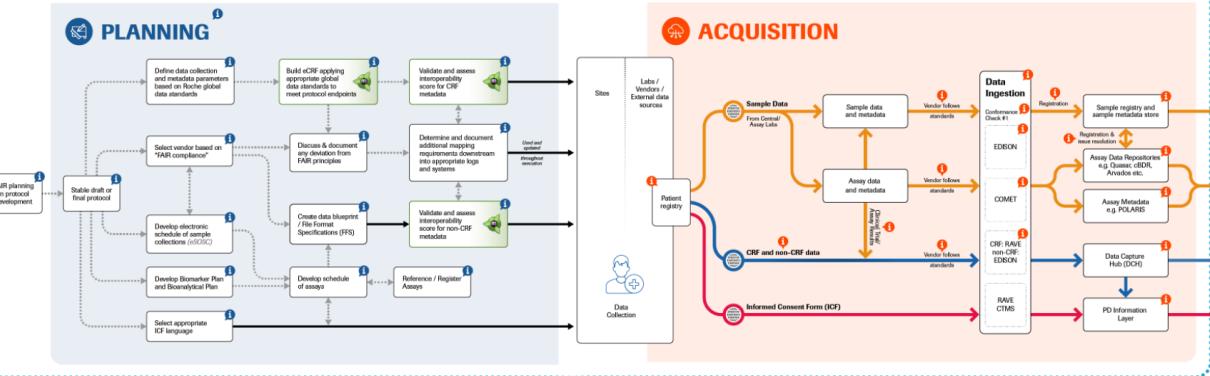


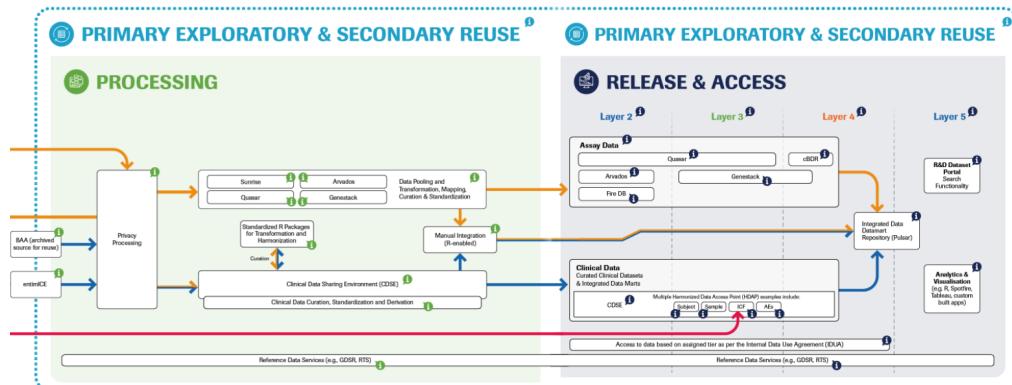
FOR ALL STUDIES



PLANNING

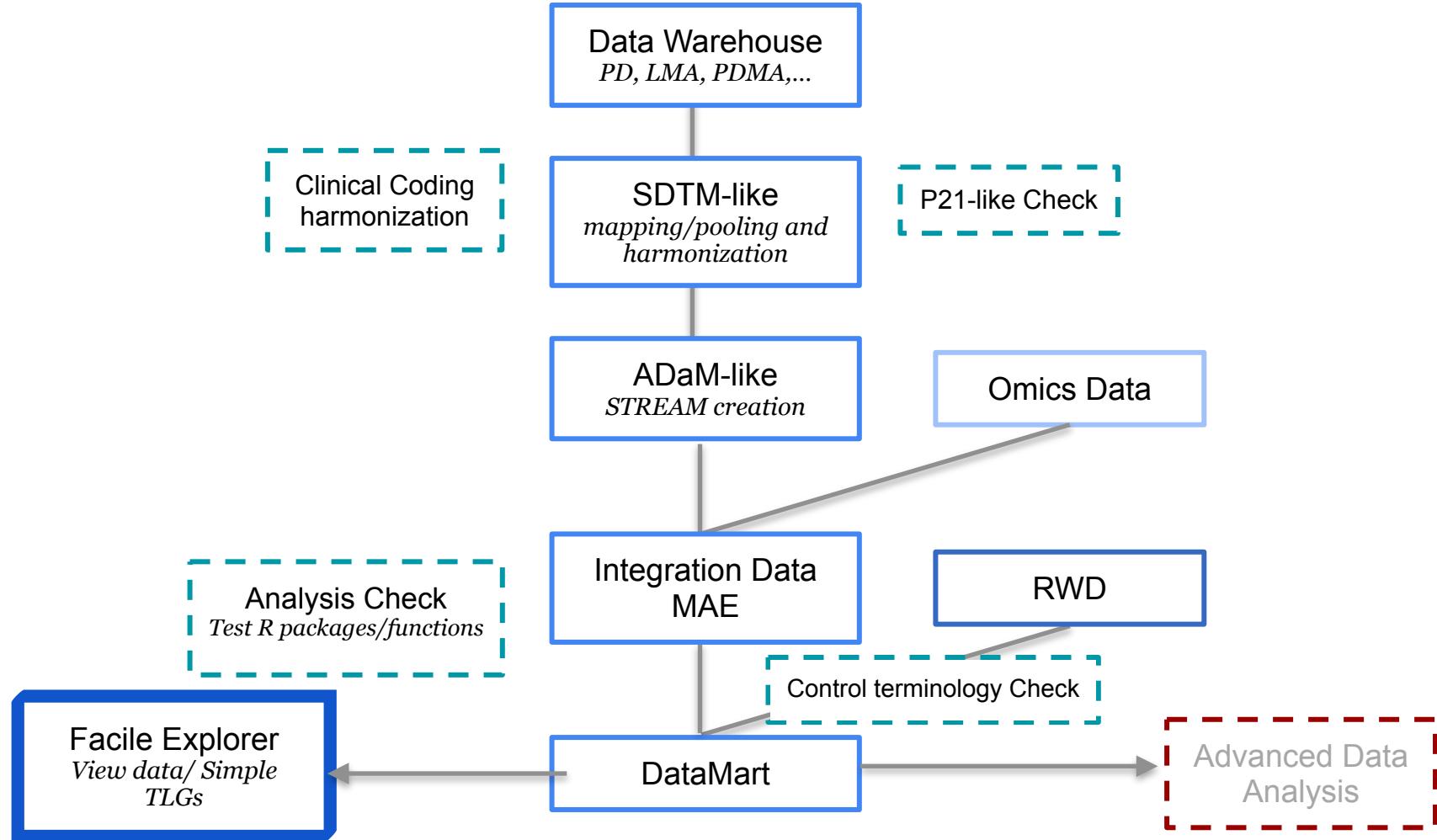
AQUISITION

FOR ALL STUDIES



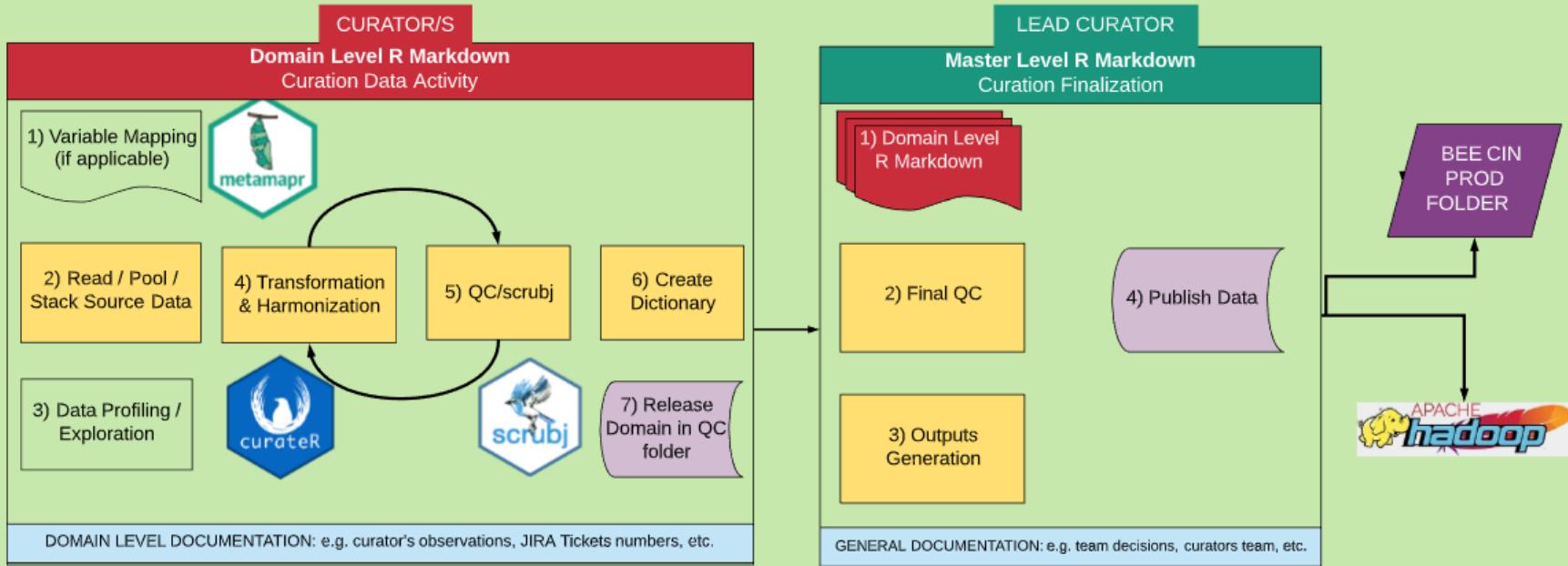
PROCESSING

RELEASE & ACCESS



Curation R Framework

Lair
SQLite



OUTPUTS/DELIVERABLES:

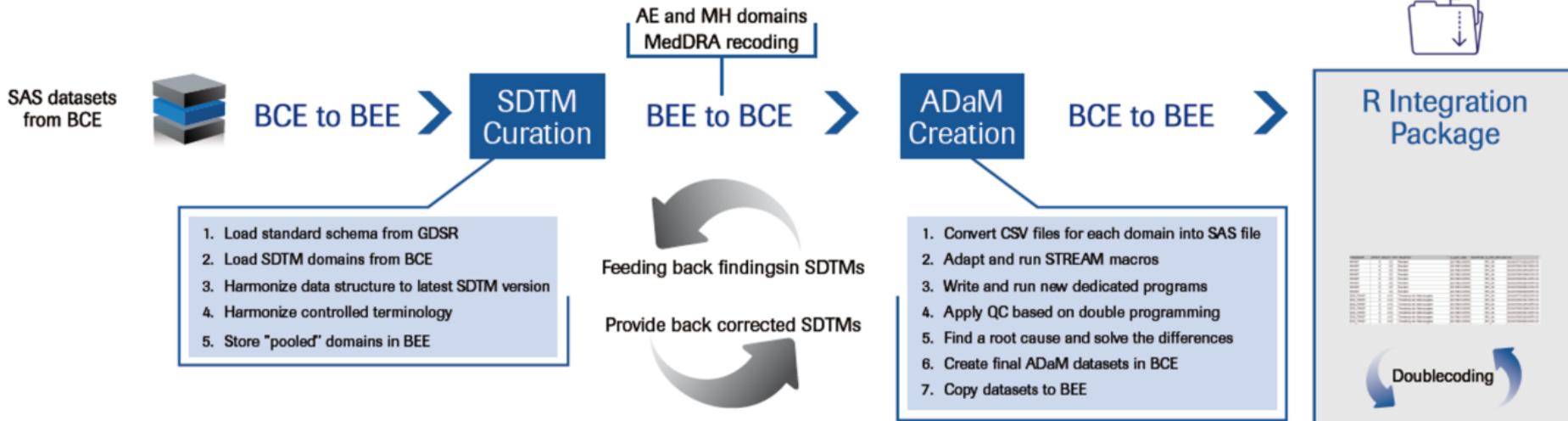
- a) curated data
- b) curated data split by studies
- c) data dictionaries
- d) params catalog (in progress)
- e) curation report/curation
- f) scrubj repot
- g) metadata (_Manifest.csv)
- ...

Curation & Integration Pipeline

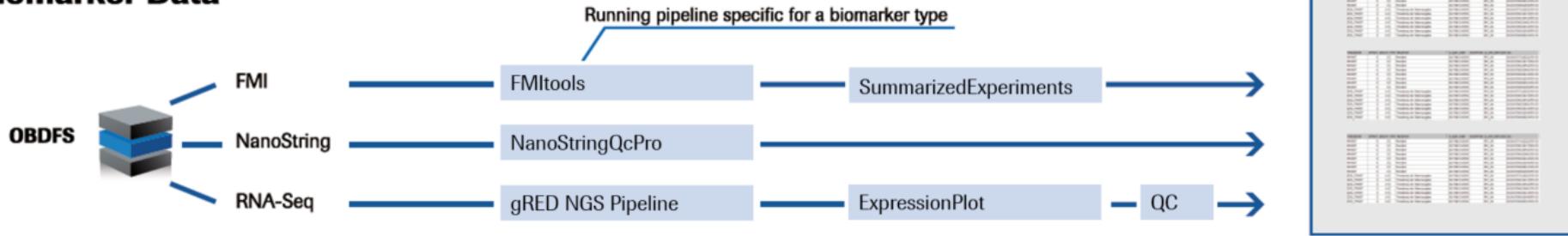
Sample Mapping Files (SMF) are produced to enable linkage of clinical and biomarker data



Clinical Data

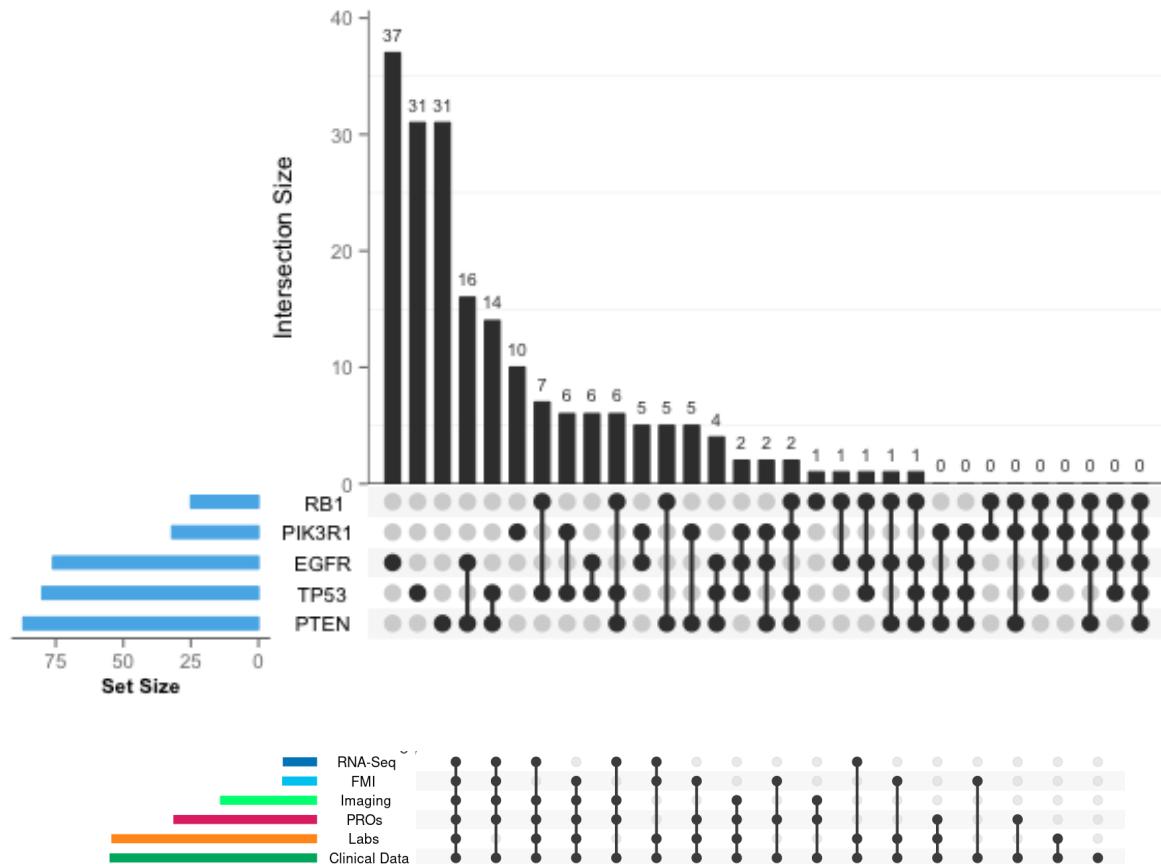
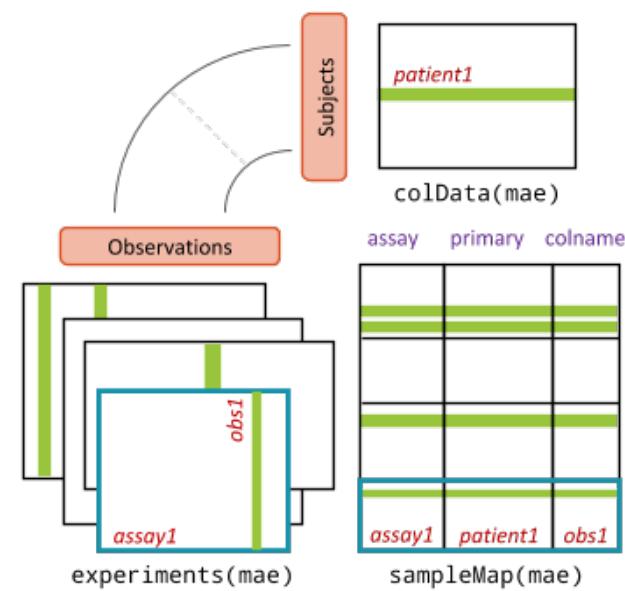


Biomarker Data



rtables Package	tern Package	teal Modules	TLGC-Catalog	Sample Apps
 <p>The rtables package is the foundation layer of NEST's tabling capabilities.</p> <p>rtables functions are used to create layouts, produce elementary statistics as well as handle many generic table control features like content pruning and sorting. You can use rtables functions to create additional simple to complex layouts to address specific study needs.</p>	 <p>The tern package is the reporting layer that sits on top of our rtables foundation.</p> <p>tern layers analytics from descriptive summaries to more complex statistics on top of the foundational rtables layouts, analytic and content controls. It is the layer that produces standard tables available in the TLG Catalog. Use tern functions to create your static TLG outputs.</p>	 <p>The teal package is developed at Roche that leverages the R Shiny package to scale development of our shiny apps.</p> <p>teal is used to develop the many teal modules that provide plug & play functionality ready for study teams to customize and deploy teal apps. It's important to note that the visualizations displayed in our teal apps are generated by tern functions. In short, teal modules leverage tern, tern in turn leverages rtables. The results are the many teal apps deployed across our portfolio enabling insights generation faster.</p>	 <p>The TLG-Catalog (TLGC) is a library of ready to use R code and teal module examples that produce our standard tables, listings and figures as seen in the STREAM catalog.</p> <p>We provide the TLGC to get you started with standard outputs for your study. To read more and access the code please visit the "TLG Catalog" chapter in agile-R here.</p>	 <p>Test Drive</p> <p>We provide these sample apps to get you started with setting up a teal application for your study. To read more and access the code behind our apps please visit the "Sample Teal Apps" chapter in agile-R here.</p>

Integration Of Multi-Omics Experiments



<https://github.com/hms-dbmi/UpSetR/>

<https://github.com/waldronlab/MultiAssayExperiment>

Roche EDIS CIT Dashboard

Introduction
Overview by Study
Data Considerations

Analyses

Selected Data
Patient Characteristics
RNA-Seq (high quality)
FMI (high quality)

Imaging
PROs
Labs (low quality)
Report

EDIS CIT Dashboard v1.1.3
Last updated 2020-12-04
Datamart v2.6.4
By dashdis v0.0.2

STUDY_ACRONYM - Study Acronym

Combine categories

Time to Event (Endpoint)

VAD_OS - Overall Survival Event (1=event, 0=censored)

Facet Plots by:

Reference Group

CAVALLI

Reference groups automatically combined into a single group if more than one value selected.

Comparison Group

Combine all comparison groups?

Plot Settings

Title

Kaplan-Meier Plot of

Table font size

Small

Large

KM Plot Time to Event Table Forest Plot

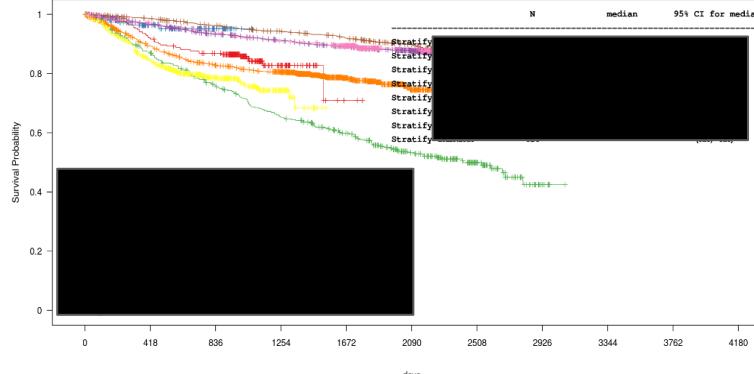
Data Dictionary

Population Selection

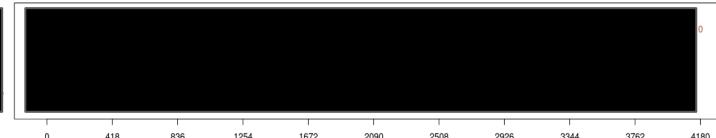


+ Add to report

Kaplan-Meier Plot of



Number of Patients at Risk



Population Selection



Quick

Custom

Quick Filtering

Reset

Select studies to include

8 items selected

Indication

DLBCL

FL

MZL

Line of therapy

1L

RR

Main Treatment

7 items selected

Age Group

<65

=>65

Sex

F

M

Baseline ECOG

0-1

2-3

Other

Missing

MYC by FISH

NEGATIVE

POSITIVE

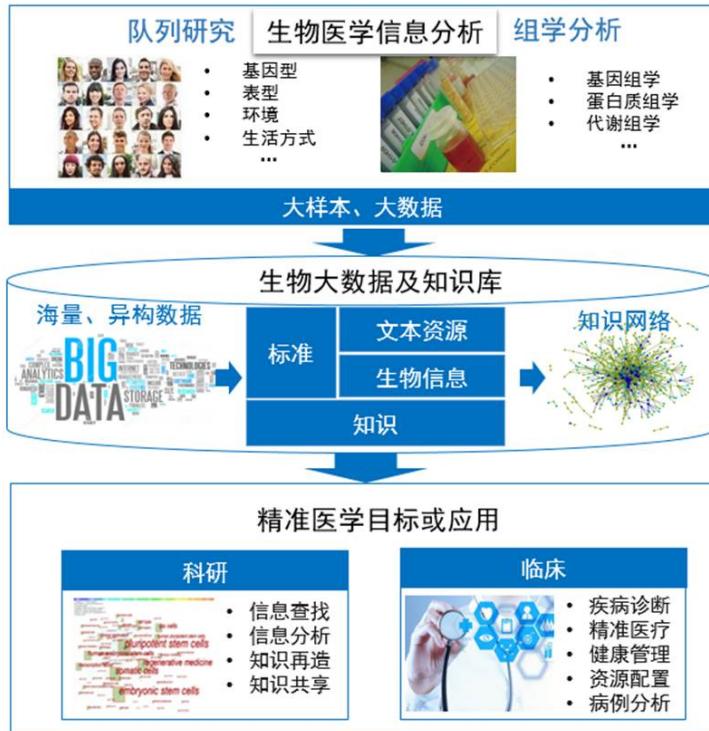
UNDETERMINED

Missing

BCL2 by FISH

NEGATIVE

POSITIVE



刘雷，博士生导师，复旦大学生物医学研究院PI，上海生物信息技术研究中心副主任。1997年毕业于美国康涅狄格大学，获细胞生物学博士学位。毕业后在美国康涅狄格大学计算机系继续博士后工作。1999年受聘于美国伊利诺伊大学组建生物信息学实验室，并担任第一任主任。2007年回国，入选中国科学院“百人计划”和上海市“浦江人才计划”。2013年受聘于复旦大学。“十一五”期间主持和参与了多个与医疗信息化相关的国家“863”项目，“十二五”承担了国家“863”“数字化医疗”项目。着重生物医学信息学研究，在组学数据的分析与挖掘，生物网络的构建与分析，生物系统的建模与模拟，医疗数据共享、整合与挖掘，个性化医疗等方面做了大量工作，取得了一系列创新性成果，研发了一系列医疗卫生软件系统，发表SCI论文60余篇，取得软件著作权20余项，申请专利6项。【详情请见个人主页】



OHDSI——观察性健康医疗数据科学与信息学

- 随着电子病历系统的普及，大量医疗相关的重要信息以电子形式存储于各种各样医疗信息系统中。经过不断积累，各种形式的电子医疗系统产生了体量庞大的临床大数据。这些数据为临床辅助诊断、药物研发、公共卫生监测评估、传染病疫情预警、个性化精准医疗等创造了巨大价值。
- 观察性健康医疗数据科学与信息学 (Observational Health Data Sciences and Informatics, OHDSI, 缩写读音同“Odyssey”或“奥德赛”，即['ɑdəsɪ]) 计划，是一个由美国哥伦比亚大学牵头，世界性的公益型非盈利研究联盟，主要研究全方位医学大数据分析的开源解决方案，旨在通过大规模数据分析和挖掘来提升临床医学数据价值，实现跨学科、跨行业的多方合作。目前，已有来自美国、加拿大、澳大利亚、英国等几十个国家地区的上百个组织机构，高校，医院和公司企业参与了OHDSI全球协作网络，如斯坦福、哈佛、杜克大学医学院，艾昆纬 (IQVIA, 原昆泰艾美仕) ，强生、诺华、甲骨文、IBM公司，拥有超过6亿人口的临床数据规模，累计协作研究发表了上百篇论文。

OHDSI 中国

- 不同的国家有不同的语言以及不同的卫生体系，这也就导致了他们医疗数据的标准不同。虽然OHDSI已经对于美国，欧洲的观察性电子病历数据开发了一些方法，标准和软件，但是具体把这些工具运用到中国的医学研究上还存在一些问题和挑战，需要从中国实际情况出发，结合中国数据标准和现状，来拓展现有OHDSI的工具和方法。同时观察性数据研究需要横跨多个知识领域，需要各种专业领域人才一起合作，现阶段中国缺少这样的人才，需要有一个组织来进行系统性，专业性的培训，让临床数据更好的来实现他的真正价值。这也是OHDSI中国成立的背景。
- OHDSI 中国于2016年12月成立，总体目标就是，利用数据科学和信息学方法，促进中国的健康医疗数据研究工作。具体来说，我们致力于推动中国健康医疗数据研究方法和应用，包括针对中国临床数据扩展和定制OHDSI的数据科学方法（如各种标准、知识库和软件工具），利用大型观察性健康医疗数据网络对全球性重要健康医疗问题展开调查研究，以及促进生物医学信息学领域的国际合作与教育培训工作。

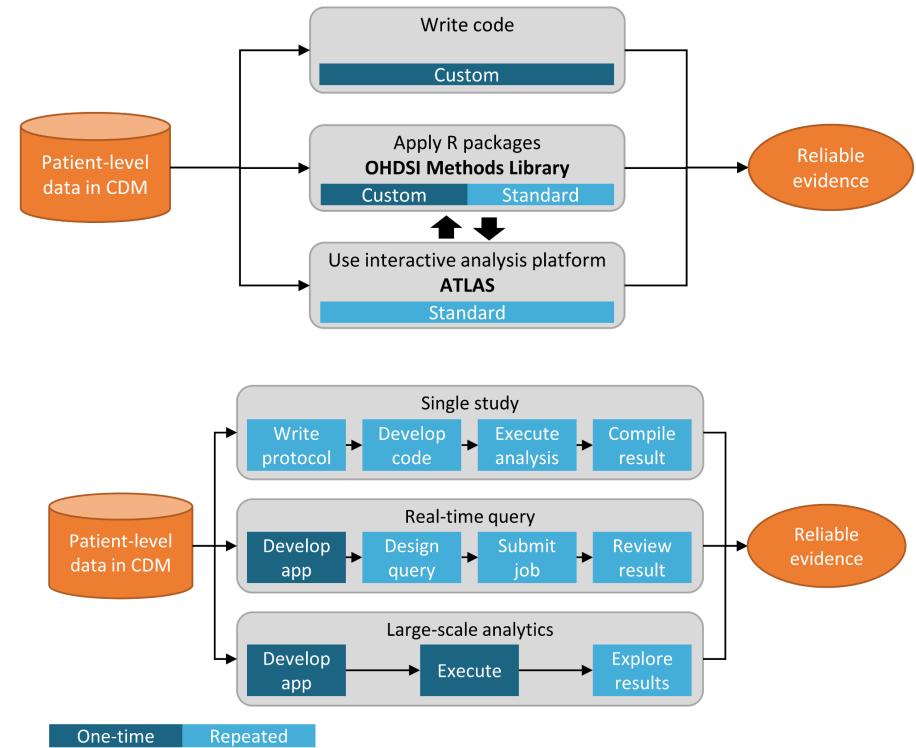
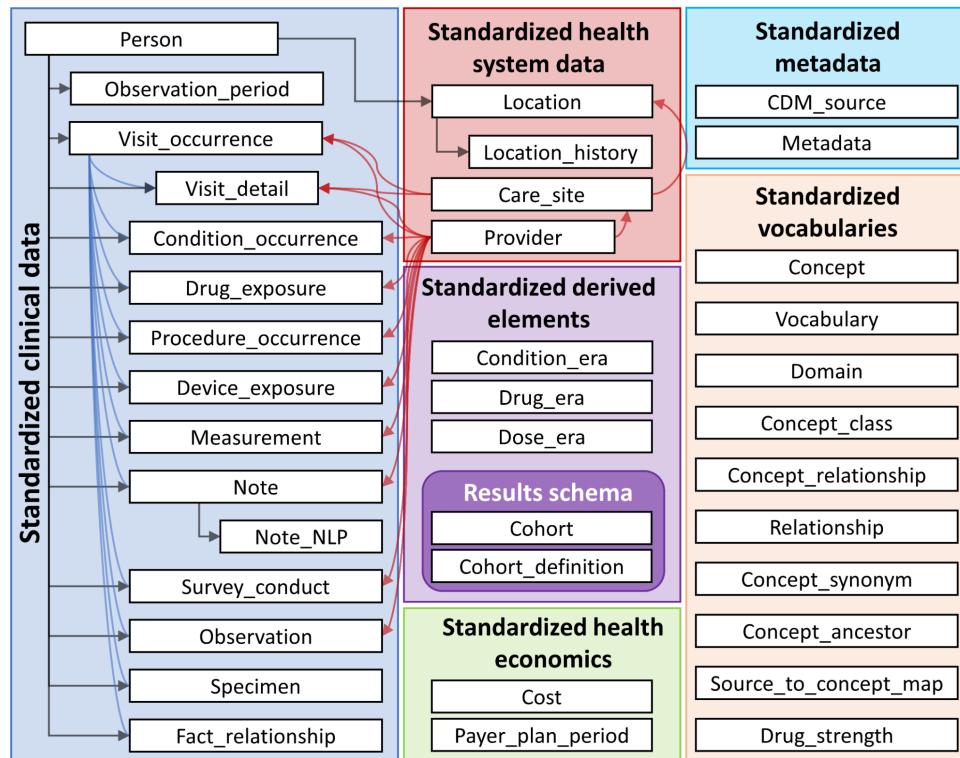
OHDSI中国核心团队

OHDSI中国主席Chair

- 德州大学休斯顿生物医学信息学院 徐华教授

标准化数据模型&术语标准CDM and vocabulary工作组-侧重于中文术语库和标准数据模型建设-负责人：

- 梅奥诊所 蒋国谦教授
- 哥伦比亚大学 翁春华教授
- 德州大学休斯顿生物医学信息学院 陶萃教授
- 复旦大学 刘雷教授
- 泰康保险 张林老师
- 南通大学 王理教授





Cohort #1770710

New users of ACE inhibitors as first-line monotherapy for hypertension

Definition Concept Sets Generation Reporting Export Messages (3)

enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

+ Add Initial Event

a drug exposure of **ACE inhibitors**

for the first time in the person's history

+ Add attribute...

with continuous observation of at least **365** days before and **0** days after event index date

Limit initial events to: **earliest event** per person.

Restrict initial events

Inclusion Criteria

New inclusion criteria

1. has hypertension diagnosis in 1 yr prior to treatment
2. Has no prior antihypertensive drug exposures in medical

Cohort Method New-user cohort studies using large-scale regression for propensity and outcome models	Self-Controlled Case Series Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.	Self-Controlled Cohort A self-controlled cohort design, where time preceding exposure is used as control.
Patient Level Prediction Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.	Case-control Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort.	Case-crossover Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control).
Empirical Calibration Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.	Method Evaluation Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.	Evidence Synthesis Combining study diagnostics and results across multiple sites.
Database Connector Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.	Sql Render Generate SQL on the fly for the various SQL dialects.	Cyclops Highly efficient implementation of regularized logistic, Poisson and Cox regression.
ParallelLogger Support for parallel computation with logging to console, disk, or e-mail.	Feature Extraction Automatically extract large sets of features for user-specified cohorts using data in the CDM.	

<https://ohdsi.github.io/TheBookOfOhdsi/OhdsiCommunity.html#the-journey-from-data-to-evidence>