

疾患の切り口から公共データベース中の NGS データを利活用する

○ 仲里 猛留*、 坊農 秀雅

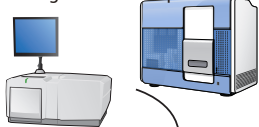
*: nakazato@dbcls.rois.ac.jp

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 (DBCLS)

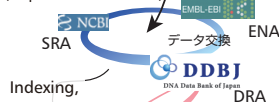


Backgrounds and motivations

Next generation Sequencer



Public Database
(Sequence Read Archive)

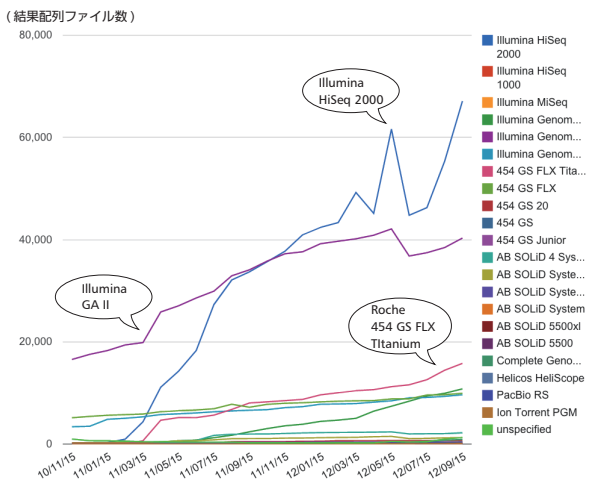
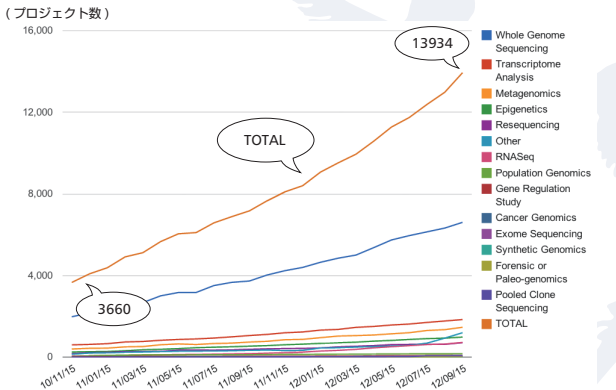


Indexing, Trend analysis



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 14000 (2012 年 10 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。

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



Species	Count	Technology	Count
<i>Homo sapiens</i>	1261	Illumina HiSeq 2000	68640
unidentified	882	Illumina Genome Analyzer II	40677
<i>Mus musculus</i>	818	454 GS FLX Titanium	15836
<i>Drosophila melanogaster</i>	284	Illumina Genome Analyzer Ix	11142
<i>Mustela putorius furo</i>	233	454 GS FLX	9961
<i>Caenorhabditis elegans</i>	198	Illumina Genome Analyzer	9847
metagenome sequence	184	AB SOLiD 4 System	2226
marine metagenome	165	unspecified	1284
<i>Arabidopsis thaliana</i>	160	AB SOLiD System 3.0	1255
unclassified phages	139	Complete Genomics	960
<i>Escherichia coli str. K-12 substr. MG1655</i>	136	454 GS 20	704
soil metagenome	127	Illumina MiSeq	449
<i>Saccharomyces cerevisiae</i>	124	PacBio RS	417
<i>Human rhinovirus A</i>	95	Helicos Heliscope	385
<i>Salmo salar</i>	84	AB SOLiD System 2.0	383
		Illumina HiSeq 1000	251
		AB SOLiD System	188
		454 GS	141
		Ion Torrent PGM	94
		454 GS Junior	88
		AB 5500 Genetic Analyzer	37
		AB 5500Xi Genetic Analyzer	20
		AB SOLiD 5500xl	12
		AB SOLiD 5500	6
		AB SOLiD 4q System	4
Total	16806 (studies)	Total	165027 (experiments)

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 Oct 3, 2012)

Big Projects relevant to diseases (top 10)

ARRA Autism Sequencing Collaboration	2257
Jackson Heart Study	2030
Allelic Spectrum Sequencing Discovery	1865
Lung Adenocarcinoma Tumor Exome Sequencing Project	1640
Prostate Cancer Genome Sequencing Project	1638
Framingham Heart Study Allelic Spectrum Sequencing Discovery	1170
Melanoma Genome Sequencing Project	1067
Swedish Schizophrenia	1067
Population-Based Case-control Exome Sequencing	981
Ciliopathies Exome Sequencing Initiative	616
Genetic Analysis of Hirschsprung Disease	616
Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations	616

Data Visualization

Frequency List (top 10)

Disease Type: ANY [Search]

Disease	疾患名	# of submission
Breast Neoplasms	乳房腫瘍	43
Prostatic Neoplasms	前立腺腫瘍	22
Disease Models, Animal	疾患モデル(動物)	20
Genetic Predisposition to Disease	遺伝的素因(疾患)	15
Disease Progression	疾病進行	14
Translocation, Genetic	転座	12
Cell Transformation, Neoplastic	腫瘍細胞形成転換	11
Lung Neoplasms	肺腫瘍	10
Staphylococcal Infections	ブドウ球菌感染症	9
Chromosome Aberrations	染色体異常	9

The sample of SRA entries relevant to disease

SRA ID	SRA Title	Disease	疾患名	PMID
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18867736
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	20876853
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18887736
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	20876853
SRA009897	In-depth characterization of the microRNA transcriptome in a leukemia progression model	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18849523
SRA050226	Validation of FLT3-TDZ as a therapeutic target in human acute myeloid leukemia	Leukemia, Myeloid, Acute	白血病-急性骨髄性	22504184
SRA029797	Exome Sequencing Identifies Somatic Mutations in Acute Myeloid Leukemia	Leukemia, Myeloid, Acute	白血病-急性骨髄性	21398634
SRA048890	GSK34725: Inhibition of the LSD1/KDMA demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia [ChIP-Seq]	Leukemia, Myeloid, Acute	白血病-急性骨髄性	22406747

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

トーゴの日シンポジウム 2012
時事通信ホール (東京)
平成 24 年 10 月 5 日

