

個別化医療に向けた ヒトゲノムバリエーションデータベース

東京大学大学院医学系研究科

徳永勝士

共同研究機関

東大附属病院、遺伝学研究所、日立製作所

目次

- ・ 背景
- ・ 目的
- ・ 構想
- ・ 研究開発の進め方
- ・ 共同研究体制
- ・ 将来展望

背景

ゲノムワイドな**SNP**タイピング技術と次世代シーケンサ技術の革新的向上と普及

新規発見の疾患関連/薬剤応答遺伝子・変異が増加



- 1) データの半永続的な管理と研究者間での情報共有
- 2) 変異と表現型との関係の体系化
 - 複数遺伝子/変異の疾患への複雑な関与
 - 同一疾患における変異部位・種類特異的症状の存在
 - 同一遺伝子/変異の複数疾患への関与
 - スタディー間の結果の不一致
(集団間差異、検出力不足、表現型定義のゆれ)

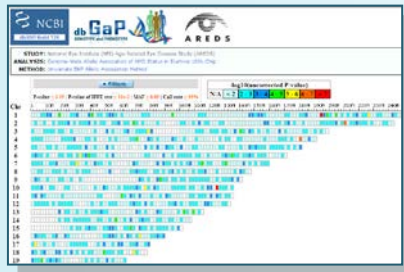
海外でのGWAS/NGSデータ共有化

1) NCBI

- dbSNP SNP情報を蓄積
- dbVAR 構造多型のデータを蓄積
- dbGaP GWAS, 次世代シーケンサー結果を含むgenotype-phenotypeに関するデータを蓄積

2) EBI

- EGA GWAS, 次世代シーケンサー結果を含むgenotype-phenotypeに関するデータを蓄積



3) Gen2Phen

- GWAS, 次世代シーケンサー結果を含むgenotype-phenotypeに関するデータを蓄積
- * 学術誌でも公的データベースへのデータ登録を要求することが増加

海外での変異データの体系化

- 1) HGMD 疾患関連遺伝子変異(生殖細胞のみ)を網羅的に収集
(有料 登録後3年を経過したデータはアカデミック無料)
- 2) NCBI
 - ・ClinVar dbSNP, dbVAR, dbGapのデータ及び、臨床検査などのサブミットデータを精査し登録したgenotypeとphenotypeの関係をDB
- 3) HGVS (human genome variation society)
 - ・LSDB (locus specific database)のリンク集
- 4) COSMIC:がん体細胞変異のDB
- 5) ICGC, TCGA Project DBがん体細胞変異の体系化とゲノム配列データアーカイブ

統合DB: GWASデータの情報共有

The image displays a collage of screenshots from three databases: GWAS, SNP, and CNV Association Reference.

- GWAS Database:** Shows a navigation menu on the left with categories like 'About GWAS Control', 'HELP/FAQ', 'ABOUT', 'ABOUT MY DATABASE', 'SEARCH & Browse', and 'Browse across the region'. The main content area includes an introductory text and a search interface.
- SNP Control Database:** Features a navigation menu and introductory text: 'The SNP control database consists of quality controls and is used in Hardy Weinberg Equilibrium test'. It includes a search bar and a 'SEARCH & Browse' section.
- CNV Association Reference:** Shows a navigation menu and introductory text: 'The copy number variation association database (CNV association DB) is a repository system and has been constructed to achieve systematic data management and information sharing of CNV data'. It includes a search bar and a 'SEARCH & Browse' section.
- GWAS Data Track:** A detailed genomic track showing association signals across chromosomes 1-22, X, and Y. It includes a legend for 'CHROMOSOME' and 'LocusZoom' scores.
- CNV Data Table:** A table with columns for 'CHROMOSOME', 'START', 'END', 'CNV ID', 'CNV TYPE', 'CNV SIZE', 'CNV SCORE', 'CNV SCORE (log10)', 'CNV SCORE (log2)', and 'CNV SCORE (log10/2)'. The table lists various CNV events across chromosomes 1-22, X, and Y.

統合化推進プログラム: ヒトゲノムバリエーションDBの構築 変異データの体系化

HLA DATABASE

Human Variation DB

Gene name: PADI4
Region: chr1:1764800-1764895
Full name: papsyl angimus domain type IV
Synonyms: PADI4-AS1, PADI4-AS2
Related disease: Rheumatoid arthritis, Type 1 diabetes

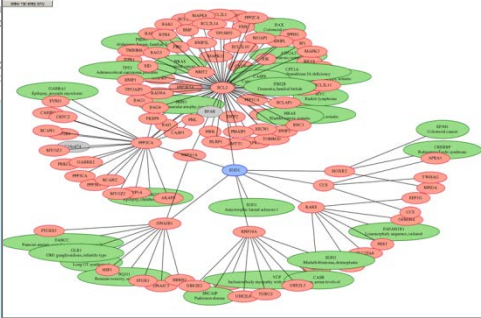
Chromosome: 6 Region: 1764801 - 1764842

Genomic position: chr1:1764809-1764811 C:3350-C G:G>T[124] ESO2569 PADI4

Genomic position	rs ID	Nil change info	Amino acid change	Gene ID	Gene name	hetero
chr1:1764809-1764811	rs1674811	C:3350-C	G:G>T[124]	ESO2569	PADI4	
chr1:1764839-1824030	rs240340	C:341-157C		ESO2569	PADI4	
chr1:1764839-1824030	rs240340	C:341-157C		ESO2569	PADI4	
chr1:1764839-1824030	rs240340	C:341-157C		ESO2569	PADI4	
chr1:1764839-1824030	rs240340	C:341-157C		ESO2569	PADI4	

集団ごとの変異の分布

疾患感受性変異の分布



Pathway上の変異

集団ごとの参照ゲノム

領域での変異情報と統計値

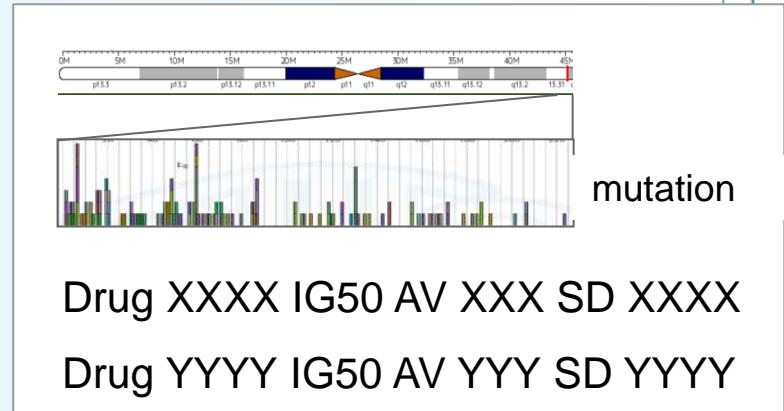
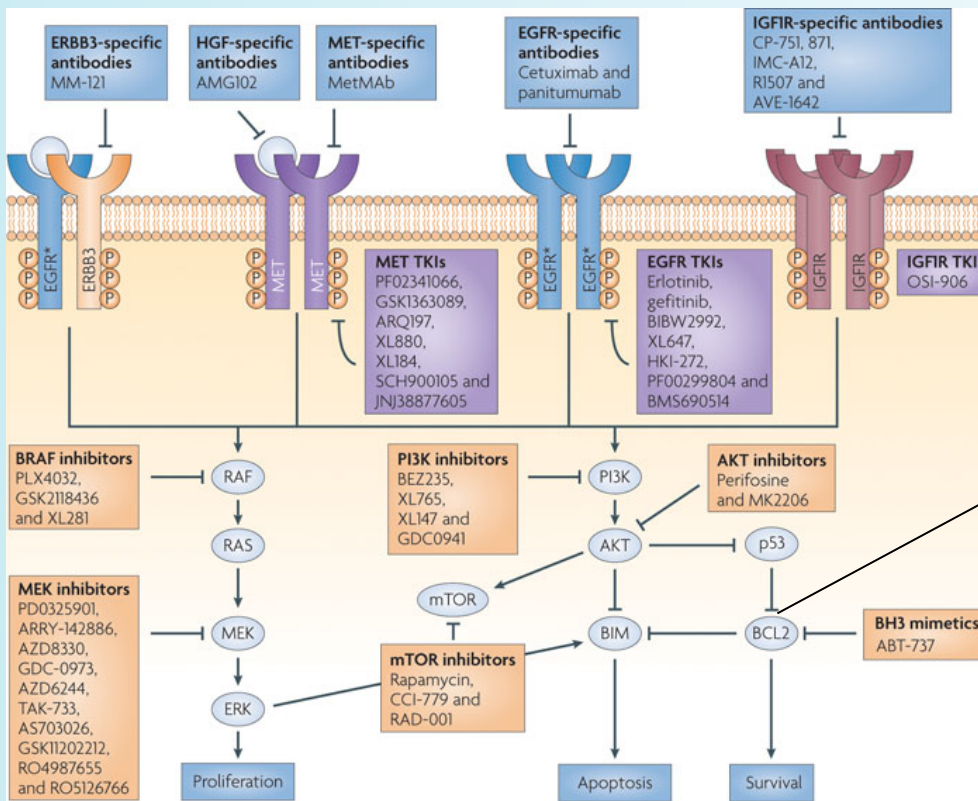
目的:

変異・疾患・薬剤応答・ウイルス耐性・臨床情報を整理・体系化し、俯瞰可能とすると共に、表現型の機序理解促進と個別化医療の礎とする

構想:

1. 次世代シーケンサーおよび、その他の解析法(GWASを含む)によって発見される多型・変異情報の預入れ・受入れ
2. 文献情報を含め過去産出された疾患・薬物応答・ウイルス耐性に関する多型・変異データの収集(統計値と各サンプルデータ)
(体細胞変異については既存DBとのリンクと一部同時表示のみ)
3. オミックスデータとの連携による疾患・薬物応答の機序理解の促進
4. 既知変異の情報(配列上の性質)を元に、新規変異の表現型を予想

パスウェイ、薬剤応答性と変異



Nature Reviews | Cancer

Pao & Chmielecki Nat. Rev. Cancer. 2010

研究の進め方

1. Human Variation DBの拡張

- ・COSMIC など体細胞変異との連携機能追加 (1年目)
- ・薬剤応答と変異の関係の解釈を容易とするパスウェイデータ、オミックスデータとの連携機能追加 (2,3年目)
- ・パーソナルゲノム情報に基づく脳疾患メカニズムの解明、ゲノム支援などからのデータ登録 (1, 2年目)
- ・新学術でシーケンスされるHLA ハプロタイプの配列決定及びDB登録 (1,2年目)
引き続き、公開DBと内部用DB(公開前の多施設情報共有)での運用 (1-3年目)

2. Human Variation DB、HLA DBのデータ充実化

- ・変異データと疾患感受性、薬剤応答性、ウイルス耐性などの表現型との関係性のデータの充実化、臨床情報の充実化 (データ受け入れと、文献からの抽出データ追加) (1-3年目)
- ・上記の公開用データは、個人を特定できないレベルとし、また、臨床情報、疾患、遺伝子での横断検索を可能とすることで、これらの関係性情報を俯瞰可能に

研究の進め方

3. GWAS-DBの拡張と充実化

- ・新学術、厚労科研など、引き続きデータの充実化とメタ解析手法の解析と適用 (1-3年目)

4. NBDCのデータアクセス委員会と連携したデータの受入れ

- ・NBDCのデータアクセスと連携し、データの受入れ、データの整理、データの再配布を行う (1-3年目)

5. 国際連携及び学会との連携

- ・EBI-EGA、およびGWAS Central、Gene2Phenとの国際連携を開始し、必要に応じてデータの交換を行えるようにする (2-3年目)
- ・人類遺伝学会とも連携し、大会、学術誌などを通して、データ登録を呼びかける (1-3年目)

保有する データ例

	名称	件数	概要
1	健常者データ	2 studies	ゲノムワイドSNPデータ: 数千人規模
2	B型肝炎	1 study	GWAS: 数千人規模
3	C型肝炎	1 study	GWAS: 数千人規模
4	原発性胆汁性肝硬変	1 study	GWAS: 数千人規模
5	自己免疫性肝炎	1 study	GWAS: 数百人規模
6	真性過眠症	1 study	GWAS: 数百人規模
7	ナルコレプシー	1 study	GWAS: 千人規模
8	変形性関節症	1 study	GWAS: 数百人規模
9	パニック障害	1 study	GWAS: 千人規模
10	パーソナリティ、気分障害	1 study	GWAS: 数百人規模
11	1型糖尿病	1 study	GWAS: 数百人規模
12	2型糖尿病	1 study	GWAS: 数百人規模
13	不育症	1 study	GWAS: 数百人規模
14	インスリン自己免疫症候群	1 study	GWAS: 数百人規模
15	スチーブンスジョンソン症候群	1 study	GWAS: 数百人規模
16	自閉症	1 study	GWAS: 数百人規模
17	多系統萎縮症	2 studies	GWAS: 数百人規模
18	脳動脈瘤	1 study	GWAS: 数百人規模
19	緑内障	1 study	GWAS: 数百人規模
20	放射線被ばく	1 study	GWAS: 数百人規模
21	肺がん	1 study	GWAS: 数百人規模
22	地域集団	2 study	GWAS: 数百人規模
23	HLA配列データ1	7 検体	HLA領域ホモ接合セルライン
24	HLA配列データ2	10検体	完全胞状奇胎
25	HLA配列データ3	1検体	HLA領域ホモ接合セルライン化前DNA
26	脾類上皮嚢腫のエクソーム解析	2検体	エクソーム解析に伴う変異データ(pool)
27	四肢突発性筋痛症のエクソーム解析	2検体	エクソーム解析に伴う変異データ(pool)
28	永久歯萌出不全のエクソーム解析	2検体	エクソーム解析に伴う変異データ(pool)
29	家族性腎臓病のエクソーム解析	4検体	エクソーム解析に伴う変異データ
30	IgA腎症のエクソーム解析	52検体	エクソーム解析に伴う変異データ (ゲノム支援)
31	慢性関節リウマチのエクソーム解析	59検体	エクソーム解析に伴う変異データ (ゲノム支援)
32	大腸癌のエクソーム解析	50検体	エクソーム解析に伴う変異データ (ゲノム支援)
33	永久歯欠如のエクソーム解析	48検体	エクソーム解析に伴う変異データ (ゲノム支援)
34	MODYのエクソーム解析	59検体	エクソーム解析に伴う変異データ (ゲノム支援)
35	脾類上皮嚢腫のエクソーム解析	8検体	エクソーム解析に伴う変異データ
36	家族性腎臓病のエクソーム解析	4検体	エクソーム解析に伴う変異データ
37	コーデン病のエクソーム解析	5検体	エクソーム解析に伴う変異データ
38	間質性腎炎のエクソーム解析	8検体	エクソーム解析に伴う変異データ
39	ウォルフラム症候群のエクソーム解析	4検体	エクソーム解析に伴う変異データ
40	卵巣癌のエクソーム解析	48検体	エクソーム解析に伴う変異データ
41	日本人健常者	400検体	エクソーム解析に伴う変異データ
42	多系統萎縮症のエクソーム	600検体	エクソーム解析に伴う変異データ (ゲノム支援)
43	日本人ゲノム全ゲノム配列	10検体	全ゲノム配列

共同研究の体制

1. 東京大学医学系研究科（研究代表） 全体の取り纏めとPJの進行
 - Human variation DBの拡張
 - 新学術”HLA進化と疾病”、厚労科研”肝炎等克服対策事業“などで産出される変異情報の登録
 - GWAS DBのデータの収集
 - 海外データベースとの連携の窓口
2. 東京大学附属病院
 - 新学術「パーソナルゲノム情報に基づく脳疾患メカニズムの解明」、新学術「ゲノム支援」、厚労科研「神経系疾患の集中的な遺伝子解析及び原因究明に関する拠点研究」で産生する変異データの登録
3. 遺伝学研究所
 - 新学術「ゲノム支援」で産出された変異データのQCを行うとともに登録を行う
 - 新学術以外で産出された変異データの収集と整理・登録
 - HLA ハプロタイプの配列決定およびDB登録
 - GWAS メタ解析手法開発と適用
4. 日立製作所中央研究所
 - Human variation DBの拡張（主にオミックスデータ連携部分）
 - GWAS , NGS データの登録
 - 文献からの変異-表現型情報の抽出と登録

将来展望

- **GWAS, NGS**データについて、**Journal** , 学会との連携の下、データのサブミッションを呼び掛け、日本人・アジア人の多型・変異情報の中心的な公共**DB**として育てていきたい。
- 大量に産生される変異情報を整理・体系化し、日本人集団に特徴的な多型・変異と疾患、薬物応答、ウイルス耐性との関係を明らかにすることが、遺伝子変異による疾患の予後予測を含めた個別化医療の実現に必要なである。