Features of functional annotation of rheumatoid arthritis susceptibility genes by Cytoscape

Chasovskikh N.Yu.

Siberian State Medical University
2, Moscow Trakt, Tomsk, 634050, Russian Federation

ABSTRACT

Aim. To evaluate the functional annotation of genes associated with rheumatoid arthritis with different parameters of the ClueGO Cytoscape tool.

Materials and methods. Genes of susceptibility to rheumatoid arthritis were extracted from publicly available database GWAS (catalog of associations of single nucleotide polymorphisms with diseases). The Gene Ontology (GO), the functional annotation of genes, was performed using Cytoscape ClueGO. The features of the functional annotation using the plugin ClueGO Cytoscape were analyzed.

Results. Depending on the initial parameters specified in the plugin, the grouping of terms according to the gene ontology was carried out with a different degree of generalization. A smaller minimum number of genes in a group allows to form a larger number of groups, which makes it possible to obtain more detailed functional characteristics.

Conclusion. The results obtained with different grouping options can be useful for further studies of genetic mechanisms of rheumatoid arthritis.

Key words: functional annotation of genes, rheumatoid arthritis, GWAS, Cytoscape.

Особенности функциональной аннотации генов предрасположенности к ревматоидному артриту при использовании Cytoscape

Часовских Н.Ю.

Сибирский государственный медицинский университет (СибГМУ)
Россия, 634050, г. Томск, Московский тракт, 2

РЕЗЮМЕ

Цель. Оценить функциональную аннотацию генов, ассоциированных с ревматоидным артритом, при разных параметрах инструмента ClueGO Cytoscape.

Материалы и методы. Гены предрасположенности к ревматоидному артриту были извлечены из публичной базы данных GWAS (каталог ассоциаций однонуклеотидных полиморфизмов с заболеваниями). Генная онтология (Gene Ontology, GO) – функциональная аннотация генов была произведена с помощью алгоритма, реализованного в Cytoscape ClueGO.

Результаты. Рассмотрены особенности выполнения функциональной аннотации с помощью плагина ClueGO при разных условиях формирования функциональных групп. В зависимости от исходных пара-

---

✉ *Chasovskikh Nataliya Yu.,* e-mail: natacch@mail.ru.

метров, задаваемых в плагине, группировка терминов согласно генной онтологии осуществляется с различной долей обобщения. Меньшее минимальное число генов в группе позволяет сформировать большее число групп, что дает возможность получить более детальную функциональную характеристику.

Заключение. Полученные варианты результатов функциональной аннотации могут быть востребованы для дальнейших исследований генетических механизмов ревматоидного артрита.

INTRODUCTION

The increase in the volume of genomic and proteomic research has led to the need to accumulate its results in specialized databases, including publicly available online resources. A huge role in the systematization and description of such data is played by gene ontology (GO, GeneOntology), a unified universal hierarchical terminology system [1]. It allows for the characterization of data into such sections as biological processes, molecular functions, and cellular components [2], in accordance with which the annotation of genes (proteins) is performed. For GO capabilities to be applicable to a specific data set, specialized tools exist, such as ClueGO Cytoscape [3], which allows simultaneous work with multiple data lists and networks.

The ability to describe the results of studies in terms of gene ontology plays an important role in the implementation of the functional characteristics of thousands of genes (for example, in microchipping). Grouping them in a certain way allows researchers to evaluate the possible contribution of the studied genes to the implementation of the physiological response or etiopathogenesis of diseases. This approach is relevant for studying the genetic factors of multifactorial (complex) pathologies [4], in particular, rheumatoid arthritis, regarding which no similar studies (based on the results of a genome-wide search) have been conducted.

The aim of this study was to evaluate the functional annotation of genes which are associated with rheumatoid arthritis with different parameters of the ClueGO Cytoscape tool.

MATERIALS AND METHODS

Rheumatoid arthritis susceptibility genes were obtained by analyzing data from an information source containing the results of genome-wide association studies (GWAS) [5], namely, associations between single nucleotide polymorphisms and diseases. To carry out functional analysis, ontology types corresponding to the GeneOntology service were selected in the ClueGO Cytoscape version 3.2.1 plugin. ClueGO allows for updates of this information, since GeneOntology is constantly updated [6].

The functional similarity of the genes was evaluated using a hypergeometric test, and the genes belonged to specific functions in terms of gene ontology. Additionally, the plugin allows users to adjust the minimum number (percentage) of genes used to form groups, by default, 3 and 4, respectively. On a positive scale from 0 to 1, the level of Cohen's kappa coefficient was established, reflecting the functional relationships between genes (0.4). In case of testing a large number of hypotheses, ClueGO allows for correction for *P* (probability of committing a type I error) using the Bonferroni and Benyamini – Hochberg methods [7]. At the same time, both *P* values for each term are presented in the Table.

Functional groups were created by iteratively merging initially defined groups based on a predetermined kappa threshold value. The program suggests choosing a "leading" term in each group according to their statistical significance and number or percentage of genes.

RESULTS AND DISCUSSION

A list of 94 genes associated with rheumatoid arthritis was formed using the GWAS resource: *ACOXL, AIRE, ALS2CR12, ANAPC4, ANO8, ANXA3, APOM, ARHGEF3, ARID5B, BAG6, BLK, BTNL2, C1orf159, C2,C5orf30, CBFB, CCL21, CCR6, CD226, CD247, CD40, CDK5RAP2, CDK6, CLYBL, CMKLR1, CTTNBP2, DDA1, DPP4, EOMES, FADS2, FAM107A, FAM98B, GATA3, GATSL3, GCH1, GRM5, GUCY1B2, HNRNPA1, IFI16, IL2RA, IL2RB, IL6R, IRF5, JAZF1, KIAA1109, KIF5A, LLGL1, MECP2, MMEL1, MOV10L1, MTF1, MTG1, MTM1, NFKBIE,*

OS9, PDE2A, PDGFA, PHF19, PHTF1, PLCL2, PLD4, PPIL4, PPM1D, PPME1, PRKCH, PTPN2, PTPN22, RABEP1, RAD51B, RASGRP1, REL, RPP14, RTKN2, SFTPD, SPTBN2, STAG1, STAT4, SUOX, SYNGR1, TEC, TIMMDC1, TNFAIP3, TNPO3, TRHDE, TYK2, UBASH3A, UBE2F, VTCN1, WDFY4, WNT10B, YDJC, ZBTB12, ZNF133, ZNF774.

The following types of ontologies were used to classify the genes: GO_ImmuneSystemProcess (immune system process), GO_Molecular Function (molecular function), GO_CellularComponent (cellular component), and GO_Biological Process (biological process). The minimum GO Level was 3, the maximum GO Level was 8.

Functional analysis of genes (for a predetermined minimum number of genes in groups 2, $p < 0.05$) revealed 8 groups of genes in accordance with the terms of gene ontology:

1) regulation of the production of interleukin (IL)-2 (includes 22 functions);

2) the signaling pathway of IL-2 (includes 15 functions);

3) antigen receptor-mediated signaling pathway (includes 12 functions);

4) production of IL-12 (includes 5 functions);

5) positive regulation of the G2 /M transition of the mitotic cell cycle (includes 2 functions);

6) regulation of neuronal synaptic plasticity (includes 2 functions);

7) positive regulation of cytotoxicity associated with natural killer cells (includes 2 functions);

8) regulation of the respiratory processes (includes 2 functions).

In addition, 6 functions that were not merged were identified:

1) the signaling pathway of IL-6;

2) chemotaxis of dendritic cells;

3) neuromuscular control of body position;

4) negative regulation of the innate immune response;

5) response to muramyl dipeptide;

6) regulation of platelet activation.

In order to determine the most informative results in terms of biological interpretation, a functional analysis of the genes associated with rheumatoid arthritis was carried out with a different minimum number of genes in groups (3 genes), $p < 0.05$.

As a result of gene grouping, 5 groups were obtained:

1) the signaling pathway of IL-2 (includes 16 functions);

2) antigen receptor-mediated signaling pathway (includes 12 functions);

3) negative regulation of cell activation (includes 11 functions);

4) production of IL-12 (includes 3 functions);

5) regulation of neuronal synaptic plasticity (includes 2 functions).

Four functions were represented in separate terms (without association):

1) negative regulation of the innate immune response;

2) respiratory gas exchange;

3) regulation of histone methylation;

4) platelet-derived growth factor receptor signaling pathway.

The results obtained indicate that depending on the initial parameters specified in the ClueGO Cytoscape plugin, the grouping of gene ontology terms associated with genes is carried out with a different degree of generalization.

In the first case (with the minimum number of genes in group 2), a larger number of groups was formed compared to those in the second case (with a minimum number of genes in group 3), which made it possible to obtain a more detailed functional characteristic.

Moreover, for some functions identified in both research options, the number of genes in the groups was smaller in the first case compared to the second one. So, for the production of IL-12, the corresponding number of genes was 5 (*CD40, CMKLR1, IRF5, NFKB1, REL*) and 6 (*CD40, CMKLR1, IRF5, NFKB1, REL, TRAF3*), respectively. For the IL-2 signaling pathway, the number of genes was 14 (*BPI, BTNL2, CCL21, CCR6, CMKLR1, GATA3, IL2RA, IL2RB, IL6R, NFKB1, PDGFA, PTGIS, PTPN2, PTPN22, RORA, SFTPD, TNFAIP3, UBASH3A)* and 19 *(BPI, BTNL2, CCL21, CCR6, CMKLR1, GATA3, IL2RA, IL2RB, IL6R, NFKB1, PDGFA, PTGIS, PTPN2, PTPN22, RORA, SFTPD, TNFAIP3, UBASH3A, VTCN1*), respectively. At the same time, for the functions of the antigen receptor-mediated signaling pathway and regulation of neuronal synaptic plasticity, the number of genes remained the same in both variants of the study.

With detailed functional annotation in the first version of the study, the presence of rheumatoid arthritis susceptibility genes in the following functions of the immune response regulation was revealed: the signaling pathway of IL-6, which is the key cytokine responsible for auto-immune inflammation [8, 9]; regulation of chemotaxis of dendritic cells; response to muramyl dipeptide (an element of the bacterial cell wall that activates both innate and acquired immunity). In addition, the affiliation of genes to the functions of positive regulation of the G2 /M transition of the mitotic cell cycle and neuromuscular control of body position was determined.

The enlarged functional groups obtained in the second case reflect the general pattern characteristic of the previous result: the participation of genes in the signa-

ling pathways of IL-2 and IL-12 was revealed. IL-2 is known to play an essential role in the development of the immune response, as it stimulates killer cells [10]. IL-12 has pronounced pro-inflammatory properties and increases the activity of natural killer cells and dendritic cells, linking the innate and acquired immunity through the combined effect [10–12]. In addition, the functional group of regulation of neuronal synaptic plasticity indicates a possible effect of genes on the process of neuronal processing of the synaptic signal.

CONCLUSION

The results obtained indicate that with rheumatoid arthritis, susceptibility genes affect not only the implementation of the immune response mediated by signaling of pro-inflammatory cytokines (interleukin-2, -6, -12) and regulation of immunocytes, but also the functions of the nervous system; in particular, synaptic signal processing and neuromuscular body position control.

To study the possible mechanisms of diseases or physiological processes, details regarding the involvement of individual signaling pathways and cellular responses may be important. To do this, it is advisable to change the minimum number of genes that are combined into a functional group towards reduction (compared to the default value in the plugin). At the same time, enlarged groups of functional gene characteristics demonstrate greater clarity when identifying general trends in biological processes.

REFERENCES

1. Blake A., Christie K.R., Dolan M.E. Gene Ontology Consortium: going forward. *Nucleic Acids Research.* 2015; 43 (1): 1049–1056. DOI: 10.1093/nar/gku1179.
2. Ashburner M., Ball C., Blake J., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000; 25 (1): 25–29. DOI: 10.1093/nar/gku1179.
3. Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W.H., Pagès F., Trajanoski Z., Galon J. ClueGO: a Cytoscape plug-into decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009; 25 (8): 1091–1093. DOI: 10.1093/bioinformatics/btp101.
4. Hindorff L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S., Manolio T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS.* 2009; 106 (23): 9362–9367. DOI: 10.1073/pnas.0903103106.
5. Welter D., MacArthur J., Morales J., Burdett T., Hall P., Junkins H., Klemm A., Flicek P., Manolio T., Hindorff L., Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research.* 2014; 42 (1): 1001–1006. DOI: 10.1093/nar/gkt1229.
6. A catalog of published genome-wide association studies. URL: http://www.genome.gov/gwastudies/.
7. Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* 1995; 57 (1): 289–300. DOI: 10.2307/2346101.
8. Udachkina E.V., Novikova D.S., Popkova T.V., Nasonov E.L. Te role of interleukin 6 in development of atherosclerosis in rheumatoid arthritis. *Modern Reumatology.* 2013; 7 (3): 25–32 (in Russ.).
9. Tanaka T., Narazaki M., Kishimoto T. IL-6 in inflammation, immunity, and disease. *Cold Spring Harb. Perspect. Biol.* 2014; 6 (10): a016295. DOI: 10.1101/cshperspect.a016295.
10. Ross S.H., Cantrell D.A. Signaling and function of interleukin-2 in T lymphocytes. *Annu. Rev. Immunol.* 2018; (36): 411–433. DOI: 10.1146/annurev-immunol-042617-053352.
11. Novikov A.A., Aleksandrova E.N., Diatroptova M.A., Nasonov E.L. The role of cytokines in the pathogenesis of rheumatoid arthritis. *Research and Practical Rheumatology.* 2010; 48 (2): 71–82 (in Russ.).
12. Vignali D.A., Kuchroo V.K. IL-12 family cytokines: immunological playmakers. *Nat. Immunol.* 2012; 13 (8): 722–728. DOI: 10.1038/ni.2366.

Author information

Chasovskikh Nataliya Yu., Dr. Sci. (Med.), Professor, Department of Medical and Biological Cybernetics, Siberian State Medical University, Tomsk, Russian Federation. ORCID 0000-0001-6077-0347.

(✉) **Chasovskikh Nataliya Yu.,** e-mail: natacch@mail.ru.