# 科学研究費助成事業 研究成果報告書



平成 27 年 6 月 2 日現在

機関番号: 32612 研究種目: 基盤研究(C) 研究期間: 2012~2014

課題番号: 24590704

研究課題名(和文)癌関連疾患予測システムの開発と疾患パスウェイの研究

研究課題名(英文)The development of cancer related disease gene prediction system and the study of

disease pathway.

### 研究代表者

満山 進 (Mitsuyama, Susumu)

慶應義塾大学・医学部・特任助教

研究者番号:30296727

交付決定額(研究期間全体):(直接経費) 4,200,000円

研究成果の概要(和文):癌関連タンパク質/遺伝子相互作用データベースCancerProView (http://cancerproview.dmb.med.keio.ac.jp)のデータとしてSMARTとNCBI Refseqからタンパク質ドメインの取得を行った。その後、疾患未同定なタンパク質について疾患関連タンパク質とのドメイン比較を行うことにより疾患の可能性を推測するプログラムの開発を行った。本研究の成果は、癌疾患の基礎研究や診断、治療などの分野で活用されると期待される。

研究成果の概要(英文): This study collected the protein domains from SMART and NCBI Refseq for the cancer related gene/protein database "CancerProView " (http://cancerproview.dmb.med.keio.ac.jp). Furthermore, this research developed another program to predict possible disease by comparing the protein domain with those of the unidentified diseases stored in CancerProView. These results would facilitate various cancer research, clinical diagnosis and treatment.

研究分野: 分子生物学

キーワード:癌 タンパク質 遺伝子 疾患 相互作用 パスウェイ データベース

### 1. 研究開始当初の背景

研究開始当初の平成 24 年の時点で、HUGO Genome Organisation) Nomenculature Committee (http://www. genename.org)には、ヒトの遺伝子記号が 32,000 登録されておりそのうちタンパク質 をコードするものは約 19,000 であった。疾 患遺伝子データベース OMIM(Online Medelian Inheritance in Man)には約 5,000 の遺伝子 が関与する疾患が登録され、疾患遺伝子変異 データベース HGMD (Human Genome Mutation Database)では、約 4,300 の単一変異疾患原 因遺伝子が検索できた。しかし、約 14,700 の遺伝子は、単一で変異が起こった場合の疾 患は不明でありその解明は重要であった。遺 伝子疾患の中でも癌は、遺伝子がコードした タンパク質が遺伝子の変異によりその機能 が変質や消滅したり、産生されたタンパク質 そのものが、ウィルスや細菌などが産生する タンパク質によりその機能が抑制されるこ とにより発症する。タンパク質の相互作用と 疾患の関係を示したデータベースには、京都 大学の KEGG や米国 BioCARTA があり BioCARTA には図として354の登録があった。疾患に関 与するものとしてタンパク質の機能ドメイ ンがあり、サンガーセンターの Pfam による と約 12,000 種類報告されていた。タンパク 質の作用機序中の遺伝子変異と疾患の発症 が関連づけられたものとして、細胞接着因子 カドへリンファミリーなどがあり、その数は 近年の研究の進展により増加している。そこ で研究代表者等は、2006 年から癌関連遺伝 子・タンパク質とその相互作用図を独自に文 献から体系的に収集しデータベース化を行 い 2008 年に CancerProView としてインター ネット上から検索可能なシステムとして公 開した (http://cancerproview.dmb.med. keio.ac.jp/php/cpv.html)。研究開始当初、 109 タンパク質パスウェイ (相互作用)、1,041 遺伝子、2,011 NCBI Refseq cDNA 配列及びア ミノ酸配列を収録し、377 疾患、180 癌関連 遺伝子/タンパク質(BRCA1 など)に関して 検索が可能であり、さらにデータの追加、機 能の追加、インターフェースの改良を行って いた。CancerProView を作成する過程で、同 じタンパク質相互作用の系内で疾患未同定 の遺伝子が多数存在した。カドヘリンファミ リーのようなタンパク質ドメインと疾患の 関連については、研究は進められているが、 未知の遺伝子変異によりタンパク質ドメイ ンが変化した場合の疾患や症状との対応付 けを予測、検索できるシステムは、まだ研究 段階でその開発は重要であった。そこで本研 究では、現在、疾患原因とされていない遺伝 子のタンパク質相互作用とそれに関与する 疾患、タンパク質ドメインなどから疾患の可 能性についての予測、探索システムの開発を 行い、疾患パスウェイを作成し疾患間の関連 性の検討を行う。

#### 2. 研究の目的

本研究では、CancerProView で収集を行った 癌関連遺伝子のタンパク質ドメインと遺伝 子変異、疾患などとの相関から疾患未同定遺 伝子の疾患予測を行い癌疾患との関連性を 明らかにする予定である。

CancerProView のデータを使用し以下の手順により原理作成、方法論の確定を行う。(図1に研究全体の概念図を示す)。

- (1) 癌関連疾患と原因遺伝子の関連付け
- (2) 癌関連遺伝子のタンパク質ドメイン構造と遺伝子変異、癌疾患との対応付け
- (3) 既知タンパク質・遺伝子相互作用、ドメイン構造、疾患の関連付けによる疾患予測モデル化
- (4) 疾患未同定遺伝子のタンパク質相互作用とドメイン構造の相似性からの疾患予測
- (5) 1, 2, 3, 4 を統合化した癌疾患予測システムの作成
- (6) 疾患パスウェイの作成 (CancerProView との統合化)

本研究は、遺伝子変異によるタンパクの機能の変化、消失を考慮し疾患の予測を行う。本研究により次世代シーケンサーなどを使用した遺伝子配列の解析結果から癌疾患と関連して原因遺伝子が未知の疾患を予測でき、診断、予防、予後の検討の役に立つと考えられる。

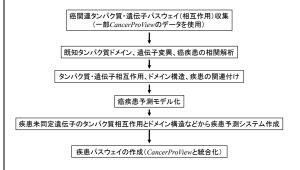


図1 研究全体の概念図

### 3. 研究の方法

# (1) 癌関連疾患と原因遺伝子の関連付け

癌関連疾患と原因遺伝子の関連付けには、 OMIM の疾患の表 morbidmap と Pubmed の文献 検索を用いる予定であった。morbidmap は項 目として疾患名、OMIM ID, OMIM 記号、座位 のデータがあり、その中の疾患名をキーワー ドとして癌疾患と OMIM 記号を抽出する予定 であったが、OMIM に Cancer, oncogene, Leukemia, tumor, syndrome などのキーワー ドで直接検索をかけても同じ結果が得られ るため morbidmap は使用しなかった。さらに OMIM の Alleric variant を参照することに同 じ結果がよって遺伝子変異を調べ、疾患との 関連性について 1759 遺伝子について疾患と の関連性をリスト化を行った。OMIM 記号は、 遺伝子記号と違うものがあるので HUGO (Human Genome Organization) Ø

Nomenclature Committee の遺伝子名のデータ ベースで 検索と確認を行った。 CancerProView に収録済みの遺伝子については、癌疾患と原因遺伝子の対応付けがされているのでそのまま使用した。遺伝子記号、疾患名の表の作成を行った。

(2) 癌関連遺伝子のタンパク質ドメイン構造と遺伝子変異、癌疾患との対応付け

癌関連遺伝子が産生するタンパク質のドメイン構造を SMART により検索を行い、検索されたタンパク質ドメインについて遺伝子変異とそれによる疾患の対応を解析を行った。対応がついたものについては、遺伝子名、疾患名、ドメイン名の表の作成を行った。既に CancerProView に収録されているものについては、遺伝子、タンパク質ドメイン、疾患についてデータ抽出を行い、それらの一覧として表の作成を行った。

(3) 既知タンパク質・遺伝子相互作用、ドメイン構造、疾患の関連付けによる疾患予測モデル化

タンパク質・遺伝子相互作用について CancerProViewやPubmed 文献などの検索を行い、データを抽出する。タンパク質・遺伝子相互作用から同じ系内にある遺伝子で、遺伝子変異による疾患が判明しているものについて、(2)で行ったタンパク質ドメインと疾患の対応付けの結果を利用して、系ごとにタンパク質ドメインと疾患をわけた。タンパク質・遺伝子相互作用、遺伝子、タンパク質ドメイン、疾患の一覧である表を作成した。

- (4) 疾患未同定遺伝子のタンパク質相互作用とドメイン構造の相似性からの疾患予測疾患原因遺伝子として同定されていない遺伝子については、タンパク質のドメイン構造から類似のドメインを持つタンパク質の疾患についてリストアップし、同じドメイン内で遺伝子変異が起きた場合それを予測疾患とした。疾患未同定の遺伝子記号、タンパク質ドメイン名、類似のドメインを持つ遺伝子記号、疾患名の表を作成を行った。
- (5) 癌疾患予測検索システムと疾患パスウェイの作成 (CancerProView との統合化)

HUGO Gene Nomenclature Committee (HGNC) のデータベースで遺伝子記号の確認を行い ながら作成をした 1758 疾患の原因遺伝子/ タンパク質のリストから遺伝子記号統合や microRNA であることが判明したことにより 消滅した 5 遺伝子を除いた 1753 遺伝子分の タンパク質のうち 796 遺伝子について SMART によりタンパク質ドメイン検索とドメイン 情報の取得を行い 1307 タンパク質について ドメイン情報を得ることができた。その結果 SMART ドメインについては、疾患との関連が 不明なものを含めて、1838 遺伝子、3259 タ ンパク質についてドメイン情報が使用にな った。その後、SMART に加えて NCBI の Refseq に記述されている 17456 遺伝子、遺伝子の転 写の多様性を含む 32354 タンパク質ドメイン のデータをデータベース上に取り込んだ。そ の後、タンパク質ドメインの詳細表示の改良を行い、SMARTドメインと NCBI の Refseqに記述されているタンパク質ドメインの両方を表示できるプログラムの開発を行った。さらに CancerProView中の疾患未同定遺伝子/タンパク質について SMART と NCBI Refseqの両方のタンパク質ドメインを使いタンパク質ドメインの比較を行うことにより疾患の推測を行うプログラムの開発を行いトップメニューの検索部分に追加を行った。

- (6) 癌関連疾患に関するタンパク質/遺伝 子相互作用図への疾患を予測ボタンの追加
- (5)で作成した CancerProView 中の疾患未同定な遺伝子/タンパク質について SMART, NCBI タンパク質ドメインの比較を行うことにより疾患の可能性を推測するプログラムの開発を行った。その後、癌関連疾患に関するタンパク質/遺伝子相互作用図について既存の 109 枚の図とそれから派生した 29 枚の図の合計 138 枚ついて疾患を推測するボタンの追加を行った。その後さらに 1 枚の相互作用図の追加を行った。
- (7) 疾患パスウェイと癌疾患予測システムのインターネット公開

本研究による研究成果をインターネット上に公開するためにネットワークの設定をサーバーに行った。

## 4. 研究成果

(1) タンパク質/遺伝子相互作用図と疾患パスウェイの表示

CancerProView のタンパク質/遺伝子相互作用図を図 2 に示す。図 2 は、(a)遺伝形質、(b)疾患名、(c)文献名、(d)遺伝子記号であり、(a)、(b)、(c)、(d)はマウスでクリックするそれぞれ(a)と(b)は 0MIM、(c)は Pubmed、(d)は Pubmed、(d)は Pubmed、(d)は Pubmed、Canceript の Puriant 表示にリンクしている。これらに加えてタンパク質ドメインの比較により予測された疾患を表示するためにボタン(PD) (e) を付け加えた。

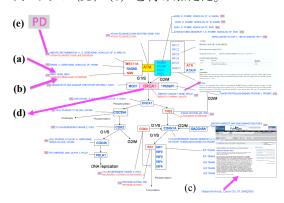


図2 タンパク質/遺伝子相互作用図

(2) タンパク質/遺伝子相互作用図からの予測された疾患の検索結果

図 2-e をクリックすると(a)遺伝子記号、(b) タンパク質ドメイン、(c) タンパク質ドメ

インと関連のある疾患名(予測疾患)の一覧が表示される(図3)。

Search Gene Symbol : MDC1

#### Disease Prediction List

Gene Symbol	Protein Domain	Disease Name
MD	Interaction	ATELOSTEOGENESIS, TYPE I
м <b>р</b> (а)	Interaction (b)	ATELOSTEOGENESIS, TYPE III (C)
MDC1	Interaction with the	BOOMERANG DYSPLASIA
MDC1	Interaction with the	IMMUNODEFICIENCY-CENTROMERIC INSTABILITY-FACIAL ANOMALIES SYNDROME
MDC1	Interaction with the	LARSEN SYNDROME, AUTOSOMAL DOMINANT
MDC1	Interaction with the	SPONDYLOCARPOTARSAL SYNOSTOSIS SYNDROME
MDC1	FHA	ATAXIA, ADULT-ONSET, WITH OCULOMOTOR APRAXIA
MDC1	FHA	ATAXIA, EARLY-ONSET, WITH OCULOMOTOR APRAXIA AND HYPOALBUMINEMIA
MDC1	FHA	BREAST CANCER, SUSCEPTIBILITY TO
MDC1	FHA	CHARCOT-MARIE-TOOTH DISEASE, AXONAL, TYPE 2A1
MDC1	FHA	LEUKEMIA, ACUTE LYMPHOBLASTIC
MDC1	FHA	LI-FRAUMENI SYNDROME 2
MDC1	FHA	NEUROBLASTOMA, SUSCEPTIBILITY TO, 1
MDC1	FHA	NIJMEGEN BREAKAGE SYNDROME
MDC1	FHA	OSTEOSARCOMA, SOMATIC
MDC1	FHA	PROSTATE CANCER, SOMATIC
MDC1	FHA	PROSTATE CANCER, SUSCEPTIBILITY TO
MDC1	Required for nuclear	ERYTHROCYTOSIS, FAMILIAL, 3
MDC1	Required for nuclear	FETAL HEMOGLOBIN QUANTITATIVE TRAIT LOCUS 5
MDC1	Required for nuclear	GONADAL DYSGENESIS, PARTIAL
MDC1	Required for nuclear	GONADAL DYSGENESIS, XY FEMALE TYPE
MDC1	Required for nuclear	SEX REVERSAL WITH PARTIAL OVARIAN FUNCTION

図3 タンパク質ドメインと関連疾患の一覧 (予測疾患一覧)(例:MDC1)

(3)遺伝子記号、タンパク質ドメイン名からの予測疾患検索と結果の表示

遺伝子記号、タンパク質ドメイン名から予測された疾患の検索を行えるように CancerProViewのトップメニューに検索ウィンドウを作成した(図 4)。(a)は遺伝記号からの検索、(b)はタンパク質ドメイン名からの検索のためのウィンドウである。

#### ♥ 5. Disease Search ♥

(a)	Gene Symbol:		Submit	Reset
(b)	Domain:	(e.g. TP53)		
		(e.g. BRCT)	Submit	Reset

図4遺伝子記号、タンパク質ドメイン名から の予測疾患検索

遺伝子記号からの検索は図 4-a のウィンドウに遺伝子記号を入力すると行える。その結果は(a)検索遺伝子の持つタンパク質ドメイン、(b)タンパク質ドメインと関連がある疾患の一覧が表示される(図 5)。

Search Gene Symbol : TP53

Disease Prediction List

Protein Domain	Disease Name		
Inten th HRM	ADRENOCORTICAL ^ * * * * * * * * * * * * * * * * * *		
Inten (a) th HRM	BREAST CANCER (b)		
Interaction with HRM	CHOROID PLEXUS PAPILLOMA		
Interaction with HRM	COLON TUMORS, CONCURRENT MULTIPLE PRIMARY		
Interaction with HRM	HEPATOBLASTOMA		
Interaction with HRM	HEPATOCELLULAR CARCINOMA		
Interaction with HRM	HISTIOCYTOMA, MALIGNANT FIBROUS		
Interaction with HRM	LI-FRAUMENI SYNDROME		
Interaction with HRM	LI-FRAUMENI SYNDROME 1		
Interaction with HRM	MULTIPLE MALIGNANCY SYNDROME		
Interaction with HRM	NASOPHARYNGEAL CARCINOMA		
Interaction with HRM	OSTEOSARCOMA		
Interaction with HRM	PANCREATIC CANCER		
Interaction with HRM	TP53 POLYMORPHISM		
Transcription activa	ADRENOCORTICAL CARCINOMA, PEDIATRIC		
Transcription activa	BREAST CANCER		
Transcription activa	CHOROID PLEXUS PAPILLOMA		
Transcription activa	COLON TUMORS, CONCURRENT MULTIPLE PRIMARY		
Transcription activa	HEPATOBLASTOMA		
Transcription activa	HEPATOCELLULAR CARCINOMA		

図 5 遺伝子記号からの予測疾患検索 (例:TP53)

タンパク質ドメイン名からの検索は図 4-b のウィンドウにタンパク質ドメイン名を入力すると行える。その結果は(a)検索タンパク質ドメインを持つ遺伝子の遺伝子記号、(b)関連がある疾患の一覧が表示される(図 6)。

Search Protein Domain : BRCT

Disease Prediction List

Gene Symbol	Disease Name		
В	BREAST CANCER		
Б (a) <sup>—</sup>	BREAST CANCER, SU (b) LITY TO		
BARDI	BREAST-OVARIAN CANCER		
BARD1	CONGENITAL CATARACTS, FACIAL DYSMORPHISM, AND NEUROPATHY		
BARD1	LEUKEMIA, ACUTE LYMPHOBLASTIC		
BARD1	LIG4 SYNDROME		
BARD1	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 1		
BARD1	MULTIPLE MYELOMA, RESISTANCE TO		
BARD1	NIJMEGEN BREAKAGE SYNDROME		
BARD1	OVARIAN CANCER, SPORADIC		
BARD1	PREMATURE CHROMOSOME CONDENSATION WITH MICROCEPHALY AND MENTAL RETARDATION		
BARD1	SEVERE COMBINED IMMUNODEFICIENCY WITH SENSITIVITY TO IONIZING RADIATION		
BRCA1	BREAST CANCER		
BRCA1	BREAST CANCER, SUSCEPTIBILITY TO		
BRCA1	BREAST-OVARIAN CANCER		
BRCA1	CONGENITAL CATARACTS, FACIAL DYSMORPHISM, AND NEUROPATHY		
BRCA1	LEUKEMIA, ACUTE LYMPHOBLASTIC		
BRCA1	LIG4 SYNDROME		
BRCA1	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 1		
BRCA1	MULTIPLE MYELOMA, RESISTANCE TO		

図6 タンパク質ドメイン名からの予測疾患 検索(例:BRCT)

(4) タンパク質ドメイン表示の改良 タンパク質のドメイン表示で SMART と NCBI Refseq のデータの両方を表示させるように 改良を行った(図7)。

Data Search
Result
Click on the gene symbol to the display of gene mutaion and polymorphisms.

	SWART FOREIGNAME					
(a)	Gene Symbol	Refseq ID	Protein ID	Protein Name	SMART Protein domain Structure	
	BRCA1	NM 007294	NP 009225	ausceptionity protein isororm i	• • • •	
				breast cancer type 1 susceptibility protein isoform 2	+	
				breast cancer type 1 susceptibility protein isoform 3		
				breast cancer type 1 susceptibility protein isoform 4	+	
	BRCA1	NM 007299	NP 009230	breast cancer type 1 susceptibility protein isoform 5	+	

The state of the s

図 7 タンパク質ドメイン表示 (例:BRCA1) (a) SMART、(b) NCBI Refseq

(5) タンパク質ドメイン詳細表示の改良 タンパク質ドメイン詳細表示(図8)で(a) タンパク質ドメインとアミノ酸配列を対比 できるようにアミノ酸配列のスケールを表 示できようにした。(b)はアミノ酸とcDNAの 表示の切り替えボタンである。(c)タンパク 質ドメイン、(d)はタンパク質ドメインのア ミノ酸配列上の位置、(e)はアミノ酸配列を 表示している。

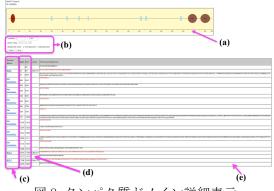


図8 タンパク質ドメイン詳細表示 (例:BRCA1、NP009225)

図 8-b で行えるタンパク質ドメイン表示のアミノ酸スケール表示(図 9-a)はアミノ酸のみ(図 9-b)とアミノ酸と cDNA の併記が行える(図 9-c)。さらにタンパク質の拡大表示も行える(図 9-d)。

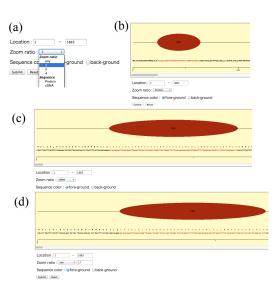


図 9 タンパク質ドメイン詳細表示の アミノ酸スケールの切り替え (例:BRCA1、NP009225)

### (6) 収録遺伝子数

本研究の結果、CancerProView に以前に集めた分を含めて表 1 のように収録することができた。

表 1 CancerProViewに収録済のデータ数

	収録数
遺伝子数	1,838 (SMART), 17,456 (NCBI)
タンパク質/cDNA数	3,259 (SMART), 32,354 (NCBI)
タンパク質/遺伝子相互作用図	139

### (7) インターネット公開

現在、テスト版( $\beta$ バージョン)として http://cancerproview. dmb. med. keio. ac. jp をアクセスすることにより閲覧できる。

#### (8) 研究総括

本研究の研究成果は、CancerProViewをアクセスすることで閲覧することができる。本研究の結果と意義は、遺伝子変異から癌関連疾患が予測できる可能性がある点にある。また、タンパク質/遺伝子相互作用図にタンパク質ドメイン比較による疾患予測ポタンをつけたことにより癌関連疾患遺伝子の予測ができ、疾患との関係(疾患パスウェイ)がわかる。それらにより診断、予防、予後の検討の役に立つと考えられる。

## 5. 主な発表論文等 〔雑誌論文〕(計 1件)

① <u>Susumu Mitsuyama</u>, Nobuyoshi Shimizu, CancerProView: a graphical image database of cancer-related genes and proteins., *Genomics*, 查読有, 100(2012), 81-92, 10.1016/j.ygeno. 2012.05.011

### 〔学会発表〕(計 7件)

- ① 満山 進, 清水信義, がん関連タンパク質 /遺伝子データベース CancerProView, 第37回日本分子生物学会年会, 2014年11 月25日, パシフィコ横浜(神奈川県・横 浜市)
- ② 満山 進,清水信義, CancerProView: がん関連疾患遺伝子/タンパク質相互作用データベース,生命医薬情報学連合大会2014,2014年10月2日~10月4日,仙台国際センター(宮城県・仙台市)
- ③ <u>満山 進</u>, 清水信義, CancerProView のさらなる機能拡張, 第73回日本癌学会学術総会, 2014年9月25日, パシフィコ横浜(神奈川県・横浜市)
- ④ <u>Susumu Mitsuyama</u>, Nobuyoshi Shimizu, New features of cancer-related protein / gene database "CancerProView",第 36 回日本分子生物学会年会,2013年12月4日,神戸ポートアイランド(兵庫県・神戸市)
- ⑤ <u>Susumu Mitsuyama</u>, Nobuyoshi Shimizu, CancerProView: The cancer-related gene/protein interaction pathway image database system. 第72回日本癌学会学術 総会, 2013年10月5日, パシフィコ横浜 (神奈川県・横浜市)
- ⑥ 満山 進, 清水信義, がん関連遺伝子/タンパク質相互作用データベース CancerProView, 第 20 回日本遺伝子診療 学会, 2013年7月20日,アクトシティ浜 松 コングレスセンター(静岡県静岡市)
- ⑦ Susumu Mitsuyama, Nobuyoshi Shimizu, Comparative analysis of the functional domains of various cancer-related proteins by CancerProView, 第35回日本分子生物学会年会,2012年12月12日,福岡国際会議場・マリンメッセ福岡(福岡県・福岡市)

# 〔図書〕(計 0件)

〔産業財産権〕

- ○出願状況(計 0件)
- ○取得状況(計 0件)

[その他]

ホームページ

http://cancerproview.dmb.med.keio.ac.jp

6. 研究組織

(1)研究代表者

満山 進 (MITSUYAMA, Susumu)

慶應義塾大学・医学部・特任助教

研究者番号:30296727