

課題番号 : 25指112
研究課題名 : 分子生物学的情報を扱う臨床研究の品質管理に関する研究
主任研究者名 : 田中紀子
分担研究者名 : 志村まり 田中康博

キーワード : omics; data management; database; clinical bioinformatics; standardization
研究成果 :

平成 26 年度の研究成果報告
(田中(紀)班)

2 年度からは、各グループ間の作業を連携的に行い、初年度で設計されたシステムに基づいて、基本的な実装を行う予定であった。以下に進捗状況と成果を報告する。

1. インフラ整備

2013 年下期に購入した解析用サーバーを外部研究者も含め研究者間で共有するため、ネットワークの設定等のインフラ整備を行った。ネットワークはパスワード認証よりも安全性の高い秘密鍵/公開鍵方式を用いた SSH プロトコルでの通信とし、解析用サーバーには SSH サーバーを経由して接続することとした。SSH サーバーはネットワークの DMZ に置かれ、外部からもアクセスが可能であるが、SSH サーバーの認証を通ったユーザーだけが解析サーバーへアクセスすることが出来る。また、解析用サーバーへのアクセスは SSH サーバーの IP アドレスからのみとし、同じネットワークにいる他ユーザーは SSH サーバー経由でなければアクセスを不可とし、セキュリティを確保している。Galaxy へのアクセスは病院および研究所からの SSH 接続が不可能ということが判明したので、SSH 接続ではなく、VPN 接続に切り替えて運用を開始した。

2. 解析用サーバー構築

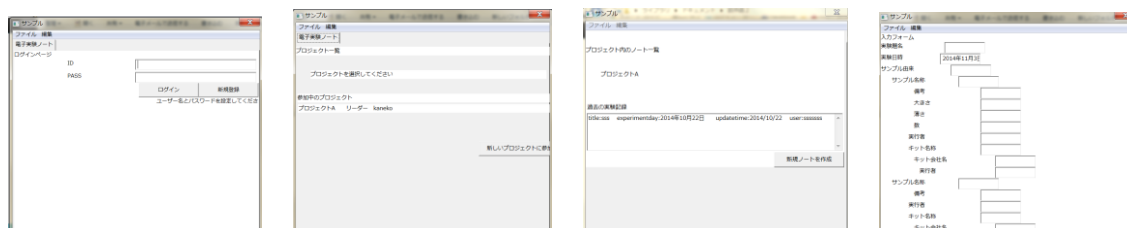
解析用サーバーは、解析データの匿名化、匿名化後データの統計解析処理および高い再現性を目的とした実験履歴を保存するシステムに利用される。このシステムは運用開始後も順次機能追加をする予定である為、新規機能やトラブルシュートの為にも試験環境が必要である。この試験環境の為のサーバー機も購入済みである。この 2 つの解析用サーバーに対し、研究履歴保存および定型処理を実行するソフトウェアである Galaxy をインストールし、その他に統計解析に用いる R 及び DNA メチル化解析及び SNP データの解析に必要なライブラリのインストールを行い、統計解析の環境構築を行った。

3. 管理機能を Galaxy に実装 (匿名化機能、SNP抽出アルゴリズム、フィルタリング、ノーマライゼーションプロセス)

田中(康)班によって開発された匿名化アルゴリズムをテスト環境への実装まで完了した。SNP抽出アルゴリズムとフィルタリング、ノーマライゼーションアルゴリズムの実装までは至らなかった。

4. 実験品質管理のためのソフトウェア開発

東京都立産業技術高等専門学校の鈴木研究室と協業し、実験ノートのソフトウェア作成を推進した。ソフトウェアの実装言語は各研究者が多く使用する Windows, Linux, Apple の各 OS へネイティブアプリケーションを作成可能なことを考慮し、python を選択した。難治性疾患研究部志村グループ松永先生へ実験ノート記載事項のヒアリングを行い、それに基づき簡易的なソフトの開発まで至った。



Subject No. : 25 指 112
Title : A study for designing quality control system for studies in clinical bioinformatics
Researchers : Noriko Tanaka, Mari Shimura, Yasuhiro Tanaka
Key word : omics; data management; database; clinical bioinformatics; standardization
Abstract :

Our goals in the second year were:

1. Set up a computer network to control the quality of clinical studies with huge biological data.
2. Construct basic system which we designed in the first year
3. Develop software to control the quality of experiments for measuring omics outcomes.

For no.1, we set the web and computing server in JCRAC data center, and set up a computer network for some terminal users in the hospital and the research institute to have access to those central servers. We have established a SSH connection to the server for the administrators and VPN connection to access Galaxy for the terminal users in the hospital and research institute .

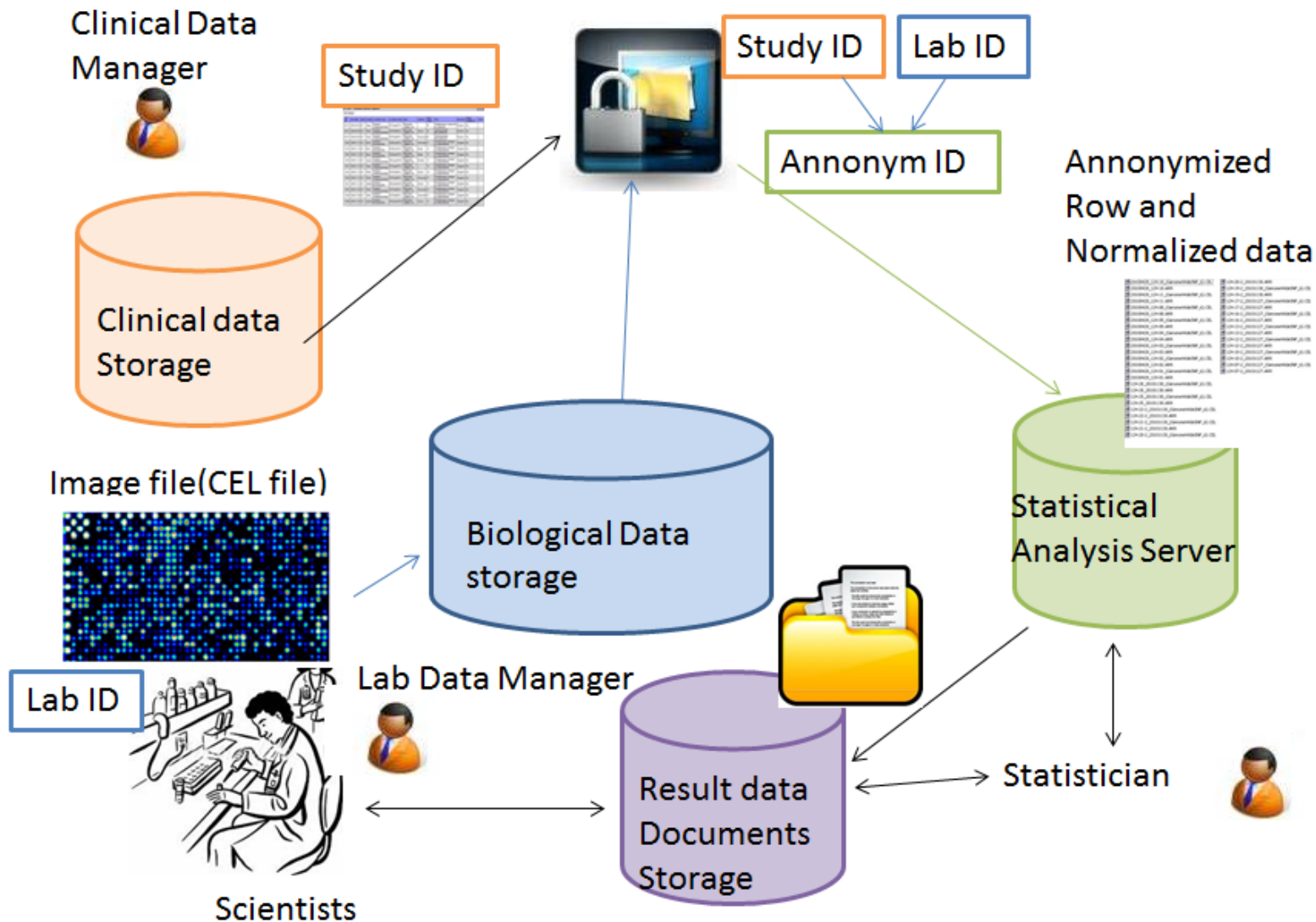
For no.2, in the first year, we choose Galaxy, which is an open, web-based platform for data intensive biomedical research. We set up the test environment for testing NCGM Galaxy customized for our purpose, QC for DNA methylation data. We tested NCGM Galaxy and related algorithms in the test environment. We developed four new algorithms to be implemented on the NCGM Galaxy. Mr.Toyooka and Mr.Mochizuki developed an algorithm to detect SNPs on the probes for Genome-Wide DNA methylation studies. Mr.Uesato developed tools to anonymize meta-files generated from omics experiments. Mr. Kurosawa developed algorithms to count peaks of multimodal distributions for filtering unexpected samples. Mr.Kurosawa, Ms.Yamazaki and Mr.Uesato developed tools for normalizing the data from Illumina Infinium assay combining with SNP and sample filtering algorithms we developed in our project.

For no.3, we designed software to control the quality of experiments for measuring DNA methylation levels, and asked to develop software to Mr.Suzuki and his students at Tokyo Metropolitan College of Industrial Technology. We use Python to develop software

研究目標-

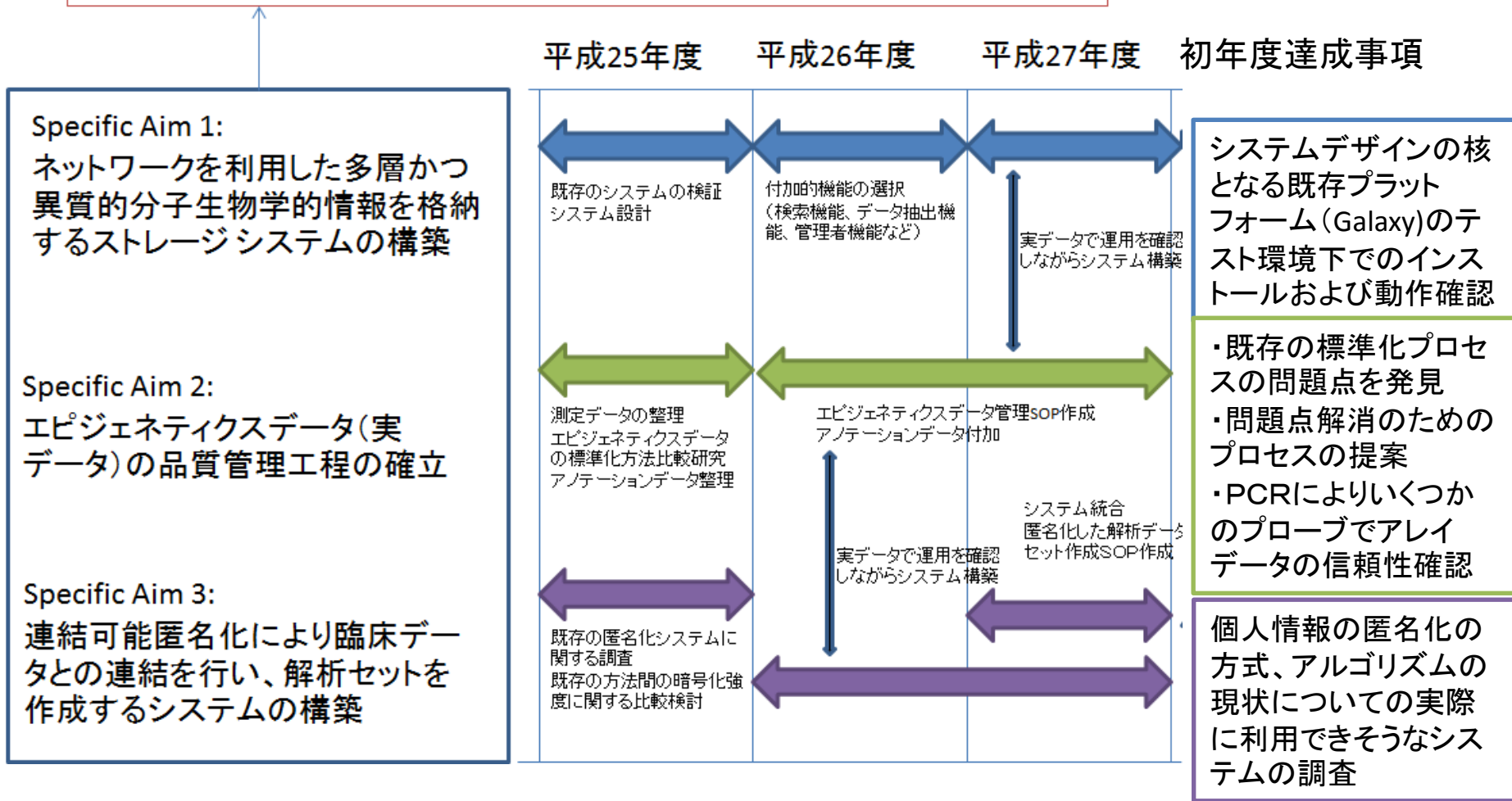
分子生物学的情報を扱う臨床研究の品質管理システムの構築

System for Anonymization



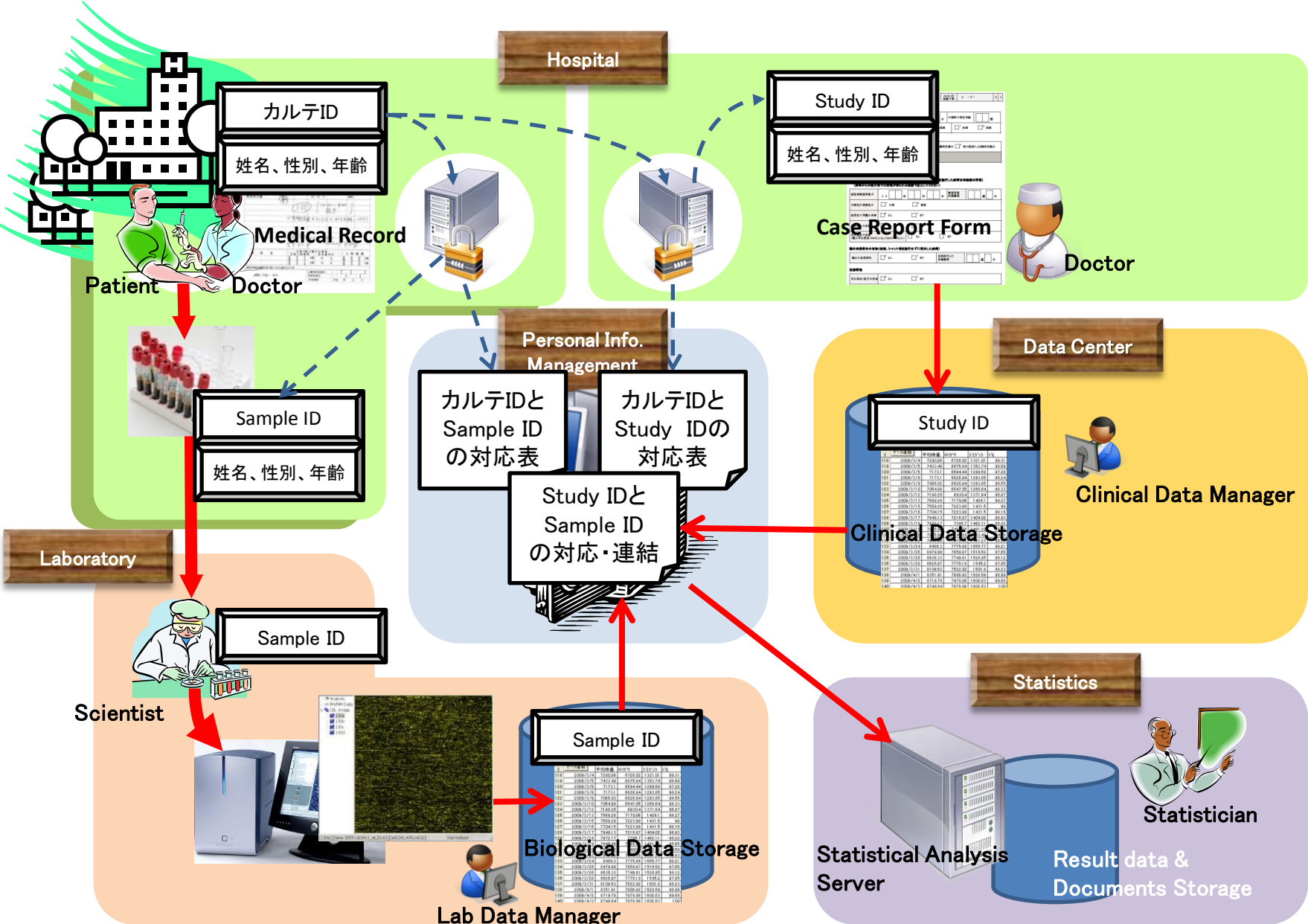
研究計画と昨年度達成内容

研究目的: 分子生物学的情報を扱う臨床研究の品質管理システムの構築

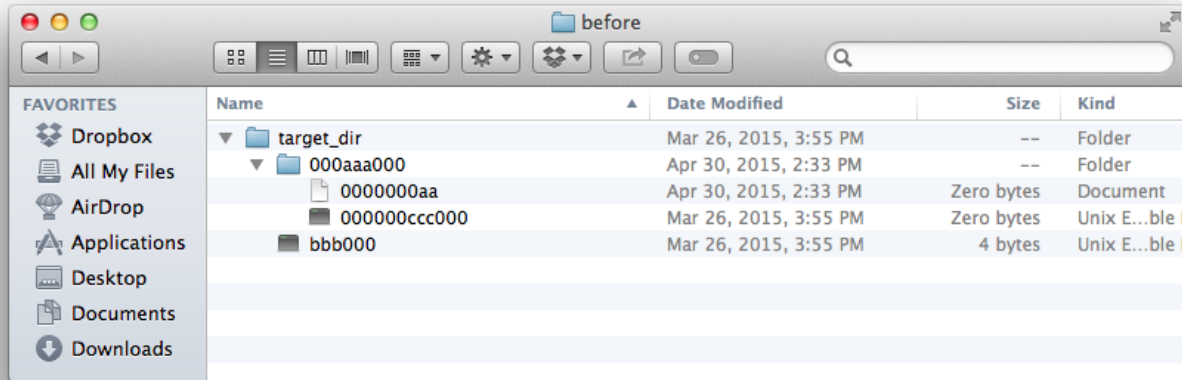


【25指112分担研究課題名】

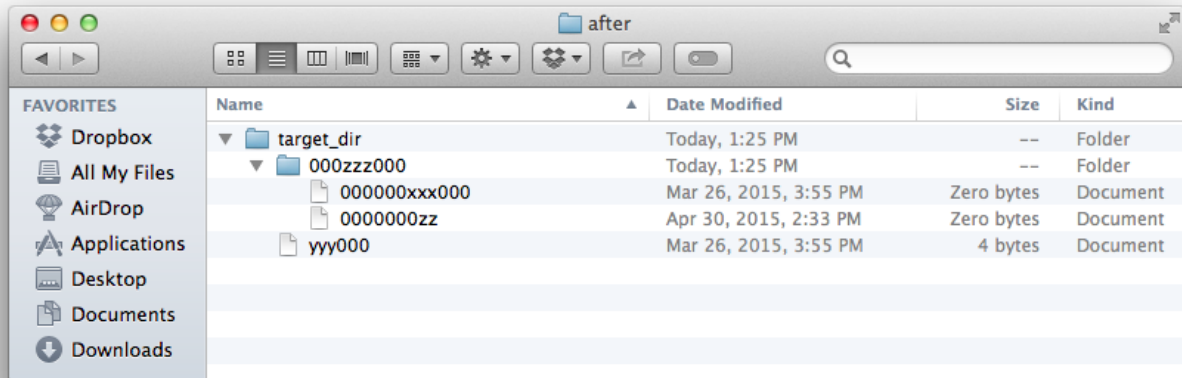
匿名化した臨床データを結合した解析セット作成システムに関する研究



2. ファイル名・ディレクトリ名の匿名化



匿名化



置換

SampleID	AnonymousID
aaa	zzz
aa	zz
bbb	yyy
ccc	xxx

ID対応表

エピジェネティクスデータの品質 管理工程の構築に関する研究

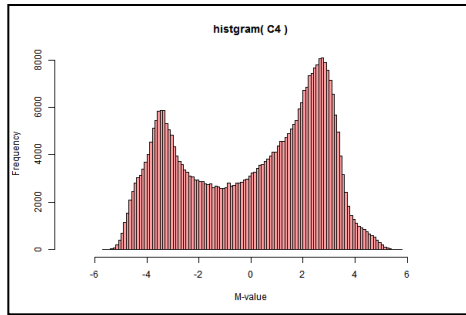
25指112分担研究

研究所難治性疾患研究室

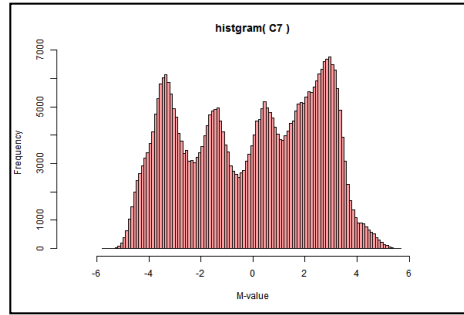
志村まり

異常なピークを示したDNAメチル化分布の解析 —B細胞性リンパ腫瘍臨床検体—

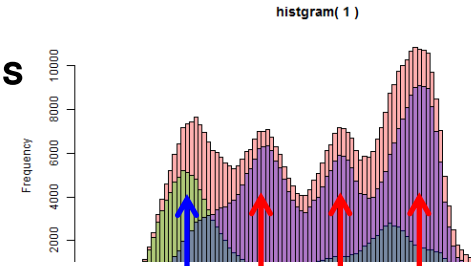
2ピーク



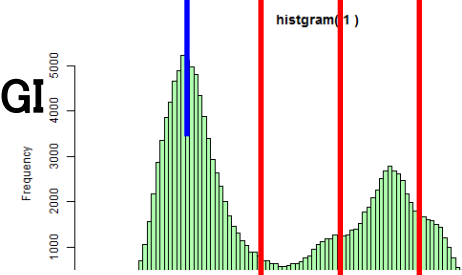
>2ピーク



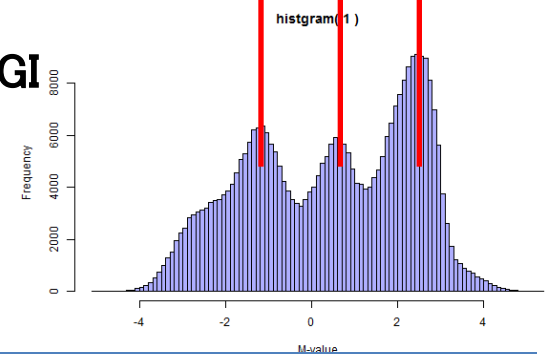
All Probes



CGI



Non-CGI



DNAメチル化分布	検体数(%)
2ピーク	6 (21.4%)
3ピーク	10 (35.7%)
>3ピーク	12 (42.9%)
合計	28

1. 40%の検体で異常な複数ピークが見られた。
2. Non-CGI領域のメチル化状態が複数ピーク形成に寄与していることが判明した。

異常なピークはB細胞性リンパ腫瘍特有？

1. RnBead workflowの適用とhistogram作成
B-cell lymphoma data (28検体)
DLBCL data (52検体)
2. TCGA DNA methylation analysis dataのDNA methylation distributionの解析 (**histogram作成終了**)
3. B-cell lymphoma (合計70検体)のPeakと臨床情報との関連の解析
survival
prognostic index factor
4. TCGA databaseのPeakと臨床情報との関連の解析(可能なら)
5. TCGA databaseのNormal vs. Tumorの解析
DNA methylation change rate

研究発表及び特許取得報告について

課題番号： 25指112

研究課題名： 分子生物学的情報を扱う臨床研究の品質管理に関する研究

主任研究者名： 田中紀子

論文発表

論文タイトル	著者	掲載誌	掲載号	年
該当なし				

学会発表

タイトル	発表者	学会名	場所	年月
FILTERING SAMPLES BASED ON BETA-MIXTURE MODEL FOR DNA METHYLATION DATA QUANTIFIED BY BISULPHITE MICROARRAYS .	Tanaka N, Kurosawa T, Inaba Y, Toyooka L, Yoshida L, Kawasaki Y.	International Biometric Conference	Florence. Italy.	2014年7月
ゲノムワイドDNAメチル化解析において、中間での高い分布の解析.	松永章弘, 豊岡理人, 吉田壘, 石坂幸人, 田中紀子, 志村まり.	第66回日本細胞生物学会.	奈良	2014年6月
メチル化アレイ測定データの分布に基づくサンプル品質管理及び人種差を考慮したプローブフィルタリングの妥当性の検討	豊岡理人 松永章弘 山崎茉莉亜 志村まり 田中紀子	日本人類遺伝学会	仙台	2013年11月
The choice of smoothing parameter and the number of permutation in estimation of multidimensional local false discovery rate based on the subset of high-dimensional genomic data.	山崎茉莉亜 豊岡理人 田中紀子	日本計量生物学会	福島	2013年5月

その他発表(雑誌、テレビ、ラジオ等)

タイトル	発表者	発表先	場所	年月日
該当なし				

特許取得状況について ※出願申請中のものは()記載のこと。

発明名称	登録番号	特許権者(申請者) (共願は全記載)	登録日(申請日)	出願国
該当なし				

※該当がない項目の欄には「該当なし」と記載のこと。
 ※主任研究者が班全員分の内容を記載のこと。