Pan-Cancer Analysis Reveals Expressional Correlations between Forkhead Box Family Members

Bokai Zhang*
22520 Mount Michael Road, Elkhorn, Nebraska, 68022, USA.

* Corresponding author. Tel.: +86 13661088966; email: Kevin120441@163.com
Manuscript submitted July, 2018; accepted December 10, 2018.
doi: 10.17706/ijbbb.2019.9.2.118-125

Abstract: The forkhead box (FOX) transcription factor protein family is responsible for a wide range of biological activities, especially in development and cell differentiation. The highly conserved gene family span from worms to mammals, and has at least 41 members in the human genome. By analyzing data from The Cancer Genome Atlas (TCGA), this study examined the correlation of expression between each human FOX members in 31 types of cancer cells in the form of heatmaps and scatterplots. The primary goal was to identify significant correlations between certain FOX family members and different types of cancers. The study identified a close expressional correlation between FOXC2 and FOXL1 genes, which exists within a cluster at 16q24.1. Other significant relations in particular types of tumor tissues were also noted.

Key words: Human FOX gene family, correlation, cancer, TCGA.

1. Introduction

The forkhead box (FOX) transcription factor family can be identified by their common DNA-binding domain named “forkhead box” or “winged helix”. Despite this common feature, members of the FOX family serve distinct biological functions. These members have been found to be responsible for functions such as cell differentiation, apoptosis and cell cycle regulation. Therefore, unregulated expression of FOX family members inevitably lead to the possibility of oncogenesis [1].

The amplification, point mutation, or translocation of the FOX family members is often observed in numerous types of cancers. Along with more sequenced whole genome data of cancers, more and more researchers have explored the role of FOX abnormalities in carcinogenesis [2].

In 2004, Katoh et al found gene clusters of the FOX family, which include FOXE3-FOXD2 locus, FOXQ1-FOXF2-FOXC1 locus, and FOXF1-FOXC2-FOXL1 locus. These clusters may indicate relationships in terms of expression in various cancer cells [3].

Other relationships may also exist outside the structural clusters. In light of the above discoveries, this study aimed to identify significant correlations of expression between FOX genes in cancer cells.

2. Methods

Forty-one known human FOX gene family members’ (Table 1)expressional level (RSEM, RNA-Seq by Expectation-Maximization, value) data from 31 types of cancer (Table 2) were acquired from The Cancer Genome Atlas (TCGA) (cancergenome.nih.gov) provisional study via cBioPortal (www.cbioportal.org) [4], [5].

Table 1. Tested FOX Gene Family Members

Symbol	Name	GeneID	Cytoband	UniProt
FOXA1	forkhead box A1	3169	14q21.1	P55317
FOXA2	forkhead box A2	3170	20p11	Q9Y261|B0ZTD4
FOXA3	forkhead box A3	3171	19q13.32	P55318|A0A024R0R3
FOXB1	forkhead box B1	27023	15q22.2	Q99853
FOXB2	forkhead box B2	442425	9q21.2	Q5VYV0
FOXC1	forkhead box C1	2296	6p25	Q12948|W6CJ52
FOXC2	forkhead box C2	2303	16q24.1	Q99958
FOXD1	forkhead box D1	2297	5q13.2	Q16676
FOXD2	forkhead box D2	2306	1p34-p32	O60548
FOXD3	forkhead box D3	27022	1p31.3	Q9UJU5
FOXD4	forkhead box D4	2298	9p24.3	Q12950
FOXE1	forkhead box E1	2304	9q22	O00358
FOXE3	forkhead box E3	2301	1p32	Q13461|A0A0A1EII5
FOXF1	forkhead box F1	2294	16q24	Q12946
FOXF2	forkhead box F2	2295	6p25.3	Q12947
FOXG1	forkhead box G1	2290/2291	14q13	P55316
FOXH1	forkhead box H1	8928	8q24.3	O75593
FOXI1	forkhead box I1	2299	5q34	Q12951|E0XEN6
FOXJ1	forkhead box J1	2302	17q25.1	Q92949|A0A024R8P1
FOXJ2	forkhead box J2	55810	12p13.31	Q9P0K8
FOXJ3	forkhead box J3	22887	1p34.2	Q9UPW0
FOXK1	forkhead box K1	221937	7p22.1	P85037
FOXK2	forkhead box K2	3607	17q25	Q01167
FOXL1	forkhead box L1	2300	16q24	Q12952|Q498Y4
FOXL2	forkhead box L2	668	3q23	P58012|Q53ZD3
FOXM1	forkhead box M1	2305	12p13	Q08050|Q53Y49|A8K591
FOXN1	forkhead box N1	8456	17q11.2	O15353
FOXN2	forkhead box N2	3344	2p22-p16	P32314
FOXN3	forkhead box N3	1112	14q31.3	O00409|A0A024R6I1
FOXN4	forkhead box N4	121643	12q24.11	Q96NZ1|A6H901
FOXO1	forkhead box O1	2308	13q14.1	Q12778
FOXO3	forkhead box O3	2309	6q21	O43524
FOXO4	forkhead box O4	4303	Xq13.1	P98177
FOXP1	forkhead box P1	27086	3p14.1	Q9H334|Q548T7|A0A0B4J2F3|Q8TEA2|A0A087X299|E9PFD3|Q8N2P0
FOXP2	forkhead box P2	93986	7q31	O15409|Q8N6B5|X5D2H2|B7ZLK5|Q8N6B6
FOXP3	forkhead box P3	50943	Xp11.23	Q9BZS1|B7ZLG1
FOXP4	forkhead box P4	116113	6p21.1	Q8IVH2
FOXQ1	forkhead box Q1	94234	6p25	Q9C009
FOXR1	forkhead box R1	283150	11q23.3	Q6PIV2
FOXR2	forkhead box R2	139628	Xp11.21	Q6PJQ5
FOXS1	forkhead box S1	2307	20q11.21	O43638

The specific list of researched genes and cancer tissues is shown in Table 2. The Spearman's rank correlation coefficient (ρ value) between each two members of the FOX family in every type of cancer was calculated, and visualized in forms of heatmaps and scatterplots. To further identify significant relationships between FOX genes, the cases for the top five Spearman's rank correlation coefficient (ρ >0.5)

were listed. Significant anti-correlation was also identified in the study.

3. Results

3.1. Identification of Expressional Correlation between FOX Members

The study produced 31 heatmaps displaying the spearman correlation between each pairs of FOX members in every studied TCGA sample groups. As an example, Fig. 1 is the heatmap for testicular germ cell cancer samples, which display the most significant correlations among all samples. When looking at individual gene pairs, the correlation coefficient between FOXC2 and FOXL1 was greater than 0.5 in 22 of the 31 cases, showing the most significant relationship. Three other pairs are also notable: FOXF1-FOXF2, FOXF2-FOXL1,

Table 2. Used TCGA Sequencing Samples

Project Name	Number of samples	Disease full name
ACC	79	adrenocortical carcinoma
BLCA	408	bladder urothelial carcinoma
BRCA	1100	breast invasive carcinoma
CESC	306	cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	36	cholangiocarcinoma
COADREAD	382	colorectal adenocarcinoma
DLBC	48	lymphoid neoplasm diffuse large B-cell lymphoma
ESCA	185	esophageal carcinoma
GBM	166	glioblastoma multiforme
HNSC	522	head and neck squamous cell carcinoma
KICH	66	kidney chromophobe
KIRC	534	kidney renal clear cell carcinoma
KIRP	291	kidney renal papillary cell carcinoma
LAML	173	acute myeloid leukemia
LGG	530	brain lower grade glioma
LIHC	373	liver hepatocellular carcinoma
LUAD	517	lung adenocarcinoma
LUSC	501	lung squamous cell carcinoma
OV	307	ovarian serous cystadenocarcinoma
PAAD	179	pancreatic adenocarcinoma
PCPG	184	pheochromocytoma and paraganglioma
PRAD	498	prostate adenocarcinoma
SARC	263	sarcoma
SKCM	472	skin cutaneous melanoma
STAD	415	stomach adenocarcinoma
TGCT	156	testicular germ cell tumors
THCA	509	lung squamous cell carcinoma
THYM	120	thymoma
UCEC	177	uterine corpus endometrial carcinoma
UCS	57	uterine carcinosarcoma
UVM	80	uveal melanoma

FOXF2-FOXS1, each displaying a coefficient greater than 0.5 in 10 of the 31 cases. The overall top five expression correlations are identified here as well, which include: FOXF1-FOXF2 in colorectal

adenocarcinoma (ρ≈0.837), FOXJ1-FOXP4 in testicular germ cell cancer (ρ≈0.826), FOXN1-FOXE1 in esophageal carcinoma (ρ≈0.811),FOXC2-FOXF2 in testicular germ cell cancer (ρ≈0.806) and FOXC1-FOXD1 in uveal melanoma (ρ≈0.799). In the clustered gene member pairs, particularly, FOXC2-FOXL1 displayed an overall significant relationship among various cancers (Fig. 2). Meanwhile, 5 other pairs showing negative relationship were identified. FOXN2-FOXP4 in testicular germ cell tumors (ρ≈-0.736), FOXJ1-FOXN2 in testicular germ cell tumors (ρ≈-0.715), FOXA3-FOXE1 in esophageal carcinoma (ρ≈-0.700), FOXM1-FOXO1 in thymoma (ρ≈-0.694) and FOXN2 -FOXP4 in uveal melanoma (ρ≈-0.687).

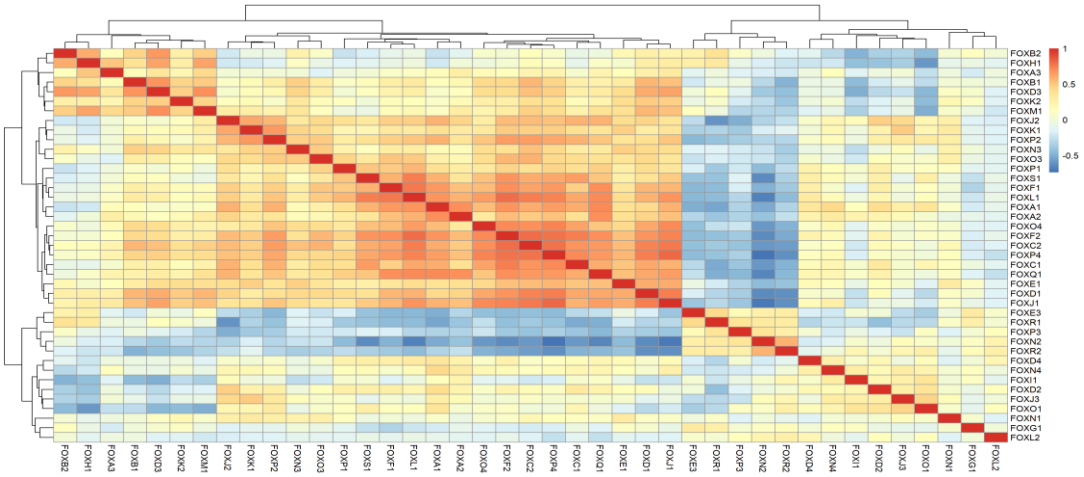

Fig. 1. The correlation heatmap for testicular germ cell cancer samples.

Both rows and columns represent the members of FOX family. Color grids represent the Spearman correlation coefficient between each pair of FOX members. Red indicates positive correlation, blue indicates negative correlation. The figure indicates that many FOX members possess expressional relationship in this particular type of cancer.

	FOXD2-FOXE3	FOXQ1-FOXF2	FOXC1-FOXQ1	FOXC1-FOXF2	FOXF1-FOXC2	FOXC2-FOXL1	FOXF1-FOXL1
ACC	0.045783517	0.488480971	0.160148809	0.304181122	0.175168888	0.503158664	0.390457644
BLCA	0.235352176	0.42163035	-0.00982748	0.094411724	0.346211774	0.473326834	0.311091138
BRCA	0.116048574	0.583787155	0.342012831	0.321850256	0.333521879	0.649621412	0.349178074
CESC	0.287612885	0.391915835	0.159688276	0.073499624	0.091389997	0.587743891	0.139406215
CHOL	0.259411227	-0.059632161	-0.10576332	-0.046223613	0.344681126	0.323257537	0.013375313
COADREAD	0.498424306	-0.031907992	0.202231414	0.111921288	0.304734783	0.312478038	0.528723264
DLBC	-0.213452591	0.44237932	-0.028915827	-0.051564578	-0.141918472	0.202756476	-0.229591837
ESCA	0.284378361	-0.032502559	-0.00056101	0.043085933	0.09912437	0.49912816	0.26364808
GBM	0.177481234	0.682137587	0.481929023	0.529413462	0.072613248	0.64491037	0.286214428
HNSC	0.356685679	0.276253657	0.060056009	0.114394609	0.026169363	0.677981775	0.030968716
KICH	-0.194726605	0.249687669	0.136430172	0.254569999	-0.02163404	0.047200366	0.243223546
KIRC	0.318424464	-0.021162116	0.047449574	0.220210337	0.4173788	0.717118799	0.276693791
KIRP	0.077688003	-0.193868711	0.126912719	0.023089436	0.453297849	0.570267951	0.451453442
LAML	0.222210134	0.371304966	0.186943156	0.295791257	-0.08057393	0.125572835	0.090859727
LGG	0.175328832	0.545475682	0.59659022	0.525586226	0.330467148	0.594088406	0.333037679
LIHC	0.287481003	0.1630255	0.09131205	0.451921956	0.653471658	0.65191524	0.694199714
LUAD	0.289606421	0.360062514	0.2374079	0.182477995	0.276800761	0.559540222	0.403669074
LUSC	0.396777513	0.550222296	0.199034568	0.219821934	0.364325969	0.772948243	0.375832526
OV	0.413560621	0.394213534	0.167761152	0.11194694	0.37780988	0.675555593	0.326766556
PAAD	0.408583022	0.213163016	0.288958634	0.109390497	0.144088046	0.166432323	-0.142296989
PCPG	0.386724212	0.355955163	0.291361868	0.589620018	0.480462715	0.751981276	0.492294504
PRAD	0.097625613	0.253358873	0.309693259	0.27847132	0.408469652	0.438700061	0.46470974
SARC	0.510100522	0.636237829	0.115887185	0.138778615	0.094454006	0.686982669	-0.054014305
SKCM	0.284590423	0.293559522	0.124899122	0.228934988	0.418318725	0.655955998	0.441668884
STAD	0.454767211	0.093307175	0.173083975	-0.024173618	0.446367092	0.364005955	0.334102128
TGCT	0.002509841	0.703736808	0.591981097	0.745365175	0.631371547	0.741739814	0.759245152
THCA	0.160055282	-0.276609047	-0.1604543	0.47732626	0.188270739	0.690299336	0.071844277
THYM	0.181915285	0.193939354	-0.168907563	0.229345289	0.404008245	0.337483168	0.34303814
UCEC	0.359093312	0.37263762	0.133325254	0.158796937	0.181032489	0.506294184	0.287264937
UCS	0.568474287	0.077067669	0.398496241	0.152126005	0.244777576	0.718697585	0.188423645
UVM	0.384241888	0.243999062	0.073699015	0.542240975	0.368865919	0.647345018	0.562720864

Fig. 2. The Spearman correlation coefficient in between members that exists in the same gene cluster in all tested cancer types.

3.2. The FOXC2-FOXL1 Relationship

Among all tested cluster pairs, only FOXC2 and FOXL1 displayed a very close positive relationship. Their expressions were positively related in all studies. The Spearman's rank correlation coefficient (ρ value) in this pair was greater than 0.7 in 6 types of tumor samples (KICH, LUSC, PCPG, TGCT, UCS, KIRC). The coefficient was higher than 0.5 in 22 out of 31 tumor types. The top 4 cases for this correlation are listed in scatterplots below (Fig. 3). The p values of each correlation were also calculated, however, some values were too small to be accurate, so the p value is shown as <2.2e-16.

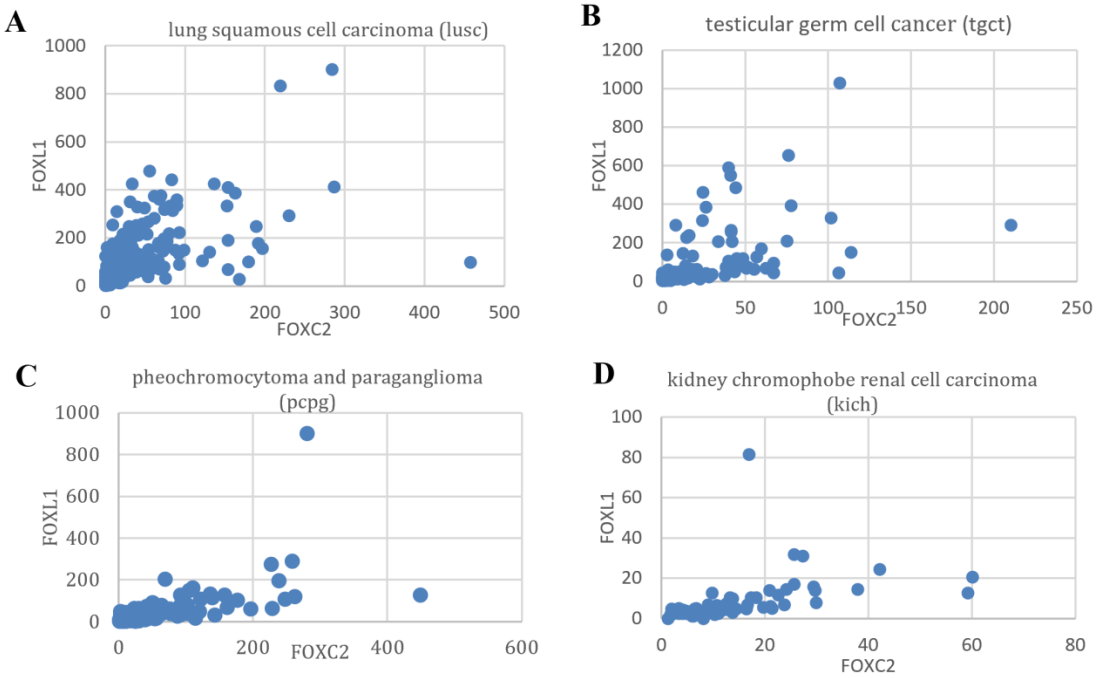

Fig. 3. Selected FOXC2-FOXL1 relationship scatterplots.

A. FOXC2 and FOXL1 showed positive correlation in lung squamous cell carcinoma, $\rho \approx 0.773$, p value<2.2e-16. B. FOXC2 and FOXL1 showed positive correlation in testicular germ cell cancer, $\rho \approx 0.742$, p value=1.66e-28. C. FOXC2 and FOXL1 showed positive correlation in pheochromocytoma and paraganglioma, $\rho \approx 0.752$, p value<2.2e-16. D. FOXC2 and FOXL1 showed positive correlation in kidney chromophobe cell carcinoma, $\rho \approx 0.789$, p value = 3.815e-15.

3.3. Other Positive Relationships

The analysis in this study also identified notable correlations between members that are not located in the same clusters. FOXF2 is located outside the FOXF1-FOXC2-FOXL1 cluster, but has showed significant correlation with FOXF1 in colorectal adenocarcinoma. Similarly, FOXF2 showed strong correlation with FOXC2 in testicular germ cell cancer (Fig. 4).

3.4. Negative Relationships

FOXA3 and FOXE1 exhibited a very strong negative correlation in esophageal carcinoma in scatterplot ($\rho \approx -0.700$) (Fig. 5), while FOXN2-FOXP4 also showed significant negative correlation in testicular germ cell tumors ($\rho \approx -0.736$), and in uveal melanoma ($\rho \approx -0.687$). The other two notable negative correlations are FOXM1 and FOXO1 in thymoma ($\rho \approx -0.694$); and FOXJ1 and FOXN2 in testicular germ cell cancer ($\rho \approx -0.715$) (Fig. 6) These findings indicates that any of these members may down regulate the other in the correlation. Like the positive relationships, future research should examine the exact relationship between FOX family

members, and may provide new possibilities for gene treatment for cancer.

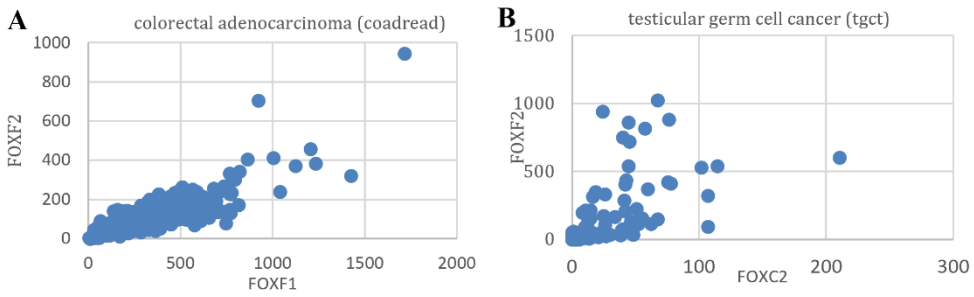

Fig. 4. Selected top positive relationships in scatterplots.

A. FOXF1 and FOXF2 showed significant positive correlation in colorectal adenocarcinoma, $\rho \approx 0.837$, p value<2.2e-16. B. FOXC2 and FOXF2 showed positive correlation in testicular germ cell cancer, $\rho \approx 0.806$. p value=4.06e-40.

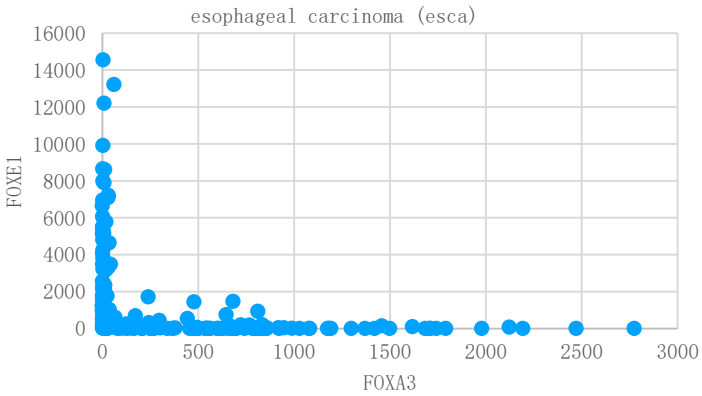

Fig. 5. FOXA3 and FOXE1 display a significant negative correlation in esophageal carcinoma, especially at high expressional levels ($\rho \approx$-0.700). p value=1.33e-28.

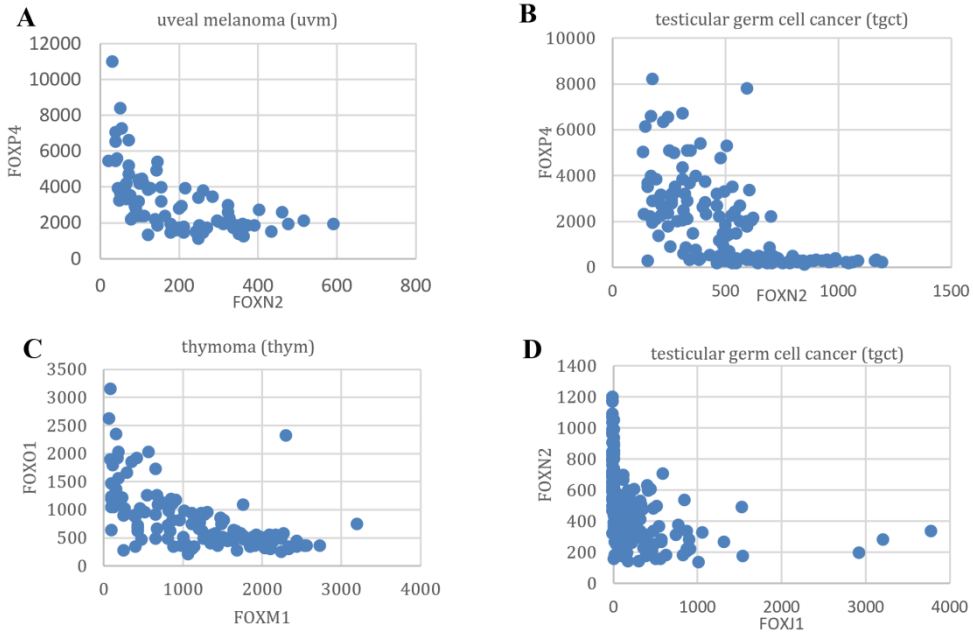

Fig. 6. Other notable negative relationships in scatterplots.

A. FOXN2 and FOXP4 showed negative correlation in uveal melanoma (ρ ≈-0.687). *p* value<2.2e-16. B. FOXN2 and FOXP4 showed negative correlation in testicular germ cell cancer(ρ ≈-0.736) *p* value<2.2e-16. C. FOXM1 and FOXO1 showed negative correlation in thymoma (ρ ≈-0.694) *p* value<2.2e-16. D. FOXN2 and FOXJ1 showed negative correlation in testicular germ cell cancer (ρ ≈-0.715) *p* value=1.01e-25.

4. Discussion

FOXC2 is involved in making critical proteins in the formation of organs and tissues before birth, such as the lungs, eyes, kidneys, cardiovascular and lymphatic systems. Insertions or deletions of FOXC2 sequence most commonly cause Lymphedema Distichiasis Syndrome [6] FOXC2 also down regulates cell-to-cell adhesion in cancer tissues, therefore increasing the possibility for cancer metastasis in malignant cancer [7]. FOXL1 has been identified as tumor suppressor responsible for the development of gastrointestinal tract. Low FOXL1 expressions tend to increase depth of invasion, lymph node metastasis, and distant metastasis [8]. As mentioned before, FOXC2 and FOXL1 exist in the same cluster together with FOXF1 at 16q24.1 [9]. The results noted in this study may be explained by their close physical relationship, as any frameshift mutation caused by insertion or deletion could affect both genes. However, other clusters identified by previous studies have not resulted in such significant correlations. Therefore, being inside a cluster does not necessarily indicate close expressional relationship. More experiments regarding this phenomenon would allow for a better understanding of how exactly these two genes interact with each other. Confirmations of the findings of this study, as well as other tumor-specific relationships may also provide alternatives on targeting genes for cancer treatment or for prognosis based on known gene functions of the FOX family members. The other pairs in the clusters did not show strong correlations, which may suggest that physical location may not be the primary reason for expression correlation.

References

[1] Myatt, S. S., & Lam, E. W. (2007). The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer*, *7*, 847-859.

[2] Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H., & Katoh, M. (2013). Cancer genetics and genomics of human FOX family genes. *Cancer Lett.*, *328*, 198-206.

[3] Katoh, M. (2004). Human FOX gene family (Review). *International Journal of Oncology*, *25*, 1495-1500.

[4] Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., & Larsson, E., *et al*. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, *6*, pl1.

[5] Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., & Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, *2*, 401-404.

[6] Uhlenhaut, N. H., & Treier, M. (2008). Transcriptional regulators in kidney disease: gatekeepers of renal homeostasis. *Trends Genet*, *24*, 361-371.

[7] Mortazavi, F., An, J., Dubinett, S., & Rettig, M. (2010). p120-catenin is transcriptionally downregulated by FOXC2 in non-small cell lung cancer cells. *Mol Cancer Res.*, *8*, 762-774.

[8] Ertao, Z., Jianhui, C., Chuangqi, C., Changjiang, Q., Sile, C., Yulong, H., Shirong, C., & Hui, W. (2016). Low level of FOXL1 indicates a worse prognosis for gastric cancer patients. *Tumour Biol.*, *37*, 11331-11337.

[9] Stankiewicz, P., Sen, P., Bhatt, S. S., Storer, M., Xia, Z., Bejjani, B. A., Ou, Z., Wiszniewska, J., Driscoll, D. J., & Maisenbacher, M. K., *et al*. (2009). Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations. *Am J Hum Genet*, *84*, 780-791.

Bokai Zhang was born in Xi'an, Shaanxi, China on Dec 22nd 2000. Currently he is a high school student at Mount Michael Benedictine High School in Omaha, NE, USA.