ImReLnc: Identifying Immune-Related LncRNA Characteristics in Human Cancers Based on Heuristic Correlation Optimization

Meihong Gao, Shuhui Liu, Yang Qi, Xinpeng Guo and Xuequn Shang*

School of Computer Science and Engineering, Northwestern Polytechnical University, Xi’an, China

Long non-coding RNAs (lncRNAs) play critical roles in cancer through gene expression and immune regulation. Identifying immune-related lncRNA (irlncRNA) characteristics would contribute to dissecting the mechanism of cancer pathogenesis. Some computational methods have been proposed to identify irlncRNA characteristics in human cancers, but most of them are aimed at identifying irlncRNA characteristics in specific cancer. Here, we proposed a new method, ImReLnc, to recognize irlncRNA characteristics for 33 human cancers and predict the pathogenicity levels of these irlncRNAs across cancer types. We first calculated the heuristic correlation coefficient between lncRNAs and mRNAs for immune-related enrichment analysis. Especially, we analyzed the relationship between lncRNAs and 17 immune-related pathways in 33 cancers to recognize the irlncRNA characteristics of each cancer. Then, we calculated the Pscore of the irlncRNA characteristics to evaluate their pathogenicity levels. The results showed that highly pathogenic irlncRNAs appeared in a higher proportion of known disease databases and had a significant prognostic effect on cancer. In addition, it was found that the expression of irlncRNAs in immune cells was higher than that of non-irlncRNAs, and the proportion of irlncRNAs related to the levels of immune infiltration was much higher than that of non-irlncRNAs. Overall, ImReLnc accurately identified the irlncRNA characteristics in multiple cancers based on the heuristic correlation coefficient. More importantly, ImReLnc effectively evaluated the pathogenicity levels of irlncRNAs across cancer types. ImReLnc is freely available at https://github.com/meihonggao/ImReLnc.

Keywords: immune-related lncRNA, pathogenicity levels, heuristic correlation coefficient, enrichment analysis, immune infiltration

1 INTRODUCTION

Cancer is a major threat to human health with high incidence and mortality (Siegel et al., 2015; Bray et al., 2018; Ferlay et al., 2019; Gharib et al., 2019). In 2018, there were approximately 18.1 million new cases and 9.6 million cancer deaths worldwide (Bray et al., 2018; Ferlay et al., 2019). Although early detection and treatment of cancer have increased the number of survivors, the results are still limited (Miller et al., 2016, 2019). Previous studies have reported that there is a link between lncRNA and the occurrence of cancer, and lncRNA plays an important role in cancer by gene expression and
immature lymphocytes can infiltrate tumor tissues are called...
filtered by the annotation file in GENECODE (v22) to obtain the expression matrix of mRNA and lncRNA.

Hepatic carcinoma-related single-cell sequencing data was downloaded from Panglaodb. We performed cluster analysis on single-cell sequencing data to obtain the expression matrix of immune and non-immune cells through the Seurat package in R software. We further analyzed the expression matrix of lncRNA in these two types of cells and obtained the expression matrix of irlncRNA and non-irlncRNA in them.

2.1.2 Immunization Data Collection

Immune-related gene lists were downloaded from Immport, which is one of the largest public databases for collecting and sharing immunology-related research resources. A total of 17 immune pathways and 1793 immune-related mRNAs were obtained for immune-related enrichment analysis, and the number of mRNAs associated with immune pathways ranged from 3 to 505 (See Supplementary Table S2, Data Sheet 1).

The immune infiltration levels of TCGA cancer samples (33 types of cancer samples) were downloaded from Timer. We obtained the infiltration levels of these cancer samples in six types of cells: B cell, T cell CD4⁺, T cell CD8⁺, Neutrophil, Macrophage, and Myeloid dendritic cell.

2.1.3 LncRNA-Disease Interactions Collection

We collected the human lncRNA-disease interactions from LncRNADisease and Lnc2Cancer. Specifically, 2665 drlncRNAs were collected from Lnc2Cancer, and 6,105 drlncRNAs were collected from Lnc2Cancer. After preprocessing (removal of duplicate items), there were 2665 drlncRNAs left in Lnc2Cancer and 5,714 drlncRNAs left in LncRNADisease.

2.2 Heuristic Correlation Optimization

LncRNA can regulate the expression of mRNA to cause cancer, and there are two regulatory patterns: direct regulation and indirect regulation (Figure 2). In the first pattern, the expression values of lncRNA and mRNA are directly related, while in the second pattern, they are partially related.

2.2.1 Direct Correlation Between mRNAs and lncRNAs

The direct correlation was computed through the method of Pearson and Spearman (Benesty et al., 2009; Sedgwick, 2014). The Pearson’s rank correlation coefficient between mRNA m and lncRNA l is defined as follows:

\[
R(m, l) = \frac{E(ml) - E(m)E(l)}{\sqrt{E(m^2) - E^2(m)} \sqrt{E(l^2) - E^2(l)}}
\]

where function \(E\) is used to calculate the mathematical expectation of variables. Similarly, the Spearman’s correlation coefficient between mRNA m and lncRNA l is defined as follows:

\[
S(m, l) = 1 - \frac{6\sum d_i^2}{r(r^2 - 1)}
\]

where \(d_i\) represents the difference between the rank of m and l, and r represents the number of cancer samples. Although Pearson correlation and Spearman correlation can effectively fit the correlation between mRNA and lncRNA in cancer samples, they have some limitations. To make the value of direct correlation more accurate, we combined these two types of correlation coefficients to represent the direct correlation coefficient as follows:

\[
D(m, l) = \frac{R(m, l) + S(m, l)}{2}
\]
The tumor purity of the sample was calculated by analyzing its mRNA expression profile using the estimate package of R software. The partial correlation coefficient between mRNA $m$ and lncRNA $l$ in sample $t$ is defined as follows:

$$P(m, l)(t) = \frac{D(m, l) - D(m, t)D(t, l)}{\sqrt{1 - D^2(m, t)}\sqrt{1 - D^2(t, l)}}$$ (4)

### 2.2.3 Heuristic Correlation Between mRNAs and lncRNAs

Both direct and partial correlations play an important role in lncRNA regulating mRNA expression. Thus, we integrated these two correlations to calculate the heuristic correlation between mRNA and lncRNA. To highlight the strong correlation and to fade the weak correlation, we defined a logistic function as follows:

$$L(x) = \frac{1}{e^{cx+d}+1}$$ (5)

where $x$ is the absolute value of the correlation coefficient. When $c$ is equal to -15 and $d$ is equal to ln (1999), $L(x)$ is approximately equal to 0 in the interval [0, 0.3], and $L(x)$ is approximately equal to 1 in the interval [0.7] (See Supplementary Figure S1, Data Sheet 1). Furthermore, we define the heuristic correlation coefficient between mRNA $m$ and lncRNA $l$ as follows:

$$H(m, l)(t) = L[\beta D(m, l) + (1 - \beta)P(m, l)(t)]$$ (6)

where $\beta$ ranges from 0 to 1 and is limited as follows:

$$\arg \min_{\beta} [1 - |\beta D(m, l) + (1 - \beta)P(m, l)(t)|]$$ (7)

### 2.3 Identification of Immune-Related lncRNAs

#### 2.3.1 Immune-Related Enrichment Analysis

The immune regulation mechanism of lncRNA was explored by analyzing the relationship between lncRNA-related mRNA and immune pathways. First, we performed a screening on the heuristic correlation matrix to obtain the rank matrix $R_{ML}$ (the filtering threshold is 0.5). As a result, $v$ lncRNA-related mRNA classes were obtained. Each class was composed of heuristic correlation coefficients between a lncRNA and the corresponding $u$ mRNAs and was sorted in descending order according to these coefficients. Then we conducted enrichment analysis for the lncRNA-related mRNA classes on 17 immune pathways using fgsea package in R software. The minimum and maximum sizes of the mRNA class were set to 1 and 5,000, respectively, and the number of permutations was set to 10,000. The enrichment analysis result was a $17 \times 10$ matrix, whose column information includes lncRNA, immune pathway, $p$-value, adjusted $p$-value, enrichment score, and five other enrichment parameters.

#### 2.3.2 Extraction of Immune-Related lncRNAs

We evaluated the relationship between lncRNA and immune pathway based on the enrichment score and $p$ value as follows (Li et al., 2020):

$$\text{IncRES}(l, w) = \begin{cases} 1 - 2p, & E(l, w) \geq 0 \\ 2p - 1, & E(l, w) < 0 \end{cases}$$ (8)

where $p$ represents the $p$-value of the enrichment result, and $E(l, w)$ represents the enrichment score of lncRNA $l$ on pathway $w$. Obviously, the range of IncRES $(l, w)$ is $[−1, 1]$, and the greater the absolute value, the higher the enrichment degree. For each type of cancer $c$, $|I_c|$ lrcRNAs were obtained by setting the absolute value of IncRES $(l, w)$ to be greater than 0.995 and setting the FDR to be less than 0.05. Finally, the set of all lrcRNAs in 33 types of cancer can be defined as follows:

$$I = \bigcup_{c=1}^{33} I_c$$ (9)

Each lrcRNA in set $I$ has several corresponding immune pathways to form lrcRNA-pathway pairs, in which the lncRNA plays an immunomodulatory role in cancer by acting on the immune pathway.

### 2.4 Pathogenicity Evaluation of Immune-Related lncRNAs

After identifying cancer-related lrcRNAs, we further analyzed these lrcRNAs to explore their pathogenicity levels.
2.4.1 Differential Expression Analysis of Immune-Related IncRNAs

Since the differential expression level of IncRNA greatly influences on the pathogenicity of IncRNA, we performed differential expression analysis of the identified irlncRNAs in 33 cancers by using the edger package of R software. Here, the fold change and the adjusted p values were used to evaluate the expression differences of irlncRNAs, and the differentially expressed irlncRNAs were acquired by setting the adjusted p-value < 0.01 and the absolute value of logFC > 1.5. \( R_{DE} \) is used to reflect the range of differential expression of irlncRNA \( l \) in set \( I \) and is defined as follows:

\[
R_{DE}(l) = N \left[ NC(l) \right]
\]

where function \( N \) is used for normalization, and \( NC(l) \) represents the number of cancers in which irlncRNA \( l \) from set \( I \) is differentially expressed. The range of \( NC(l) \) is \([0, 24]\), and the upper limit depends on the number of cancers (with tumor and normal samples) for which we performed differential expression analysis. \( R_{FC}(l) \) is used to reflect the intensity of differential expression of irlncRNA \( l \) in set \( I \) and is defined as follows:

\[
R_{FC}(l) = N \left[ \frac{\sum_{c=1}^{NC(l)} FC(c, l)}{NC(l)} \right]
\]

where function \( N \) is used for normalization, and \( FC(c, l) \) represents the fold change of the expression value of irlncRNA \( l \) in cancer \( c \).

2.4.2 Pathogenicity Prediction of Immune-Related IncRNAs

Because the number of cancers related to IncRNA-pathway pairs and the differential expression level of IncRNA have a great influence on the pathogenicity level of IncRNA, we merged these two kinds of information to evaluate the pathogenicity level of irlncRNA \( l \) in set \( I \) as follows:

\[
P_{score}(l) = \frac{\sum_{w=1}^{17} N[NC(l, w)] + R_{DE} + R_{FC}}{19}
\]

where function \( N \) is used for normalization, and \( NC(l, w) \) represents the number of cancers in which IncRNA \( l \) regulates immune pathway \( w \).

3 RESULTS

We proposed a computational method, ImReLnc, to identify irlncRNAs and predict their pathogenicity levels across cancer types (Figure 1). The immunological and pathogenic properties of the identified irlncRNAs have been verified in this section. In particular, the experiment was performed on R 3.4.1 with Ubuntu 20.04.2, the CUDA version was 11.3, and the memory was 8 × 11019 M.

3.1 Immune-Related IncRNAs and Pathways

Through steps one to three in Figure 1, we comprehensively analyzed the heuristic correlation between mRNA and IncRNA of 33 cancers and the enrichment of IncRNA in 17 immune pathways. As a result, we obtained a series of IncRNA-pathway pairs (Figure 3) for each cancer. We found that the number of irlncRNA involved and the enriched immune pathways differed between cancers, and the degree of enrichment of cancers in some immune pathways tended to be similar. We further calculated the number of cancer-related irlncRNAs (Figure 4A) and immune pathways (Figure 4B). It was found that the number of cancer-related irlncRNAs ranged from 6.75 to 11.21 (after log2). Specifically, the number of UCEC-related irlncRNAs was the least at 6.75 (after log2), and the number of CHOL-related irlncRNAs was the highest at 11.21 (after log2). As for the number of cancer-related immune pathways, they ranged from 11 to 17. It should be noted that among 33 types of cancers, two types (6.1%) were enriched in 11 pathways, one type (3%) was enriched in 13 pathways, two types (6.1%) were enriched in 14 pathways, eight types (24.2%) were enriched in 15 pathways, thirteen types (39.4%) were enriched in 16 pathways, and the remaining seven types (21.2%) were enriched in 17 pathways. We found that the number of cancer types enriched in more than 16 pathways accounted for 60.6% of the total cancer types, and the number of cancer types enriched in more than 15 pathways accounted for 84.8% of the total cancer types. This indicates that most of the pathways are involved in the immune regulation process of cancer.

Compared with cancers with a small number of IncRNA-pathway pairs, cancers with a large number of IncRNA-pathway pairs are more likely to be enriched in the immune pathway, and for the convenience of description, we refer to these cancers as immune-preferred cancers in the following text. We analyzed the enrichment of cancers in immune pathways to find immune-preferred cancers (Figure 5A). Cholangiocarcinoma (CHOL), Kidney Chromophobe (KICH), Testicular Germ Cell Tumors (TGCT), Thyroid carcinoma (THCA), and Thymoma (THYM) were found to have more obvious immune enrichment, thus they are the so-called immune-preferred cancers. Studies have shown that immunotherapy plays an important role in the treatment of these five types of cancer (Schott et al., 2001; Drake and Stein, 2018; Hogdall et al., 2018; Jakopic et al., 2020; Kalavská et al., 2020). For these cancers, immunotherapy can achieve better prognostic effects, and our findings can provide a reference for their immunotherapy.

Cancers also have a preference for their enriched immune pathways. They will be mainly enriched in some immune pathways, which we call highly enriched immune pathways, while their enrichment in other pathways will be very limited. We analyzed the distribution of IncRNA-pathway pairs on the immune pathways to discover highly enriched immune pathways in cancers (Figure 5B). Cancers were mainly enriched in 5 immune pathways (Antigen Processing and Presentation, Antimicrobials, Cytokine Receptors, Cytokines, and Natural Killer Cells Cytotoxicity), and the average enrichment ratio of IncRNA-pathway pairs related to these pathways was 0.790 832.3. Specifically, 18 of 33 (more than half) cancers were enriched in these five immune pathways by more than 0.790 832.3 (Figure 6). Obviously, these five pathways are highly enriched immune pathways for those 33 cancers, and these pathways are active in
the immune regulation process of the cancers. Antigen processing and presentation is involved in the decomposition and presentation of antigen proteins in immune regulation (Vyas et al., 2008). Antimicrobials are related to the life activities of bacteria (Van Harten et al., 2018). Cytokine Receptors and Cytokines play a signal transduction role in immune regulation (O’Shea et al., 2019). As for Natural Killer Cell Cytotoxicity, it is related to the immune regulation of Natural Killer Cell (Topham and Hewitt, 2009). The functions mentioned above are related to the life activities of cancer cells, thus the high enrichment of these five immune pathways is closely related to the occurrence of cancer.

3.2 The Distribution of Immune-Related lncRNAs
Cancers are enriched in immune pathways with a series of lncRNA-pathway pairs, and each pathway is related to a variety of lncRNAs which are the irlncRNAs to be recognized.
These lncRNAs are involved in the immune regulation process of cancer. We analyzed the characteristics of cancers and immune pathways related to them (See Supplementary Table S1, 2). It was found that the total number of lncRNAs was 7,038, the number of cancers related to lncRNA ranged from 1 to 32, and the number of immune pathways associated with lncRNA ranged from 1 to 15. For the number of cancers and pathways in these ranges, we further explored the distribution of lncRNAs on them (See Supplementary Figure S2, Data Sheet 1). Firstly, for the distribution of different numbers of cancers, we found that lncRNAs related to only one type of cancer accounted for 0.417 448 1 of the total lncRNAs, lncRNAs related to two types of cancer accounted for 0.209 718 7 of the total lncRNAs, lncRNAs related to three types of cancer accounted for 0.114 237 of the total lncRNAs, and the remaining 4–32 types of cancer-related lncRNAs accounted for 0.258 596 2 of the total lncRNAs. Then, for the distribution of different numbers of immune pathways, we found that lncRNAs associated with one to four immune pathways accounted for 0.638 107 4 of the total lncRNAs, and the remaining 0.361 892 6 of lncRNAs were related to 5–15 pathways. Overall, more than half of lncRNAs played a regulatory role in several types of
cancer, and these regulatory roles involved multiple immune pathways.

The immunomodulatory effect in immune-preferred cancers is quite active, and we further analyzed irlncRNAs related to these cancers (See Supplementary Table S3). There were 34 irlncRNAs related to all five immune-preferred cancers, and another 5,097 irlncRNAs were related to one to four of the five immune-preferred cancers. The highly enriched immune pathway plays an important role in the immune regulation of cancers. Thus we obtained irlncRNAs related to the highly enriched pathways (See Supplementary Table S4). There were 1,264 irlncRNAs related to all five highly enriched immune pathways, and another 5,473 irlncRNAs were related to one to four of the five highly enriched immune pathways. Obviously, most irlncRNAs (0.627 558) were involved in one of the five immune-preferred cancers, and they were almost evenly distributed in the five pathways. This was significantly different from the distribution of irlncRNAs in total cancers and total immune pathways due to the immune preference of the cancers and the high enrichment of the immune pathways.

### 3.3 The Expression of Immune-Related IncRNAs

Since the immune regulation process is performed by immune cells, the expression of irlncRNA in immune cells should be higher than that in other cells. We compared the expression levels of irlncRNA and non-irlncRNA in immune and non-immune cells on single-cell sequencing data from the PanglaoDB database.  
1) We normalized the single-cell sequencing data of hepatic carcinoma.  
2) The IncRNA count data was extracted.  
3) The irlncRNA count data and non-irlncRNA count data were extracted.  
4) We compared the expression of irlncRNA and non-irlncRNA in immune and non-immune cells, respectively (Figures 7A–C). It was found that the average expression of irlncRNA in immune cells was higher than that of non-irlncRNA, while in non-immune cells, their difference was slight (Figure 7A). The distributions of irlncRNA and non-irlncRNA in immune cells and non-immune cells further confirmed this conclusion (Figures 7B,C). To further explore the expression characteristics of irlncRNA, we analyzed the correlation between IncRNA expression and the level of immune infiltration in cancers. We first estimated the immune infiltration level of cancer samples through the TIMER online tool (Newman et al., 2015; Becht et al., 2016; Li et al., 2016; Aran et al., 2017; Finotello et al., 2019). Then, we calculated the correlation between IncRNA expression and the level of immune infiltration to find out infiltration-related IncRNAs (infrlncRNAs). Finally, we compared the distribution of our method (ImReLnc) and ImmLnc on irlncRNA and infrlncRNA. Specifically, we calculated the irlncRNA rate (IR) in all IncRNAs, the irlncRNA rate in infrlncRNA (IRINF), the infrlncRNA rate (INFR) in all IncRNAs, the infrlncRNA rate in irlncRNA (INFRI) for ImReLnc and ImmLnc (See Supplementary Table S3, S4, Data Sheet 1). It was found that the IRINF/IR and IRINF/IR of ImReLnc were significantly higher than ImmLnc in all 33 cancers (See Supplementary Figure S3, Data Sheet 1). Therefore, the irlncRNAs recognized by ImReLnc have more significant immune properties than those recognized by ImmLnc.

In order to show that the heuristic correlation coefficient is better than the direct or partial correlation coefficient, we performed the following analysis on the colon cancer data set. First, we calculated three correlation coefficients: direct correlation coefficient, partial correlation coefficient, and heuristic correlation coefficient. Then, we performed immune-related enrichment analysis based on the above correlation coefficients. Afterward, we screened the enrichment results and obtained three sets of irlncRNAs. Finally, we performed Cox regression analysis on the three groups of irlncRNAs. Specifically, six highly pathogenic irlncRNAs RP5-884M6.1, RP11-742B18.1, HOTAIR, AC004988.1, CTD-2357A8.3, and...
GS1-600G8.5, were found to be associated with cancer prognosis (Table 1). Among them, RP5-884M6.1, RP11-742B18.1, and HOTAIR were found to be related to the occurrence of colon cancer by previous studies (Zhou et al., 2019; Paredes et al., 2020; Gong et al., 2021). The prognostic analysis related to the three correlation calculations is shown in Figure 8. We found that the prognostic effects of lncRNA calculated by direct correlation and partial correlation were not much different. The prognostic effect

| LncRNA          | Univariate                  | Multivariate             | Source          |
|-----------------|-----------------------------|--------------------------|-----------------|
|                 | HR (95% CI for HR) | p.value | Coef | z     | p     |              |
| HOTAIR          | 0.97 (0.95–0.99)  | 0.026   | 5.79E-03 | 1.914 | 0.00158 | PMID: 34630664 |
| AC004988.1      | 1 (1–1)                | 0.00019 | -4.88E-04 | -1.638 | 0.00166 | -              |
| RP5-884M6.1     | 1 (1–1)                | 0.0019  | 1.83E-03 | 2.92  | 0.00356 | PMID: 32983990 |
| CTD-2357A8.3    | 0.97 (0.95–0.99) | 0.0072  | -0.0012 | -1.387 | 0.01535 | -              |
| GS1-600G8.5     | 0.99 (0.95–1)       | 0.012   | 1.19E-02 | 1.154  | 0.02929 | -              |
| RP11-742B18.1   | 1 (1–1)                | 0.00098 | -3.10E-02 | -2.045 | 0.04093 | PMID: 31516583 |

**Figure 8** Kaplan-Meier curves of immune-related lncRNAs from different correlation calculations. (A) Direct correlation calculation in the training set. (B) Direct correlation calculation in the testing set. (C) Direct correlation calculation in the total set. (D) Partial correlation calculation in the training set. (E) Partial correlation calculation in the testing set. (F) Partial correlation calculation in the total set. (G) Heuristic correlation calculation in the training set. (H) Heuristic correlation calculation in the testing set. (I) Heuristic correlation calculation in the total set.
of lncRNA calculated by heuristic correlation was better than the former two. This indicates that heuristic correlation calculation has advantages in identifying immune-related prognostic characteristics in cancer.

### 3.4 The Pathogenicity Level of Immune-Related lncRNAs

Having proved that ImReLnc can effectively identify irlncRNA, we next explored the application of ImReLnc in assessing the pathogenicity level of irlncRNA. We first compared the relationship between irlncRNAs from disease databases (LncRNADisease2.0 and Lnc2Cancer2.0) and irlncRNAs recognized by ImmLnc. Specifically, the top 500 irlncRNAs (determined by the number of cancers involved in irlncRNA) were used for analysis, and we found that the irlncRNA recognized by ImReLnc accounted for a higher proportion in the disease database (Table 2). Besides, we analyzed the distribution of pathogenicity levels of irlncRNAs in ImReLnc. Two methods were used to divide the data set where the sorted irlncRNAs (descending order) were located. The first method was to evenly divide irlncRNAs into 6 data sets according to the number of cancers involved in irlncRNA (Table 3), and the second method was to divide irlncRNAs into 5 data sets evenly according to the level of pathogenicity of irlncRNAs (Table 4). In Table 3, the average pathogenicity level of irlncRNAs in LncRNADisease2.0 and Lnc2Cancer2.0 is higher than or approximately equal to that of all identified irlncRNAs, especially in Dataset 1. This indicates that the calculated pathogenicity level of irlncRNA is credible, and irlncRNA is particularly pathogenic within the range of high pathogenicity levels. Similarly, in the range of high pathogenicity levels in Table 4, the mean pathogenicity level of irlncRNAs in LncRNADisease2.0 and Lnc2Cancer2.0 is higher than the mean pathogenicity level of all identified irlncRNAs. This further shows that the calculated pathogenicity level of irlncRNA is credible, and its pathogenicity within the interval of high pathogenicity level is quite apparent.

### 4 DISCUSSION

Long non-coding RNA plays an essential role in cancer via gene expression and immune regulation (Hu et al., 2013; Chen et al., 2017), and the identification of irlncRNA is of great significance to the diagnosis and treatment of cancer. This study implemented an R program, ImReLnc, to identify irlncRNAs and analyze their pathogenicity levels across cancer types. Firstly, ImReLnc calculated heuristic correlation based on direct correlation and

---

**Table 2**: The comparison of the relationship between irlncRNA and drlncRNA in ImmLnc and ImReLnc. The data set in the table is divided equally according to the number of irlncRNAs. IrlncRNAs refer to immune-related lncRNAs, drlncRNA refers to disease-related lncRNAs.

| Dataset | Distribution of ImReLnc | Distribution of ImmLnc |
|---------|-------------------------|------------------------|
|         | Count | LncRNADisease2.0 | Lnc2Cancer2.0 | Count | LncRNADisease2.0 | Lnc2Cancer2.0 |
| Total   | 500   | 110            | 111           | 500   | 106            | 109           |
| Dataset1| 100   | 29             | 25            | 100   | 27             | 22            |
| Dataset2| 100   | 27             | 25            | 100   | 28             | 25            |
| Dataset3| 100   | 30             | 31            | 100   | 25             | 29            |
| Dataset4| 100   | 14             | 11            | 100   | 15             | 20            |
| Dataset5| 100   | 10             | 15            | 100   | 11             | 13            |

**Table 3**: The pathogenicity distribution of immune-related lncRNA (irlncRNA). The data set is divided equally according to the number of irlncRNAs.

| Pathogenicity level of IrlncRNA | Total | Dataset1 | Dataset2 | Dataset3 | Dataset4 | Dataset5 | Dataset6 |
|---------------------------------|-------|----------|----------|----------|----------|----------|----------|
| All identified irlncRNAs       | 7,038 | 1,173    | 1,173    | 1,173    | 1,173    | 1,173    | 1,173    |
| Min                             | 0.028 | 0.345    | 0.298    | 0.263    | 0.231    | 0.197    | 0.028    |
| Max                             | 0.558 | 0.558    | 0.345    | 0.298    | 0.263    | 0.231    | 0.197    |
| Mean                            | 0.271 | 0.399    | 0.320    | 0.280    | 0.247    | 0.215    | 0.164    |
| In LncRNADisease2.0             |       |          |          |          |          |          |          |
| Count                           | 613   | 202      | 111      | 102      | 83       | 61       | 74       |
| Count/Count                     | 0.087 | 0.172    | 0.094    | 0.087    | 0.053    | 0.052    | 0.063    |
| Min1                            | 0.108 | 0.346    | 0.299    | 0.263    | 0.232    | 0.197    | 0.108    |
| Max1                            | 0.558 | 0.558    | 0.345    | 0.298    | 0.262    | 0.231    | 0.197    |
| Mean1                           | 0.308 | 0.415    | 0.326    | 0.279    | 0.248    | 0.219    | 0.161    |
| In Lnc2Cancer2.0                |       |          |          |          |          |          |          |
| Count2                          | 666   | 181      | 115      | 114      | 93       | 77       | 86       |
| Count2/Count                    | 0.094 | 0.154    | 0.098    | 0.097    | 0.079    | 0.066    | 0.073    |
| Min2                            | 0.091 | 0.345    | 0.299    | 0.263    | 0.231    | 0.197    | 0.091    |
| Max2                            | 0.558 | 0.558    | 0.345    | 0.298    | 0.261    | 0.231    | 0.196    |
| Mean2                           | 0.295 | 0.411    | 0.321    | 0.280    | 0.249    | 0.215    | 0.163    |
partial correlation to provide a ranking score for lncRNA-related mRNA class. Then, ImReLnc performed immune-related enrichment analysis based on the ranking score mentioned above to obtain irlncRNAs. Besides, ImReLnc conducted differential expression analysis of these irlncRNAs. Finally, the pathogenicity levels of the irlncRNAs were estimated by their immune and differential expression characteristics across cancer types.

We compared the expression levels of irlncRNAs in immune cells and non-immune cells, which showed that the expression of lncRNAs in immune cells was higher than that in non-immune cells (Figure 7). Besides, we analyzed the correlation between lncRNA expression and the level of immune infiltration. We found that the IR was significantly lower than the IRINF, and the INFRI was also significantly lower than the INFRI, which indicated that the immunomodulatory effect of the identified irlncRNAs was significant (See Supplementary Table S3, Data Sheet 1). Furthermore, for the irlncRNAs in each cancer identified by ImReLnc, we compared them with those identified by ImmLnc (See Supplementary Table S5, Data Sheet 1). We found a significant overlap between the irlncRNAs recognized by the two methods, which further demonstrated the reliability of ImReLnc in recognizing irlncRNAs. Finally, we explored the application of ImReLnc in evaluating the pathogenicity level of irlncRNAs. For the pathogenicity levels of irlncRNAs estimated by ImReLnc, we compared them with the drlncRNAs in the disease database (Table 3 and Table 3), and the results showed that drlncRNAs had a significantly higher proportion in the highly pathogenic irlncRNA interval. We also compared the relationship between lncRNA (from the LncRNA Disease and Lnc2Cancer disease databases) and irlncRNA (identified by ImmLnc and ImReLnc), and we found that the lncRNA recognized by ImReLnc accounted for a higher proportion in the disease database (Table 2). This indicates that our calculation of the pathogenicity levels of irlncRNAs is credible.

lncRNA FEZF1-AS1 was found to have the highest pathogenic level in cancer and was related to the occurrence of GBM, LUAD, HNSC, LGG, LUSC, PRAD, SKCM, BLCA, LIHC, READ, TGCT, THYM, ACC, MESO, UCS, and CHOL. We infer that lncRNA FEZF1-AS1 plays an important regulatory role in these 16 cancers. For GBM, LUAD, PRAD, LIHC, READ, and UCS, studies have shown that lncRNA FEZF1-AS1 plays a crucial regulatory role in them (Luo et al., 2020; Liu et al., 2018; Zhu et al., 2019; Wang et al., 2018; Bian et al., 2018; Zhang and Li, 2018). As for the other ten types of cancer, their cancer tissues are similar to or adjacent to the cancer tissues of the above six types of cancer (See Supplementary Table S6, Data Sheet 1). Studies have shown that cancers with similar original tissues may share lncRNA regulatory mechanisms (Zhang et al., 2016; Li et al., 2020). The original tissue of LGG is similar to GBM, the original tissue of LUSC and MESO is similar to LUAD, and the original tissue of CHOL is similar to LIHC. Therefore, like in their similar tissues, lncRNA FEZF1-AS1 may also play an immunoregulatory role in LGG, LUSC, MESO, and CHOL. The original tissue of THYM and HNSC is adjacent to GBM, the original tissue of ACC is adjacent to GBM, the original tissue of BLCA and TGCT is adjacent to PRAD, and the original tissue of SKCM is adjacent to UCS. These cancers may be caused by the metastasis of their adjacent tissues. Thus they have a similar lncRNA regulatory mechanism as their adjacent tissues. Among them, ACC and SKCM are relatively far from their neighboring tissues, which may be caused by long-distance metastasis of advanced cancer. Overall, lncRNA FEZF1-AS1 is highly pathogenic and plays a crucial immunomodulatory role in cancer, which can provide a necessary reference for cancer treatment.

In summary, ImReLnc accurately identified irlncRNAs in cancers based on heuristic correlation optimization. More importantly, ImReLnc effectively assessed the pathogenicity level of irlncRNAs based on their immune and differential expression characteristics. However, due to the limitation of data acquisition, ImReLnc can only identify irlncRNAs in 33 cancers. In addition, we identified genes related to lncRNA expression and further investigated the enriched immune pathways, but it is still challenging determine how specific lncRNA affects gene expression. In the future, we hope to perform irlncRNA recognition for other types of cancer and explore the targets of these irlncRNAs.
DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: (1) TCGA: https://portal.gdc.cancer.gov/, (2) GENCODE: https://www.gencodegenes.org/, (3) PanglaoDB: https://panglaodb.se/index.html, (4) TIMER: http://cistrome.dfci.harvard.edu/TIMER/, (5) LncRNAProfile: http://www.rnanut.net/lncrnadisease/, (6) Lnc2Cancer2.0: http://www.bio-bigdata.net/lnc2cancer2.0/index.html, (7) Immport: https://www.immport.org/home.

AUTHOR CONTRIBUTIONS

MG designed and implemented the method. MG, SL, YQ and XG wrote this manuscript. All authors read and approved the final manuscript.

REFERENCES

Anderson, N. M., and Simon, M. C. (2020). The Tumor Microenvironment. Curr. Biol. 30, R921–R925. doi:10.1016/j.cub.2020.06.081

Aran, D., Hu, Z., and Butte, A. J. (2017). Xcell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. Genome Biol. 18. doi:10.1186/s13059-017-1349-1

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2018). LncRNADisease 2.0: an Updated Database of Long Non-coding RNA-Associated Diseases. Nucleic Acids Res. 47, D1034–D1037. doi:10.1093/nar/gky905

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an Updated Database of Long Non-coding RNA-Associated Diseases. Nucleic Acids Res. 47, D1034–D1037. doi:10.1093/nar/gky905

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. Genome Biol. 17, 218–220. doi:10.1186/s13059-016-1070-5

Benesty, I., Chen, J., Huang, Y., and Cohen, I. (2009). "Pearson Correlation Coefficient," in Noise Reduction in Speech Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–4. doi:10.1007/978-3-642-00296-0_5

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long Noncoding Rna and Cancer: a New Paradigm. Cancer Res. 77, 3965–3981. doi:10.1158/0008-5472.can-16-2634

Bian, Z., Zhang, J., Li, M., Feng, Y., Wang, X., Zhang, J., et al. (2018). LncRNA- FEZ1P-AS1 Promotes Tumor Proliferation and Metastasis in Colorectal Cancer by Regulating PKM2 Signaling. Clin. Cancer Res. 24, 4808–4819. doi:10.1158/1078-0432.ccr-17-2967

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a Cancer J. clinicians 68, 394–424. doi:10.3332/caac.21492

Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene Regulation in the Immune System by Long Noncoding Rnas. Nat. Immunol. 18, 962–972. doi:10.1038/ni.3771

Chi, Y., Wang, D., Wang, J., Yu, W., and Yang, J. (2019). Long Non-coding Rna in the Pathogenesis of Cancers. Cells 8, 1016. doi:10.3390/cells8091015

Coussens, L. M., Zitvogel, L., and Palucka, A. K. (2013). Neutralizing Tumor-Promoting Chronic Inflammation: A Magic Bullet? Science 339, 286–291. doi:10.1126/science.1232227

Drake, C. G., and Stein, M. N. (2018). The Immunobiology of Kidney Cancer. J. Clin. Oncol. 36, 3547–3552. doi:10.1200/jco.2018.79.2468

Esteller, M. (2011). Non-coding Rnas in Human Disease. Nat. Rev. Genet. 12, 861–874. doi:10.1038/nrg3074

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., et al. (2019). Estimating the Global Cancer Incidence and Mortality in 2018: Globocan Sources and Methods. Int. J. Cancer 144, 1941–1953. doi:10.1002/ijc.31937

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of Rna-Seq Data. Genome Med. 11, 34–40. doi:10.1186/s13073-019-0638-6

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., et al. (2021). GenoCine 2021: Nucleic Acids Res. 49, D916–D923. doi:10.1093/nar/gka1087

Fridman, W., Galon, J., Dieuosme Jean, M., Cremer, L., Fisson, S., Damotte, D., et al. (2011). "Immune Infiltration in Human Cancer: Prognostic Significance and Disease Control," in Cancer Immunology and Immunotherapy. Editor G. Dranoff (Berlin, Heidelberg: Springer Berlin Heidelberg), 1–24.

Gao, M., Guo, Y., Xiao, Y., and Shang, X. (2021). Comprehensive Analyses of Correlation and Survival Reveal Informative Lncrna Predictive Signatures in Colon Cancer. World J. Surg. Oncol. 19, 1–15. doi:10.1186/s12957-021-02196-4

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., et al. (2018). Lnc2Cancer v2.0: Updated Database of Experimentally Supported Long Non-coding RNAs in Human Cancers. Nucleic Acids Res. 47, D1028–D1033. doi:10.1093/nar/gky1096

Ghahri, E., Nazemalhosseini-Mojariad, E., Baghdar, K., Nayeri, Z., Sadeghi, H., Rezasoltani, S., et al. (2021). Identification of a Stool Long Non-coding Rnas Panel as a Potential Biomarker for Early Detection of Colorectal Cancer. J. Clin. Lab. Anal. 35, e23601. doi:10.1002/jcla.23601

Ghahri, E., Anarak, F., Baghdar, K., Gavrielid, P., Sadeghi, H., Nasrabadi, P. N., et al. (2019). Investigating the Diagnostic Performance of Hottip, Pvt1, and uca1 Long Noncoding Rnas as a Predictive Panel for the Screening of Colorectal Cancer Patients with Lymph Node Metastasis. J. Cell Biochem. 120, 14780–14790. doi:10.1002/jcb.28739

Gibb, E. A., Brown, C. J., and Lam, W. L. (2011). The Functional Role of Long Non-coding Rna in Human Carcinomas. Mol. Cancer 10, 38–17. doi:10.1186/1476-4598-10-38

Gong, N., Li, F., Wang, Y., Li, Z., Wang, W., Gong, T., et al. (2021). Hoxc11 Positively Regulates the Long Non-coding Rna Hotair and Is Associated With Poor Prognosis in colon Adenocarcinoma. Exp. Ther. Med. 22, 1–6. doi:10.3892/etm.2021.10745

Hanahan, D., and Coussens, L. M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell 21, 309–322. doi:10.1016/j.cccr.2012.02.022

Hogdall, D., Lewinska, M., and Andersson, J. B. (2018). Desmoplastic Tumor Microenvironment and Immunotherapy in Cholangiocarcinoma. Trends Cancer 4, 239–255.

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T. M., Muljo, S. A., et al. (2013). Expression and Regulation of Intergenic Long Noncoding Rnas during T Cell Development and Differentiation. Nat. Immunol. 14, 1190–1198. doi:10.1038/ni.2712

FUNDING

Thanks for the support of the National Natural Science Foundation of China (grant numbers 61772426, U1811262).

ACKNOWLEDGMENTS

We would like to thank all members of the laboratory for their valuable discussions and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.792541/full#supplementary-material
Zhou, W., Pan, B., and Liu, L. (2019). Integrated Bioinformatics Analysis Revealing Independent Prognostic Long Non-coding RNAs DNAH17-AS1 and RP11-400N13.2 and Their Potential Oncogenic Roles in Colorectal Cancer. *Oncol. Lett.* 18, 3705–3715. doi:10.3892/ol.2019.10730

Zhu, L. F., Song, L. D., Xu, Q., and Zhan, J. F. (2019). Highly Expressed Long Non-coding Rna Fezf1-As1 Promotes Cells Proliferation and Metastasis through Notch Signaling in Prostate Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 23, 5122–5132. doi:10.26355/eurrev_201906_18176

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gao, Liu, Qi, Guo and Shang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.