

Research on subject profile of stem cell based on knowledge graph

Introduction

With the expansion of research fields and rapid increase of scientific literatures, knowledge structures of subjects become more and more complex. Subject profile can reveal the multi-level and multidimensional knowledge structures of subjects like user profile characterizes users. As a new knowledge organization technology, knowledge graph can be used to construct subject profile.

Stem cell and regenerative medicine is the frontier of life science, and it is expected to become the third way to treat diseases after drug therapy and surgery. Based on multi-source scientific literatures and information, a domain knowledge graph of stem cell was constructed. Based on the knowledge graph, we are designing and constructing subject profile of stem cell, which is of great significance to help decision-makers and researchers to grasp the overall research of stem cell at different levels and perspectives.

Methodology

1. *Constructing stem cell knowledge graph*

First, the “produce, study, research and clinical” scientific literatures and information of stem cell are integrated from multiple information sources, such as papers, patents, reports, clinical trials, drugs, products, projects, etc. Then, knowledge entities are extracted and annotated from these literatures and information using SemRep tool and UMLS and the relationship among those knowledge entities are calculated by data mining techniques. Finally, we linked these knowledge entities and literatures to construct stem cell knowledge graph, which achieved multi-dimensional knowledge retrieval and navigation. The knowledge graph can also provide triple data for knowledge discovery or other applications.

2. *Profiling the stem cell subject*

Based on the knowledge graph, we are planning to construct a multi-dimensional and fine-grained subject profile of stem cell from the perspectives of subject knowledge structure, development trend, research topics, key drugs and major diseases, etc. This research is ongoing.

Research Progress

At present, we are constructing stem cell subject profile from four perspectives:

1. *Stem cell knowledge structure*

From these perspectives of scientific instruments, experimental animals, methods and techniques, cells, organs, diseases, and genes etc., more than 20,000 knowledge entities were mining from the scientific literatures and information. And the literatures were tagged with these knowledge entities. At present, there are more than 2.2 million tag data in the knowledge graph. The tag clustering is being applied to enrich the tagging system as Fig. 1, which help to show the stem cell knowledge structure more clearly.

2. *Research front*

Based on the knowledge graph, 22 frontier topics of stem cell were mined by combining scientometrics indicators and text mining technology. The research focus analysis at different levels, such as global research, China and Chinese Academy of Sciences, are ongoing.

3. *Profile of scientific output*

From these perspectives of projects, papers, patents, and news, etc., the scientific output of scientists, institutions, and topics of stem cell were profiling.

4. *Key drugs and major diseases*

Combining clinical trials and papers of stem cell knowledge graph, the key drugs and major diseases for hematopoietic stem cells were found (Fig. 2).

Conclusions

Subject profile and knowledge graph are helpful for knowledge mining and domain knowledge discovery and can provide deeper knowledge services. This study is supported by the Informationization Special

Project of Chinese Academy of Sciences “E-Science Application for Knowledge Discovery in Stem Cells” and ongoing.

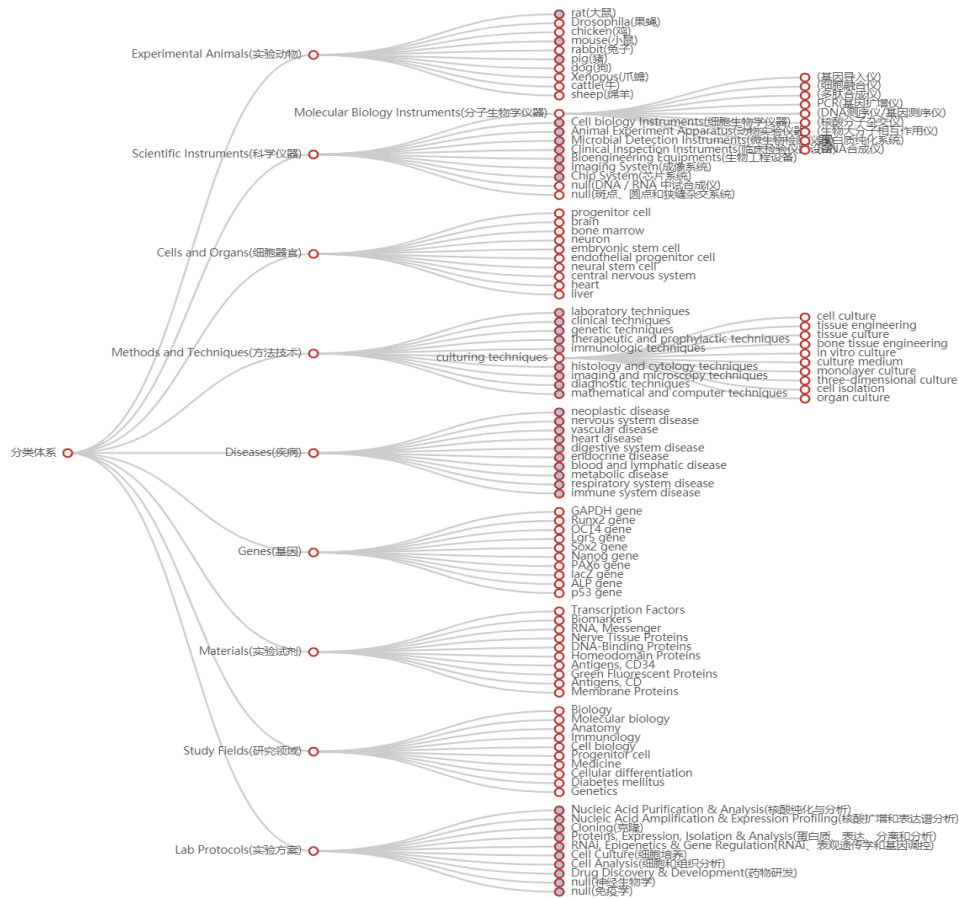


Fig 1. Part of the tagging system of stem cell subject profile

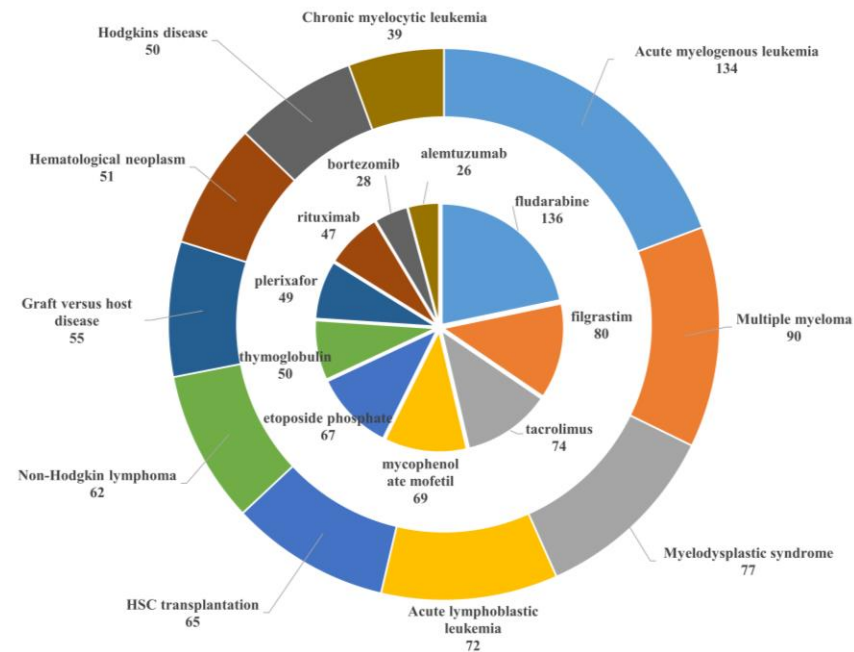


Fig 2. key drugs and major diseases of hematopoietic stem cells