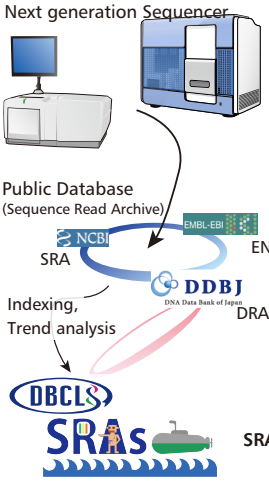
Database Center for Life Science (DBCLS)
Research Organization of Information and Systems (ROIS)

Functional interface for quick access to disease-relevant NGS data

次世代シーケンサによる疾患に関連するデータを使い倒すために

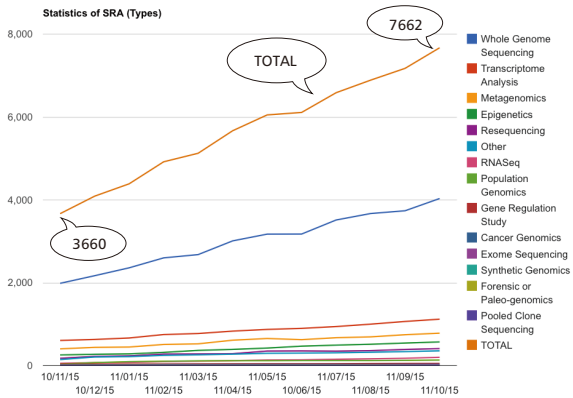
○Takeru Nakazato*, Tazro Ohta, Akinori Yonezawa, Hidemasa Bono
仲里 猛留　大田 達郎　米澤 明憲　坊農 秀雅
*: nakazato@dbcls.rois.ac.jp
Database Center for Life Science (DBCLS), Res. Org. of Info. and Systems (ROIS)
情報・システム研究機構 ライフサイエンス統合データベースセンター

Backgrounds and motivations



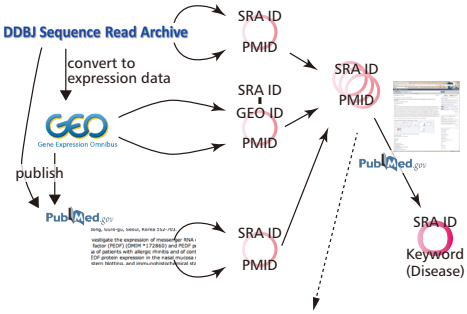
SRAs: the survey of read archives
http://sra.dbcls.jp/

ここ最近、次世代シーケンサ (NGS) による成果が次々に発表されている。マイクロアレイのデータが GEO に登録されるのと同様に、NGS データも公共データベースである Sequence Read Archive (SRA) に登録され、日米欧の 3 局でデータ交換がなされている。その数は、プロジェクト数で 8300 (2011 年 12 月現在) に及んでいる。DBCLS では、登録データに対して、目次作成、データの傾向分析を行い、NGS データの検索サイトを構築、提供している。今回、NGS データに対する 1 つの切り口として疾患に注目し、データの整理と、その傾向を分析した。また、NGS データを疾患から検索するインターフェースを構築した。

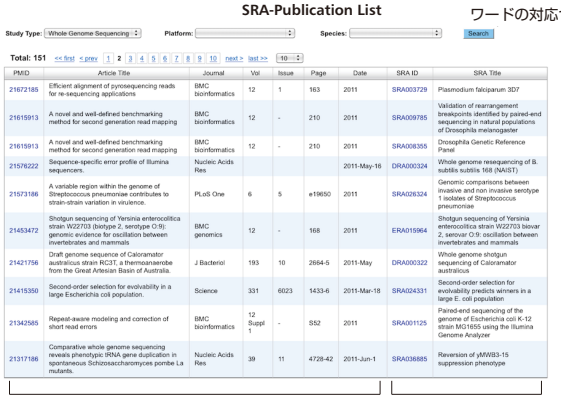


※ シーケンサ別、生物種別の傾向については、上記 SRAs サイトをご覧いただくか、発表者  にお尋ねください。

Methods



SRA の各エントリから参考文献の PMID を抽出し、SRA ID-PMID ペアを作成した。一部の NGS データは発現値に変換され、GEO データベースに収められている。そこで GEO からも文献情報を収集した。また、逆に SRA や GEO の ID が記載された文献からも SRA ID-PMID の収集を行った。その後、文献に付与された MeSH キーワードを抽出し、SRA と疾患キーワードの対応づけを行った。



Results and Discussions

※ デモしますので発表者  までお気軽に。
http://sra.dbcls.jp/

Statistics (as of Dec 11, 2011)

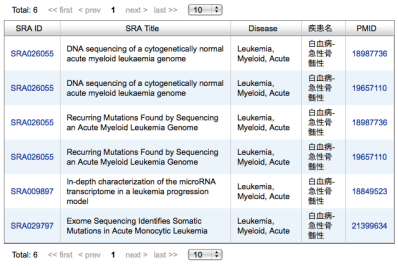
Big Projects relevant to diseases (top 10)

The Cancer Genome Atlas Project at NCI/NHGRI	30167
ARRA Autism Sequencing Collaboration	5222
Towaeds an Genomic Understanding of Myeloma	894
Genomic Sequencing of Head and Neck Cancer	883
Whole Exome Sequencing in Early-Onset Myocardial Infarction	621
High-throughput semi-quantitive analysis of insertional mutations in heterogeneous tumors	379
NCI Cancer Genome Characterization Initiative (CGCI)	266
MicroRNA sequence and expression analysis in breast tumors by deep sequencing	245
Frequent mutations of ubiquitin mediated proteolysis pathway in clear cell renal cell carcinoma	219
Frequent mutations of genes in transitional cell carcinoma of the bladder	219
...	(experiments)

Data Visualization

Frequency List (top 10)

The sample of SRA entries relevant to disease



Tree view of diseases with SRA entries

Conclusions

- 公共データベース SRA に登録された次世代シーケンサによるデータを疾患の切り口から整理した。
- 整理したデータを頻度別、性質別に見られるようにし Survey of Read Archives (SRAs) より閲覧可能にした。 http://sra.dbcls.jp/
- GEO に収載された「発現データ」も含め、実データにさらにアクセスしやすくする整理が課題

第 34 回 日本分子生物学会年会
パシフィコ横浜
平成 23 年 12 月 13 日～16 日