/fused between nuclei.

V. CONCLUSION

In this study, we aimed to identify alternative biomarkers for early diagnosis of Facioscapulohumeral Muscular Dystrophy (FSHD) by focusing on genes downstream of the disease-causing gene DUX4, particularly long non-coding RNAs (lncRNAs). Through a comprehensive analysis of bulk RNA-seq and single-nucleus RNA-seq data, we gained insights into the expression patterns and functions of these genes in FSHD pathology. Our findings indicate that both protein-coding genes and lncRNAs downstream of DUX4 were significantly upregulated in FSHD-Hi individuals compared to healthy controls (FSHD-Lo), suggesting their potential involvement in disease progression. Notably, lncRNAs exhibited distinct advantages over protein-coding genes as potential biomarkers for FSHD diagnosis. They demonstrated higher specificity for disease-associated cell populations and exhibited closer

associations with pathways related to muscle development and immune response. Moreover, co-expression analysis revealed low levels of co-expression between lncRNAs and protein-coding genes within the same nucleus or cell. However, in FSHD-Hi individuals with more severe pathology, the co-expression levels were higher, suggesting a potential collaborative role in mediating FSHD pathology. These findings contribute to our understanding of FSHD and provide valuable insights for the development of diagnostic strategies and potential therapeutic interventions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Zihan Liao conducted the research; Zihan Liao analyzed the data; Zihan Liao and Virgia Wang wrote the paper; all authors had approved the final version.

REFERENCES

[1] R. Tawil and S. M. V. D. Maarel, "Facioscapulohumeral muscular dystrophy," *Muscle Nerve*, vol. 34, no. 1, pp. 1–15, 2006. pmid:16508966

[2] R. J. L. F. Lemmers, R. Tawil, L. M. Petek, *et al.*, "Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2," *Nature Genetics*, vol. 44, no. 12, pp. 1370–1374, 2012.

[3] N. Jansz, K. Chen, J. M. Murphy, *et al.*, "The epigenetic regulator SMCHD1 in development and disease," *Trends in Genetics*, vol. 33, pp. 233–43, 2017.

[4] L. Snider, L. N. Geng, R. J. Lemmers, *et al.*, "Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene," *PLoS Genet.*, vol. 6, no. 10, e1001181, 2010.

[5] D. Bosnakovski, Z. Xu, E. J. Gang, *et al.*, "An isogenic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies," *EMBO J.*, vol. 27, no. 20, pp. 2766–2779, 2008. pmid:18833193

[6] R. J. Lemmers, P. J. V. D. Vliet, R. Klooster, *et al.*, "A unifying genetic model for facioscapulohumeral muscular dystrophy," *Science*, vol. 329, pp. 1650–1653, 2010.

[7] A. Tassin, D. Laoudj-Chenivesse, C. Vanderplanck, *et al.*, "DUX4 expression in FSHD muscle cells: How could such a rare protein cause a myopathy?" *J. Cell. Mol. Med.*, vol. 17, no. 1, pp. 76–89, 2013. pmid:23206257

[8] S. Jiang, K. Williams, X. Kong, *et al.*, "Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei," *PLoS Genetics*, vol. 16, no. 5, e1008754, 2020.

[9] M. E. Ritchie, B. Phipson, D. Wu, *et al.*, "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, e47, 2015.

[10] A. Conesa and M. J. Nueda, "maSigPro: Significant gene expression profile differences in time course gene expression data," R package version 1.68.0, 2022.

[11] P. S. T. Russo, G. R. Ferreira, L. E. Cardozo, *et al.*, "CEMiTool: A bioconductor package for performing comprehensive modular co-expression analyses," *BMC Bioinformatics*, vol. 19, no. 56, pp. 1–13, 2018.

[12] Y. Zhou, B. Zhou, L. Pache, *et al.*, "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, pp. 1–10, 2019.

[13] H. Wickham, W. Chang, and M. H. Wickham, "Package 'ggplot2'," *Create Elegant Data Visualizations Using the Grammar of Graphics*, version 2.1, pp. 1–189, 2016.

[14] M. Perry, "Heatmaps: Flexible heatmaps for functional genomics and sequence features," R package version 1.20.0, 2022.

[15] T. I. Jones, J. C. J. Chen, F. Rahimov, *et al.*, "Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis," *Hum. Mol. Genet.*, vol. 21, no. 20, pp. 4419–4430, 2012.