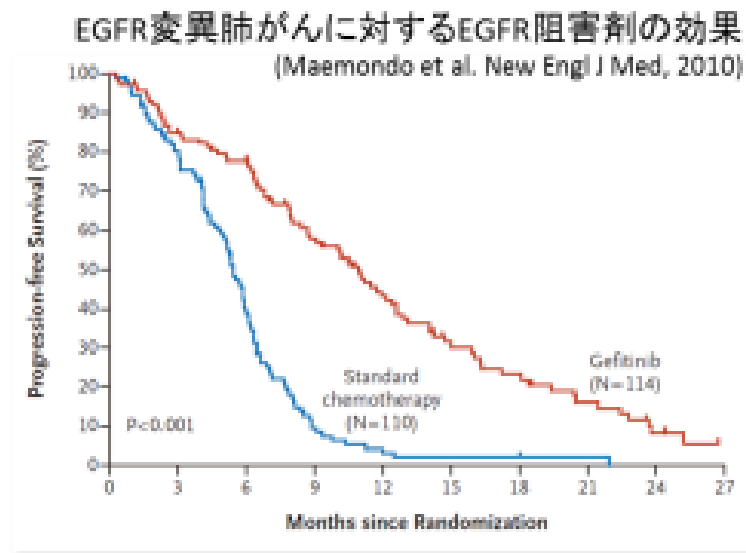
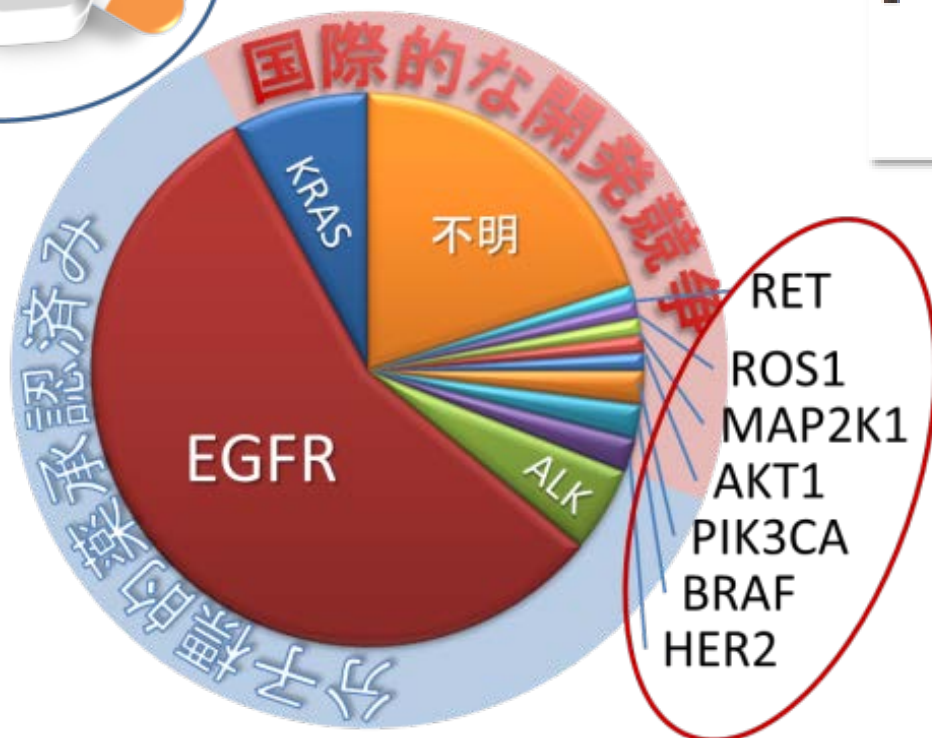


疾患ヒトゲノム変異の 生物学的機能注釈を目指した 多階層オミクスデータの統合

東京大学新領域創成科学研究科
菅野純夫

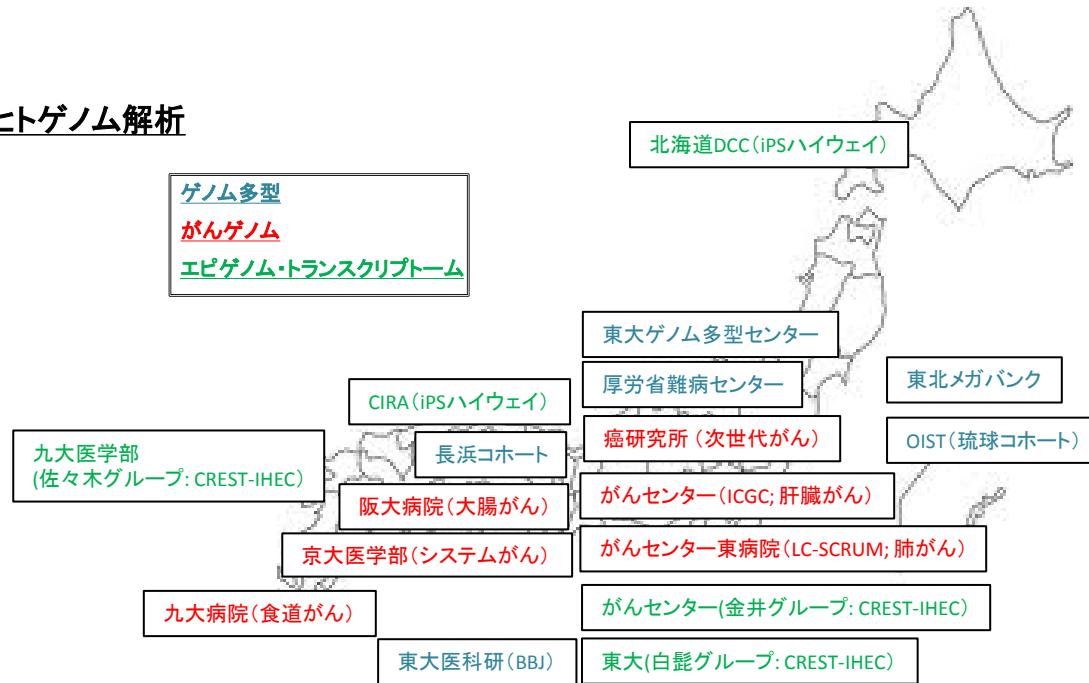
創薬スクリーニングが急速に進行している。



肺腺がんのドライバー変異

新規バイオマーカーの探索/セルアッセイによるスクリーニングにゲノム情報が強力

全国に展開するヒトゲノム解析



ゲノムデータは急速に蓄積している

ヒトオミクスデータ推定蓄積量

ゲノム多型(WGS/WES): >2000人

がんゲノム(WGS/WES/Target Seq): >1000症例

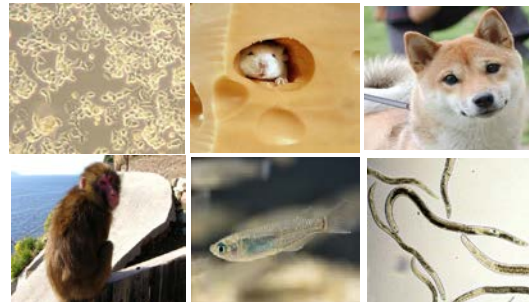
トランスクリプトーム(RNA Seq): >1000例

エピゲノム(BS/ChIP Seq): <100例

+培養細胞+PDX+モデル系:>5000例

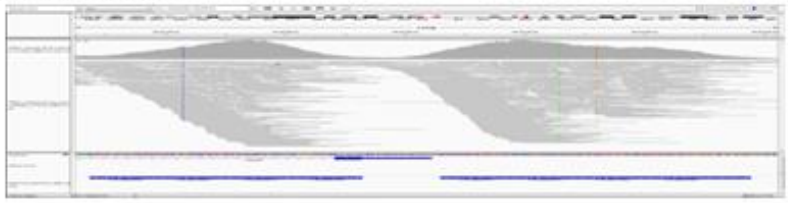
+マウス等モデル生物: ???例

+個別研究者の蓄積するオミクス情報: ???例

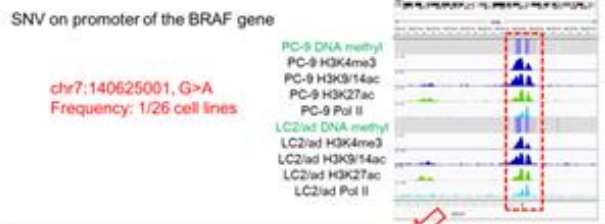


データ統合が目指すヒトゲノム臨床応用研究(がんへの応用を例に)

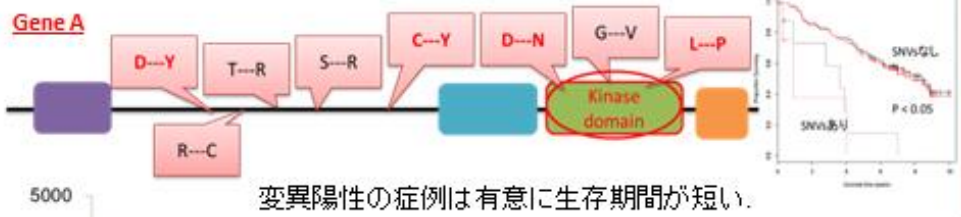
WGS/WES解析



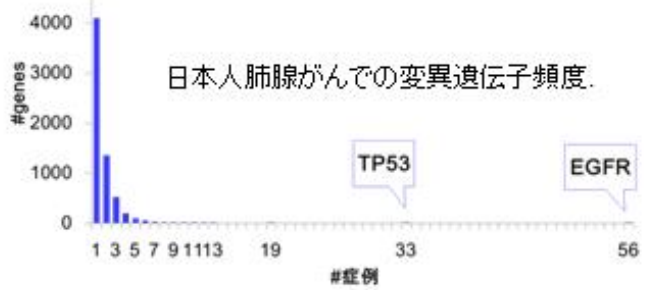
Regulatory SNVsの解析



Coding SNVsの解析例

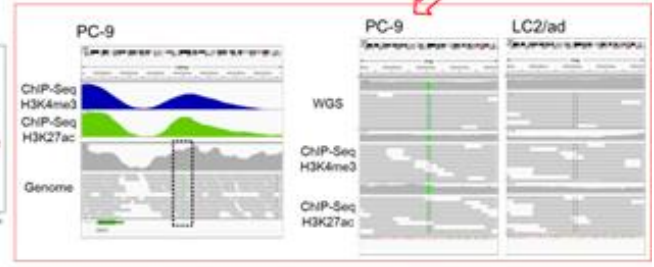


変異陽性の症例は有意に生存期間が短い。



日本人肺腺がんでの変異遺伝子頻度。

・症例間で変異遺伝子が重複することは例外的な遺伝子を除いて、まれ

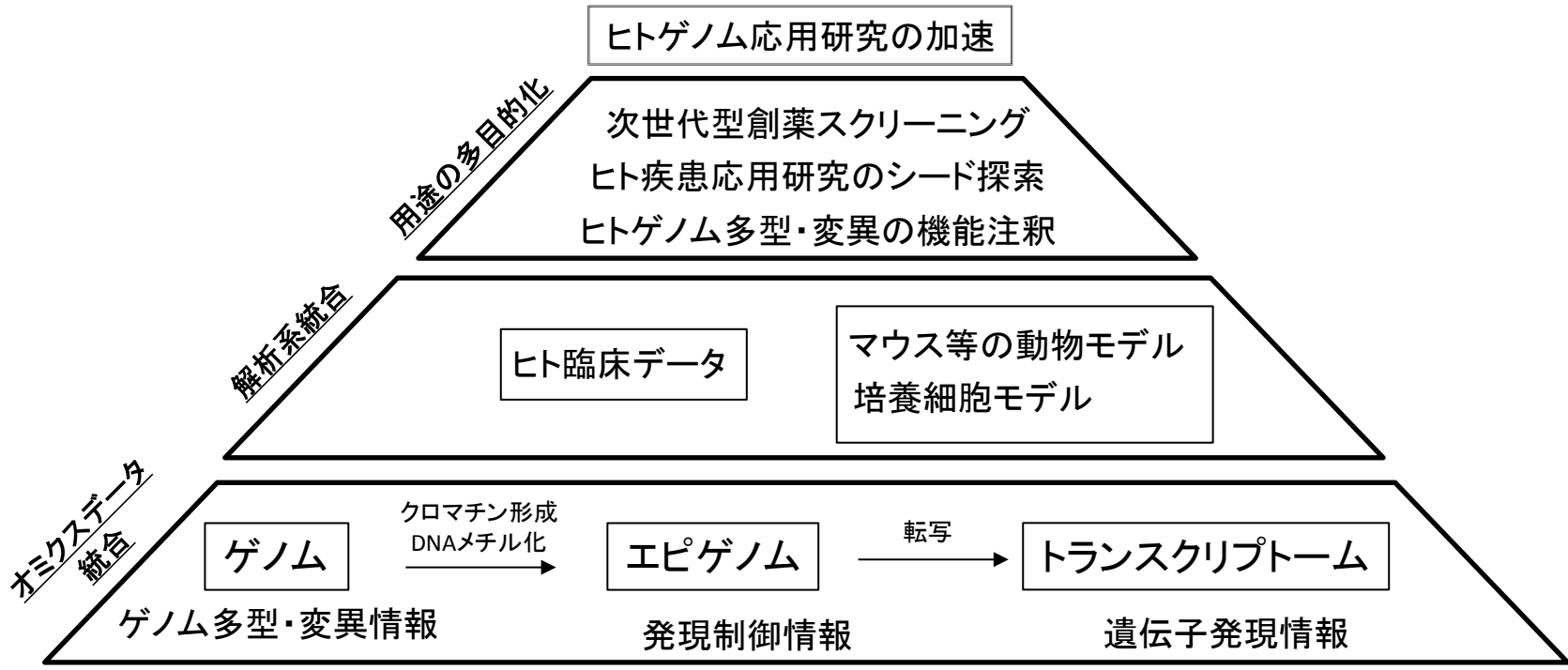


・Passenger変異<->Driver変異の区分が困難

・Regulatory SNPについての情報が圧倒的に不足

ゲノム情報が創薬ゲノミクス・臨床応用へ直結しない

データ統合が加速するヒトゲノム臨床応用研究



国内外に急速に蓄積するオミクスデータ

ゲノムデータベース(標準多型)

国内

東北メディカル・メガバンク機構
TOHOKU MEDICAL MEGABANK ORGANIZATION

難病情報センター
次世代遺伝子解析装置を用いた難病の原因究明、治療法開発プロジェクト

エピゲノムデータベース (未公開)

海外

IHEC
International Human Epigenome Consortium

ROADMAP epigenomics PROJECT

統合データベース

海外

ENCODE

データバンク

国内

NBDC
National Bioscience Database Center

DBJ

海外

Japanese Genotype-phenotype Archive

SRA
The Sequence Read Archive (SRA) stores raw sequencing data from the next generation sequencing (NGS) systems: Applied Biosystems SOLiD System, Illumina HiSeq/TruSeq, etc.

ゲノムデータベース(疾患変異)

海外

NHLBI Exome Sequencing Project (ESP)
Exome Variant Server

International Cancer Genome Consortium

The Cancer Genome Atlas
Understanding genomics to improve cancer care

データ産生

国内

文部科学省科学研究費新学術領域研究
ゲノム支援

EPOC
Epigenome Profiling Research in Cancer Therapeutics
利用・促進施設研究センター

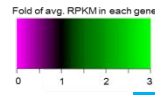
P-DIRECT
文部科学省
次世代がん研究シーズ戦略的育成プログラム
Project for Development of Innovative Research on Cancer Therapeutics

CREST/ITEL
国際ヒトエピゲノムコンソーシアム 日本チーム

海外

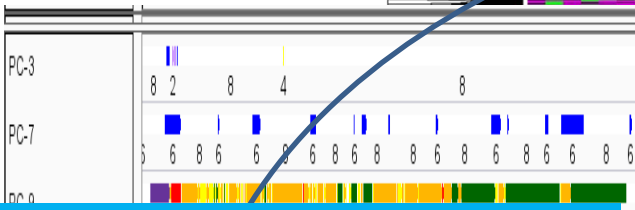
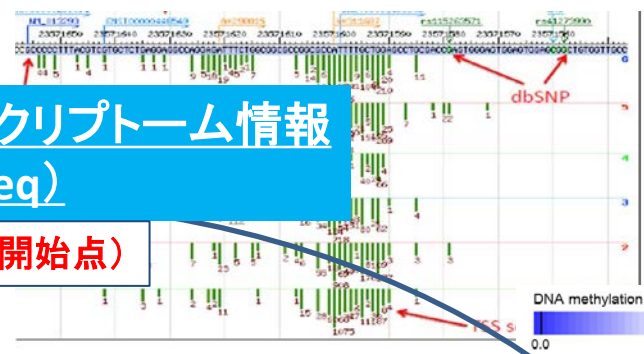
ENA
European Nucleotide Archive

ヒト応用研究を志向したオミクス情報の統合 (EGFR遺伝子を例に)



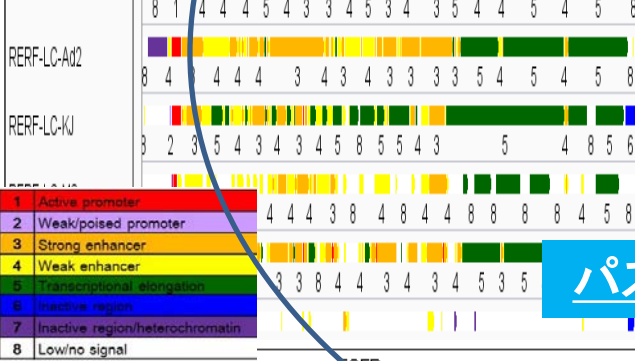
**転写開始点/トランスクリプトーム情報
(TSS/RNA Seq)**

(発現量と転写開始点)

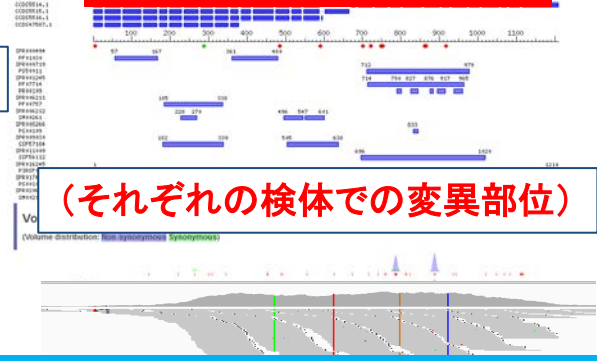


クロマチン情報 (ChIP Seq)

(ChrHMMパターンで示すヒストン修飾)



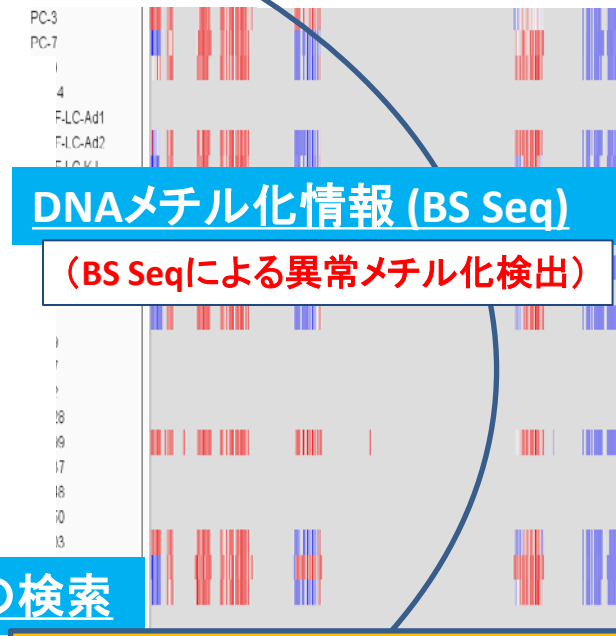
**ヒトゲノム
変異情報の統合**



(それぞれの検体での変異部位)

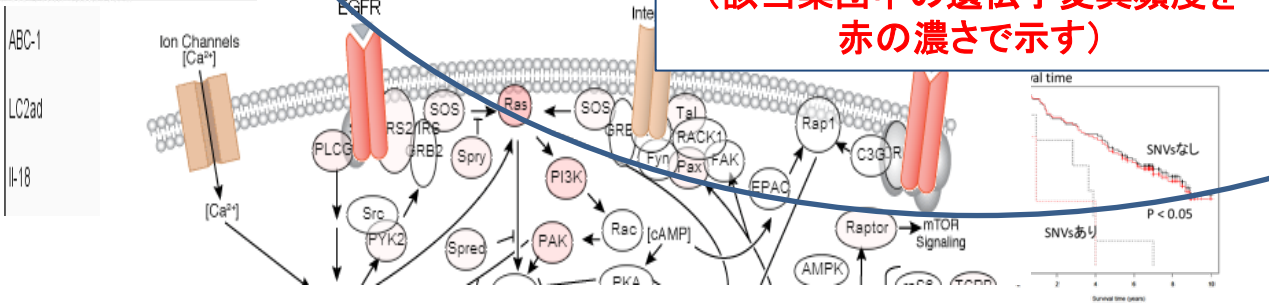
DNAメチル化情報 (BS Seq)

(BS Seqによる異常メチル化検出)



パスウェイマップ (文献情報) からの検索

(該当集団中の遺伝子変異頻度を
赤の濃さで示す)



モデル系とのさらなる統合

研究開発体制と役割

菅野: 研究の統括/医科研スパコンの運用

(雇用予定)

鈴木: データの加工とデータベース設計

+ 研究員 3名

+ 技術補佐員 2名

土原: ヒト疾患応用研究を志向した検索システムの実装

河野: データベースの構築、オントロジーの整備



データ統合対象

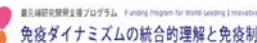
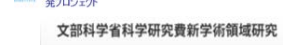
	名称	件数 (保有しているもの)	概要
1	ヒトがんゲノムデータ	300	肺腺・小細胞がん、食道がん、胃がん、大腸がんの全ゲノム/エキソームデータ
2	ヒト正常ゲノムデータ	500	申請者らの収集した正常日本人全ゲノム/エキソームデータの頻度情報
3	ヒトエピゲノム、トランスクリプトームデータ	300	肺腺がん、大腸がん: 正常肝臓(予定)、正常大腸(予定)、iPS/ES細胞
4	ヒト培養細胞オミクスデータ	30 (300計測点)	肺腺がん、大腸がん(予定)、胃がん(予定)、初代培養正常細胞・PDX(予定)
5	マウスオミクスデータ	10 (400計測点)	正常組織、初代培養マクロファージ、B細胞等の免疫担当細胞(FIRST審良P)
6	* その他のモデル動物データ	100 (500計測点)	「ゲノム支援」で収集、公開予定のデータ

用いるすべてのデータは申請者らがデータ産出、1次データ加工を実際に行ったものである。

*** モデル系から得られたオミクスデータに焦点。**

**特に培養細胞40種類:
ゲノム-エピゲノム-トランスクリプトームのオミクス
データが同一の材料から得られたもの**

**比較ゲノムブラウザー
によりヒトゲノムと対応づけ**



メインサーバー: 医科研ヒトゲノムセンタースパコンシステム

構成名	機種名	Rpeak (Gflops)	CPU	OS	CPU コア 数	メモ リ/ node	ディ スク 容量
分散メモリ型 Type-A	Hitachi HA8000/RS210	150,141	Intel Xeon X5675 (3.06GHz)	Red Hat Enterprise Linux	144	144 GB	-
分散メモリ型 Type-B	Hitachi HA8000- tc/HT225		AMD Opteron 6276 (2.3GHz)		16,128	32 GB	-
共有メモリ型	SGI UV 100	1,361	Intel Xeon E7 8837 (2.66GHz)		128	2 TB	58 TB
ディスクア レイ	-	-	-	-	-	-	2 PB

既存のサーバーの利用: 最低限の設備投資での計画遂行

統合に用いるオミクスデータの多く

=

申請者らが収集・加工したデータ

文部科学省科学研究費新学術領域研究
ゲノム支援

CREST
iHEC
International Human Epigenome Consortium, iHEC, Team Japan (supported by CREST, JST)
科学技術振興機構・戦略的創造研究推進事業「エピゲノム研究に基づく診断・治療へ向けた新技術の創出」
国際ヒトエピゲノムコンソーシアム 日本チーム

Trial Center

最先端研究開発支援プログラム
免疫ダイナミズム

難病情報センター
次世代遺伝子解析装置を用いた難病の原因究明、治療法開発プロジェクト

申請者らが自身によるメタデータの付加、オントロジーの整備が可能

申請者らが運用する既存のデータベース・検索システムの転用

基本的な枠組みは構築済み！データは加工済み！！

現有するデータは初年度内での整備・公開を目指す



最終年度の高度検索システムの開発へと初年度から注力

検索例1

SNV on promoter of BRAF

chr7:140625001, G>A

Frequency: 1/26 cell lines

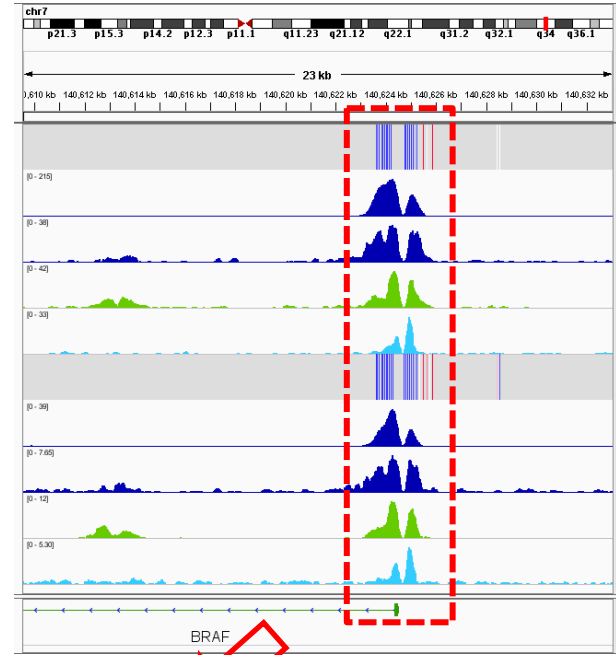
＝疾患ゲノムのその座標で“何が起きているのか”を網羅的に検索

このゲノム変異はエピゲノム、トランスクリプトームに変化を与えない。

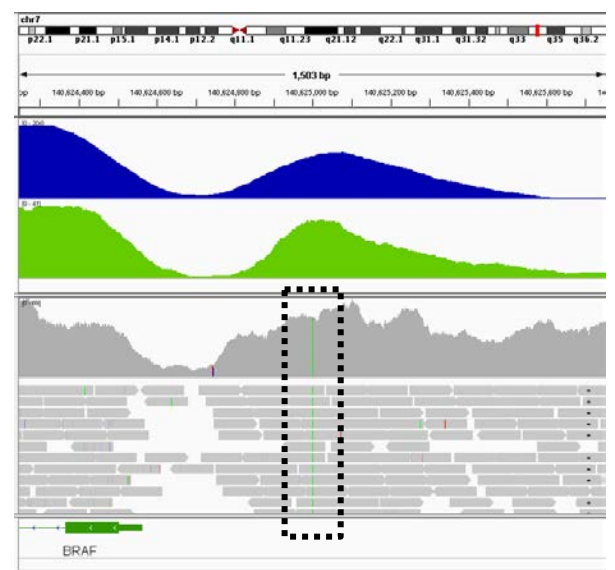


中立変異の可能性が高い？

- PC-9 DNA methyl
- PC-9 H3K4me3
- PC-9 H3K9/14ac
- PC-9 H3K27ac
- PC-9 Pol II
- LC2/ad DNA methyl
- LC2/ad H3K4me3
- LC2/ad H3K9/14ac
- LC2/ad H3K27ac
- LC2/ad Pol II

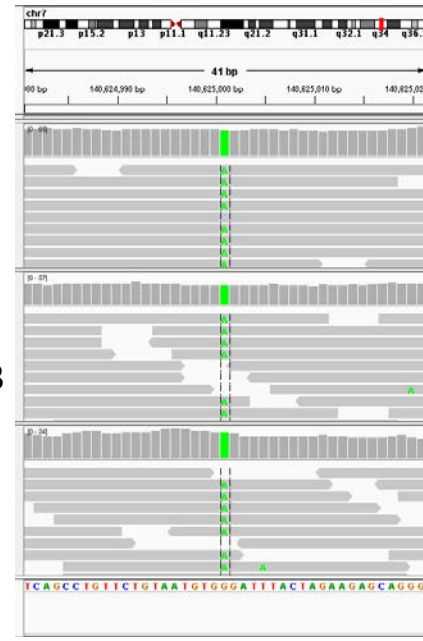


PC-9



ChIP-Seq
H3K4me3
ChIP-Seq
H3K27ac
Genome

PC-9



WGS
ChIP-Seq
H3K4me3
ChIP-Seq
H3K27ac

LC2/ad



ヒト疾患ゲノム統合DB: DB-MGS(仮称)

ヒトゲノム・エピゲノム・トランスクリプトームデータの統合
 ヒト疾患ゲノム変異(特にregulatory SNV)への機能的注釈
 オントロジー、メタデータの整備
 培養細胞、モデル生物オミクスデータの活用
 パターン検索システムの開発と実装



オミクスデータ統合が加速するヒトゲノム臨床応用研究

＝疾患ゲノムのその座標で“何が起きているのか”を網羅的に検索