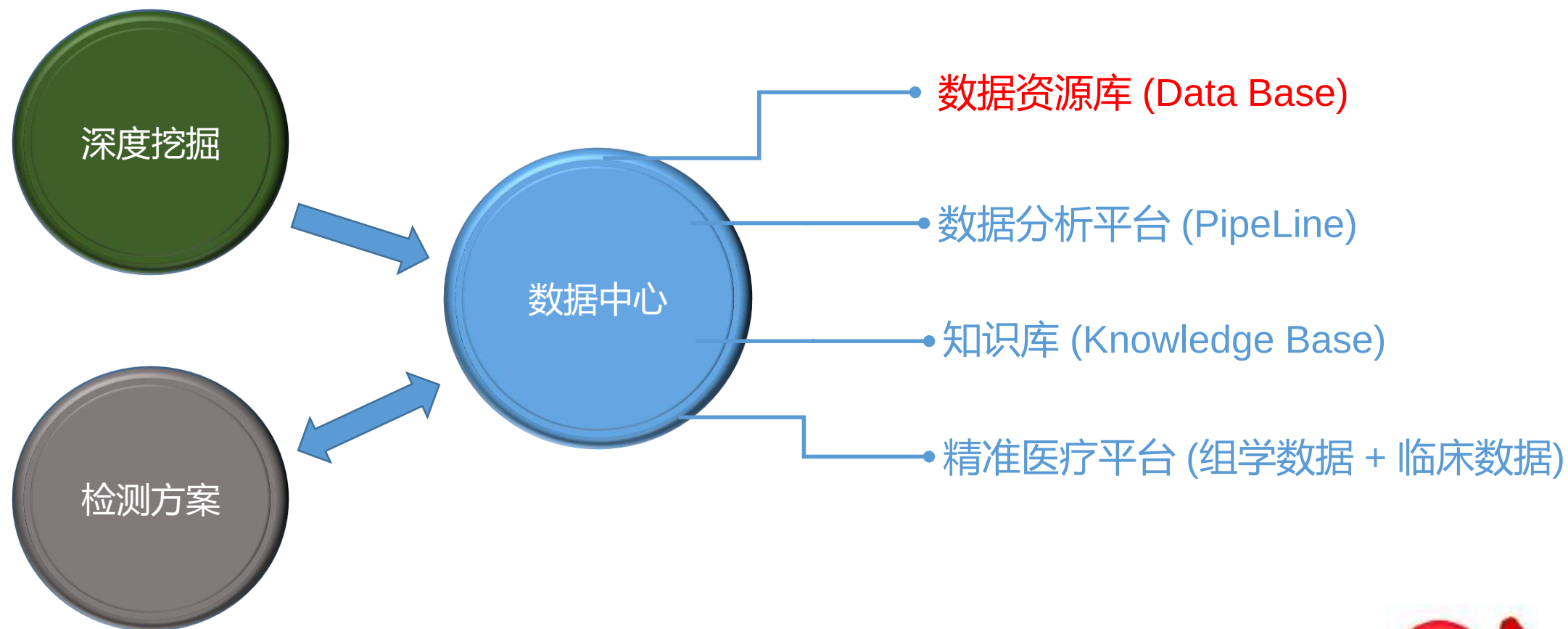


组学大数据平台与精准医疗



- 荧光定量PCR、基因芯片、SNP分型、二代测序



乳腺癌组学数据分析与可视化平台BCIP

平台简介

建立了以基因为中心的乳腺癌数据分析平台。

分析处理了来自TCGA、metabric、GEO三大数据库中的30个数据集的数据，包含9000多个组织样本。样本的临床数据包括癌症分型、分期、是否绝经、预后、ER+ / -、PR+ / -、Her2+ / -、P53突变、年龄等。

方便生物医学工作者，对关注的基因进行检索，从差异表达分析、生存分析、共表达分析、KEGG代谢通路等多个层次进行分析并可视化展示。

辅助识别乳腺癌的调控和驱动基因，找到乳腺癌研究和治疗的潜在的生物标志物。

案例成果

网址：<http://omics.bmi.ac.cn/bcancer/>
文章：BCIP: a gene-centered platform for identifying potential regulatory genes in breast cancer[J]. Scientific Reports, 2017, 7.

DOI: [doi:10.1038/srep45235](https://doi.org/10.1038/srep45235) 影响因子: 4.259

PMID: 28327601

文章发表于2017年Scientific Reports

SCIENTIFIC REPORTS

OPEN
BCIP: a gene-centered platform for identifying potential regulatory genes in breast cancer

Jiag Wu¹, Shuofang Hu¹, Yuesen Chen¹, Jinyang Li¹, Yan Zhang¹, Xue-Yue Qiang Shi¹, Ningling Xiao¹ & Xiaomeng Peng¹

Breast cancer is a disease with high heterogeneity. Many causes on homogeneity and progression are still elusive. It is critical to identify genes that play important roles in the progression of tumors, especially for tumors with gene progression such as breast cancer. BCIP is a gene-centered platform. To facilitate the identification of potential regulatory genes in breast cancer, we present the Breast Cancer Integrative Platform (BCIP, <http://omics.bmi.ac.cn/bcancer/>). BCIP maintains multi-omic data collected with strict quality standards and processed with various normalization methods, including gene expression profiles from RNA-seq and TSS-normalized samples, copy number variations, methylation profiles, 3D genome data, miRNA-NR-target interactions, co-expressed genes, KEGG pathways, and transcription factor (TF) gene regulatory networks. This platform provides a user-friendly interface integrating comprehensive and flexible analysis tools on differential gene expression, copy number variations, and survival analysis. The platform characteristics of BCIP is that users can perform a variety of complex analysis with single or combined clinical features, including subtypes, histological grades, pathological stages, metastatic status, lymph node status, ER/PR/HER2 status, TNM invasion status, menopausal status, age, tumor size, therapy response, and prognosis. BCIP will help to identify regulatory or driver genes and candidate biomarkers for further research in breast cancer.

Breast cancer is a frequently diagnosed cancer and is the leading cause of cancer death among females worldwide. In 2016, 1.7 million cases were diagnosed and 521,000 deaths occurred in 2016, accounting for 25% of the total cancer cases and 17% of all the cancer deaths among females¹. Breast cancer is a heterogeneous disease with a high degree of diversity in morphology, histology, pathological features, and genomic alterations such as gene mutations and structural variations². Advances in high-throughput sequencing and other technologies have enabled breast cancer researchers to identify different prognostic indicators and therapeutic responses³⁻⁵. A number of studies have focused on the complex interactions and dysregulation of gene expression in particular breast cancer subtypes⁶⁻⁸.

In fact, breast cancer (BRCA) has high heterogeneity and the most common and the most complex is breast ductal carcinoma⁹. This is a high-grade, heterogeneous disease with a poor genetic outcome and thus lacks effective targeted therapy¹⁰. However, several studies based on TSS have identified potential candidate genes that may be potential drug targets for the treatment of BRCA¹¹⁻¹³. For example, BRCA has been characterized as an oncogenic driver in breast cancer¹⁴. The breast cancer (BRCA) data analysis platform, integrating genomic data with multiple clinical and therapy information, is critical to investigate the underlying mechanisms of breast cancer. BCIP is a platform for breast cancer research and clinical research. This study suggests that BCIP profiles are important resources for the study of breast cancer and its treatment. BCIP is a platform for breast cancer research and clinical research. This study suggests that BCIP profiles are important resources for the study of breast cancer and its treatment. BCIP is a platform for breast cancer research and clinical research. This study suggests that BCIP profiles are important resources for the study of breast cancer and its treatment.

乳腺癌数据库平台网站



BCIP

Breast Cancer Integrative Platform

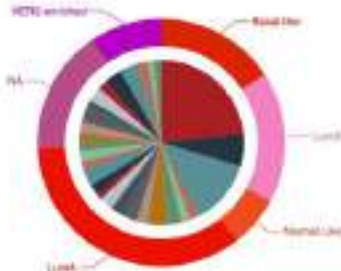
Introduction



Breast cancer is a malignant tumor that develops from breast tissue. The most common symptom of breast cancer is typically a lump in the breast. The primary types of breast cancer are ductal carcinoma and lobular carcinoma.

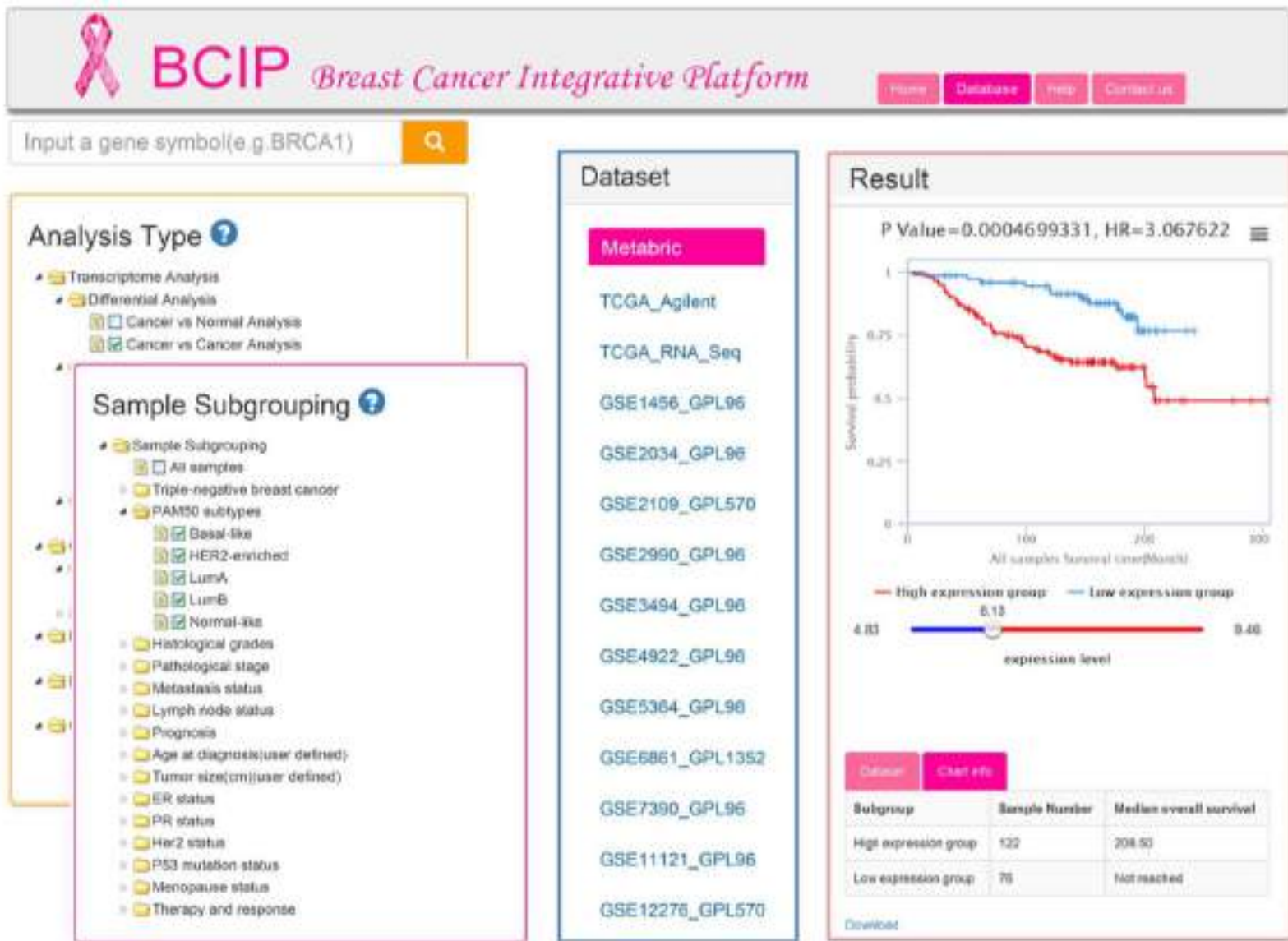
Home Database Help Contact us

Summary of The Platform



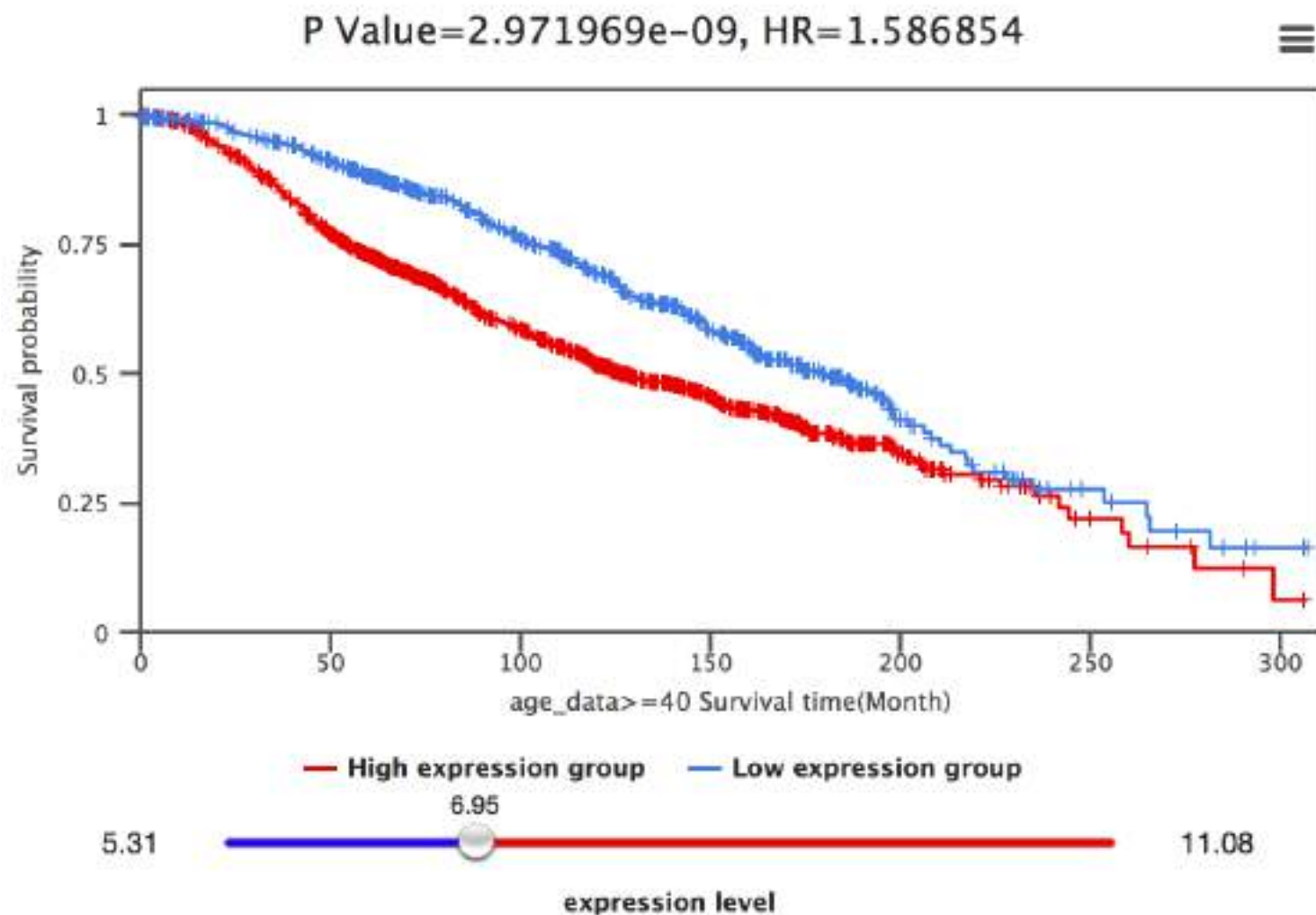
临床特征抽提

- 15个临床特征
- 三阴/非三阴型
- PAM50型
- 组织学分级
- 病理分期
- 转移状态
- 淋巴结转移
- ER + / -
- PR + / -
- Her2+ / -
- TP53突变
- 是否绝经
- 年龄
- 肿瘤大小
- 疗效
- 预后

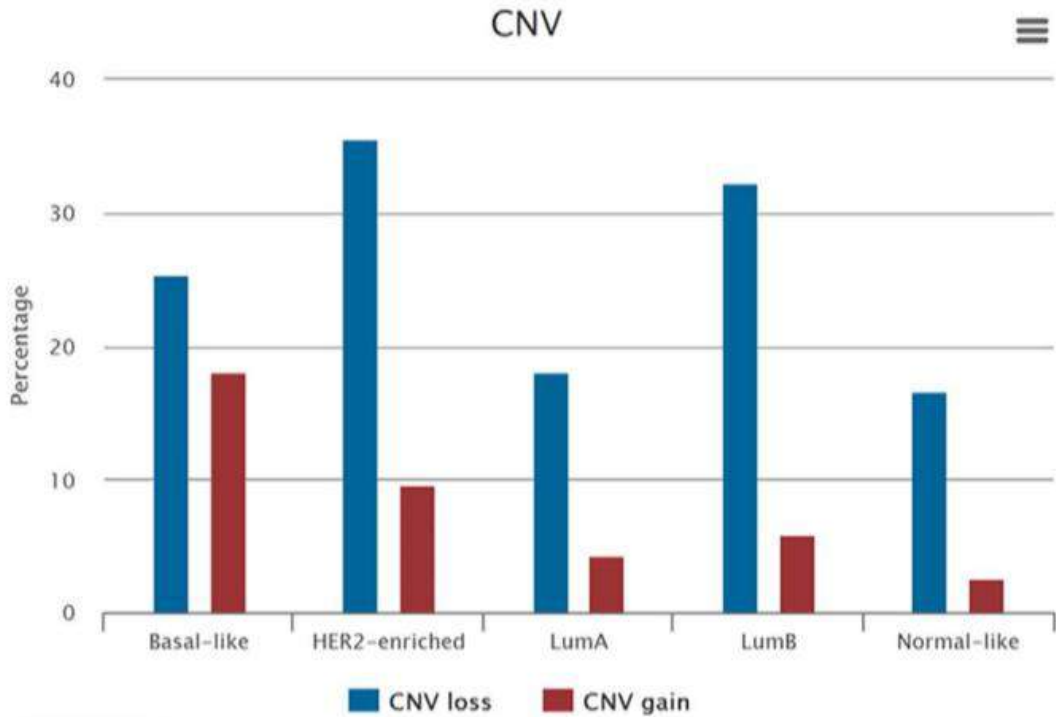


生存分析

- MELK的过量表达与较差预后相关
- 表明MELK与基底样乳腺癌相关



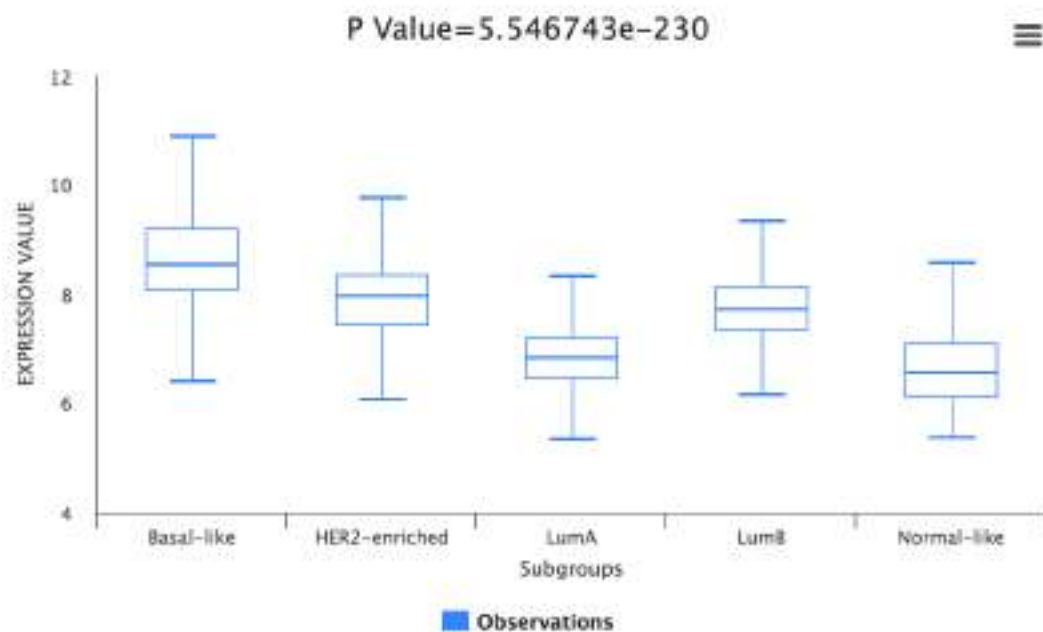
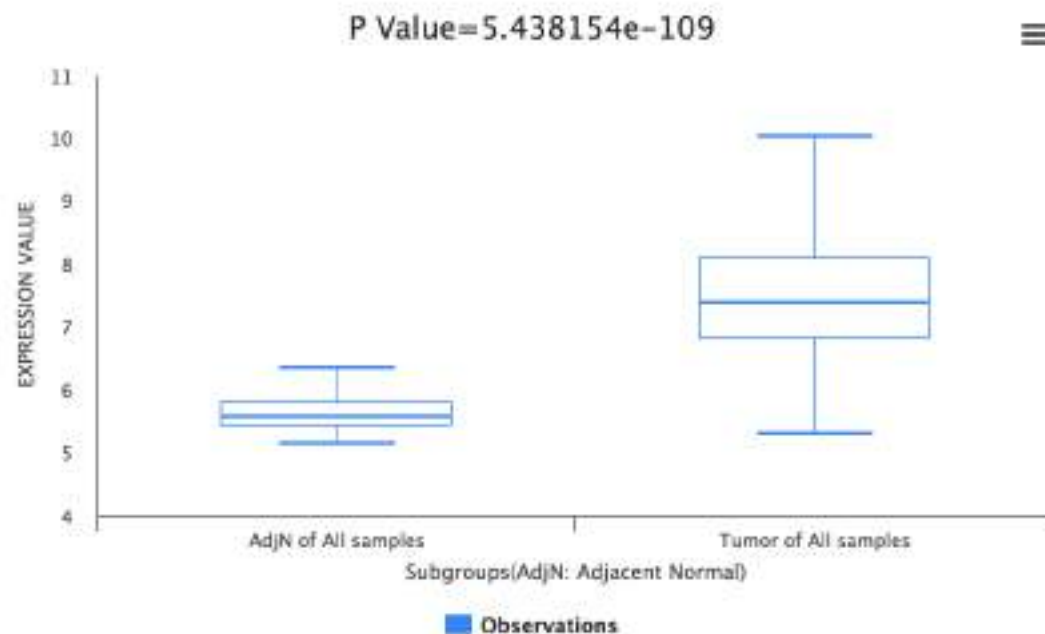
- 在 METABRIC 数据集 PAM50亚型中拷贝数减少和增加的百分比情况



Dataset	Chart info	
Subgroup	CNV loss	CNV gain
Basal-like	84(25.38%)	60(18.13%)
HER2-enriched	85(35.56%)	23(9.62%)
LumA	129(18.04%)	31(4.34%)
LumB	158(32.24%)	29(5.92%)
Normal-like	33(16.58%)	5(2.51%)

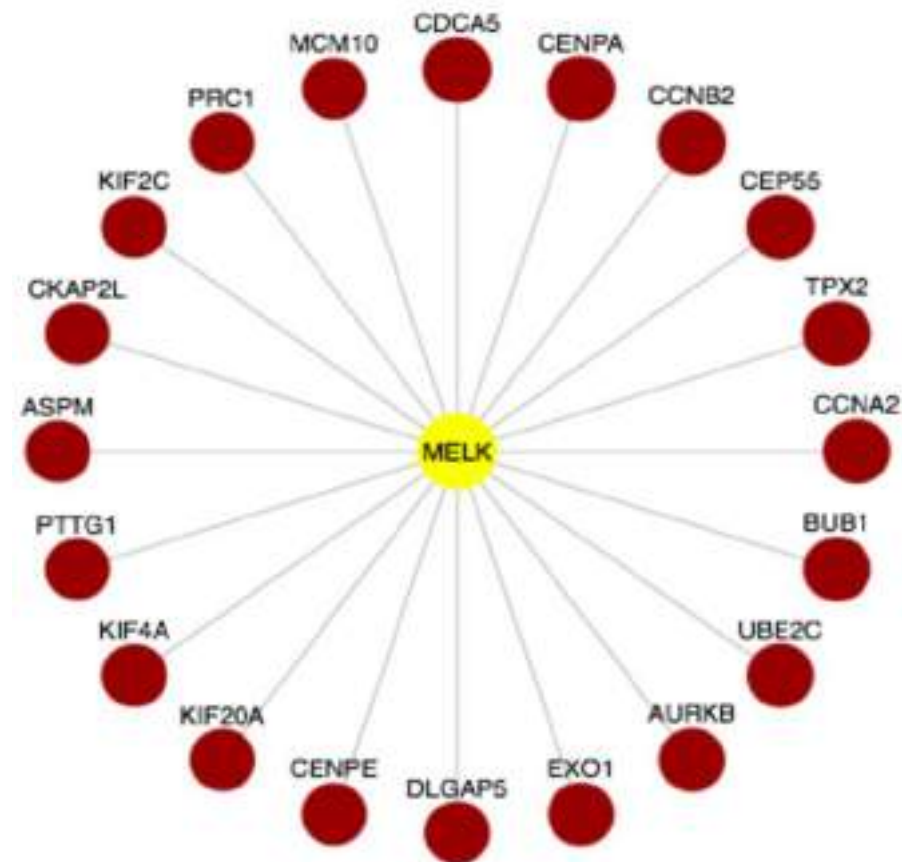
差异表达分析

- 肿瘤组织相比于周围正常组织，MELK的表达量要高出许多
- PAM50型乳腺癌中的基底样乳腺癌，MELK的表达量最高。



共表达分析

- 分析MELK影响基底样乳腺癌的机理
- 在基底样乳腺癌的METABRIC数据集中，MELK与包括CDCA5,TPX2和CEP55在内的78个基因共表达。
- 一些研究已经阐述了TPX2和CEP55是参与乳腺癌转移、侵袭、增殖和扩散的关键分子。CDCA5也被报道在肺癌中起关键作用，并可作为口腔鳞细胞癌的治疗靶点。
- 这些结果都可以作为挖掘MELK在乳腺癌中的潜在功能和机制的有用线索。
- 肿瘤组织相比于周围正常组织，MELK的表达量要高出许多



Top 20 genes correlated with the query gene

All samples

miRNA靶相互作用分析

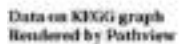
- 发现hsa-miR-193b-3p and hsa-miR-372-5p与miRNA靶相互作用有关

miRBase Accession	miRNA	Target Gene	Strong Evidence			Less Strong Evidence				Reference PMID
			Reporter Assay	Western Blot	qPCR	Microarray	NGS	pSILAC	Other	
MIMAT0002819	hsa-miR-193b-3p	MELK				✓				20304954
MIMAT0019745	hsa-miR-4668-5p	MELK					✓			22012620
MIMAT0026484	hsa-miR-372-5p	MELK					✓			22012620
MIMAT0026484	hsa-miR-372-5p	MELK					✓			23824327
MIMAT0004687	hsa-miR-371a-5p	MELK					✓			22012620
MIMAT0004687	hsa-miR-371a-5p	MELK					✓			23824327
MIMAT0030018	hsa-miR-7703	MELK					✓			23824327
MIMAT0014977	hsa-miR-3115	MELK					✓			23824327
MIMAT0016886	hsa-miR-4252	MELK					✓			23824327
MIMAT0021116	hsa-miR-5186	MELK					✓			23824327

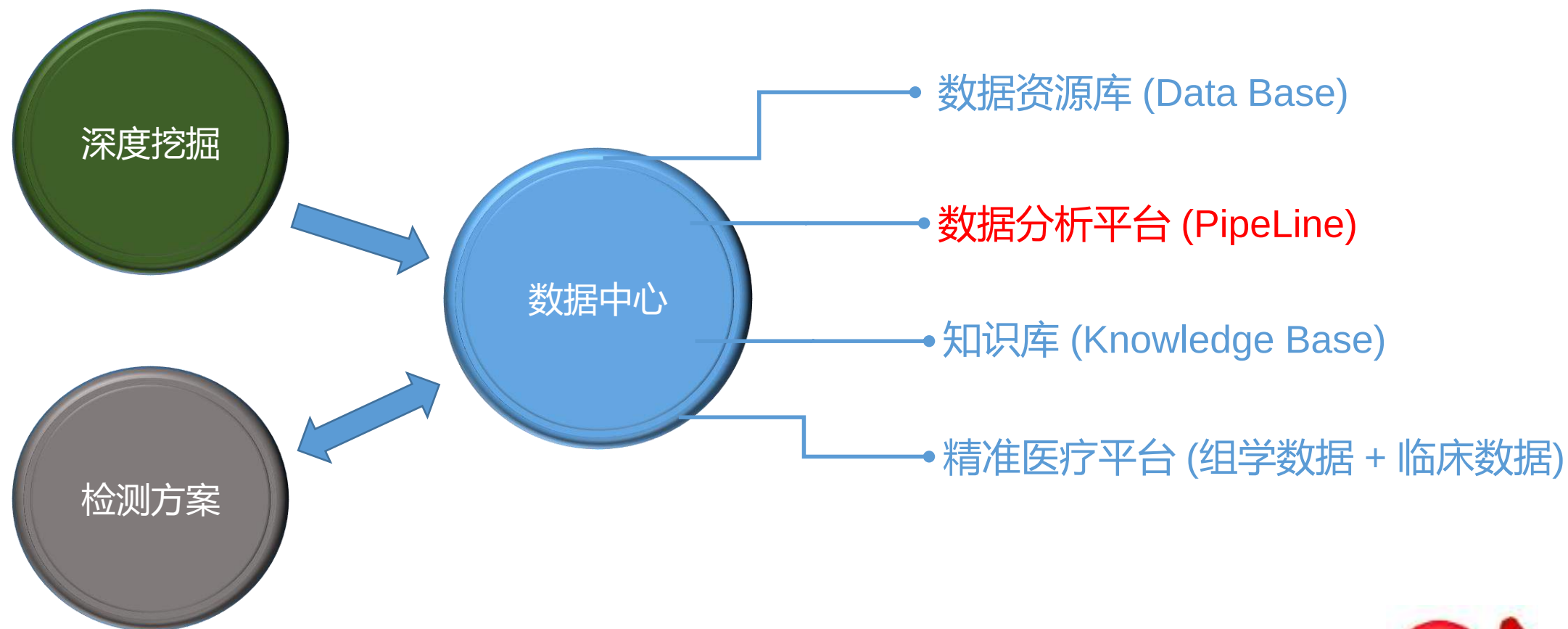
10 | Page 1 of 7 | Displaying 1 to 10 of 69 items

Data were collected from miRTarBase (Nucleic Acids Res. 2016; 44, D239–D247.).

KEGG通路分析



组学大数据平台与精准医疗

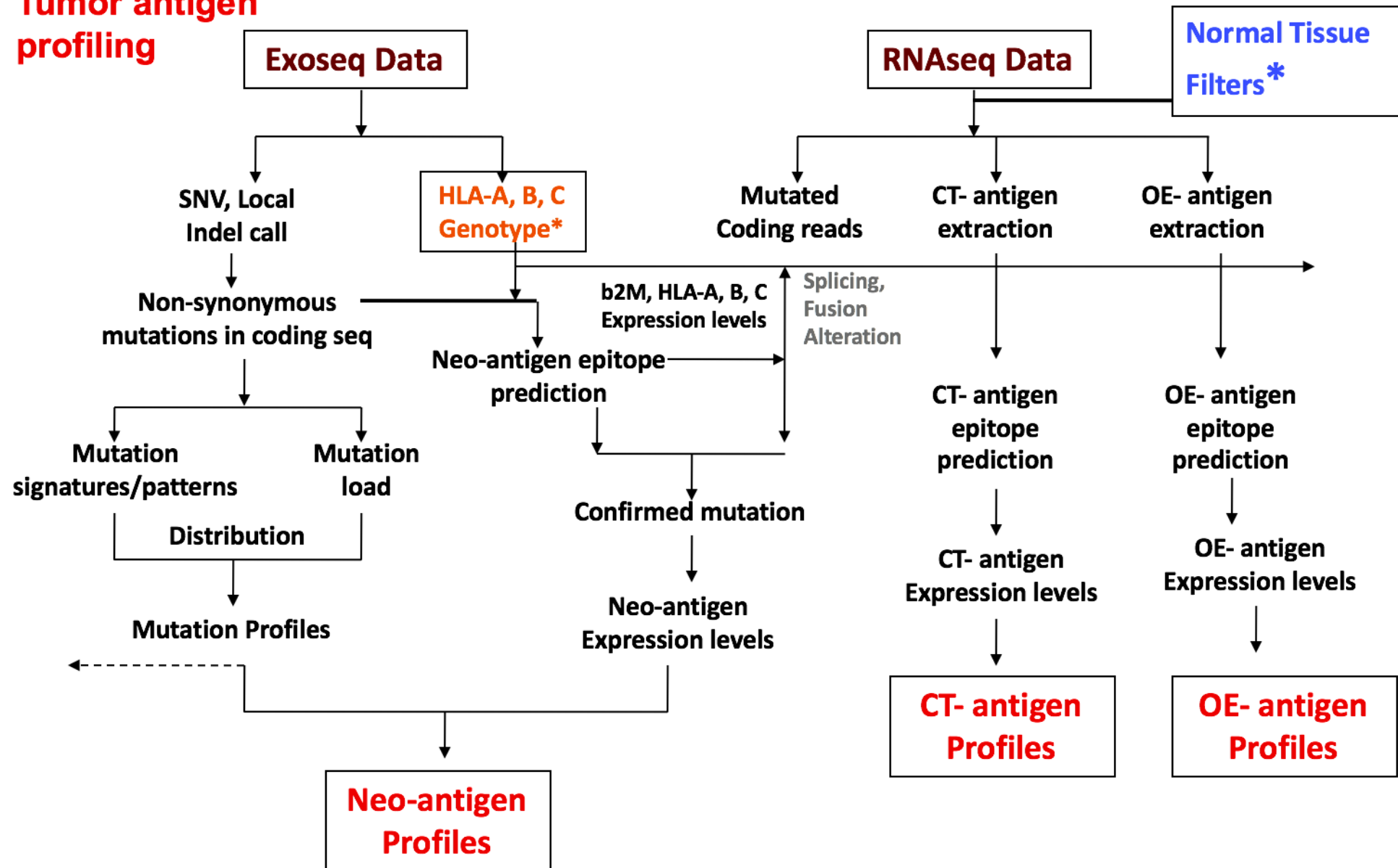


- 荧光定量PCR、基因芯片、SNP分型、二代测序

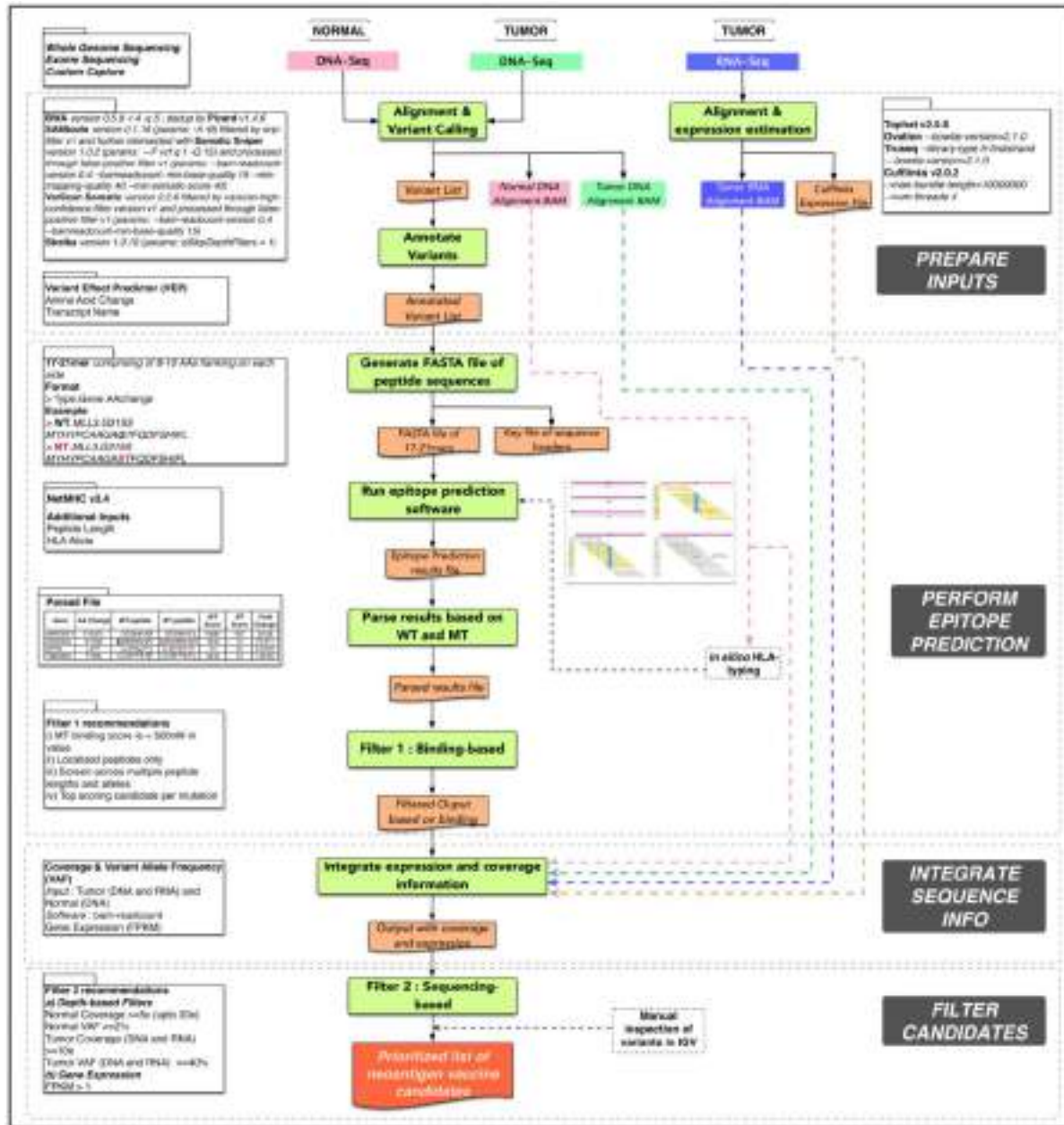


Integration of Exoseq and RNAseq data for tumor antigen profiling

Tumor antigen profiling



pipeline pVAC-Seq (personalized Variant Antigens by Cancer Sequencing)



输入数据的准备
(全基因组与全外显子组测序)
BWA;SAMtools;VarScan
somatic;Strelka;Tophat;Ovation;TruSeq;Cufflinks
Variant Effect Predictor VEP

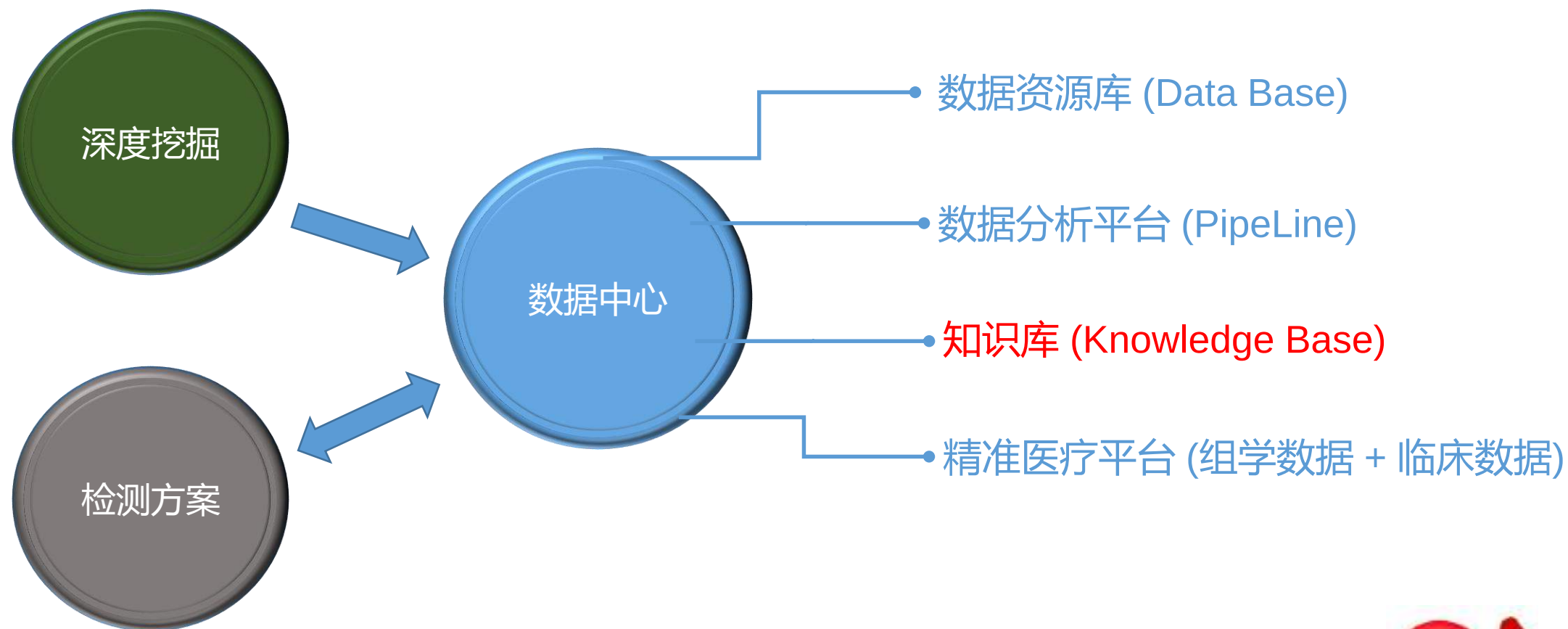
抗原表位预测
FASTA文件生成
运行抗原预测软件NetMHC
结果解析

整合测序信息
Coverage & Variant Allele Frequency (VAF)

候选抗原的过滤
深度过滤
基因表达

Hundal J, Carreno B M, Petti A A, et al. pVAC-Seq: A genome-guided in silico, approach to identifying tumor neoantigens[J]. Genome Medicine, 2016, 8(1):11.

组学大数据平台与精准医疗



- 荧光定量PCR、基因芯片、SNP分型、二代测序



人体自免疫的抗原数据库平台 AAgAtlas1.0

平台简介

第一个系统搜集描绘人体自免疫抗原的数据库。

文本挖掘与人工校验相结合的方法构建了相关的数据库旨在为基础与转化研究提供一个全面的自免疫抗原数据集。

最终确定了1126自身抗原基因，涵盖了肿瘤、心血管疾病和自身免疫病等1071种人类相关疾病，构建了第一个全面的人类自身抗原数据库（AAgAtlas1.0）。对肝癌相关自身抗原开展初步生物信息学分析发现这些抗原基因参与了细胞周期、细胞凋亡、基因表达和免疫系统等多个重要的生物学过程，表明了这些蛋白在肝癌发生发展中可能具有重要的作用。

案例成果

网址：<http://biokb.ncpsb.org/aagatlas/>
文章：AAgAtlas 1.0: a human autoantigen database
Nucl. Acids Res. first published online October 23, 2016
DOI: [10.1093/nar/gkw946](https://doi.org/10.1093/nar/gkw946) 影响因子：10.162

文章发表于2016年Nucleic Acids Research



人体自免疫抗原的数据库平台网站



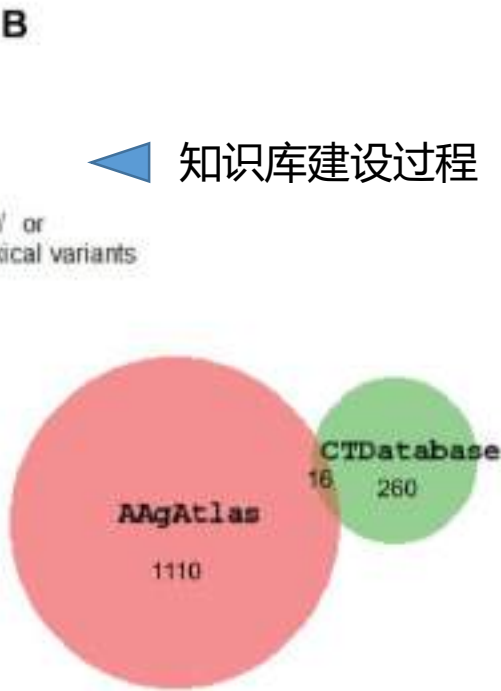
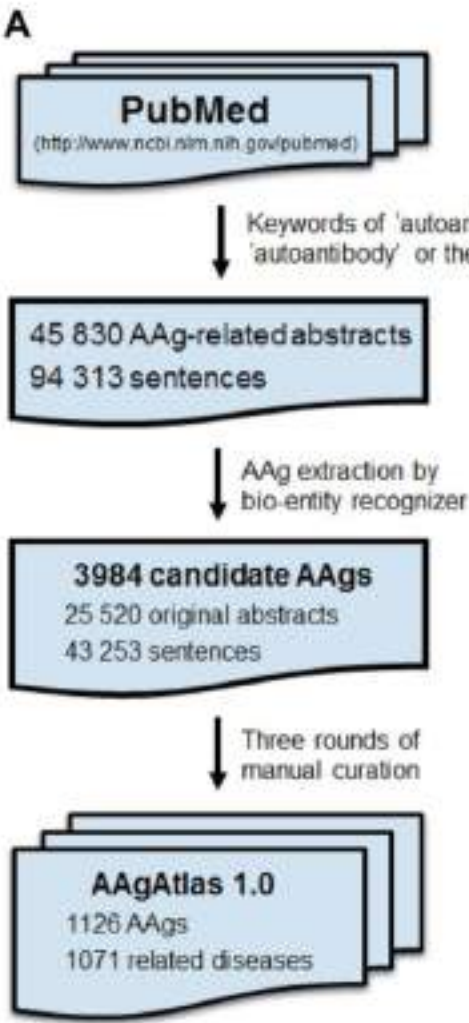
人体自免疫的抗原数据库平台 AAgAtlas1.0

人类自抗原分类

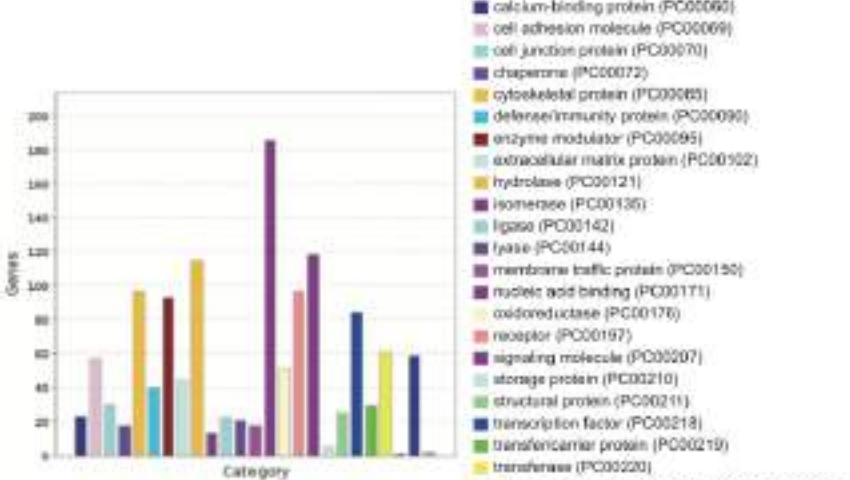
Classification of 1126 human autoantigens

文本挖掘
机器学习
人工校验
.....

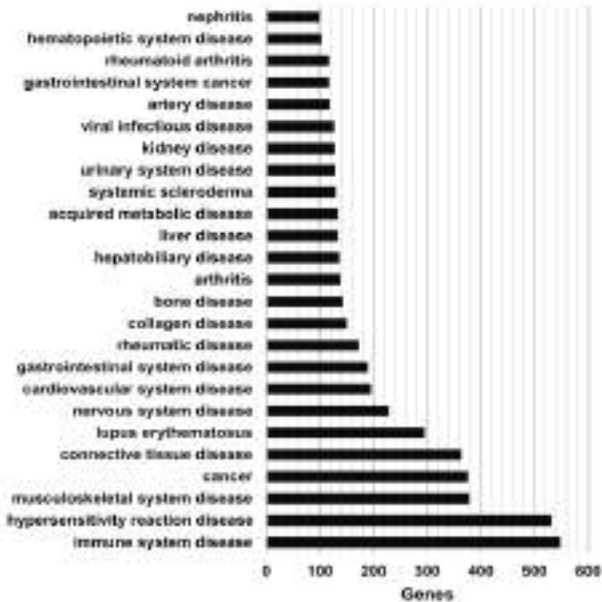
自抗原知识库
.....



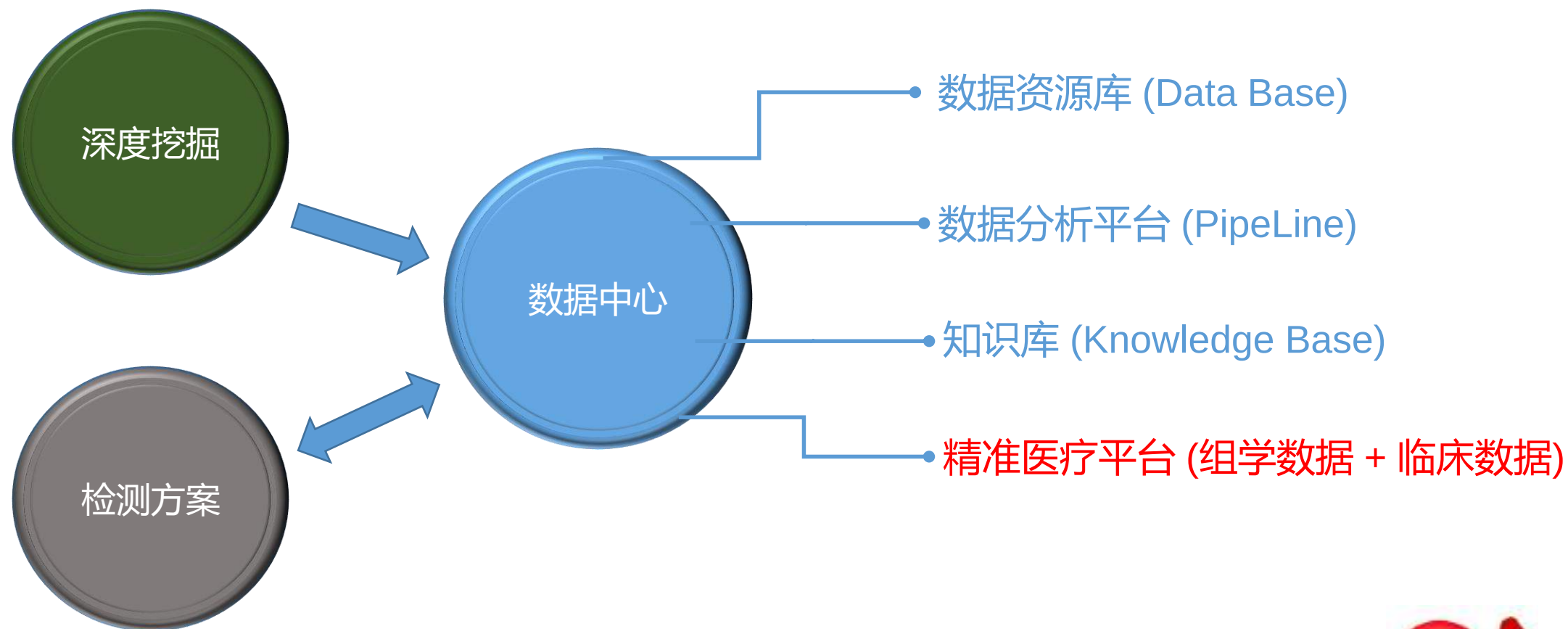
与自抗原有关的疾病



Major diseases related to human autoantigens



组学大数据平台与精准医疗



- 荧光定量PCR、基因芯片、SNP分型、二代测序





ISM

谢谢批评指正!