Using *fanyi* to reduce language barriers, promote information absorption, and communication

Difei Wang[1,2], Guannan Chen[3], Xiao Luo[2], Li Zhan[2], Shuangbin Xu[2,4], Qianwen Wang[2], Junrui Li[1,2], Rui Wang[2], Shaodi Wen[2,5], Guangchuang Yu[2,*]

[1] Nanfang Hospital, Southern Medical University, Guangzhou, 510515, China

[2] Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China

[3] Wu Lien-The Institute & Department of Microbiology, Harbin Medical University, Harbin, 150081, China

[4] Microbiome Medicine Center, Department of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou 510515, China

[5] Department of Oncology, The Affiliated Cancer Hospital of Nanjing Medical University & Jiangsu Cancer Hospital, Nanjing, 210009, China

[*] Correspondence: Guangchuang Yu gcyu1@smu.edu.cn

Abstract

Understanding the latest research advancements and sharing one's own achievements is extremely important for young researchers. However, language barriers restrict the absorption of knowledge and career development. To address these issues, we have developed the fanyi package, which utilizes AI-driven online translation services to accurately translate between multiple languages. This helps researchers rapidly capture information, promotes the preservation of a multilingual environment, and ultimately fosters innovation in the academic community. Interpreting genes of interest (such as mutant genes) contributes to proposing molecular mechanisms, but accessing such information typically requires tedious manual retrieval. The fanyi package also provides tools for retrieving gene information from NCBI, enhancing the accessibility of gene information. Being able to quickly and efficiently obtain these pieces of information, coupled with translation capabilities, can help researchers better understand the information. The fanyi package is freely available via https://github.com/YuLab-SMU/fanyi. The translation functions are not limited to biomedicine research, they can be applied to any other domain as well.

Introduction

English, as the first language of scientific research, plays an important role in disseminating scientific research results. However, this also poses challenges for non-native English-speaking researchers in writing and talking about their research. Many graduate students from non-English speaking countries face more or less obstacles to using English in the academic environment. Early-career researchers may find it challenging to master scientific writing in English in order to effectively convey the significance of their work, which can pose an extra obstacle when aiming to publish their work and advance their careers. In scientific writing, researchers often rely on various tools or language editing services. The application of large language models such as ChatGPT in academic writing has sparked considerable debate and remains controversial [1]. This language barrier is also the reason why many local media and social platforms use their local language to interpret scientific results. With the development of artificial intelligence, text translation has become quite accurate. A large number of students use translation software to convert English materials into their language when reading scientific research papers and also use translation software to translate their domestic language into English when writing papers.

In biomedical research, a common scenario is for researchers to search for one term or gene after another, obtain explanations for the term or gene, and then use translation tools to translate this information into the local language, to quickly capture relevant information for interpretation. This is usually cumbersome and requires a lot of time. For example, when conducting single-cell analysis, the interpretation of marker genes in different cell clusters often requires the aforementioned method. To interpret these genes, as mentioned above, repeating operations for each gene requires a significant amount of time investment. Many researchers choose not to do this and instead directly use the genes for functional enrichment analysis [2]. Only the results of enrichment analysis were used to selectively search for certain genes. Although it saves time, it is also very likely to miss some important genes.

In response to this situation, we have developed the R package, fanyi, which means “translation” in Chinese Pinyin. It can utilize online AI translation services to translate between multiple languages. At present, this package can support users to use Baidu, Bing, Youdao, Volcengine, Caiyun, Tencent, and ChatGLM translation services. With this package, users can translate the text using R language, which is widely used in multiple fields. For example, R is a necessary Swiss army knife for omics data analysis and visualization. With the help of fanyi, analysis results can be directly translated into the local language, and descriptions of functions, pathways, phenotypes, and diseases will become easier to understand. The fanyi package is freely available at https://cran.r-project.org/package=fanyi and https://github.com/YuLab-SMU/fanyi.

Retrieving gene information can be considered a daily task, such as searching for marker genes to better interpret cell types. To better serve researchers in the biomedical field with this translation tool, we have implemented automatic retrieval of gene descriptions. This allows for batch retrieval of analysis results, such as a list of differentially expressed genes, genes with high mutations in tumors, and marker genes for different cell types, etc. By automatically retrieving descriptions and summaries of genes from previous studies, it will greatly save the user time. Moreover, through translation, language barriers can be eliminated. Enable researchers to spend more time thinking

about scientific questions, while fostering a multilinguistic culture in the science community.

Results

Fanyi retrieves marker gene information to assist in interpreting results

One of the main goals of biological research is to identify molecular mechanisms. After obtaining genes of interest, a significant amount of time is typically spent on searching and summarizing the functions of these genes to provide clues for proposing mechanistic hypotheses. The fanyi package can batch retrieve summary descriptions of genes from the NCBI database, greatly simplifying this step. Here, we use a single dataset from the Seurat example [3], containing 2,700 Peripheral Blood Mononuclear (PBMC) cells. After identifying the marker genes of all the cell clusters, we converted the gene symbols to ENTREZ gene IDs using clusterProfiler [2] and then queried gene information using the gene_summary function provided in the fanyi package.

```
require(Seurat)
pbmc.markers <- FindAllMarkers(pbmc, only.pos = TRUE)

library(dplyr)
top2 <- pbmc.markers |>
    group_by(cluster) |>
    filter(avg_log2FC > 1) |>
    arrange(p_val_adj, by_group=TRUE) |>
    slice_head(n = 2) |>
    ungroup()

top2$gene <- sub("\\.\\d", "", top2$gene)
g <- clusterProfiler::bitr(top2$gene,
              fromType = 'SYMBOL',
              toType = 'ENTREZID',
              OrgDb = 'org.Hs.eg.db')

library(fanyi)
gs <- gene_summary(g$ENTREZ)
```

As demonstrated in Figure 1, the information output by FindAllMarkers can be integrated with the output of gene_summary, and all these pieces of information can be exported as a table to help the user understand the marker genes. Combining expression levels, p values, and detailed gene function descriptions of marker genes, researchers can better understand and interpret the cell types and functions of different cell clusters.

gene	avg_log2FC	p_val_adj	description	summary
Naive CD4 T				
CCR7	2.397366	1.312702e-83	C-C motif chemokine receptor 7	The protein encoded by this gene is a member of the G protein-coupled receptor family. This receptor was identified as a gene induced by the Epstein-Barr virus (EBV), and is thought to be a mediator of EBV effects on B lymphocytes. This receptor is expressed in various lymphoid tissues and activates B and T lymphocytes. It has been shown to control the migration of memory T cells to inflamed tissues, as well as stimulate dendritic cell maturation. The chemokine (C-C motif) ligand 19 (CCL19/ECL) has been reported to be a specific ligand of this receptor. Signals mediated by this receptor regulate T cell homeostasis in lymph nodes, and may also function in the activation and polarization of T cells, and in chronic inflammation pathogenesis. Alternative splicing of this gene results in multiple transcript variants. [provided by RefSeq, Sep 2014]
LDHB	1.206019	5.137444e-108	lactate dehydrogenase B	This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post-glycolysis process. Alternatively spliced transcript variants have been found for this gene. Recent studies have shown that a C-terminally extended isoform is produced by use of an alternative in-frame translation termination codon via a stop codon readthrough mechanism, and that this isoform is localized in the peroxisomes. Mutations in this gene are associated with lactate dehydrogenase B deficiency. Pseudogenes have been identified on chromosomes X, 5 and 13. [provided by RefSeq, Feb 2016]
CD14+ Mono				
LGALS2	5.651790	0.000000e+00	galectin 2	The protein encoded by this gene is a soluble beta-galactoside binding lectin. The encoded protein is found as a homodimer and can bind to lymphotoxin-alpha. A single nucleotide polymorphism in an intron of this gene can alter the transcriptional level of the protein, with a resultant increased risk of myocardial infarction. [provided by RefSeq, Jul 2008]
S100A8	6.645056	0.000000e+00	S100 calcium binding protein A8	The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. S100 genes include at least 13 members which are located as a cluster on chromosome 1q21. This protein may function in the inhibition of casein kinase and as a cytokine. Altered expression of this protein is associated with the disease cystic fibrosis. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jan 2016]

Figure 1. Table of marker gene information. As a demonstration, only the top two marker genes belonging to two cell clusters (Naïve CD4 T and CD14+ Mono) are displayed.

By combining information/knowledge, fanyi helps create easy-to-understand infographics

Infographics contain data visualization such as bar charts and descriptive text representing information, making the conveyed information clearer and more quickly understood. The gene description and summary information queried by the gene_summary function can be used in data visualization to provide a more informative graph.

```
library(ggsc)
library(ggplot2)
library(ggfun)
library(aplot)

d <- gs[1:4, ]
d$name <- factor(d$name, levels = d$name)
p <- ggplot(d, aes(x=1, y=3)) +
    geom_text(aes(label=str_wrap(summary, 75)), hjust=0) +
    xlim(1,2) +
    facet_wrap(~name, ncol=1) +
    theme_noaxis()+
    xlab(NULL) + ylab(NULL)

p2 <- sc_violin(pbmc, d$name, ncol=1) +
    theme(legend.position='none',
        axis.text.x = element_text(angle=60, hjust=1)) +
```

```
    xlab(NULL) + ylab(NULL)

pa <- insert_right(p2, p, width=1.5)
```

The above source codes employ the in-house developed package, ggsc, to visualize the expression distribution of the selected marker genes among different cell clusters as violin plots and display the gene summary side by side to help the user interpret the results as demonstrated in Figure 2A.

```
pb <- sc_dim(pbmc) +
  sc_dim_geom_label(geom = ggrepel::geom_text_repel, color = "black",
                    bg.color = "white") +
    theme(legend.position = "none")

pc <- sc_feature(pbmc, features='CD3E', density = TRUE) +
    labs(subtitle = gs$description,
         caption=str_wrap(gs$summary, 180)) +
    scale_color_viridis_c() +
    theme(legend.position = c(.95, .2))
```

We further use ggsc to visualize the pbmc dataset. Figure 2B illustrates the cell clustering results and Figure 2C visualizes the expression of the CD3E gene in different cells as a density plot. With the gene information provided, we can display the gene description as a plot subtitle and the gene summary information as a plot caption. The information provided in Figure 2C helps us to capture the idea of how the CD3E gene functions and to interpret the expression distribution among different cell clusters.

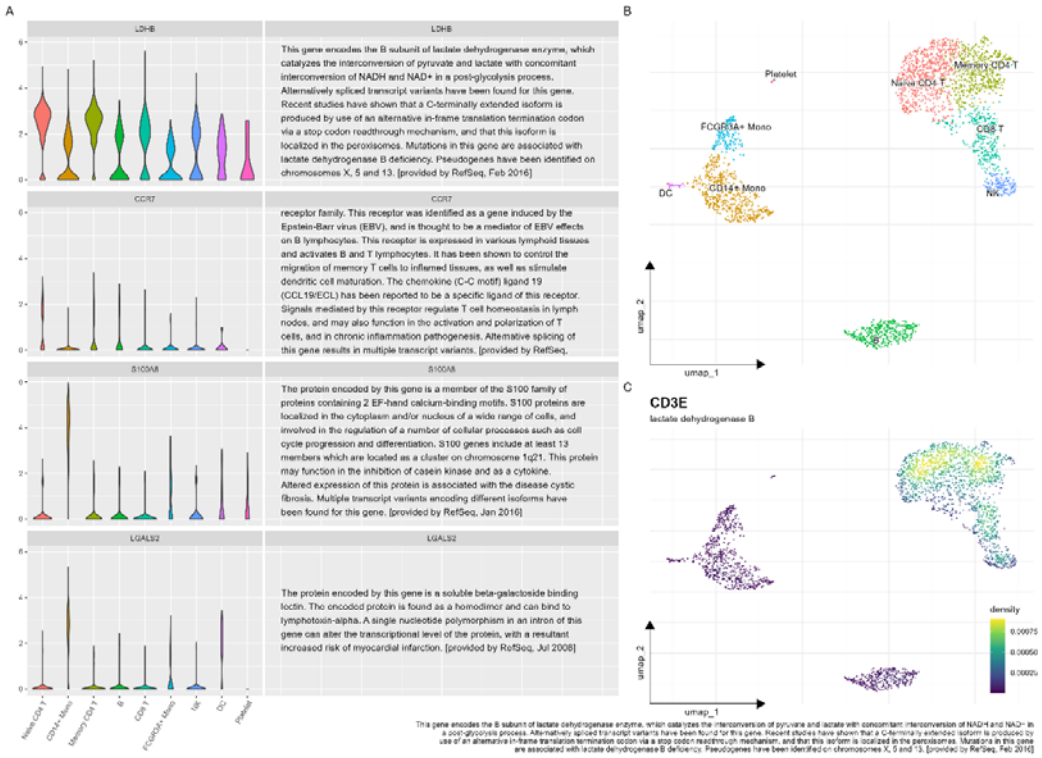


Figure 2. Use gene information to create an informative graph that contains a visualization of data and biological knowledge. (A) Violin graphs display gene expression distributions among different cell clusters with gene summary information side by side to explain the gene functions. (B) Cell type annotation on a UMAP plot. (C) Density plot to visualize the expression levels of the CD3E gene across different cells on a UMAP with the information of full gene name and summary of gene function.

Fanyi uses AI-driven translation services to help researchers quickly digest information

The language barrier makes it difficult for professional content to be quickly understood and absorbed, while also hindering researchers from effectively expressing and conveying information. The fanyi package, with the assistance of multiple AI-driven translations, can effectively perform batch and automated conversion of information across multiple languages. In this example, we used an RNA-Seq dataset of four human airway smooth muscle cells (GEO: GSE52778) [4]. This dataset can be accessed via the Bioconductor package, airway.

```
library(DESeq2)
library(dplyr)
library(clusterProfiler)
library(fanyi)
library(ggtree)

data(gse, package='airway')
```

```
dds <- DESeqDataSet(gse, design = ~donor + condition)
dds <- DESeq(dds)
res <- results(dds)

res2 <- as.data.frame(res)
res2$ID <- rownames(res)
res2 <- res2[!is.na(res2$log2FoldChange), ]
de <- arrange(res2, padj) |>
    group_by(sign(log2FoldChange)) |>
    slice(1:20) |>
    pull(ID)

mat <- assay(rlog(dds))[de, ]
mat <- mat - rowMeans(mat)

de <- sub("\\.\\d+", "", de)
rownames(mat) <- de
d <- as.data.frame(mat)
k <- kmeans(t(mat), 2)
d <- d[, names(sort(k$cluster))]
```

As shown in the above code, we employed DESeq2 [5] to identify differentially expressed genes. We extracted 20 genes from up-regulated and down-regulated genes respectively. These 40 genes were further used to demonstrate how fanyi can help interpret the results.

```
g <- bitr(de, 'ENSEMBL', c('ENTREZID', 'SYMBOL', 'GENENAME'), 'org.Hs.eg.db')
g$gn2 <- translate(g$GENENAME)

hc <- hclust(dist(mat))
p <- ggtree(hc)

p <- p %<+% g
p <- p + geom_tiplab(aes(label=SYMBOL), align=T)
p2 <- gheatmap(p, d, offset=3, colnames_angle=90, colnames_position="top",
 colnames_offset_y=2)

pg <- p2 +
    geom_tiplab(aes(label = gn2),
                align=TRUE, offset=13, linetype=NA) +
    hexpand(.6) + vexpand(.05) +
    theme(legend.position = c(.1, .9))
```

We used bitr function to convert gene IDs and used the translate function to convert the full gene name into Chinese. We then applied the hierarchical clustering method to cluster the genes based on the expression profiles. The clustering tree was visualized using ggtree [6–8] with gene symbols and then gheatmap was utilized to visualize the tree with the expression matrix [9]. Finally, we displayed the Chinese version of the full gene names to help users better understand the functions of these differential genes (Figure 3A). With fanyi, users can translate information into their local language with more than two hundred languages supported and display the information with the English version side by side. We no longer need to translate professional vocabulary manually. It saves a lot of time and helps users to quickly capture the information when looking at the figure.

As demonstrated in Figure 3A, the genes can be divided into two clades, we applied cuttree function to separate the genes into two clusters and then employed clusterProfiler [2,10] to characterize biological processes regulated by these two gene clusters as demonstrated in Figure 3B. In fanyi, we provide a helper function, translate_ggplot, to help users translate words and sentences in a ggplot graph. As demonstrated in Figure 3C, the biological process terms presented in Figure 3B are translated into Chinese with only one line of function call.

```
x <- cutree(hc, k=2)
gc <- split(g$ENTREZID, x[g$ENSEMBL])
bp <- compareCluster(gc, fun=enrichGO, ont="BP", OrgDb='org.Hs.eg.db')

dp1 <- dotplot(bp)
dp2 <- translate_ggplot(p1, 'y')
```

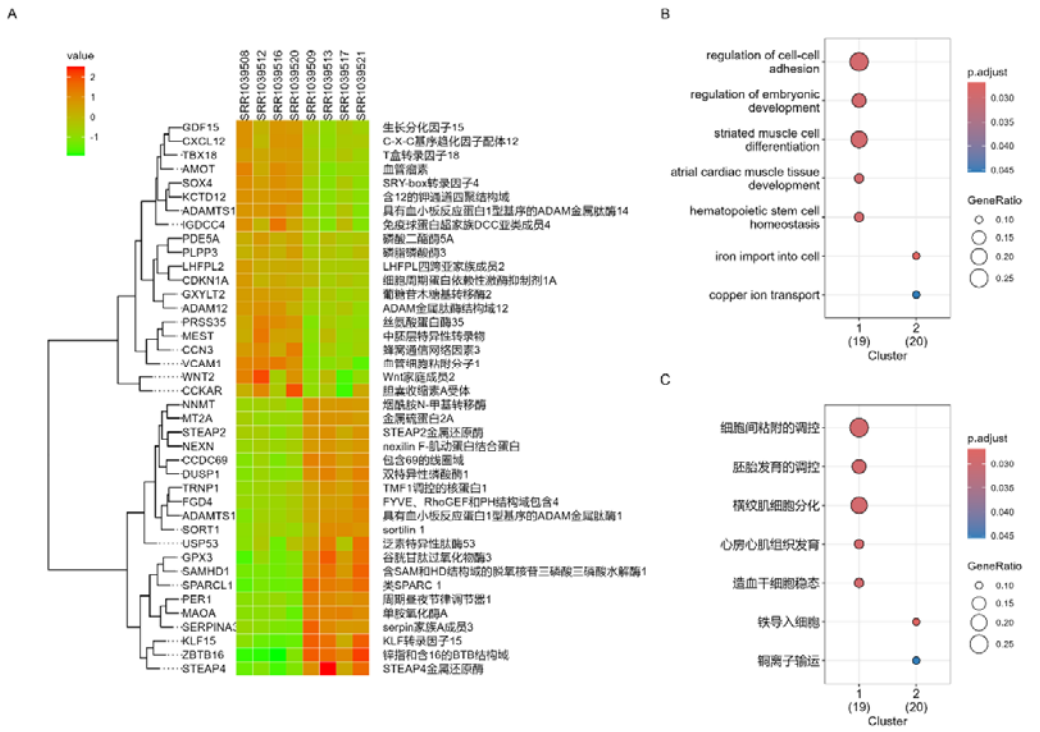


Figure 3. Translated information in the local language helps users quickly capture and understand the information presented. (A) A heatmap visualizes differential expressed genes across different samples with gene symbols and translated full gene names. (B) Functional characterization of differentially expressed genes with different expression patterns. (C) A translated version of (B).

Discussion

Language barrier has serious consequences for science and science education, including inequality for under-represented communities, difficulties in disseminating non-English knowledge, and challenges in uptaking knowledge in educational settings [11]. Language shapes the way we think [12]. People speaking different languages tend to think in different ways, which can bring new perspectives and ideas. The diversity of languages helps foster innovation and creativity, which is exactly what the academic community needs.

The fanyi package helps lower the language barrier by employing various AI-driven online translation services. In particular, information can be translated into the local language to help quickly uptake the information, non-English-language information can be translated into English to increase the visibility of the content, and existing resources and new discoveries can be disseminated in multiple languages to communicate effectively to non-English speakers. This will promote the preservation of a multilingual environment in the academic community.

The fanyi package presented here can also help retrieve and interpret biological information. It not only makes the gene information accessible but also saves researchers a significant amount of time.

Querying gene information can be used to quickly create a summarized table with rich information for marker genes, differential expression genes, highly mutated genes, etc., to help users digest how the genes work. The information query by the package can also be utilized to annotate the plot to create an informative graph with more or less self-explanation.

Acknowledgments

This study was supported by the National Natural Science Foundation of China [32270677]. The manuscript was drafted with the help of the fanyi package. We appreciate the feedback and support from fanyi users.

References

1. Birhane, A., Kasirzadeh, A., Leslie, D., and Wachter, S. (2023). Science in the age of large language models. Nat. Rev. Phys. *5*, 277–280. 10.1038/s42254-023-00581-4.

2. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation, 100141. 10.1016/j.xinn.2021.100141.

3. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. 10.1016/j.cell.2021.04.048.

4. Himes, B.E., Jiang, X., Wagner, P., Hu, R., Wang, Q., Klanderman, B., Whitaker, R.M., Duan, Q., Lasky-Su, J., Nikolos, C., et al. (2014). RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. PloS One *9*, e99625. 10.1371/journal.pone.0099625.

5. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.

6. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36. 10.1111/2041-210X.12628.

7. Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. Curr. Protoc. Bioinforma. *69*, e96. 10.1002/cpbi.96.

8. Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T.T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. iMeta *1*, e56. 10.1002/imt2.56.

9. Yu, G., Lam, T.T.-Y., Zhu, H., and Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. Mol. Biol. Evol. *35*, 3041–3043.

10.1093/molbev/msy194.

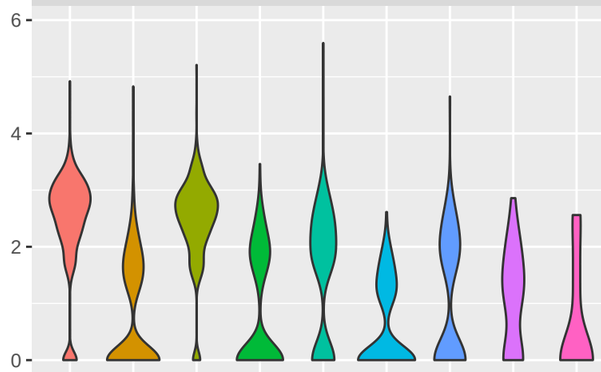
10. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287. 10.1089/omi.2011.0118.

11. Amano, T., Rios Rojas, C., Boum II, Y., Calvo, M., and Misra, B.B. (2021). Ten tips for overcoming language barriers in science. Nat. Hum. Behav. *5*, 1119–1122. 10.1038/s41562-021-01137-1.

12. Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. Cognit. Psychol. *43*, 1–22. 10.1006/cogp.2001.0748.

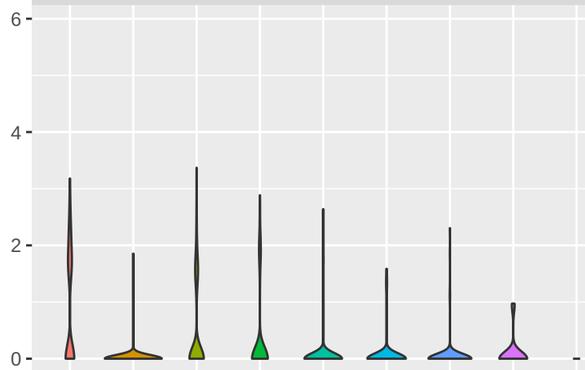
gene	avg_log2FC	p_val_adj	description	summary
Naive CD4 T				
CCR7	2.397366	1.312702e-83	C-C motif chemokine receptor 7	The protein encoded by this gene is a member of the G protein-coupled receptor family. This receptor was identified as a gene induced by the Epstein-Barr virus (EBV), and is thought to be a mediator of EBV effects on B lymphocytes. This receptor is expressed in various lymphoid tissues and activates B and T lymphocytes. It has been shown to control the migration of memory T cells to inflamed tissues, as well as stimulate dendritic cell maturation. The chemokine (C-C motif) ligand 19 (CCL19/ECL) has been reported to be a specific ligand of this receptor. Signals mediated by this receptor regulate T cell homeostasis in lymph nodes, and may also function in the activation and polarization of T cells, and in chronic inflammation pathogenesis. Alternative splicing of this gene results in multiple transcript variants. [provided by RefSeq, Sep 2014]
LDHB	1.206019	5.137444e-108	lactate dehydrogenase B	This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post-glycolysis process. Alternatively spliced transcript variants have been found for this gene. Recent studies have shown that a C-terminally extended isoform is produced by use of an alternative in-frame translation termination codon via a stop codon readthrough mechanism, and that this isoform is localized in the peroxisomes. Mutations in this gene are associated with lactate dehydrogenase B deficiency. Pseudogenes have been identified on chromosomes X, 5 and 13. [provided by RefSeq, Feb 2016]
CD14+ Mono				
LGALS2	5.651790	0.000000e+00	galectin 2	The protein encoded by this gene is a soluble beta-galactoside binding lectin. The encoded protein is found as a homodimer and can bind to lymphotoxin-alpha. A single nucleotide polymorphism in an intron of this gene can alter the transcriptional level of the protein, with a resultant increased risk of myocardial infarction. [provided by RefSeq, Jul 2008]
S100A8	6.645056	0.000000e+00	S100 calcium binding protein A8	The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. S100 genes include at least 13 members which are located as a cluster on chromosome 1q21. This protein may function in the inhibition of casein kinase and as a cytokine. Altered expression of this protein is associated with the disease cystic fibrosis. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jan 2016]

A



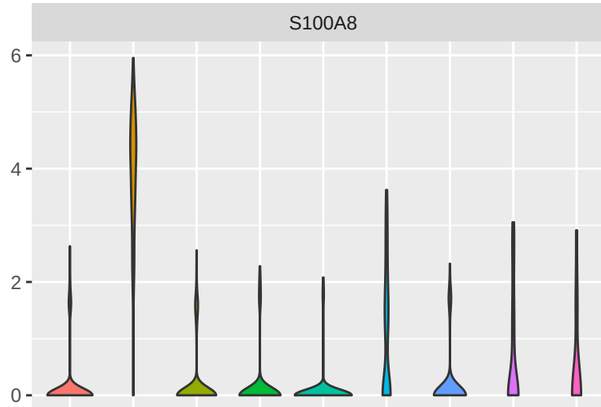
LDHB

This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post−glycolysis process. Alternatively spliced transcript variants have been found for this gene. Recent studies have shown that a C−terminally extended isoform is produced by use of an alternative in−frame translation termination codon via a stop codon readthrough mechanism, and that this isoform is localized in the peroxisomes. Mutations in this gene are associated with lactate dehydrogenase B deficiency. Pseudogenes have been identified on chromosomes X, 5 and 13. [provided by RefSeq, Feb 2016]



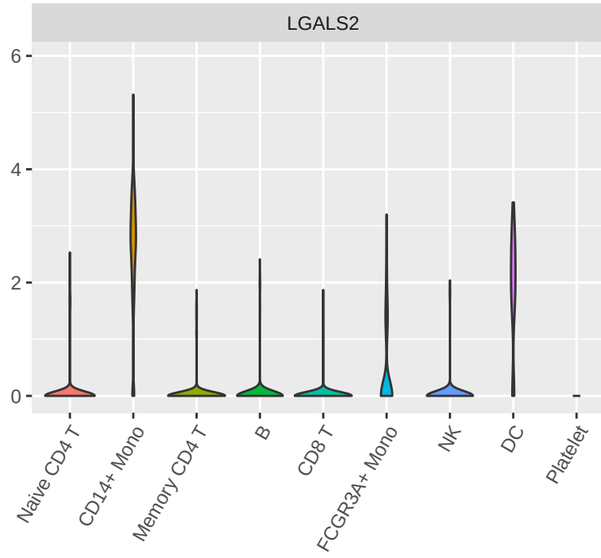
CCR7

receptor family. This receptor was identified as a gene induced by the Epstein−Barr virus (EBV), and is thought to be a mediator of EBV effects on B lymphocytes. This receptor is expressed in various lymphoid tissues and activates B and T lymphocytes. It has been shown to control the migration of memory T cells to inflamed tissues, as well as stimulate dendritic cell maturation. The chemokine (C−C motif) ligand 19 (CCL19/ECL) has been reported to be a specific ligand of this receptor. Signals mediated by this receptor regulate T cell homeostasis in lymph nodes, and may also function in the activation and polarization of T cells, and in chronic inflammation pathogenesis. Alternative splicing of this gene results in multiple transcript variants. [provided by RefSeq,



S100A8

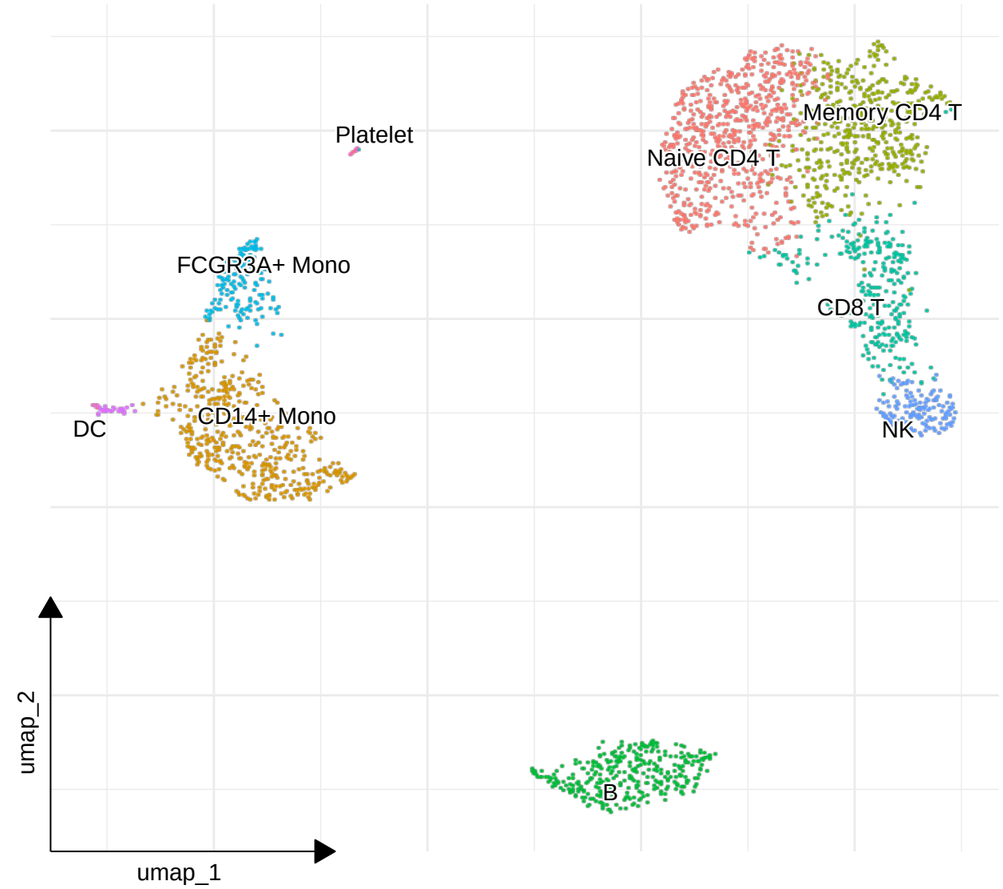
The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF−hand calcium−binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. S100 genes include at least 13 members which are located as a cluster on chromosome 1q21. This protein may function in the inhibition of casein kinase and as a cytokine. Altered expression of this protein is associated with the disease cystic fibrosis. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jan 2016]



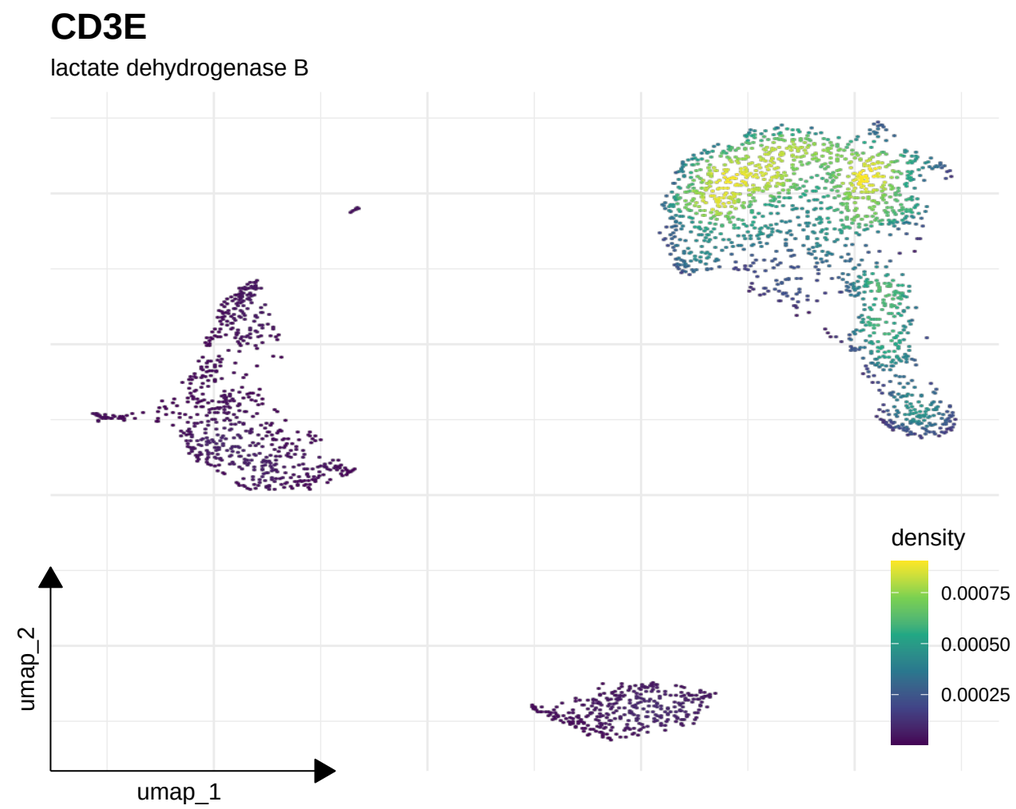
LGALS2

The protein encoded by this gene is a soluble beta−galactoside binding lectin. The encoded protein is found as a homodimer and can bind to lymphotoxin−alpha. A single nucleotide polymorphism in an intron of this gene can alter the transcriptional level of the protein, with a resultant increased risk of myocardial infarction. [provided by RefSeq, Jul 2008]

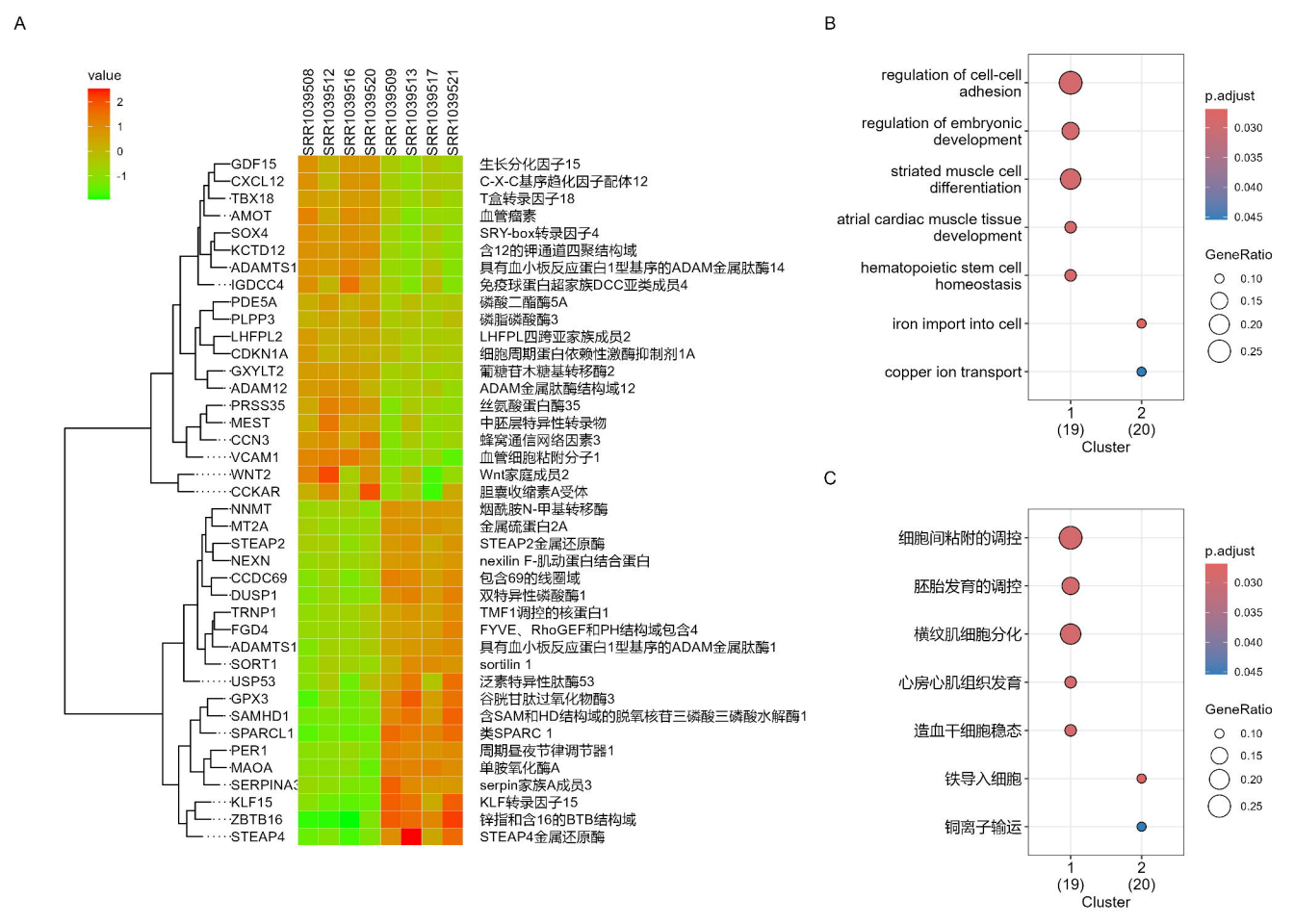
B



C



This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post−glycolysis process. Alternatively spliced transcript variants have been found for this gene. Recent studies have shown that a C−terminally extended isoform is produced by use of an alternative in−frame translation termination codon via a stop codon readthrough mechanism, and that this isoform is localized in the peroxisomes. Mutations in this gene are associated with lactate dehydrogenase B deficiency. Pseudogenes have been identified on chromosomes X, 5 and 13. [provided by RefSeq, Feb 2016]


A
value
2
1
0
-1
SRR1039508
SRR1039512
SRR1039516
SRR1039520
SRR1039509
SRR1039513
SRR1039517
SRR1039521
GDF15 生长分化因子15
CXCL12 C-X-C基序趋化因子配体12
TBX18 T盒转录因子18
AMOT 血管瘤素
SOX4 SRY-box转录因子4
KCTD12 含12的钾通道四聚结构域
ADAMTS1 具有血小板反应蛋白1型基序的ADAM金属肽酶14
IGDCC4 免疫球蛋白超家族DCC亚类成员4
PDE5A 磷酸二酯酶5A
PLPP3 磷脂磷酸酶3
LHFPL2 LHFPL四跨亚家族成员2
CDKN1A 细胞周期蛋白依赖性激酶抑制剂1A
GXYLT2 葡糖苷木糖基转移酶2
ADAM12 ADAM金属肽酶结构域12
PRSS35 丝氨酸蛋白酶35
MEST 中胚层特异性转录物
CCN3 蜂窝通信网络因素3
VCAM1 血管细胞粘附分子1
WNT2 Wnt家庭成员2
CCKAR 胆囊收缩素A受体
NNMT 烟酰胺N-甲基转移酶
MT2A 金属硫蛋白2A
STEAP2 STEAP2金属还原酶
NEXN nexilin F-肌动蛋白结合蛋白
CCDC69 包含69的线圈域
DUSP1 双特异性磷酸酶1
TRNP1 TMF1调控的核蛋白1
FGD4 FYVE、RhoGEF和PH结构域包含4
ADAMTS1 具有血小板反应蛋白1型基序的ADAM金属肽酶1
SORT1 sortilin 1
USP53 泛素特异性肽酶53
GPX3 谷胱甘肽过氧化物酶3
SAMHD1 含SAM和HD结构域的脱氧核苷三磷酸三磷酸水解酶1
SPARCL1 类SPARC 1
PER1 周期昼夜节律调节器1
MAOA 单胺氧化酶A
SERPINA3 serpin家族A成员3
KLF15 KLF转录因子15
ZBTB16 锌指和含16的BTB结构域
STEAP4 STEAP4金属还原酶
B
regulation of cell-cell adhesion
regulation of embryonic development
striated muscle cell differentiation
atrial cardiac muscle tissue development
hematopoietic stem cell homeostasis
iron import into cell
copper ion transport
1 (19)
2 (20)
Cluster
p.adjust
0.030
0.035
0.040
0.045
GeneRatio
0.10
0.15
0.20
0.25
C
细胞间粘附的调控
胚胎发育的调控
横纹肌细胞分化
心房心肌组织发育
造血干细胞稳态
铁导入细胞
铜离子输运
1 (19)
2 (20)
Cluster
p.adjust
0.030
0.035
0.040
0.045
GeneRatio
0.10
0.15
0.20
0.25