



开放科学
(资源服务)
标识码
(OSID)

基于 Citespace 的基因芯片领域知识图谱分析

孙劲松 郑彦宁 袁芳

中国科学技术信息研究所 北京 100038

摘要: 随着后基因组时代的到来, 基因芯片技术因其高通量、高效率以及高自动化等特点成为了最具发展潜力的高端生物技术之一。知识图谱提供了一种从海量文本数据中抽取结构化知识的手段, 目前广泛应用于多个领域。论文以 Web of Science 平台的 SCIE 数据库为数据来源, 利用 Citespace 知识图谱软件, 采用内容分析法和词频分析法, 分别从研究时序、研究热点、研究主题、知识结构、研究机构等多个方面对基因芯片领域的研究进展、研究现状以及研究前沿进行分析。结果表明, 基因芯片文献数量呈先增后减的趋势; 研究内容呈阶段性变化; 突现词分布较为广泛; 关键词聚类特征较为明显; 文献共被引聚类维度较多; 机构合作较为密切; 哈佛大学等是主要的高产机构。

关键词: Citespace; 基因芯片; 知识图谱; 可视化

中图分类号: G35

Knowledge Mapping of Gene Chip Based on Citespace

SUN Jingsong ZHENG Yanning YUAN Fang

Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract: With the advent of the post-genome era, gene chip technology has become one of the most promising high-end biotechnologies due to its high throughput, high efficiency and high automation. The knowledge map provides a means of extracting structured knowledge from massive text data and is currently used in many fields. The article uses the SCIE database of Web of Science platform as the data source, applying content analysis method and word frequency analysis method by using the Citespace knowledge mapping software, to analyze the research progress, research status and research frontiers in the field of gene chip from the perspectives of research time series, research hotspots, research topics, knowledge structure, research institutions, etc. The results showed that the number of gene chip literature increased first and then decreased; the research content showed a phase change; the emergence of prominent words was more extensive; the keyword clustering characteristics

作者简介: 孙劲松 (1994-), 硕士研究生, 研究方向: 前沿领域分析与专利分析, E-mail: sunjs2017@istic.ac.cn; 郑彦宁 (1965-), 研究馆员, 研究方向: 竞争情报、情报学理论方法; 袁芳 (1989-), 博士后, 研究方向: 产业竞争情报研究。

were more obvious; the literature was cited more clustering dimensions; the organization cooperation was closer Harvard University and other major high-yield institutions.

Keywords: Citespace; Gene Chip; knowledge map; visualization

引言

伴随着人类基因组计划的实施,越来越多的动植物以及微生物的基因组序列得以测定,如何对大量的遗传信息进行快速高效地检测和分析显得尤为重要,基因芯片技术就是顺应这项计划兴起的产物^[1]。基因芯片,又称DNA芯片,是指将大量探针分子固定于支持物上后与标记的样品分子进行杂交,通过检测每个探针分子的杂交信号强度进而获取样品分子的序列信息^[2],由于这个过程构成的二维DNA探针阵列与计算机的电子芯片十分相似,因此被称为基因芯片^[3]。由于基因芯片具有信息量大、操作性强、灵敏度高等优点,目前广泛应用于疾病诊断、药物筛选、司法鉴定、环境保护等多个领域^[4]。基因芯片技术综合了多种现代高精尖技术,是分子生物学、微电子、计算机等多门学科结合的结晶,预计将成为新一代分子诊断试剂开发的主流,被专家誉为“诊断行业的终极产品”^[5]。

知识图谱,是通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论和方法与计量学引文分析、共现分析等方法相结合,利用可视化的手段来呈现科学知识的结构、规律及分布情况^[6]。知识图谱从语义角度出发,通过对概念、实体及其关系进行描述,有助于更好地组织和管理海量信息,从而挖掘出

数据背后隐藏的有价值的知识^[7]。本研究采用Citespace软件对基因芯片领域的科研文献数据进行知识图谱分析,进而了解该领域的研究进展、研究现状以及未来的研究前沿。

1 数据来源及处理

1.1 数据来源

论文数据来源于Web of Science(简称WOS)核心合集集中的Science Citation Index Expanded(简称SCIE)数据库。检索式为:TS=((“gene chip*”) or (“genechip*”) or (“gene array*”) or (“genearray*”) or (“DNA chip*”) or (“DNAchip*”) or (“DNA array*”) or (“DNAarray*”) or (“oligonucleotide* chip*”) or (“oligonucleotide*array*”) or (“oligonucleotide* array*”) or (“genome* chip*”) or (“genome*chip*”) or (“genome* array*”)),时间跨度为所有年份(1900-2019年)。将检索结果精炼为article,共检索到12916条论文数据,检索时间为2019年3月17日。

1.2 数据处理

将WOS下载得到的数据导入到Citespace中,根据论文的实际情况,将时区设置为1985-2019年,时间切片为1年,阈值为Top50,连线强度选择Cosine,网络剪裁方法为Minimum

Spanning Tree 和 Pruning sliced networks。根据研究目的选择不同的节点类型进行分析,运行 CiteSpace 后点击可视化,即可得到相应的知识图谱。

2 分析与讨论

2.1 研究时序分析

通常认为,移动平均趋势线可以消除实际数据值的随机波动,从而有效地预测未来数据的发展趋势^[8]。通过对某一领域历年的论文数量进行统计,利用双周期移动平均趋

势线进行拟合,依照年份绘制文献数量时序分布图,有助于了解该研究领域的发展历程。如图 1 所示,可以将基因芯片的研究过程大致划分为四个阶段:

第一阶段(1989-1997年)为萌芽期,1989年美国罗彻斯特大学的作者 WU 等发表了第一篇关于基因芯片的论文 *Fitness reduction associated with the deletion of a satellite DNA array*, 此时基因芯片研究开始进入萌芽时期。但受当时理论认识水平的限制,9年间基因芯片领域发表论文数量均未超过百位数,个别年份数量为0,年均发文量仅为13篇。

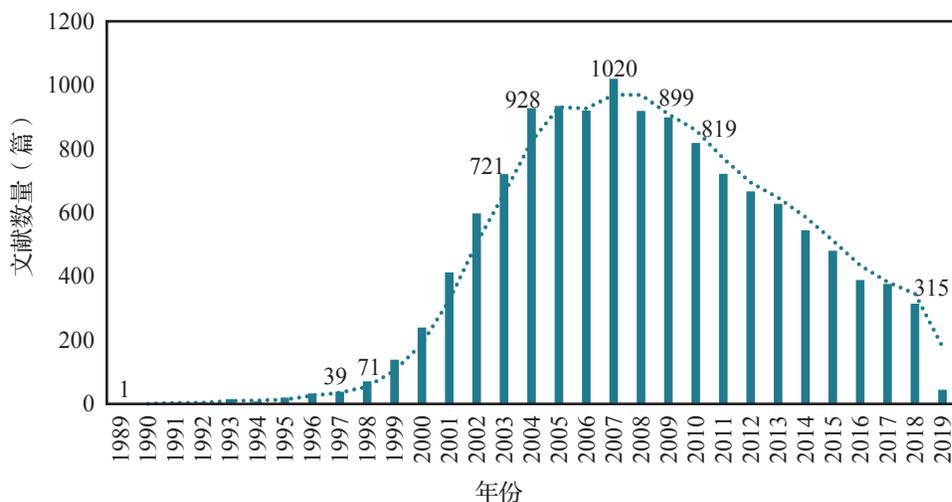


图 1 文献数量时序分布

第二阶段(1998-2003年)为发展期,基因芯片的文献数量以逐年加速的方式激增,国内外学者对基因芯片的研究兴趣逐渐增强。阶段起点1998年发文量为71篇,阶段终点2003年发文量为721篇,约为1998年的10.2倍,6年内年均发文量为363篇,与萌芽期相比有了很大程度的增长。将第一阶段与第二阶段结合起来看,基因芯片研究呈持续增长的发展态势。

第三阶段(2004-2009年)为平稳期,基因芯片的文献数量呈波动发展的状态,6年间均保持在较高水平,其中2007年达到峰值1020篇。尽管2009年文献数量略有回落,但仍高于前两阶段任一年份的发文量。这一阶段的年均发文量为936篇,约等于第二阶段年均发文量的2.6倍。

第四阶段(2010-2019年)为成熟期,基因

芯片的文献数量呈逐步下降的趋势,2018年基因芯片发文量是2010年的0.38倍。究其原因,可能在于基因芯片的大部分主题在此前研究中均已有所探讨,并在主要问题的研究上基本达成了共识,进一步深入研究的空间较为有限。虽然发文量与前一阶段相比有所下降,但与早期相比仍处于较高的数量水平。

为了进一步揭示基因芯片领域的研究发展脉络,在文献数量时序分析的基础上构建了关键词 Time-zone 图谱进行分析。Time-zone 图谱是将相同时间内的节点集合在相同的时区中,这里的相同时间对于关键词而言就是它们首次出现的时间。如图2所示,随着时间的推移,Time-zone 图谱中节点数量先增后减,节点之间的连线先密后疏,与历年文献数量的发展历程基本对应,每年都对应不同的研究内容。1989-1997年,关于基因芯片的一些研究开始起步,其中 sequence (序列)是最早的研究内容之一,从1993年开始, hybridization (杂交)、 messenger RNA (信使核糖核酸)、 oligonucleotide array (寡核苷酸阵列)、 gene expression (基因表达)、 microarray (微阵列)、 cancer (癌症)等词语相继出现。1998-2003年,随着文献数量的骤增,爆发了大量新的研究话题,许多与基因芯片相关的词开始出现,例如 protein (蛋白质)、 cell (细胞)、 apoptosis (细胞凋亡)、 in vitro (体外)、 differentiation (分化)、 transcription factor (转录因子)、 carcinoma (肿瘤)等,研究内容逐渐丰富,这一阶段节点最大且节点数量最多,节点间的连线最为密切。2004-2009年,节点数量较前减少,节点间的连线较前分散,此阶段研究较为多元化,研究内容包括 oxida-

tive stress (氧化应激)、 comparative genomic hybridization (比较基因组杂交)、 colorectal cancer (结肠直肠癌)、 arabidopsis thaliana (拟南芥)等。2010-2019年,开始进入基因芯片研究的深化发展阶段,主要是对以往的研究进行深入探讨,一些新的关键词也随之出现,例如 biomarker (生物标志物)、 angiogenesis (血管生成)、 transcriptome (转录组)、 microRNA (微型核糖核酸)、 signaling pathway (信号通路)等。可见,随着基因芯片领域的不断完善,相关的研究和探索在广度和深度上都有了进一步的发展。

2.2 研究热点分析

通常认为,关键词是一篇文献核心内容的浓缩和提炼,具有一定的高度概括性、可量性和直接相关性,对高频关键词进行分析有助于了解相关领域的研究热点^[9]。在关键词共现图谱中,节点的大小代表了关键词的频次,节点越大,频次越高。节点之间的连线代表了关键词的共现关系,连线越粗,节点之间关系越紧密,同时出现在同一篇文章中的可能性越大^[10]。如图3所示,图谱中共包含了235个节点,1471条连线,网络密度为0.0535, Modularity $Q=0.3867>0.3$, Mean Silhouette $=0.539>0.5$,说明该图谱的结构较为合理,基因芯片领域热点话题的凝聚度较高。图3中节点最大的关键词是 gene expression (基因表达),说明基因芯片领域的研究成果中 gene expression (基因表达)出现的频次最高。此外, microarray (微阵列)、 expression (表达)、 oligonucleotide array (寡核苷酸阵列)和 identification (鉴定)等关键词的节点也较大,频次均高达1000以上。

酒酵母)、oligonucleotide array (寡核苷酸阵列)、escherichia coli (大肠杆菌)、induction (诱导)、DNA array (脱氧核糖核酸阵列)、polymorphism (多态性)等在1991-2011年不同年份间成为突发性关键词,其中hybridization(杂交)突现强度最高,这与基因芯片的主要原理

相符。此外,arabidopsis thaliana (拟南芥)、inflammation (炎症)、disease (疾病)、proliferation (增殖)、microRNA (微型核糖核酸)等是最近几年的突发性关键词,根据可视化理论可知,这些关键词相关的主题很可能是基因芯片最近几年的研究前沿。

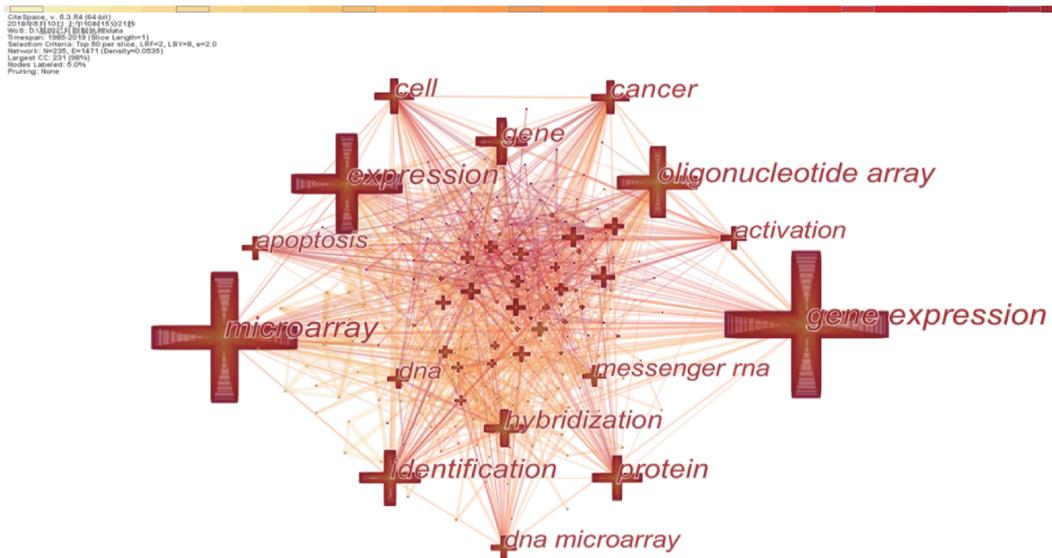


图3 关键词共现图谱

表1 主要关键词频次与中心度分布

序号	频次排序	频次	中心度排序	中心度
1	gene expression	2212	gene expression	0.15
2	microarray	1912	microarray	0.15
3	oligonucleotide array	1213	oligonucleotide array	0.15
4	protein	861	hybridization	0.14
5	hybridization	778	sequence	0.14
6	cell	717	cancer	0.12
7	cancer	707	DNA	0.11
8	apoptosis	498	messenger RNA	0.08
9	activation	480	cell	0.07
10	messenger RNA	471	apoptosis	0.06

为了进一步对基因芯片领域的前沿趋势进行总结,将Burst值排名前十五的关键词归

纳如下:(1)saccharomyces cerevisiae (酿酒酵母)、escherichia coli (大肠杆菌)等底

盘细胞以及 *arabidopsis thaliana* (拟南芥) 等被子植物可能是基因芯片领域未来的重点研究对象。(2) *sequence* (序列)、*oligonucleotide array* (寡核苷酸阵列)、*DNA array* (脱氧核糖核酸阵列)、*microRNA* (微型核糖核酸) 等可能是基因芯片领域未来的研究主题。(3)

hybridization (杂交)、*induction* (诱导)、*polymorphism* (多态性)、*proliferation* (增殖) 等可能是基因芯片领域未来的重点研究方向。

(4) *inflammation* (炎症)、*disease* (疾病) 等医学研究可能是基因芯片未来的重点应用领域。

表 2 主要突现关键词分布

序号	关键词	突现强度	开始年	结束年	突变点 (1991-2016年)
1	sequence	27.6555	1991	2002	
2	hybridization	58.6429	1993	2002	
3	saccharomyces cerevisiae	46.0010	1996	2004	
4	density oligonucleotide array	46.8935	1997	2003	
5	oligonucleotide array	25.2446	1997	2004	
6	escherichia coli	28.8380	1998	2004	
7	induction	25.2929	2000	2004	
8	DNA array	29.3089	2001	2004	
9	classification	27.4121	2003	2006	
10	polymorphism	24.7928	2007	2011	
11	arabidopsis thaliana	39.3639	2008	2015	
12	inflammation	27.0158	2009	2016	
13	disease	32.3035	2010	2016	
14	proliferation	25.1393	2011	2015	
15	microRNA	29.7577	2013	2016	

2.3 研究主题分析

在 2.2 节的基础上, 为了进一步归纳出基因芯片领域的主题内容, 本研究采用 LLR 算法对关键词进行聚类分析^[15], 如图 4 所示, 通过对给定的文献关键词进行分类聚合, 将内容联系密切的关键词聚合为文献簇, 排除其中类别较小的聚类, 选取 7 个具有代表性的聚类如下: #0 “mutation detection”、

#1 “transcriptional activator”、#2 “gene expression”、#3 “balanced karyotype”、#4 “insulin sensitivity”、#5 “trypanosoma brucei”、#6 “human papillomavirus”。其中, 聚类 0 主要表示基因突变检测, 用于对肺癌、乳腺癌和结直肠癌等肿瘤疾病的早期筛查、诊断及预后; 聚类 1 主要是指转录激活因子, 用于激活基因表达或促进基因转录; 聚类 2 主要

表示的是基因表达的相关研究；聚类 3 主要是用于染色体平衡易位核型分析；聚类 4 主要是从遗传角度对胰岛素敏感性进行研究；聚类 5

主要是对布氏锥虫病的诊断和治疗；聚类 6 主要针对的是人类乳头瘤病毒，即 HPV 病毒的研究。

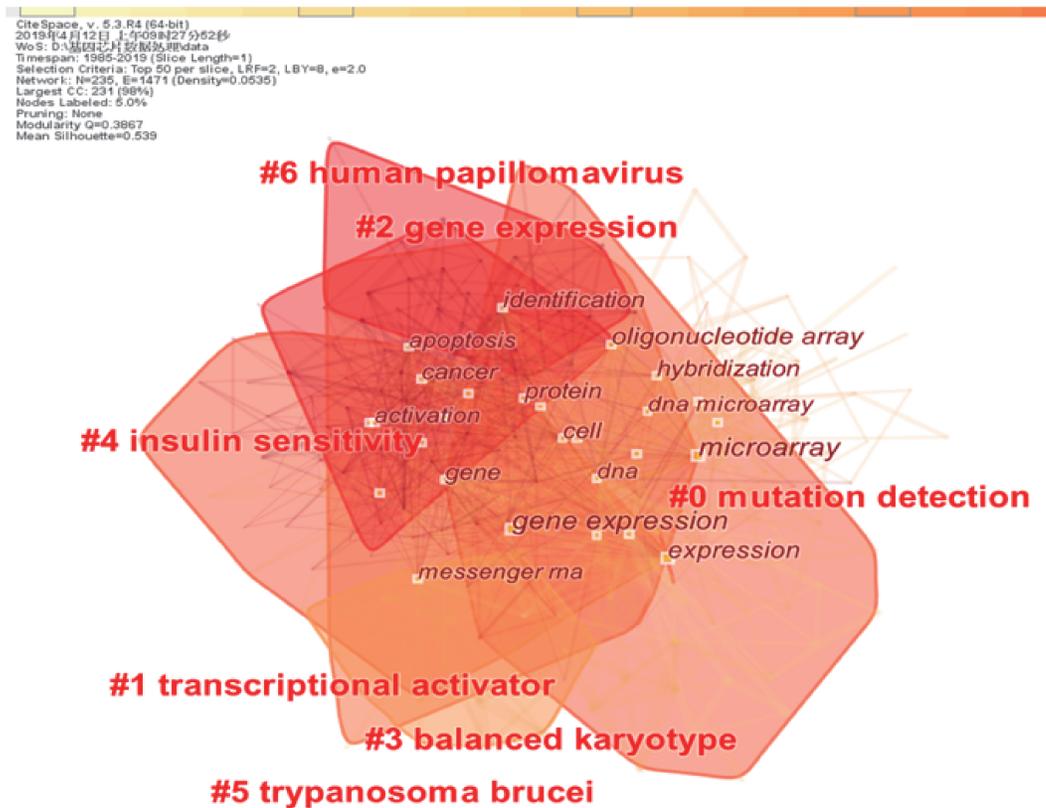


图 4 关键词聚类图谱

为了更加直观地反映出基因芯片领域不同时间段研究主题的演进与发展趋势，本研究对该领域进行 Time-Line 分析，进而了解该领域 1985-2019 年间各类文献的时间跨度以及某一特定聚类研究的兴起、繁荣及衰落过程，从而探索该领域的动态演化路径^[16]。Time-Line 图谱是将相同聚类的文献放置在同一水平线上，文献的时间置于视图的最上方，越向右时间越近。如图 5 所示，关于 #0 “mutation detection” 的研究主要集中在 1991-2008 年间，关于 #1 “transcriptional activator” 的研究从 1993 年持续至今，

关于 #2 “gene expression” 的研究从 1995 年持续至今，关于 #3 “balanced karyotype” 的研究从 1992 年持续至今，关于 #4 “insulin sensitivity” 的研究从 1999 年持续至今，关于 #5 “trypanosoma brucei” 的研究主要集中在 1993-2000 年间，关于 #6 “human papillomavirus” 的研究从 1998 年持续至今。从统计结果可以看出，关于 #0 “mutation detection” 这一类别的研究开始得最早，除了 #0 “mutation detection” 和 #5 “trypanosoma brucei” 以外，其他主题的研究一直在推进。

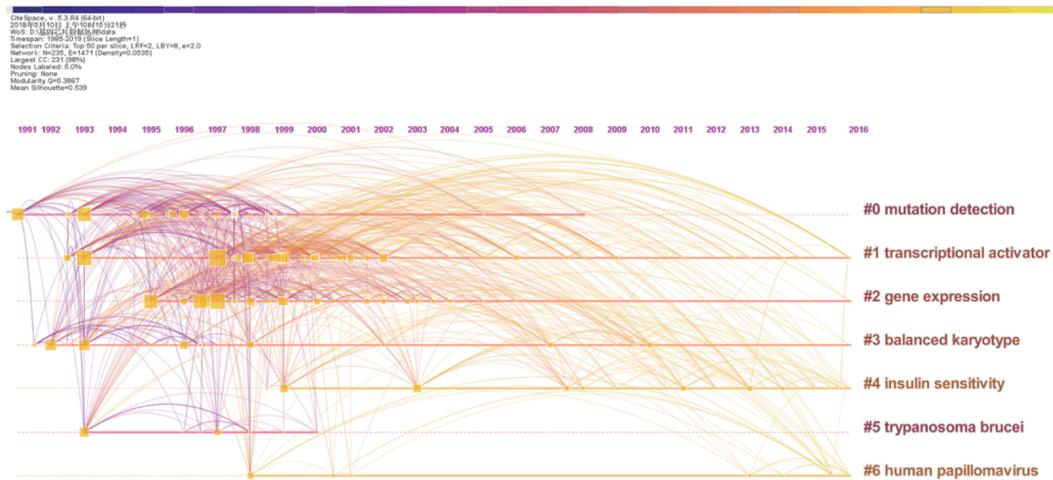


图 5 关键词 Time-Line 图谱

2.4 知识结构分析

文献共被引是指两篇或两篇以上文献共同出现在同一篇施引文献的参考文献目录中，则这两篇或两篇以上文献形成共被引关系^[17]，它体现了科学知识的累积性、连续性、继承性以及学科之间的交叉性和渗透性。文献的共被引频次越高，则表明彼此在研究主题、理论和方法上的相关性越紧密^[18]。此外，学术界还通常利用文献的被引频次来评估文献的学术价值和水平^[19]。由图 6 和表 3 可知，LOCKHART DJ 等于 1996 年发表的 *Expression monitoring by hybridization to high-density oligonucleotide arrays*、LI C 等于 2001 年发表的 *Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection* 以及 IRIZARRY RA 等于 2003 年发表的 *Exploration, normalization, and summaries of high density oligonucleotide array probe level data* 等文献节点较大，被引频次均高达 500 以上，分别位居前三，表明他们在文献中提出的相关概念与理论以及

对基因芯片应用的探索，为后续学者在该领域的研究提供了重要依据。

为了进一步了解基因芯片领域的知识结构，对共被引文献进行聚类分析^[20]，如图 7 所示，基因芯片领域的研究主要围绕 13 个维度进行，根据聚类节点的大小选取前 10 个主要标识词进行分析。

(1) 关于 affymetrix genechip (affymetrix 基因芯片) 的研究。affymetrix 公司是全球销量最高的基因芯片厂家，以完备的芯片设计、可靠的分析结果以及强大的生物信息学分析能力，成为了基因芯片领域的行业标准，其所生产的基因芯片具有很强的影响力。

(2) 关于 prenatal diagnosis (产前诊断) 的研究。与传统的超声检查以及入侵式细胞学诊断胎儿染色体异常相比，比较基因组杂交等技术具有很高的分辨率，不仅能检查出染色体上基因片段微小的缺失和重复，而且不需要进行细胞培养，可以直接用来检测羊水和绒毛膜绒毛样品，既节约了时间，又提高了产前诊断的准确性。

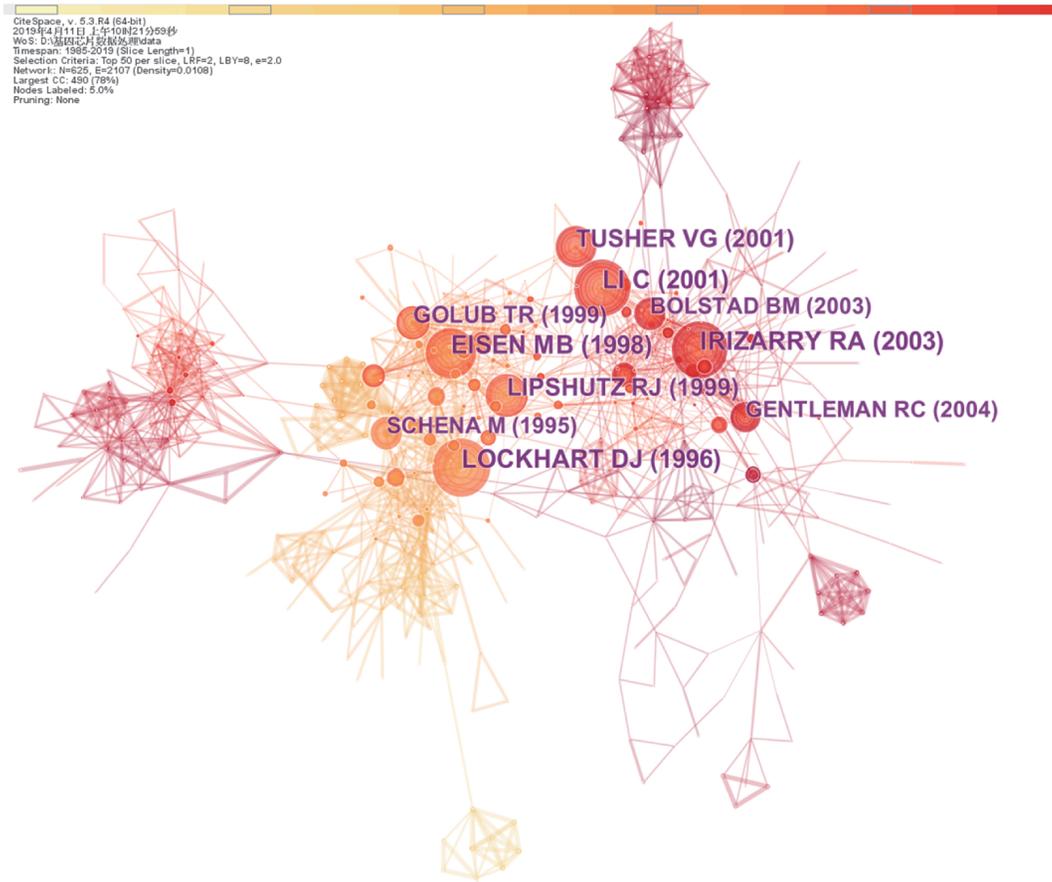


图 6 文献共被引图谱

表 3 主要共被引文献

序号	文献	第一作者	年份	被引频次
1	Expression monitoring by hybridization to high-density oligonucleotide arrays	LOCKHART DJ	1996	531
2	Model-based analysis of oligonucleotide arrays:Expression index computation and outlier detection	LI C	2001	512
3	Exploration, normalization, and summaries of high density oligonucleotide array probe level data	IRIZARRY RA	2003	500
4	Cluster analysis and display of genome-wide expression patterns	EISEN MB	1998	455
5	High density synthetic oligonucleotide arrays	LIPSHUTZ RJ	1999	393
6	Significance analysis of microarrays applied to the ionizing radiation response	TUSHER VG	2001	371
7	Molecular classification of cancer:Class discovery and class prediction by gene expression monitoring	GOLUB TR	1999	327
8	A comparison of normalization methods for high density oligonucleotide array data based on variance and bias	BOLSTAD BM	2003	296
9	Bioconductor:open software development for computational biology and bioinformatics	GENTLEMAN RC	2004	294
10	Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray	SCHENA M	1995	293

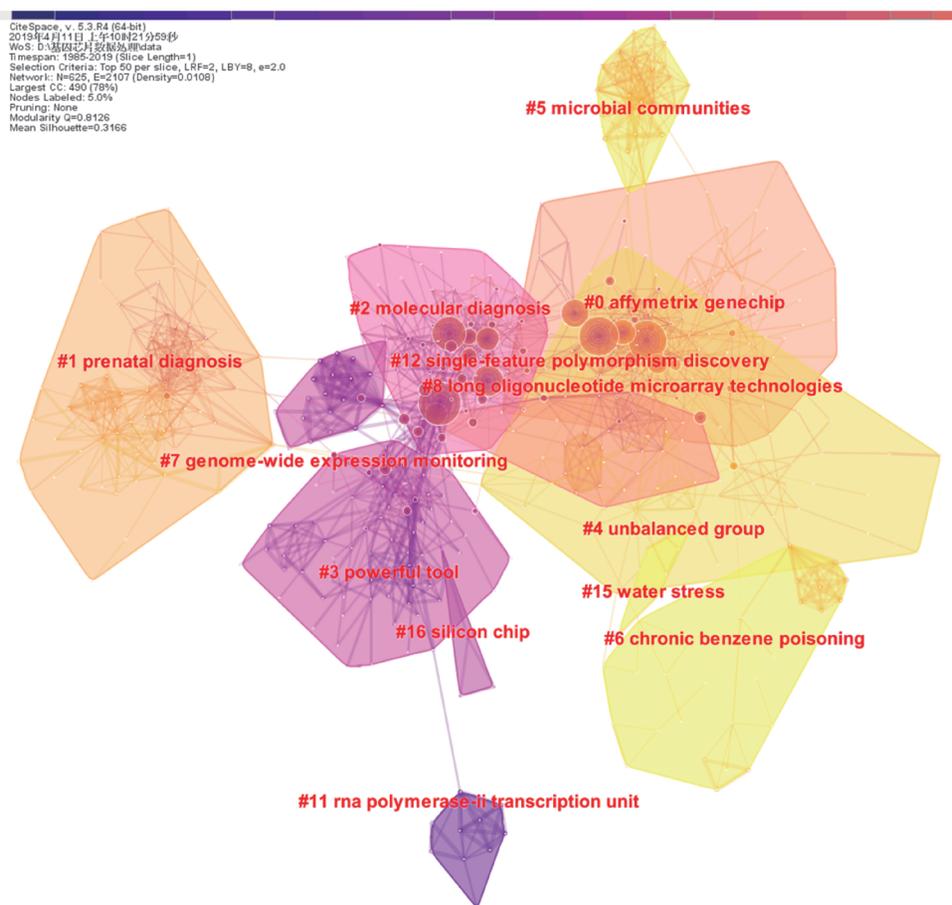


图7 文献共被引聚类图谱

(3) 关于 molecular diagnosis (分子诊断) 的研究。基因芯片的产生为分子生物学的研究提供了一个巨大的技术平台, 对鉴别疾病的分子亚型、获得药物的分子机制等多个方面都起到了重要作用。

(4) 关于 powerful tool (强有力工具) 的研究。在基因芯片得到广泛应用的同时, 对其所得数据的处理方法也受到了越来越多的关注。基因芯片的数据分析, 就是把原始数据按照一定的标准精简、归类, 然后从归类数据中寻找有实际生物学意义的过程。在数据分析过程中, 相关软件的开发和应用也提供了很好的助力, 例如 Genesist、AMADA、TIGR 等, 其中 TIGR

是使用较广的一种软件。

(5) 关于 unbalanced group (不平衡组) 的研究。平衡易位是一种常见的染色体结构异常, 利用全基因组单体型连锁分析技术, 使用基因芯片获得全基因组 SNP 基因分型, 实现正常型胚胎和易位型胚胎的精准区分, 从而帮助染色体平衡易位患者生育健康婴儿。

(6) 关于 microbial communities (微生物群落) 的研究。传统的微生物培养鉴定方法耗时较长, 难以满足快速检测的需求, 利用基因芯片技术可以在短时间内筛选出不同微生物的特异靶基因, 然后对微生物群落的生理状态、活动、组成等进行检测。此外, 基因芯片还可

应用于细菌的耐药性检测，这一点是传统培养方法无法实现的。

(7) 关于 chronic benzene poisoning (慢性苯中毒) 的研究。细胞内信号传导通路是一个复杂的网络，涉及多个基因群的相互作用，苯的毒性与致癌作用原因复杂，利用基因芯片技术检测苯中毒患者细胞信号传导相关基因的差异表达，有助于探讨慢性苯中毒的发病机理。

(8) 关于 genome-wide expression monitoring (全基因组表达监测) 的研究。利用基因芯片高通量的特性，可以在很短时间内测定不同功能状态、不同组织部位的基因表达，从而得到特异的全基因组表达谱。

(9) 关于 long oligonucleotide microarray technologies (长链寡核苷酸芯片技术) 的研究。与传统的 cDNA 芯片相比，长链寡核苷酸芯片具有成本低、特异性好等优点，目前正逐步被实验室和市场采纳。在制备长链寡核苷酸芯片的过程中，需要对片基的处理和点样液的选择进行改良，从而使芯片的固定效率、灵敏度、特异性以及信噪比等得到优化。

(10) 关于 single-feature polymorphism discovery (单一特征多态性发现) 的研究。例如，单核苷酸多态性是由于单个核苷酸改变而导致的核酸序列多态，是研究人类家族和动植物物种遗传变异的重要依据，被广泛应用于群体遗传学和疾病相关基因的研究，在药物基因组学、诊断学和生物医学研究中起着重要作用。

2.5 研究机构分析

通过对机构进行合作网络分析，有助于了解相关领域的研究力量布局^[21]。在机构合

作网络中，节点的大小代表了机构的发文量，节点间连线的粗细代表了机构之间的合作密切程度^[22]。图 8 中共包含了 342 个节点，1459 条连线，网络密度为 0.025，说明基因芯片领域各机构之间的合作联系较为紧密。在 1985-2019 年间，基因芯片的研究成果共涉及 6866 所研究机构。结合图 8 和表 4 可知，哈佛大学的发文量高达 332 篇，是基因芯片领域发文量最多的机构，占据着绝对优势地位。此外，美国国家癌症研究所、德克萨斯大学、斯坦福大学和华盛顿大学（西雅图）等也都是基因芯片领域硕果累累的知名研究机构，发文量均高达百篇以上，具有较强的科研实力和领域影响力。由表 4 可知，发文量位居前十的高产机构中，美国机构多达 8 所，日本和中国各占 1 所，可见美国在基因芯片领域的研究力量较为强劲。

3 结语

(1) 从研究时序来看，基因芯片论文数量经历了萌芽期（1989-1997 年）- 发展期（1998-2003 年）- 平稳期（2004-2009 年）- 成熟期（2010-2019 年）四个阶段，发文量总体呈先增后减的趋势，2007 年达到峰值 1020 篇，目前已基本进入成熟阶段。与历年文献数量的发展历程类似，基因芯片领域的研究发展脉络也大致经历了四个时期，目前已进入深化发展阶段，研究内容包括 biomarker（生物标志物）、angiogenesis（血管生成）、transcriptome（转录组）、microRNA（微型核糖核酸）、signaling pathway（信号通路）等。

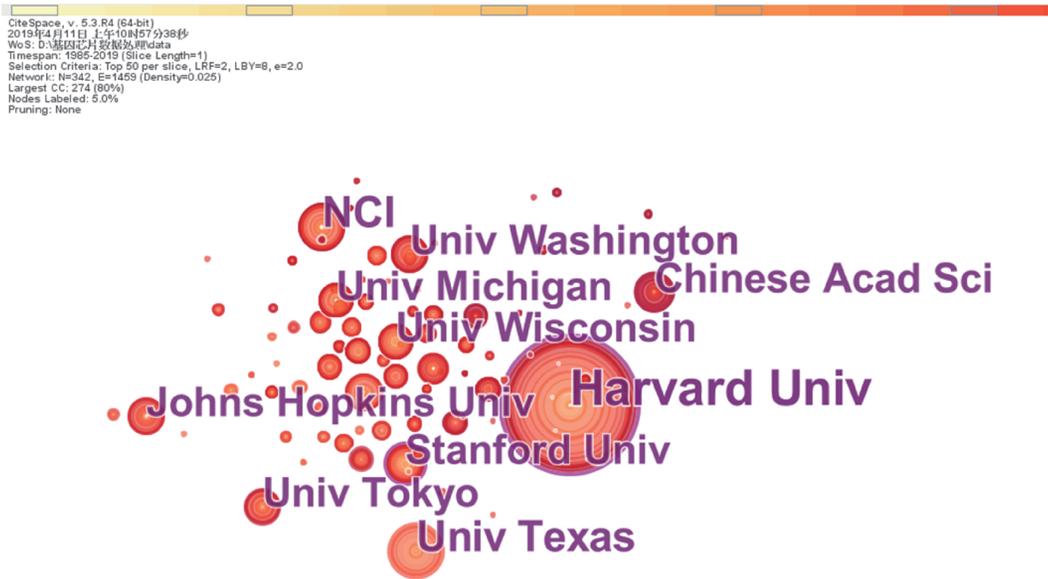


图 8 机构合作网络图谱

表 4 主要机构发文量分布

序号	机构英文名称	机构中文名称	发文量（篇）
1	Harvard Univ	哈佛大学	332
2	NCI	美国国家癌症研究所	143
3	Univ Texas	德克萨斯大学	139
4	Stanford Univ	斯坦福大学	134
5	Univ Washington	华盛顿大学（西雅图）	130
6	Univ Tokyo	东京大学	123
7	Univ Wisconsin	威斯康星大学	121
8	Univ Michigan	密歇根大学	118
9	Johns Hopkins Univ	约翰·霍普金斯大学	116
10	Chinese Acad Sci	中国科学院	115

(2) 从研究内容来看，基因芯片领域的研究热点主要集中在 gene expression（基因表达）、microarray（微阵列）、oligonucleotide array（寡核苷酸阵列）、hybridization（杂交）、

cancer（癌症）等方面。研究前沿主要体现在 arabidopsis thaliana（拟南芥）、inflammation（炎症）、disease（疾病）、proliferation（增殖）、microRNA（微型核糖核酸）等方面。对

其进一步归纳可知,基因芯片领域的研究主题主要包括 mutation detection (基因突变检测)、transcriptional activator (转录激活因子)、gene expression (基因表达)、balanced karyotype (染色体平衡易位核型分析)、insulin sensitivity (胰岛素敏感性)、trypanosoma brucei (布氏锥虫病)、human papillomavirus (人类乳头瘤病毒)等。知识结构主要包括 affymetrix genechip (affymetrix 基因芯片)、prenatal diagnosis (产前诊断)、molecular diagnosis (分子诊断)、powerful tool (强有力工具)、unbalanced group (不平衡组)、microbial communities (微生物群落)、chronic benzene poisoning (慢性苯中毒)、genome-wide expression monitoring (全基因组表达监测)、long oligonucleotide microarray technologies (长链寡核苷酸芯片技术)、single-feature polymorphism discovery (单一特征多态性发现)等。

(3)从研究主体来看,基因芯片领域的6866所研究机构之间合作联系较为密切,其中哈佛大学、美国国家癌症研究所、德克萨斯大学、斯坦福大学和华盛顿大学(西雅图)等机构属于中坚力量,哈佛大学贡献最多,在发文量上占据着绝对优势。此外,基因芯片领域的前十所高产机构基本以美国力量为主,可见美国在基因芯片领域的研究实力不容小觑。

参考文献

- [1] Bacher U, Shumilov E, Flach J, et al. Challenges in the introduction of next-generation sequencing (NGS) for diagnostics of myeloid malignancies into clinical routine use[J]. Blood Cancer Journal, 2018, 8(11):113.
- [2] Alhadrami H A. Biosensors: classifications, medical applications, and future prospective[J]. Biotechnology and Applied Biochemistry, 2018, 65(3):497-508.
- [3] Saliba J, Daou A, Damiati S, et al. Development of microplatforms to mimic the in vivo architecture of CNS and PNS physiology and their diseases[J]. Genes, 2018, 9(6):285.
- [4] Lizama B N, Palubinsky A M, McLaughlin B. Alterations in the E3 ligases parkin and chip result in unique metabolic signaling defects and mitochondrial quality control issues[J]. Neurochemistry International, 2018(117):139-155.
- [5] 有小娟, 鲁珍. 基因芯片技术及其在微生物检测中的应用[J]. 西安文理学院学报(自然科学版), 2017, 20(2):94-96, 116.
- [6] 姜国哲, 王兰成. 基于知识图谱的网络舆情知识组织方法研究[J]. 情报理论与实践, 2019, 42(1):58-64.
- [7] AMiner. 人工智能之知识图谱[R]. 北京:清华大学人工智能研究院, 北京智源人工智能研究院, 清华-工程院知识智能联合研究中心, 2019:2-9.
- [8] 李馨旖, 周清. 基于移动平均线改进的多元非线性回归模型[J]. 首都师范大学学报(自然科学版), 2017, 38(3):9-12.
- [9] Belvaux G, Wolsey L A. Bc-prod: A specialized branch-and-cut system for lot-sizing problems[J]. Management Science, 2000, 46(5):724-738.
- [10] 于果鑫. 国内军校图书馆研究路径及热点探析——基于 CiteSpace 的计量分析[J]. 现代情报, 2018, 38(8):144-153.
- [11] 肖香龙, 李信, 高寒, 等. 基于关键词共现的学科领域研究空白 (Research Gaps) 发现[J]. 情报工程, 2018, 4(6):37-50.
- [12] 孙冰, 徐晓菲, 苏晓. 技术扩散主路径及核心企业的识别研究——以手机芯片专利引文网络为例[J]. 情报学报, 2019, 38(2):201-208.
- [13] 李韵婷, 郑纪刚, 张日新. 国内外智库影响力研究的前沿和热点分析——基于 CiteSpace V 的可视化计量[J]. 情报杂志, 2018, 37(12):78-85.
- [14] 肖获昱. 基于 CiteSpace 的图书馆智库服务研究可

- 视化分析[J]. 图书馆工作与研究, 2018(11):94-99.
- [15] 梁怀新. 基于 Citespace 的我国国家安全研究知识图谱分析[J]. 情报杂志, 2019(6):23-29.
- [16] 李杰, 陈超美. Citespace: 科技文本挖掘及可视化(第二版)[M]. 北京: 首都经济贸易大学出版社, 2017:158-159.
- [17] 蔡小飞, 张菲菲. 基于 CiteSpace 的国内档案研究热点分布研究[J]. 宁波大学学报(教育科学版), 2019, 41(1):58-63.
- [18] 崔春舜, 孙晓北, 杨渊, 等. 基于 CiteSpace 的医学人工智能研究热点与前沿探测[J]. 中国数字医学, 2018, 13(10):8-10.
- [19] 黄亚江, 张书鸣, 梁云, 等. 基于 CiteSpace 的国内 BIM 研究演化过程、热点与主题探析[J]. 项目管理技术, 2019, 17(1):12-19.
- [20] 胡泽文, 李玉平, 张静, 等. 大数据研究前沿、热点与合著模式的图谱分析[J]. 情报工程, 2018, 4(4):34-49.
- [21] Huang Y, Ma J, Porter A L, et al. Analyzing collaboration networks and developmental patterns of Nano-Enabled Drug Delivery (NEDD) for brain cancer[J]. Beilstein Journal of Nanotechnology, 2015, 6(6):1666-1676.
- [22] 周静梅, 黄颖, 孔小梅, 等. 我国图书情报学领域机构合作网络分析[J]. 情报工程, 2017, 3(6):103-115.