中国科学院成都文献情报中心
Chengdu Library and Information Center, CAS

中国科学院广州生物医药与健康研究院
GUANGZHOU INSTITUTES OF BIOMEDICINE AND HEALTH, CHINESE ACADEMY OF SCIENCES

Research on subject knowledge discovery based on stem cell knowledge graph

Zhengyin Hu[1], Zhiqiang Zhang[1], Ning Yang[1], Xiaochu Qin[2]

[1] Chengdu Library and Information Center (CLIC), CAS

[2] Guangzhou Institutes of Biomedicine and Health (GIBH), CAS

2019.10.17， Atlanta

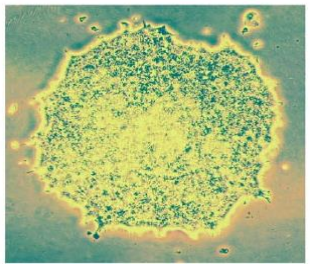
Outline

- 1. Research background
- 2. Introduction to stem cell knowledge graph
- 3. Cases of subject knowledge discovery

Stem Cell and Regenerative Medicine (1/2)

History of Human Medicine

1)drug therapy 2)surgical treatment 3)regenerative medicine

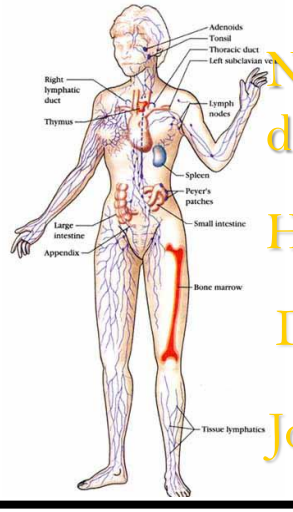


pluripotent stem cells

differentiation

functional cells

transplant



Neurological diseases

Heart diseases

Diabetes

Joint injury

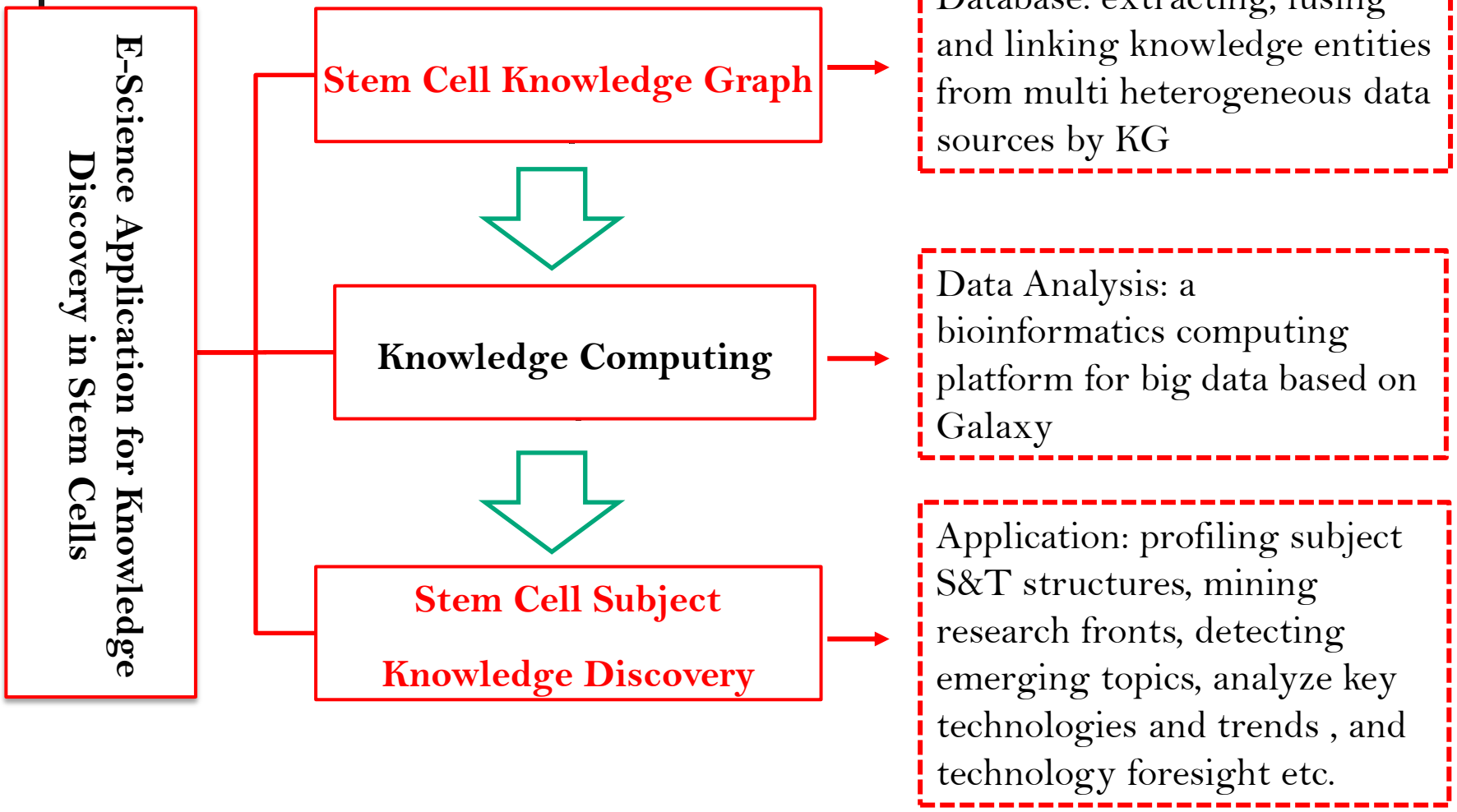
Stem cells and RM are new trends for Human Medicine

Stem Cell and Regenerative Medicine (2/2)

Since 2000, the research achievements of stem cells and RM have been selected 11 times world's ten largest S&T breakthrough in Science magazine.

- In 2012, Prof. Yamanaka of Kyoto University and Prof. John Gordon of Cambridge University won the Nobel prize in physiology and medicine for the research about stem cells and RM.
- In 2016, "Organ Repair and Reconstruction (applications of Stem Cell)" was chosen as one of 60 major breakthroughs of Chinese Academy of Sciences' 13th Five-Year Plan.
  - In 2018, CAS launched an Informationization Special Project of "**E-Science Application for Knowledge Discovery in Stem Cells**" to support the stem cells and RM Strategic Leading STI Program.

Overview of the Project

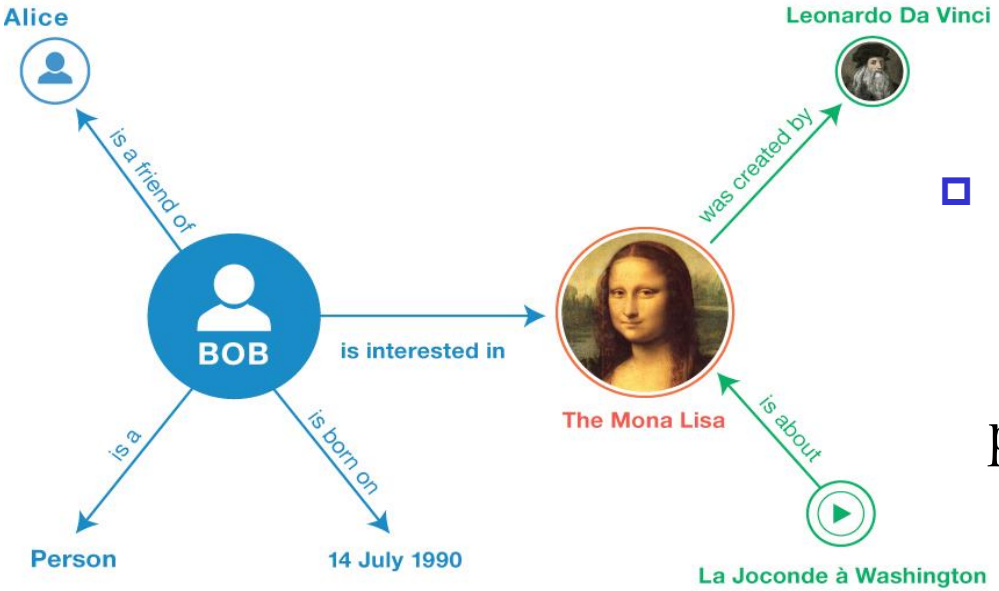


Outline

- 1. Research background
- **2. Introduction to stem cell knowledge graph**
- 3. Cases of subject knowledge discovery

What is Knowledge Graph (1/2)

- We are surrounded by entities, which are connected by relations. So we need a structured and formal representation of knowledge . KG are a natural way to represent entities and their relationship and can be managed efficiently.



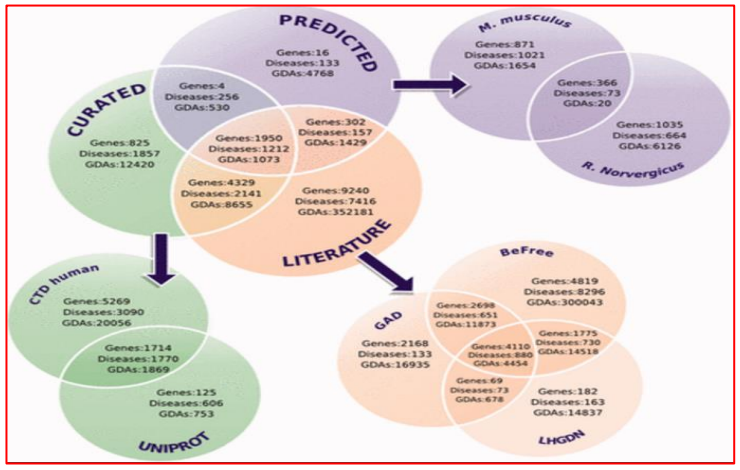
- KG is derived from semantic web and was formally proposed by Google in 2012.

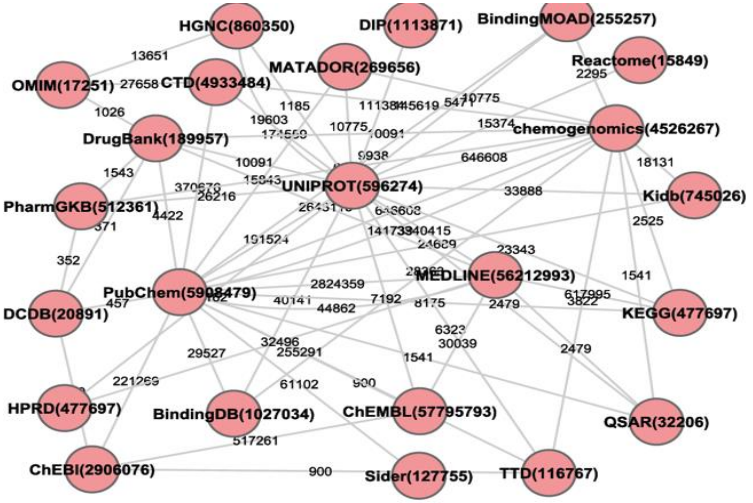
http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/example-graph.jpg

What is Knowledge Graph (2/2)

KG Uses a knowledge representation formalism (e.g.,RDF, Subject-Predicate-Object **SPO triples**) to implement semantic descriptions of entities and their relationships.

- Entities: real world objects (e.g., ACTB gene, CD34 positive cell, Zebrafish) and concepts (e.g., gene, cell, lab aminal)
- Relationships: descript relations among entities with a well defined meaning (e.g., TREATS)





Heterogeneous data of stem cell KG

Data sources: Papers, Patents, Projects, Clinical Trials, Drugs,...

S&T Instruments

Lab animals

Diseases

Cells

Organs

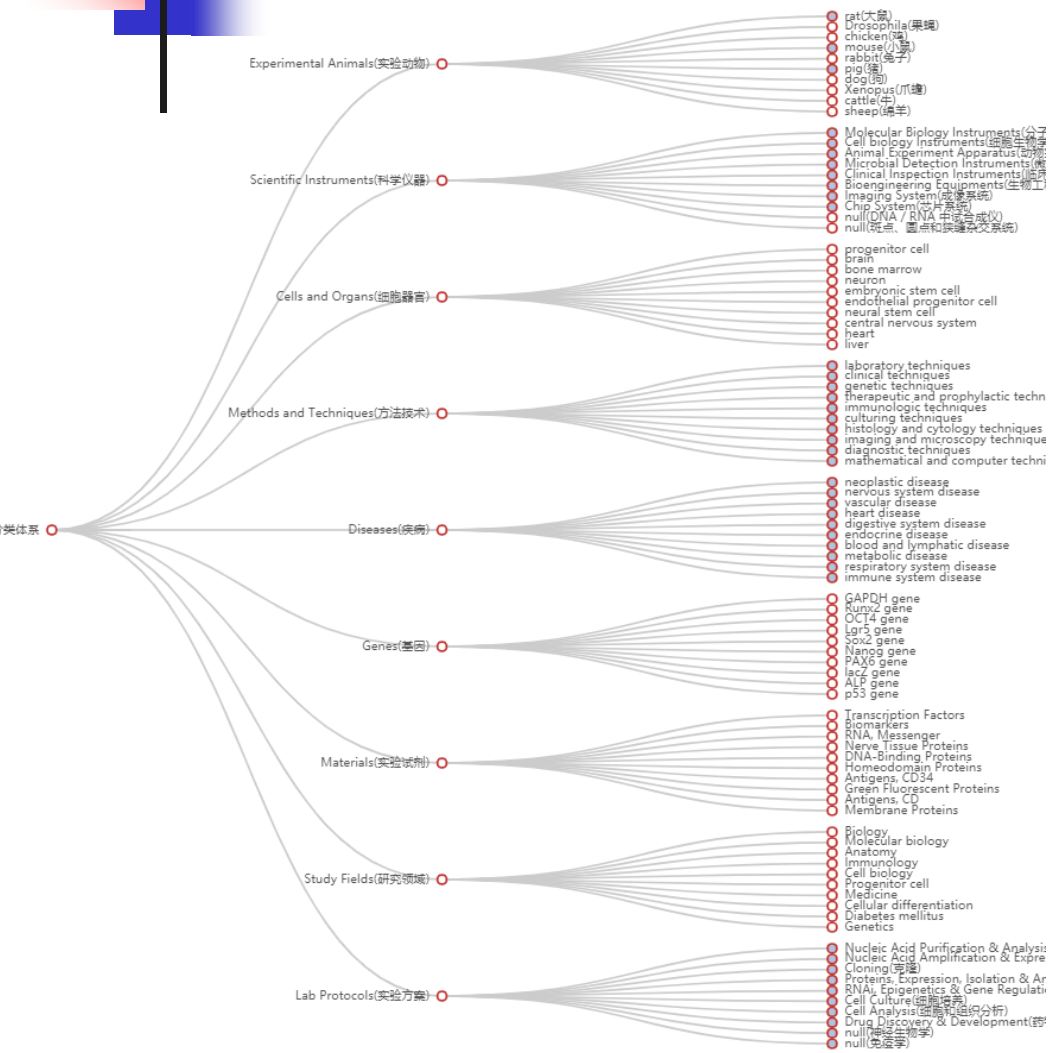
Genes

Drugs

Methods

Entities

Overview of stem cell KG



Basic Data

More than 400,000 records, from 12 kinds of S&T information such as patents, papers, etc.

Knowledge Entities

From the perspectives of scientific instruments, lab animals, methods, cells, organs, diseases, genes, mining 20,000+ knowledge entities.

Relationship

Based on SPO, more than 2,000,000 linkages and relationship.

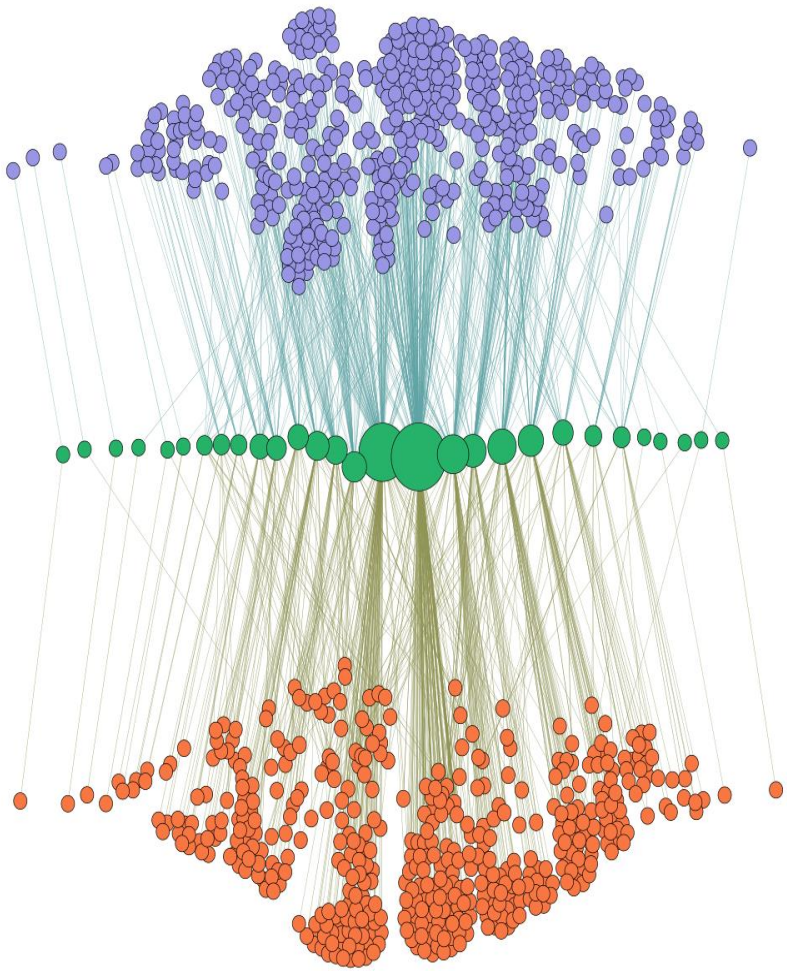
Output Profiling

Profiling the S&T outputs for institutions, scientists and topics.

The Relationship in stem cell KG

- Literatures - Literatures
  - citation, output, co-author, co-topics, co-keywords, etc.
- Literatures – Knowledge entities
  - useMethod, utilizeInstrument, useAnimal, cureDisease …
- Knowledge entities – Knowledge entities
  - ISA, AFFECTS, CAUSES, COMPLICATES, CONVERTS_TO, DIAGNOSES, LOCATION_OF, PREVENTS, STIMULATES, TREATS, … (from UMLS Semantic Network)
  - Bind, Pathway, hasGO, hassideffect, hassubstructure, …(from other Ontology)

Entities and relationship based on SPO triples



细胞器官：

oligodendrocyte | macrophage | microglia | white matter | Schwann cell

方法技术：

in situ hybridization

疾病：

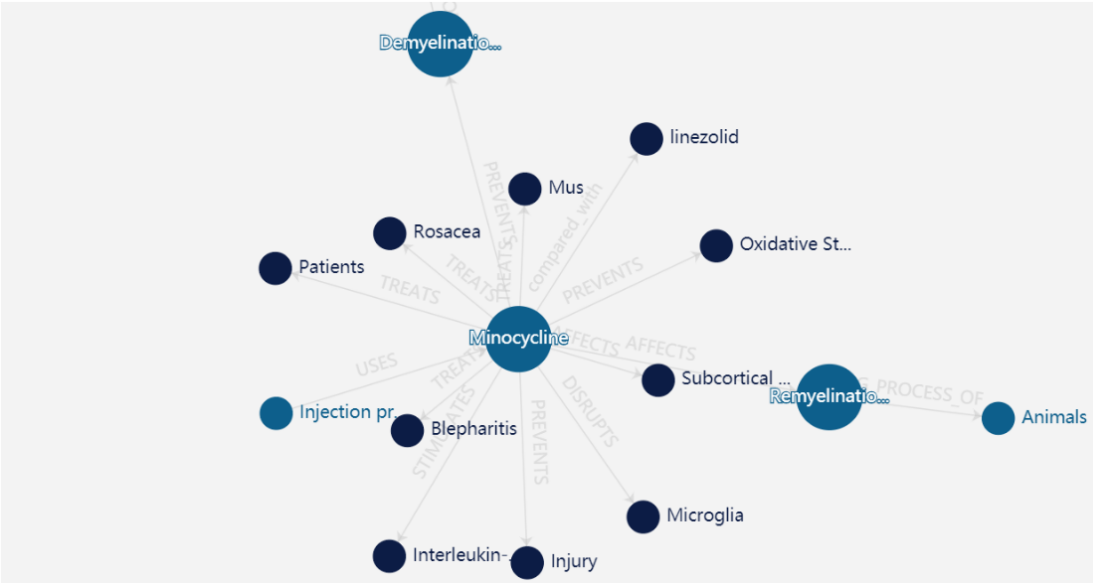
Encephalomyelitis

实验试剂：

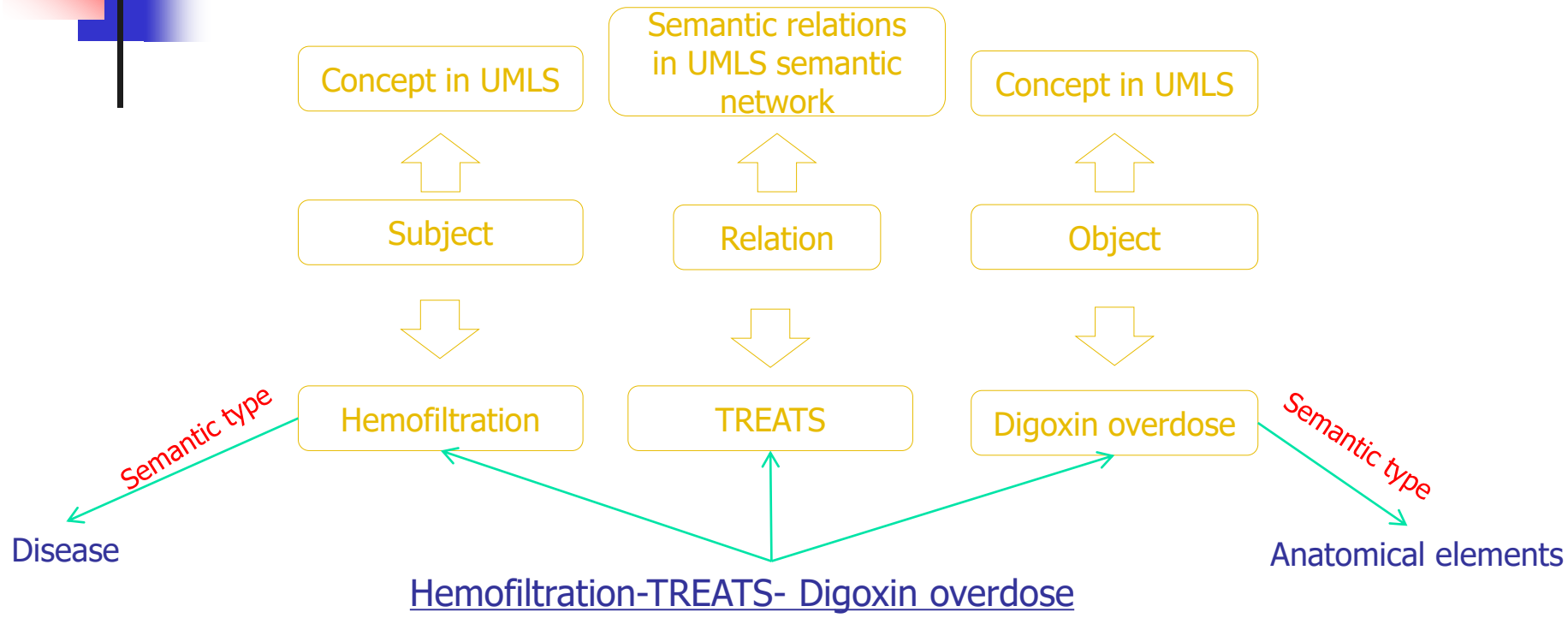
Bromodeoxyuridine(D001973) | Anti-Bacterial Agents(D000900) | CD11b Antigen(D039481) | Ethidium(D004996) | Minocycline(D008911)

实验方案：

细胞染色方案



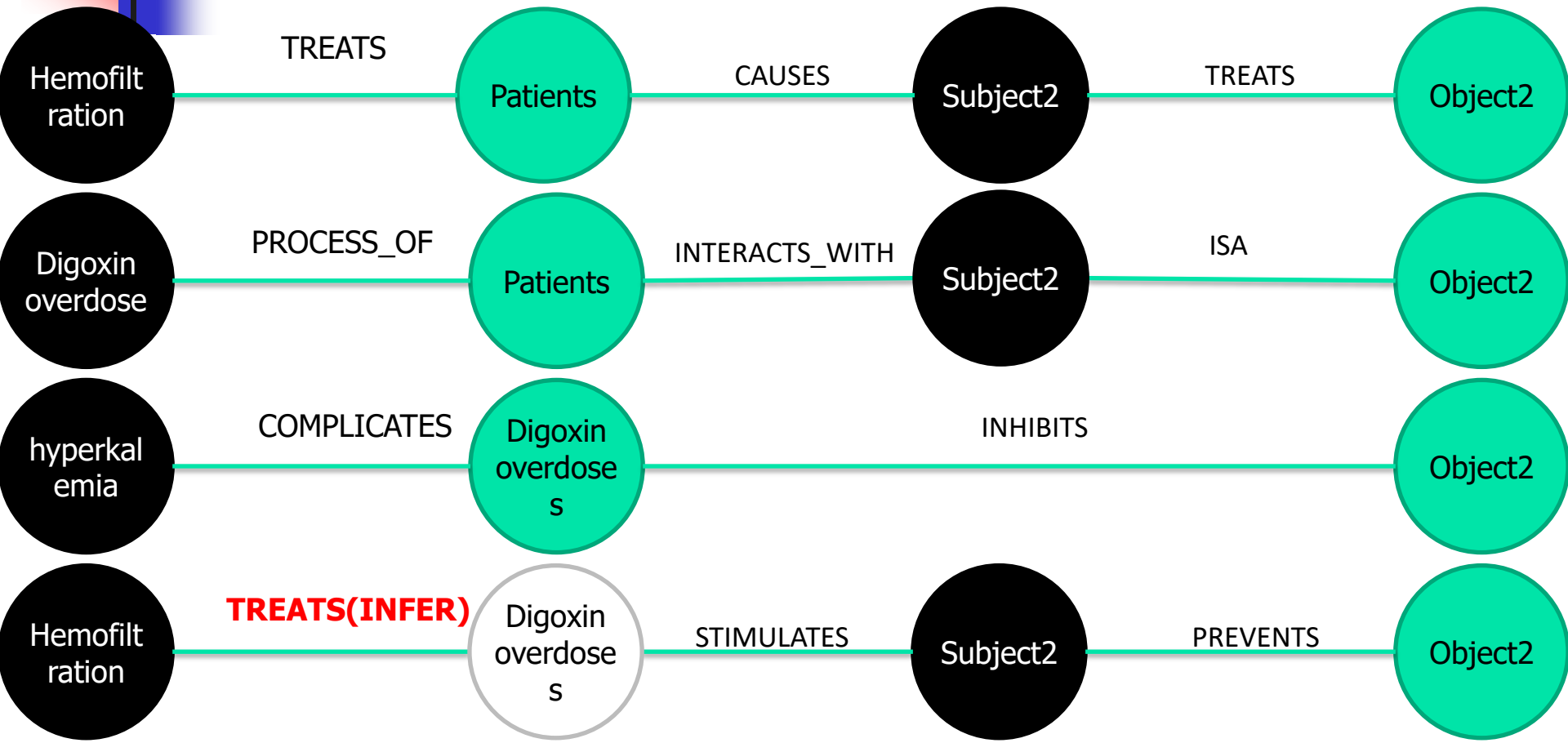
SPO triple sample



"We used hemofiltration to treat a patient with digoxin overdose that was complicated by refractory hyperkalemia."

1.Hemofiltration-TREATS-Patients
2.Digoxin overdose-PROCESS_OF-Patients
3.hyperkalemia-COMPLICATES-Digoxin overdose
4. Hemofiltration-TREATS(INFER)-Digoxin overdose

SPO is important for KG



Knowledges Entities in stem cell KG

Statistics	152 Scientific instrument	68 laboratory animals	68 Experimental scheme	14041 Experimental reagent	2628 gene	689 disease	274 Cells organ	2588 Method and technology	630 Immunotherapy	2224733 Knowledge relevance

疾病

- Disease(74)
- Infarction(35)
- Sclerosis(31)
- Inflammation(25)
- Diabetes Mellitus(24)
- Vascular Diseases(19)
- Leukemia(18)
- Heart Failure Congestive(16)
- Obesity(14)
- behavioral and mental disorders(14)
- Infection(12)
- Coronary Disease(11)
- Hypertension(10)

实验试剂

- Biomarkers(D015415)(65)
- Transcription Factors(D014157)(63)
- RNA Messenger(D012333)(60)
- Antigens CD34(D018952)(52)
- Homeodomain Proteins(D018398)(50)
- Nerve Tissue Proteins(D009419)(47)
- DNA-Binding Proteins(D004268)(46)

科学仪器

- PCR(基因扩增仪)(2)
- 生物芯片(2)
- 细胞分析仪(2)

实验方案

- DNA Extraction Protocols(DNA提取方案)(8745)
- Agarose Protocols(琼脂糖方案)(4544)
- cDNA Protocols(cDNA方案)(4496)
- Protein Purification Protocols(蛋白纯化方案)(4494)
- Microplate Protocols(显微术方案)(4479)
- mRNA Protocols(mRNA方案)(4477)
- BrdU方案(4474)

实验动物

- mouse(小鼠)(13786)
- rat(大鼠)(4224)
- Drosophila(果蝇)(1142)
- chicken(鸡)(478)
- zebrafish(斑马鱼)(387)
- rabbit(兔子)(351)
- pig(猪)(308)
- Wistar rat(Wistar鼠)(244)
- Xenopus(爪蟾)(195)
- cat(猫)(149)
- dog(狗)(143)

基因

- p53 gene(11)
- GAPDH gene(10)
- Oct4 gene(9)
- Nanog gene(8)
- Sox2 gene(8)
- PAX6 gene(6)

Stem cell KG

Epithelial vimentin plays a functional role in mammary gland

Emilia Peuhu, Reetta Virtakoivu, Anja Mai, Anni Warri, Johanna Ivaska

Development, Issue: 22, Volume: 144, Pages: 4103-4113. | 2017-11-15

Knowledge lens:

patents | scholarly works | funds | experiments

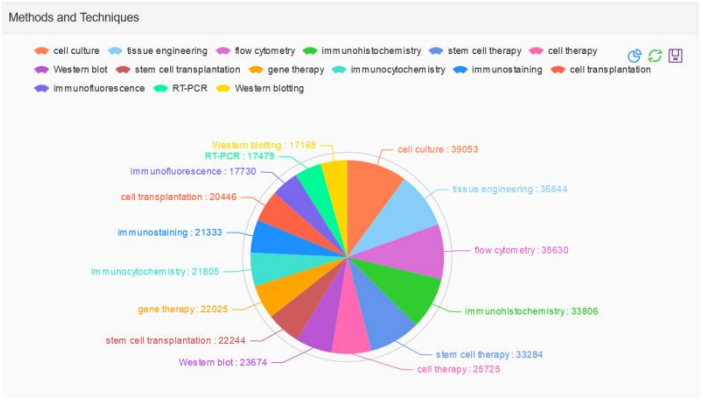
Tags

Linking

Stem Cell KG

Visualization

Navigation



干细胞领域知识发现平台

STEM CELL INNOVATION ENGINE

提升科研信息化水平
支撑研究所科技创新

- 融合多源信息 打通数据孤岛
- 挖掘知识关联 放大数据价值
- 集成知识计算 促进知识发现

-- 知识图谱数据统计 --

科学仪器TOP 10

仪器名称	数量
蛋白质纯化系统	4643
多肽合成仪	460
血液电解质分析仪	376
生长曲线分析仪	325
Chip System	304
DNA测序仪/基因测序仪	279
Clinical Inspection Instruments	209

实验动物TOP 10

动物名称	数量
mouse	86
rat	34
Drosophila	22
chicken	18
zebrafish	17
rabbit	17
pig	17

实验方案TOP 10

实验方案	数量
mRNA Protocols	4477
BrdU方案	4474
Membrane Protein Protocols	4461
SDS PAGE Protocols	4461
Primary Cell Protocols	4454
Cell Culture	4449

实验试剂TOP 10

实验试剂名称	数量
Transcription Factors(D014157)	2889
Biomarkers(D015415)	2877
RNA Messenger(D012333)	2681
Nerve Tissue Proteins(D009419)	2259
DNA-Binding Proteins(D004268)	1698
Homeodomain Proteins(D018398)	1466
Antigens CD34(D018952)	1461

基因TOP 10

基因名称	数量
GAPDH gene	132
Runx2 gene	101
Nanog gene	97

疾病TOP 10

疾病名称	数量
Infarction	655
Sclerosis	649
Diabetes Mellitus	438

SCIE 干细胞领域知识发现平台
STEM CELL INNOVATION ENGINE

提升科研信息化水平
支撑研究所科技创新

- 融合多源信息 打通数据孤岛
- 挖掘知识关联 放大数据价值
- 集成知识计算 促进知识发现

知识棱镜

- 科学仪器
- 实验动物
- 实验方案
- 实验试剂
- 基因
- 疾病
- 细胞器官
- 方法技术
- 免疫治疗
- 期刊来源
- 关键词
- 文献类型
- 机构
- 作者

论文

我的订阅

标题 | 文献类型 | 作者 | 结果中检索

论文 专利 期刊 产品 项目 法规 专家 实验 临床试验 资讯 机构 硕博论文

共找到60200条结果 排序: 相关度

Neural stem/precursor cells for the treatment of ischemic stroke.

MARCO BACIGALUPPI
JOURNAL OF THE NEUROLOGICAL SCIENCES, Issue: , Volume: 265, Pages: 73-77. | 2008-02-15
知识棱镜:
专利 论文 项目 实验 ★收藏 打开全文

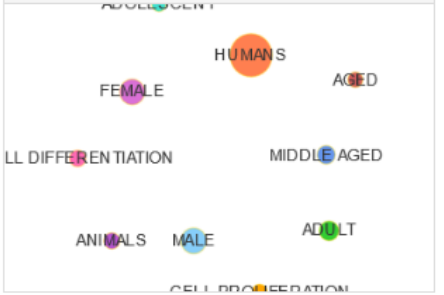
Expression profiles of inka2 in the murine nervous system

YUMI IWASAKI, TAKAHITO YUMOTO, SHIN ICHI SAKAKIBARA
GENE EXPRESSION PATTERNS, Issue: , Volume: 19, Pages: 83-97. | 2015-01-09
知识棱镜:
专利 论文 项目 实验 ★收藏 打开全文

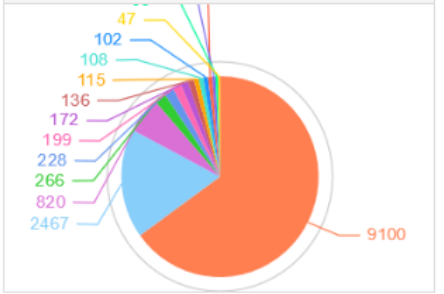
The Missing Niche for Spermatogonial Stem Cells: Do Blood Vessels Point the Way?

GUNAPALA SHETTY, MARVIN L MEISTRICH
CELL STEM CELL, Issue: 4, Volume: 1, Pages: 361-363. | 2007-01-10
知识棱镜:

关键词云图



实验动物



A Continuous Molecular Roadmap to iPSC Reprogramming through Progression Analysis of Single-Cell Mass Cytometry

CellStemCell, Issue: 3, Volume: 16, Pages: 323-337. | 2015-03-05
Patent citations Scholarly citations Reference Count:

Eli R. Zunder;Ernesto Lujan;Yury Goltsev;Marius Wernig;Garry P. Nolan

j.stem.2015.01.015

摘要

To analyze cellular reprogramming at the single-cell level, mass cytometry was used to simultaneously measure markers of pluripotency, differentiation, cell-cycle status, and cellular signaling throughout the reprogramming process. Time-resolved progression analysis of the resulting data sets was used to construct a continuous molecular roadmap for three independent reprogramming systems. Although these systems varied substantially in Oct4, Sox2, Klf4, and c-Myc stoichiometry, they presented a common set of reprogramming landmarks. Early in the reprogramming process, Oct4highKlf4high cells transitioned to a CD73highCD104highCD54low partially reprogrammed state. Ki67low cells from this intermediate population reverted to a MEF-like phenotype, but Ki67high cells advanced through the M-E-T and then bifurcated into two distinct populations: an ESC-like NanoghighSox2highCD54high population and a mesendoderm-like NanoglowSox2lowLin28highCD24highPDGFR-αhigh population. The methods developed here for time-resolved, single-cell progression analysis may be used for the study of additional complex and dynamic systems, such as cancer progression and embryonic development.

文献类型:
Resource

出版商:
Elsevier

发表时间:
2015-03-05

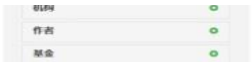
基金信息

实验动物:
sheep(绵羊) salamander(蝾螈)

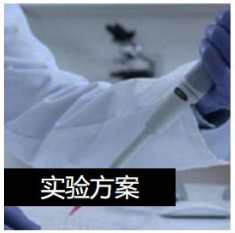
科学仪器:
Clinical Inspection Instruments(临床检验仪器设备)

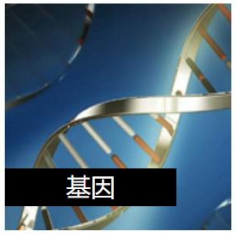
免疫治疗:
Colony-forming unit

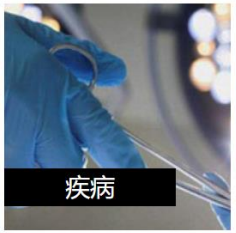
实验方案:
Protein Purification Protocols(蛋白纯化方案) DAPI方案



干细胞 STEM CELL 知识导航







- 论文
- 专利
- 产品
- 项目
- 实验
- 临床试验
- 资讯
- 硕博论文

论文的实验方案知识导航如下：

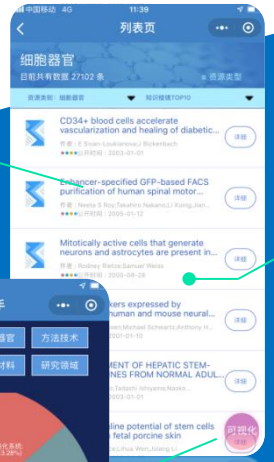
按字顺浏览： 3(1) A(4) B(1) C(9) D(4) E(1) G(2) I(1) L(1) M(3) N(4) P(7) R(3) S(3) T(3) W(1) c(1) m(1) s(1) 免(4) 凋(1) 分(2) 干(2) 染(1) 标(1) 流(1) 神(1) 细(2)

Protein Purification Protocols(蛋白纯化方案)	Primary Cell Protocols(原代细胞方案)	Protein Isolation Protocols(蛋白分离方案)
Protein Expression Protocols(蛋白表达方案)	Proteins Expression Isolation & Analys	Pluripotent Stem Cell Protocols(多能干细胞方案)
PCR Protocols(PCR方案)		

Stem cell KG WeChat APP

打开微信
扫一扫
搜一搜

资源详细信息



资源列表

可视化展示

资源检索

用户中心

released at the 11th Guangzhou International Conference on Stem Cell and Regenerative Medicine

Outline

- 1. Research background
- 2. Introduction to stem cell knowledge graph
- **3. Cases of subject knowledge discovery**

Discover popular diseases and novel relationship

Objective

- Discover popular diseases and latent relations of diseases in hematopoietic stem cell (HSC) using a hybrid approach combining text mining (KG) and scientometrics indicators.

Method

- 1. Construct SPO Semantic Network from stem cell KG (papers)
- 2. Detect Community from the SPO network
- 3. Discover popular diseases and latent relations in HSC

Construct SPO Semantic Network(1/2)

HSC papers and the corresponding SPO predications of each paper are retrieved from the stem cell KG from 2009-2018, which include 9, 576 research articles and 78, 249 SPO predications.

- Then, SPO triples are cleaned by term clumping process which includes general cleaning and pruning sub-processes. General cleaning will remove some common academic/scientific subjects or objects such as “cells,” “organ”. Some predicates such as “LOCATION_OF,” “PART_OF” that reflect hierarchy or position relationship and are meaningless for mining the relations of diseases will also be removed. The pruning process helps with further cleaning by discarding the very low frequency subjects, predicates, objects or SPO.

Construct SPO Semantic Network(1/2)

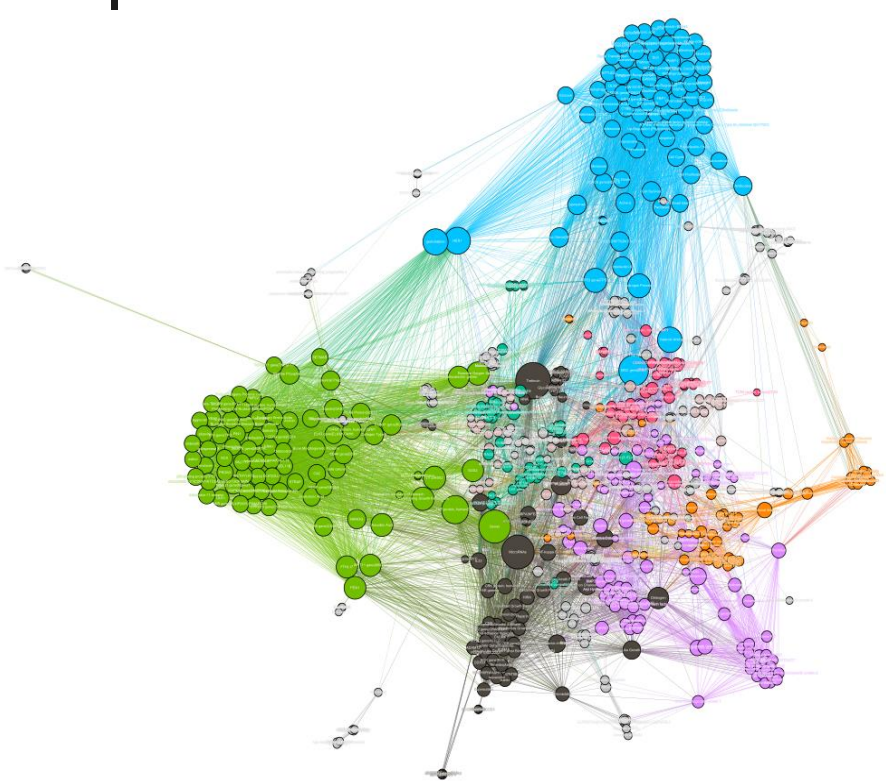
Predication_name	Predication_type
AFFECTS	Function（功能）
AUGMENTS	Effect（效果）
CAUSES	Effect（效果）
COMPLICATES	Function（功能）
DIAGNOSES	Function（功能）
different_than	Effect（效果）
DISRUPTS	Function（功能）
higher_than	Effect（效果）
INHIBITS	Function（功能）
lower_than	Effect（效果）
PREVENTS	Function（功能）
STIMULATES	Function（功能）
TREATS	Cure(治疗)

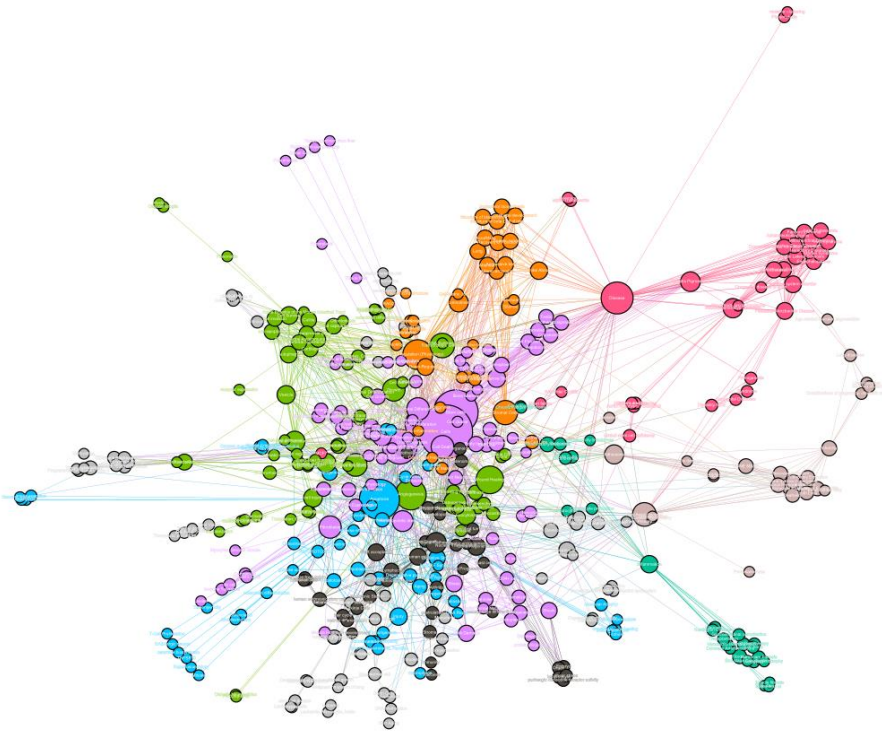
Detect Community based on Guided Local Search

Algorithm 1 The Guided Local Search algorithm

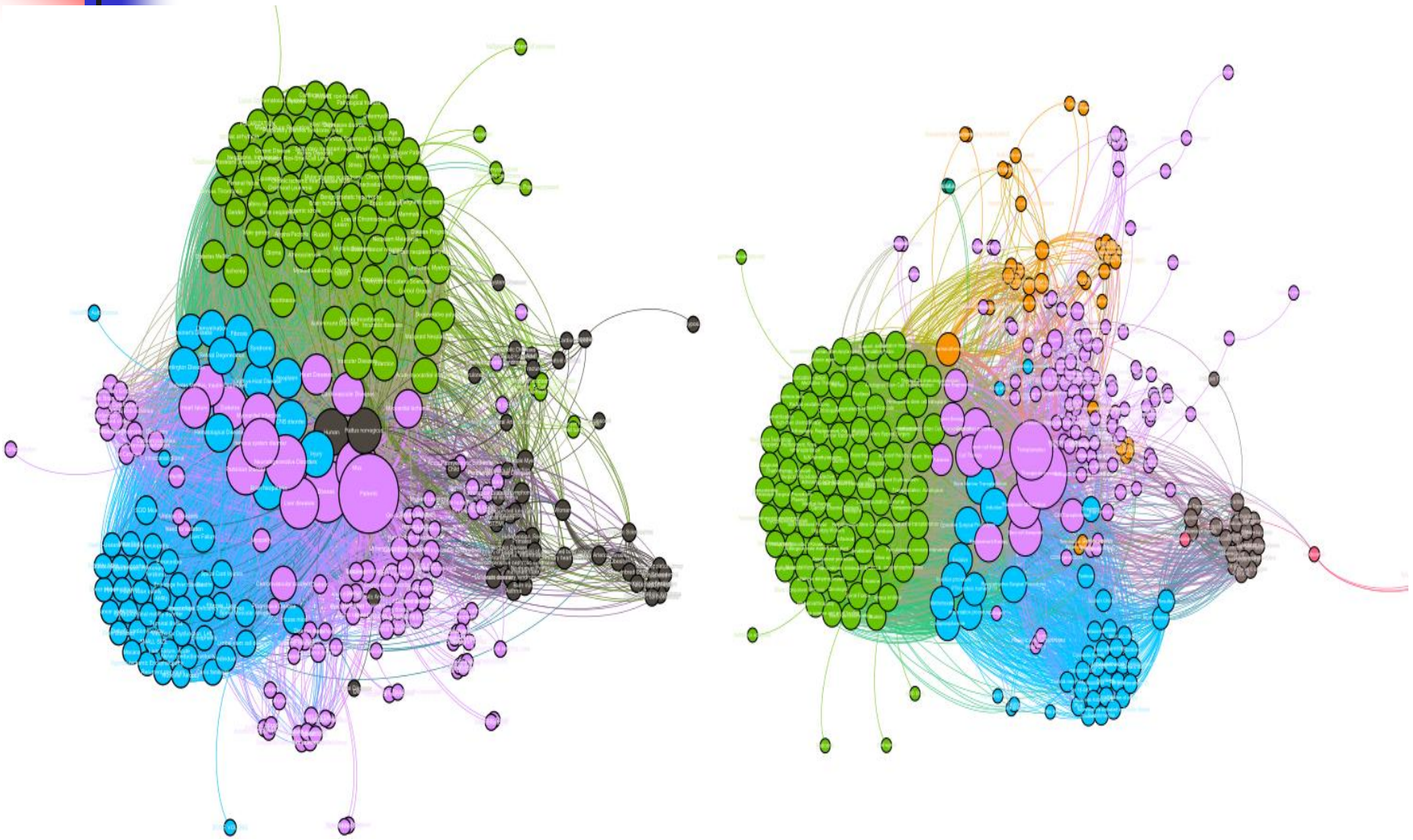
1: **Input**: SPO-based semantic relation network G
2: **Output**: Sequence of communities
3: $P = \{x^1, \ldots, x^p\} \leftarrow$ Random_Initialization (P)
4: **repeat**
5: $x^i \leftarrow$ Local_Search (x^i)
6: randomly choose an individual x^i from P and a vertex w
7: find $w \in C_i$ in m individuals and $w \in C_j$ in $p - m$ individuals
8: $\Delta T(w, C_i, C_j) \leftarrow \sum_{k=1}^{m} \Delta Q(w, C_i^k, C_j^k)$
9: **if** $\Delta T(w, C_i, C_j) \geq 0$ **then**
10: $w \in C_j$ in m individuals
11: **end if**
12: update $\{x^1, \ldots, x^p\}$
13: **until** the modularity function value is not improved any more

Subject and Object communities of *effect* group



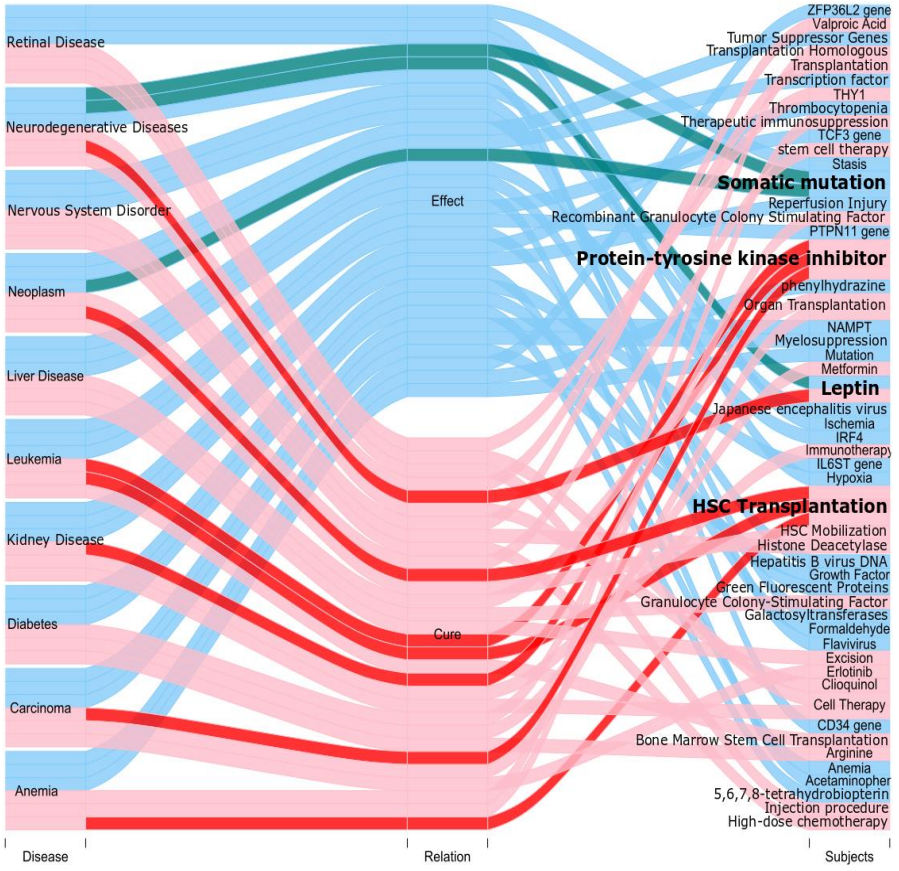


Subject and Object communities of cure group



Top 10 popular diseases and latent relations in HSC

- According to the communities detected, some popular diseases and latent relations are discovered. The left in figure show the popular diseases and the right show some entities related to these diseases. The linkages between them are two types of relations, *Effect* and *Cure*. Some strong linkages are colored with red or deep blue, which show that the entities are very close to the corresponding diseases.



- Some latent relations can also be found by knowledge inference. For example, “Protein-tyrosine kinase inhibitor” has strong *Cure* relation with “leukemia”, “kidney disease”, and “Somatic mutations” has strong *Effect* relation with “neoplasm” and “neurodegenerative diseases.

Summary

- KG can link, integrate and fuse heterogeneous data from multiple sources with more details in a simple manner, which can support more effective biomedical knowledge discovery.
- The stem cell KG can help to effectively and accurately discover the popular diseases and latent relations in HSC, which are conducive to new knowledge discovery and inspire innovation in HSC.
- Further, we will conduct a comprehensive analysis from papers, patents and CT.

Thank you for attention!

huzy@clas.ac.cn