

上海交通大学医学院



学者介绍
Author introduction



黄金艳 博士

副研究员、博士生导师

ORCID ID: 0000-0002-8053-0209

HUANG Jin-yan

Ph.D

Associate Professor, Doctoral Supervisor

ORCID ID: 0000-0002-8053-0209

黄金艳 (1980—), 上海交通大学医学院附属瑞金医院、上海血液学研究所、医学基因组学国家重点实验室副研究员和博士生导师。2007 年于同济大学获博士学位。2009—2011 年于法国巴黎国家测序中心 (CNG) 和人类多态性研究中心 (CEPH) 从事生物信息学的博士后研究工作。2012—2014 年于哈佛大学公共卫生学院从事高通量测序数据分析相关的博士后研究工作。现任中国病理生理学会实验血液学专业委员会青年委员会委员, *PLoS One*、*Bioinformatics*、*Journal of Nutrigenetics and Nutrigenomics* 等国际期刊审稿专家。

• 研究主要集中在基因组、转录组和表观遗传等第二代测序数据的生物信息学分析和医学大数据等交叉学科的研究, 包含生物信息分析工具开发和白血病数据分析应用研究 2 个方面。以第一作者或通信作者在 *Nature Communication*、*PNAS*、*Bioinformatics*、*Nucleic Acids Research*、*Cell Research*、*EBioMedicine* 等国际知名期刊发表论文 10 余篇, 另以共同作者在 *Nature*、*Nature Genetics*、*British Medical Journal* 等著名期刊发表论文 30 余篇。2014 年于法国巴黎获吉尔—卡恩学术奖。2015 年入选上海高校青年东方学者岗位计划、上海市浦江人才计划, 2016 年入选上海市教育委员会高峰高原学科建设计划。

HUANG Jin-yan born in 1980, received his Ph.D from Tongji University in 2007 respectively. He engaged in bioinformatics postdoctoral research in Center National deGénotypage (CNG) and the Centre d'Etude du Polymorphisms Humain (CEPH) in Paris, France from 2009 to 2011. Then he went to Harvard School of Public Health to do postdoctoral research from 2012 to 2014, which focused on the high-throughput sequencing data analysis. Dr. HUANG is currently as an associate professor and doctoral supervisor in State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine. He is a member of Youth Committee of the Experimental Blood Committee of the Chinese Society of Pathophysiology. Meanwhile, he has served as a peer reviewer for several journals, such as *PLoS One*, *Bioinformatics*, *Journal of Nutrigenetics and Nutrigenomics*, etc.

• Dr. HUANG's main research focuses on the interdisciplinary research of bioinformatics analysis of next-generation sequencing data such as genome, transcriptome, and epigenetics, as well as medical big data, including the development of bioinformatics analysis tools and the application in leukemia data analysis. He has published several papers as the first or corresponding author in some famous journals, including *Nature Communication*, *PNAS*, *Bioinformatics*, *Nucleic Acids Research*, *EBioMedicine*, *Cell Research*, etc. Besides, he is the co-author of more than 30 articles in the top journals, including *Nature*, *Nature Genetics*, *British Medical Journal*, etc. Dr. HUANG has won the "Gilles-Kahn Academic Award" in 2014 in Paris. Also, he was enrolled into "Shanghai Youth Eastern Scholar" and "Shanghai Pujiang Plan" in 2015, and "Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support" in 2016.



论著·基础研究

急性淋巴细胞白血病基因融合与突变知识库的构建

严天奇¹, 陈立伟², 朱勇梅², 李剑峰², 代雨婷^{2,3}, 崔舒雅², 姜璐², 陈冰², 黄金艳²

1. 上海交通大学系统生物医学研究院, 系统生物医学教育部重点实验室, 上海 200240; 2. 上海交通大学医学院附属瑞金医院, 上海血液学研究所, 医学基因组学国家重点实验室, 上海 200025; 3. 上海交通大学生命科学技术学院, 上海 200240

[摘要] 目的· 建立急性淋巴细胞白血病 (acute lymphoblastic leukemia, ALL) 基因融合与突变知识库, 以辅助临床基因检测。

方法· 通过对文献进行文本挖掘, 收集 ALL 相关的基因融合与突变注释信息。基于 NodeJS 平台的 Express 框架和 MySQL 数据库系统的服务端开发环境, 构建知识库服务网站。**结果**· 通过对文献进行文本挖掘和人工手动矫正, 共收集了 246 条 ALL 相关融合和突变基因词条, 每项词条包含生物学性状、临床相关性解释、临床指导意见、靶向药物和化疗药物等多个生物学和临床相关条目, 建立起一个便于管理的 ALL 相关知识库。**结论**· 该知识库除了包含 ALL 相关或潜在相关癌症基因基本信息外, 还着重整理了临床相关的注释信息, 为 ALL 的临床基因检测及后续的精准医疗提供参考。

[关键词] 急性淋巴细胞白血病; 基因融合; 基因突变; 知识库

[DOI] 10.3969/j.issn.1674-8115.2018.09.005 **[中图分类号]** R733.7 **[文献标志码]** A

Construction of a knowledge database of gene fusion and mutation in acute lymphoblastic leukemia

YAN Tian-qi¹, CHEN Li-wei², ZHU Yong-mei², LI Jian-feng², DAI Yu-ting^{2,3}, CUI Shu-Ya², JIANG Lu², CHEN Bing², HUANG Jin-yan²

1. Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China;

2. State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China;

3. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[Abstract] **Objective**· To construct a database of fusion and mutation gene annotations for acute lymphoblastic leukemia (ALL) to assist genetic testing. **Methods**· ALL related gene annotations were collected by mining medical literature. The web server of the database was constructed based on Express framework, a NodeJS web application framework, and MySQL as the server-side development environment. **Results**· A total of 246 ALL-associated fusion and mutation gene entries were collected through programmed and manual text mining, including biological characteristics, clinical characteristics, clinical directions, target drugs and chemotherapeutic drugs. A web server of the ALL related gene knowledge database was established for convenient data management. **Conclusion**· In addition to the basic information about ALL related or potentially related genes, the database also involves clinical information which can be a reference tool for precision medicine of ALL.

[Key words] acute lymphoblastic leukemia; gene fusion; gene mutation; knowledge database

急性淋巴细胞白血病 (acute lymphoblastic leukemia, ALL) 是血液系统中淋巴细胞异常增殖引起的一类恶性肿瘤, 其中儿童占所有 ALL 患者的近 80%^[1]。ALL 患者体内造血干细胞在分化发育为淋巴细胞的过程中, 由于多种因素阻滞于某个发育阶段并恶性增殖, 进而严重影响正常的生理功能。按细胞类型分, ALL 可划分为急性 B 淋巴细胞白血病 (B-ALL) 和急性 T 淋巴细胞白血病 (T-ALL)。随着 ALL 各种组学的研究不断深入, 从基因到细胞层面的分子标志物不断被纳入 ALL 的诊断标准。得益于建立

在 ALL 分子分型基础上的诊断治疗策略, ALL 患者的生存时间和生存质量得到了显著提高, 特别是儿童 ALL 患者的长期生存率已经从最早不到 10% 提高到了近 90%^[2]。

ALL 在基因组水平上最常见的异常包括基因突变 (gene mutation) 和基因融合 (gene fusion)。它们不仅是 ALL 分型的重要依据, 也是治疗策略和预后判断的重要标志。在上海血液学研究所先前关于 B-ALL 的研究工作中, 对 92 例成人和 111 例儿童 B-ALL 病例进行了全景式的测序数据分析, 发现了未报道的 105 个重现性基因突变、29

[基金项目] 国家自然科学基金 (81570122, 81770205); 上海市教育委员会高峰高原学科建设计划 (20161303) (National Natural Science Foundation of China, 81570122, 81770205; Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support, 20161303)。

[作者简介] 严天奇 (1993—), 男, 硕士生, 电子信箱: tianqi_yan@sjtu.edu.cn。

[通信作者] 黄金艳, 电子信箱: jinyan@shsmu.edu.cn。



个融合基因，并深入探索了 8 个表达谱特征组对应的遗传学特征^[3]。同时，在后续关于 T-ALL 的研究工作中，分析了 130 例患者的测序数据，发现了未曾报道的 6 个基因突变和 18 个融合基因，也进一步探索了 3 个表达谱亚组各自的分子标志物^[4]。Lilljebjorn 等^[5] 对 195 例 B-ALL 样本的测序数据进行了分析研究，发现了 27 种新的融合基因，还发现在 *IGH-DUX4* 和 *ERG-DUX4* 的融合亚组中 *DUX4* 基因的过表达常伴随着 *ERG* 的缺失突变。Qian 等^[6] 在对 231 例 ALL 儿童病例的分析中，发现了 31 种新融合基因，并在 *EP300/CREBBP-ZNF384* 融合亚组中发现 *CLCF1* 和 *BLTA* 基因的表达上调。Liu 等^[7] 对 264 例 T-ALL 样本进行了基因组整合分析，发现了 53 个新驱动基因；并对突变基因与 ALL 分组特征进行了研究，发现 *NRAS/FLT3* 突变与幼稚型 T-ALL、*JAK3/STAT5B* 突变与 *HOXA1* 失调的 ALL、*PTPN2* 突变与 *TLX1* 失调的 T-ALL、*PIK3RI/PTEN* 突变与 *TAL1* 失调的 ALL 等相关。

以上 ALL 全景式基因组及转录组的研究工作，为深入研究 ALL 致病机制及分子分型提供了基础，为临床基因融合及突变检测提供了理论基础，同时也说明传统的检测手段不能充分检测到相应的融合基因与突变。检测出融合基因或突变，仅仅是检测工作的第一步，更重要的是对检测结果的准确解读。ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) 是美国国立卫生研究院 (National Institutes of Health, NIH) 维护的疾病变异知识库，它对大量疾病相关的基因突变都做了注释，但是并未收录融合基因的相关知识^[8]。DisGeNET (<http://www.disgenet.org/>) 是一个收集整理其他公共数据库信息，并计算出疾病与基因突变关联度的知识库^[9]。The Atlas of Genetics and Cytogenetics in Oncology and Haematology (<http://atlasgeneticsoncology.org/>) 数据库成立于 1997 年，致力于收集完善肿瘤疾病相关的基因、细胞遗传学和临床研究等各方面的研究信息^[10]。虽然已经有了如上这么多综合性的注释知识库，但对于高异质性的 ALL 疾病，特别是在基因检测方面，仍然需要一个更有针对性的临床基因检测相关的知识库。

因此，构建一个以大队列基因组学数据分析结果为基础，涵盖 B-ALL 及 T-ALL 部分未曾引起重视的潜在驱动突变及融合基因的知识库，整合临床治疗及用药信息，对于 ALL 的临床检测具有重要的现实意义。本研究构建了一个基于 Web 服务的 ALL 知识库管理查询系统，实现了信息多人协同编辑，同时联合一线临床医生、基础研究人员和生物信息分析人员一同收集修订词条内容，以尽量确

保注释信息的科学性。

1 材料与方法

1.1 基因注释信息收集

在实验室前期的工作中，我们收集了 1 285 例 B-ALL 的测序数据^[3, 5-6, 11-15] 和 557 例 T-ALL 的测序数据^[4, 6-7, 16-17]，并通过分析找到了大量的高频基因融合和突变。基于这些数据分 析 和 COSMIC (<https://cancer.sanger.ac.uk/cosmic/>)^[18] 数据库中的 ALL 热点突变，本知识库初期纳入了 91 条融合基因和 190 条突变基因。这些基因基本涵盖了目前在 ALL 临 床 上 遇 到 的 绝 大 部 分 高 重 现 性 的 融 合 / 突 变 受 累 的 基 因，如 *SETD2*、*ABL1*。

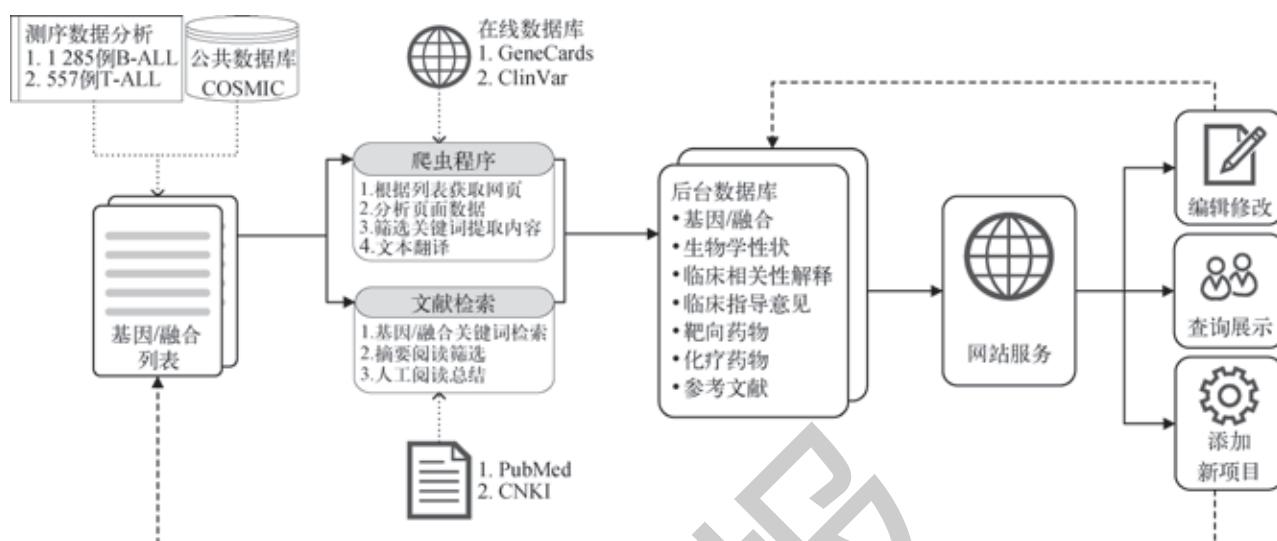
对每个融合 / 突变基因，我们设计了 7 个信息字段：基因名、生物学性状、临床相关性解释、临床指导意见、靶向药物、化疗药物、参考文献。在仔细研究数据源的特征后，通过编写爬虫程序（网络内容抓取程序），获得每项词条的大部分的注释信息，其他部分则通过人工文献检索提取。每项的生物学性状包括基因编码的蛋白质、编码蛋白的功能、相关的疾病以及突变情况等信息，主要通过爬虫程序从公共数据库，如 GeneCards 数据库 (<http://www.genecards.org/>) 和 ClinVar 中抓取，然后由人工整理。临床相关解释、临床指导意见、靶向药物及化疗药物等信息主要通过检索 PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) 和 CNKI (<http://www.cnki.net/>) 等平台的文献获得。临床相关解释包括基因突变的位点、融合基因的断点、涉及的白血病亚型、流行病学情况以及对预后的影响。临床指导意见包含推荐治疗方案以及治疗效果（图 1）。

1.2 知识库服务设计

知识库所有信息的存储管理采用 MySQL (<https://www.mysql.com/>) 数据库管理系统。项目收集的相关信息分别存储在不同的主表单中，并且还提供一个临时数据表，用于存储临时编辑信息，以实现多人协同编辑。同时，为了方便管理不同编辑员的操作权限，我们还设计了一个用户权限管理数据表。

知识库构建的底层数据库，通过 Web 网站的方式向终端用户提供查询、录入及修订等功能。Web 前端页面以 Bootstrap (<http://getbootstrap.com/>) 作为框架构建，服务后台使用基于 JavaScript 语言的 NodeJS (<https://nodejs.org/en/>) 平台开发（图 2）。

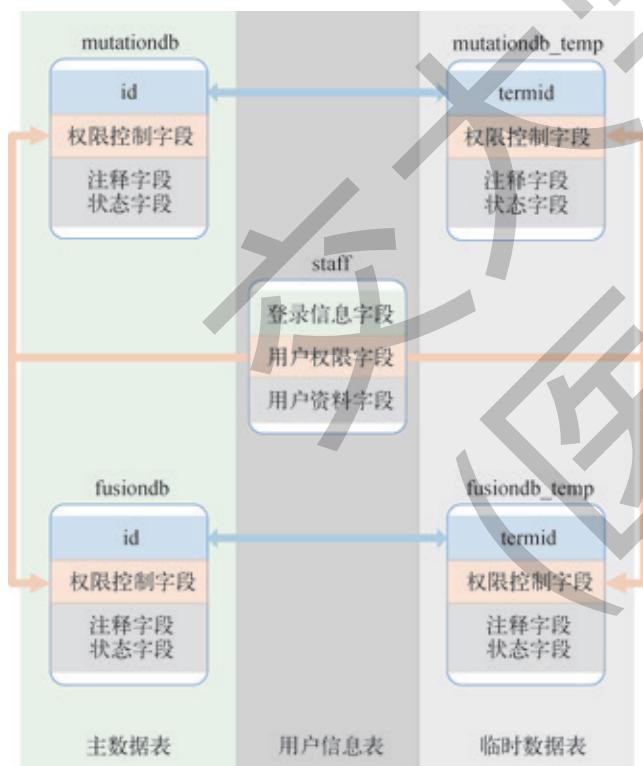




注：首先通过数据分析和查询公共数据库，筛选出需注释的知识库基因和融合列表。再以此列表作为检索条件，通过自动化爬虫程序抓取公共数据库中的相关信息，同时通过人工检索阅读的方式查阅总结文献信息。然后将收集的基因注释信息导入后台数据库的7个信息字段。最后以此知识库构建网站服务，为用户提供修改、查询和新增新项目的功能界面。

图1 基因注释数据收集流程

Fig 1 Data collection process of gene annotation



注：知识库的数据表可分成数据信息记录表和权限控制表2种。其中信息记录表包括：主数据表mutationdb、fusiondb，和对应的临时数据表mutationdb_temp、fusiondb_temp。主数据表信息可由用户查询检索。临时数据表存储编辑人员提交的修订信息，并在审核后更新主数据表记录。主数据表和临时数据表分别通过id字段和termid字段映射关联。权限控制表即用户信息表staff。用户信息表staff用于验证用户登录信息，并控制用户对各词条的编辑修改操作。用户信息表staff通过用户权限字段与各信息表中的权限控制字段相关联，用以控制用户对各条记录的修改访问。

图2 知识库设计

Fig 2 Design of the MySQL database

2 结果

2.1 数据信息

本知识库共完成了155个突变基因词条和91个融合基因词条的相关资料的收集整理工作。由于白血病基因的复杂性，知识库中不少词条不仅仅见于ALL，在其他类型白血病中也频繁出现，如慢性淋巴细胞白血病（chronic lymphocytic leukemia, CLL）、慢性髓性白血病（chronic myeloid leukemia, CML）、急性髓性白血病（acute myeloid leukemia, AML）。此外，还有44条词条未发现与白血病有关的研究报道，但相关研究都表明其与血液疾病相关。在靶向药物和化疗药物信息部分，由于药物开发要求基因作用机制研究得非常透彻而且开发周期极为漫长，所以这部分只有32个词条有较为完整的注释，其他都处于空缺状态。例如本知识库中收录的ALL的经典分型Ph染色体阳性组分子标志基因融合BCR-ABL1，其部分信息字段如下。

生物学性状：BCR基因定位于22q11.23，编码的蛋白具有Ser/Thr激酶活性，是p21rac的GTP酶活化蛋白。ABL1原癌基因位于9q34.12，编码一个非受体酪氨酸激酶，参与细胞分裂、黏附、分化、应激等多个细胞代谢过程。该蛋白含3个SRC同源（SH）结构域，即SH1、SH2、SH3，其中SH1具有酪氨酸激酶活性，SH2和SH3调控酪氨酸激酶活性，SH3起负调控作用。另外，ABL1蛋白还有3个核定位信号结构域，3个DNA结合结构域和1个F-肌动蛋白结合结构域。已经发现ABL1基因在白



血病细胞中能与多个基因发生融合, 最常见的即染色体易位 t(9;22)(q34;q11) 形成的 BCR-ABL1 融合基因。该融合基因因断裂点不同可产生 3 种融合蛋白, 分别为 p210、p190 和 p230, 三者的 ABL 激酶活性均显著升高。

临床相关性解释: t(9;22)(q34;q11) 形成 BCR-ABL1 融合基因, 可见于 95% 以上的 CML, 3% ~ 5% 的儿童 B-ALL, 成人 B-ALL 的发生率随患者发病年龄增大而升高, 分别为 14% (20 ~ 29 岁)、24% (30 ~ 39 岁)、33% (40 ~ 49 岁)、26% (≥ 50 岁)。BCR-ABL1 融合基因罕见于 AML 和 T-ALL。BCR-ABL1 阳性的 B-ALL 中, 约 50% 成人和 80% 儿童表达 p190 融合蛋白, 其余表达 p210 融合蛋白。患者初发时白细胞计数升高常伴中枢神

经系统受累。免疫表型多为 CD10⁺, 有时为 CD19⁺CD10⁻。BCR-ABL1 阳性的 B-ALL 患者预后差, 酪氨酸激酶抑制剂联合强化疗可提高此类患者的生存率。

本知识库现已为上海血液学研究所的基因检测研发项目提供部分知识的注释内容, 后期进一步完善后, 将为临床检测及诊断提供直接服务。

2.2 知识库网站

本项目 Web 网站主要设计了 4 个模块: 登录验证模块、用户管理模块、数据信息管理模块、数据查询模块, 以实现为授权用户提供在线合作编辑和检索浏览等功能。网站的部分页面展示见图 3。



注: A. 登录页面; B. 词条概览管理页 (①新建编辑词条项; ②词条项内容编辑或用户编辑权限管理); C. 词条编辑页 (③变更词条项的当前状态); D. 词条项编辑授权页 (④通过下拉框选择用户, 点击添加用户并确认即可授权用户编辑词条); E. 用户添加页 (⑤添加用户前需检查是否已存在同名用户; ⑥用户组、权限级别和活动状态选择); F. 词条检索结果。

图 3 ALL 知识库网站页面

Fig 3 Webpages of the ALL database

2.2.1 登录验证模块 未经验证的用户通过网址进入本知识库的任何页面都会跳转到登录页面 (图 3A)。按照页面提示输入用户名、密码及验证码登录之后, 就能访问知识库内容了。在用户登录的同时, 服务器后台会记录此次登录时间和 IP 地址, 并从 staff 表中取出登录账户的相关权限信息。在权限信息有效期内, 访问行为通过与访问内容的权限预设值比较实现层级权限控制。

2.2.2 数据信息管理模块 数据信息管理模块实现了用户在管理后台编辑授权、修改编辑、提交修改、审核评定的协同编辑工作。编辑员等级的用户可以对其授权编辑的

词条信息进行编辑修改; 当提交编辑或修改信息后, 词条状态即发生变化, 以通知管理员进行审核评定。管理员等级的用户可以直接对词条信息进行修改, 或者审核由编辑员提交的修改信息。如果审核通过, 主数据表信息即根据修改信息进行更新, 否则重新通知编辑员修改该项。此外, 管理员还可对每个词条的授权编辑用户进行管理, 减少或增加某个词条的编辑员。图 3B ~ D 分别展示了管理员在管理后台对已有词条项的管理页面、词条项内容的编辑页面、词条项的权限管理页面。

2.2.3 用户管理模块 用户管理模块是主要为网站管理



员设计的功能模块。网站管理员登录网站管理后台之后，通过权限控制，可以浏览普通用户无法接触的用户管理模块。通过此模块可以查看所有用户的最近登录时间及IP地址，创建新用户，停用已有用户，编辑已有用户的基本信息及访问权限等。图3E展示了添加新用户的界面。

2.2.4 数据查询模块 数据查询模块是本知识库的前台功能模块，所有登录用户都可通过此模块检索某个ALL相关的基因或者融合。根据用户搜索的是基因还是融合，页面会根据数据库中的记录向用户给出选择提示框。如果数据库中有对应的记录，便会展示知识库中的记录信息(图3F)。

3 讨论

生物相关知识库的构建是生物信息学发展的一大基石，生物信息学分析结果的解释离不开知识库信息的注释。在异质性较高的恶性肿瘤研究领域，从基因组到转录组再到蛋白组以及相关药物研究等方面，已经相继建立了大量的生物信息学数据，如The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>) 和 European Genome-phenome Archive (EGA, <https://ega.crg.eu/>) 等。此外，The Atlas of Genetics and Cytogenetics in Oncology and Haematology 数据库除了收集大量实体瘤的信息外，更将白血病单独列出。目前，国内有刘辉等^[19]对318个相关

基因构建的冠心病基因注释数据库(CADgene)，潘小勇等^[20]构建的基因表达量和疾病的关联数据库(OUGene)，葛树建等^[21]构建的阿尔茨海默病临床患者信息的数据平台等众多医学数据库。

本项目从临床医生的使用角度出发，构建ALL相关基因知识库，为ALL的精准医疗做出了补充。知识库中词条各字段的信息如前文中基因融合BCR-ABL1所示：生物学信息注释了词条涉及基因的一些基本信息，特别是其编码蛋白的各功能结构域信息；临床相关性解释包含与该词条相关的流行病学研究结果以及一些典型的分子表型特征。

与目前国内的一些疾病临床医学数据库不同，本知识库更注重收集突变/融合的临床研究资料，而不是每个患者的疾病相关信息。这种归纳性的研究资料无法满足临床医生对特定患者的诊断参考需求，但却更适用于患者确诊后的治疗方案选择。与国内外癌症基因研究型数据库相比，本知识库有临床研究报道的词条，收集整理了临床指导意见等信息，并经由临床医生修订编辑，因而在临幊上更具有实用价值。此外，本知识库构建的数据后台管理应用还提供了完善的词条增添、修订等功能，为以后的数据扩充、更新提供了方便。本知识库上线测试使用后，除收到一些关于词条内容的修改意见外，还有不少关于网站功能的意见，如搜索结果的模糊匹配、词条审阅后的消息反馈等。根据这些反馈意见，我们将继续完善、丰富知识库的内容与功能。

参·考·文·献

- [1] Terwilliger T, Abdul-Hay M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update[J]. *Blood Cancer J*, 2017, 7(6): e577.
- [2] Xu H, Shu Y. Insights of acute lymphoblastic leukemia with development of genomic investigation[J]. *Methods Mol Biol*, 2018, 1754: 387-413.
- [3] Liu YF, Wang BY, Zhang WN, et al. Genomic profiling of adult and pediatric B-cell acute lymphoblastic leukemia[J]. *EBioMedicine*, 2016, 8: 173-183.
- [4] Chen B, Jiang L, Zhong ML, et al. Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia[J]. *Proc Natl Acad Sci U S A*, 2018, 115(2): 373-378.
- [5] Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia[J]. *Nat Commun*, 2016, 7: 11790.
- [6] Qian M, Zhang H, Kham SK, et al. Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP[J]. *Genome Res*, 2017, 27(2): 185-195.
- [7] Liu Y, Easton J, Shao Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia[J]. *Nat Genet*, 2017, 49(8): 1211-1218.
- [8] Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence[J]. *Nucleic Acids Res*, 2018, 46(D1): D1062-D1067.
- [9] Piñero J, Bravo À, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants[J]. *Nucleic Acids Res*, 2017, 45(D1): D833-D839.
- [10] Huret JL, Ahmad M, Arsaban M, et al. Atlas of genetics and cytogenetics in oncology and haematology in 2013[J]. *Nucleic Acids Res*, 2013, 41(Database issue): D920-D924.
- [11] Churchman ML, Low J, Qu C, et al. Efficacy of retinoids in IKZF1-mutated BCR-ABL1 acute lymphoblastic leukemia[J]. *Cancer Cell*, 2015, 28(3): 343-356.
- [12] Gu Z, Churchman M, Roberts K, et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia[J]. *Nat Commun*, 2016, 7: 13331.
- [13] Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia[J]. *N Engl J Med*, 2014, 371(11): 1005-1015.
- [14] Roberts KG, Morin RD, Zhang J, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia[J]. *Cancer Cell*, 2012, 22(2): 153-166.
- [15] Zhang J, McCastlain K, Yoshihara H, et al. Dereulation of DUX4 and ERG in acute lymphoblastic leukemia[J]. *Nat Genet*, 2016, 48(12): 1481-1489.
- [16] Seki M, Kimura S, Isobe T, et al. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia[J]. *Nat Genet*, 2017, 49(8): 1274-1281.
- [17] Yasuda T, Tsuzuki S, Kawazu M, et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults[J]. *Nat Genet*, 2016, 48(5): 569-574.
- [18] Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution[J]. *Nucleic Acids Res*, 2017, 45(D1): D777-D783.
- [19] Liu H, Liu W, Liao Y, et al. CADgene: a comprehensive database for coronary artery disease genes[J]. *Nucleic Acids Res*, 2011, 39(Database issue): D991-D996.
- [20] Pan XY, Shen HB. OUGENE: a disease associated over-expressed and under-expressed gene database[J]. *Science Bulletin*, 2016, 61(10): 752-754.
- [21] 葛树建, 林炜炜, 丛琳, 等. 基于Web构建山东省阿尔茨海默病临床数据库[J]. *山东大学学报(医学版)*, 2015, 53(10): 42-45.