

## 仕様書

1. 件名：ヒト疾患および体質個人差に関する遺伝子変異・多型及び症例の病状、体質、生活習慣の情報の入力
2. 概要：ヒト遺伝子疾患変異研究会データベース作成委員会では、遺伝子疾患データベース *MutationView* を開発しインターネットで公開してきた (<http://mutview.dmb.med.keio.ac.jp>)。本業務は、*MutationView* のデータ拡充を図り、ヒト疾患および体質個人差に関するゲノム多様性情報の統合データベース (HGV-DB)を作成するために、ヒト疾患および体質個人差に関する遺伝子変異・多型、遺伝子の表現型、疾患の症状と変異の相関などのデータ入力を行うものである。

補記：本データベースの性格上、昨今、迅速・大量にデータが産出されている次世代 DNA シーケンサーによる塩基配列データのゲノム上へのマッピング情報や変異検索情報などの実験データのデータベース構築を業務内容とするものではなく、発表済み論文からの文字データとしての変異情報および疾患の臨床情報を抽出して、*MutationView* で表示可能な形式に変換し、格納することが作業内容となる。そのために、ゲノムについてのみならず、遺伝子疾患・症状についての情報を対象として取り扱うものである。

3. 数量：1,550 件

### 4. 作業内容

#### 4.1. 提供データ

請負業者には、キュレーターがヒト遺伝子に関する情報を記載した図表や医学論文から、変異に関する記述と、症例の症状、体質、生活習慣に関する記述を抽出し Excel シート等に転記したデータを提供する。

#### (1) 変異に関する記述

- ・ 変異名
- ・ ゲノム上および cDNA 上での変異・多型の位置
- ・ DNA の変化の実体
- ・ その他：上記以外に、レファレンスとなる遺伝子全配列、変異があるタンパク機能ドメイン、などの情報が含まれる。また、変異名が、アミノ酸に基づいたものしか記されて

いない場合は、塩基変化に基づく名称を、追加する必要がある。逆（塩基変化に基づく名称のみの場合は、アミノ酸変化に基づく名称を補完）も同様となる。

(2) 症状、体質、生活習慣に関する記述

- ・ その変異を保持する症例／個体の疾患名あるいは個人差のタイプ
- ・ 症状／個人差の内容記
- ・ 年齢・性別・人種等の情報
- ・ その他：上記以外に、文献上の患者 ID や、収載されていた文献情報、家系性の有無、zygosity、近親婚の有無、発症年齢などの情報、が含まれる。さらに、症状に関する具体的な記述が対象となるが、何が含まれるか…については疾患によって異なる。通常、その疾患についての検査データや、ある症状の有無や程度のような多岐に渡るデータが含まれる。複数症状を呈する疾患の場合は、それぞれの症状について、入力することが必要となる。

4.2. Mutation View データの入力

(1) 変異基本情報の作成作業

請負業者は、提供データ「4.1. (1) 変異に関する記述」に対して、定型の記述（※）が為されているものに関しては、Mutation View データ変換プログラム（既存のプログラム）を実行するものとする。このプログラム実行過程で、以下の変異基本情報が作成される（テキストファイルとして出力される）。

補記：※定型の記述とは、変異の遺伝子上の位置、変化前／後の塩基およびアミノ酸などが明記されているものを指す。キュレーターから抽出される転記データは、文献の記載のままなので定型の記述が完備していない場合もある。

・ 変異種別

種別	記号
Nonsense	N
Missense	M
Deletion	D
Frameshift Deletion	FD
Large Deletion	LD
Insertion	I
Frameshift Insertion	FI
Large Insertion	LI
Indel	ID
Frameshift Indel	FID

種別	記号
Large Indel	LID
Duplication	DU
Triplet Repeat	TR
Translocation	TL
Splicing	S
Regulation	R
Promoter	PR
Polymorphism	P
Complex	C
SNP	SNP

- ・ 変異位置
- ・ 欠失する塩基数（塩基置換の場合は、置換する塩基数）
- ・ 挿入する塩基（塩基置換の場合は、変化後の塩基）
- ・ *MutationView* 規定変異名 の例

#### 1) DNA 変化に基づく変異名の命名

##### Substitution

c.512A>G

c. [変異位置] [変異前の塩基] > [変異後の塩基]

##### Deletion

c.535delG

c. [欠失開始位置の塩基番号] \_ [欠失終了位置の塩基番号※] del [欠失塩基※]

※欠失塩基が 5bp 以上の場合塩基数。欠失塩基が 1bp の場合、欠失終了位置の塩基番号は省略。

##### Insertion

c.124\_125insCC

c. [挿入される位置の前の塩基番号] \_ [挿入される位置の後の塩基番号] ins [挿入塩基※]

※挿入塩基が 5bp 以上の場合塩基数。

##### Indel

c.168\_169delinsGGT

[欠失開始位置の塩基番号] \_ [欠失終了位置の塩基番号※] delins [挿入塩基※]

※挿入塩基が 5bp 以上の場合塩基数。欠失塩基が 1bp の場合は、欠失終了位置の塩基番号は省略

##### Polymorphism、SNP

c.1041T>C

c. [変異位置] [変異前の塩基] > [変異後の塩基]

##### Large Deletion

DelEx3-4

DelEx3

※nomenclature では、塩基位置で Deletion（上述）と同様に記述することとしているが、現在の MutationView データでは、削除範囲の明瞭化のために、エキソン記述による表現法を用いている。

### Splicing

c.607+1G>A

c. [変異位置] [変異前の塩基] > [変異後の塩基]

※ここでは塩基置換型を挙げる。Deletion 等によるものは、該当の記述法に従う。

### 2) アミノ酸変化に基づく変異名命名

#### Missense/Nonsense

p.Gln368X

p. [変異前のアミノ酸] [コドン番号] [変異後のアミノ酸]

#### Deletion

p.Gln8del

p. [変異前のアミノ酸] [コドン番号] del

#### Frameshift Deletion

p.Arg8GlyfsX16

p. [変異前のアミノ酸] [コドン番号] [変異後のアミノ酸] fsX [停止コドンとなるアミノ酸の相対位置]

### ・ 変異の結果 の例

#### Missense

AGC-CGC at ntd 2362 -> Ser-Arg at codon 788

#### Nonsense

CGA-TGA at ntd 2413 -> Arg-Stop at codon 805

#### Splicing

G-T at ntd 902-1(IVS9)

G-T at ntd 5971(Ex42)

### **Deletion**

6-bp del GAGACA at ntd 8710

### **Frameshift Deletion**

4-bp del AAAG at ntd 1024-> frameshift at codon 343

### **Insertion**

3-bp ins GAA at ntd 1021

### **Frameshift Insertion**

4-bp ins GACA at ntd 1142 -> frameshift at codon 381

### **Indel**

GAA-CCC at ntd 7793

### **Frameshift Indel**

GAAA-AT at ntd 8978 -> frameshift at codon 2993

- ・ 国際学会規定変異名 (<http://www.hgvs.org/mutnomen/> 参照)

### (2) データ入力作業

請負業者は、(1) の処理で得られた変異基本情報を、提供データ「4.1. (2) 症状に関する記述」と併せて、変異や多型を有する症例（個体）あるいは家系ごとに、*MutationView* 既定の様式の表に入力するものとする。この際、入力データの不備、矛盾及び国際学会規定に準じているか等を確認し、不明点があれば、キュレーターに照会するものとする。

なお、1 症例（あるいは 1 家系）の情報を 1 件として数えるものとする。

※補記1：キュレーターが転記したエクセルファイルのデータは、文献の記述の抽出であるため、本データベースの DBMS/表示システム *MutationView* の統計機能であるクラシファイ機能で統合・分類するために、*MutationView* に合わせた規格化が必要となる。例えば、(A) 検査値データなどの場合：症状（あるいは検査値）の重篤度の評価の付加、(B) 発症年齢などについても、Juvenile という表記のデータと、実年齢のデータが、データ元の論文により混在するので、評価データを付加する必要がある。(C) 症状記載についても、用語の統一が図られていない場合が多いため、統一化の実施、をおこなう必要がある。これらの必要性のために、請負業者には、仕様書 9. その他 (1) 応札条件で記す実績があることを求める。

※補記2：上で記したように、この症状についてのデータが最も多岐に渡り、データ量も膨大になる。また、他のデータベースにはない特有の機能（グラフィカルインターフェースによるクラシファイ）を有効に用いるためのデータとなり、このデータベース中での重要度も高い。そのため、データ構築のためには、医師やキュレーターのような特殊な専門性は求めないが、遺伝子疾患・症状についての相当の知識を有することが必要となる。

※補記3：1件あたりの平均情報量は、症例についての拡張情報（疾患についての検査値などの情報）の量によって千差万別であり、この程度という基準的なデータを示す事が困難（扱う疾患によっては、例示したものを遙かに超える分量となる可能性もある）で、業務を容易なものに見誤る可能性もあるので、ここでは示さないものとする。

#### 5. 納入期限

発注後五ヶ月以内とする。

#### 6. 納入場所

慶應義塾大学

#### 7. 納入方法

CD-ROM等の記憶媒体により納品すること。

#### 8. 検収

*MutationView*テストサーバにデータをアップロードし、*MutationView*アプレット上に入力データを表示し、表示の遺伝子構造上で位置情報が正しく入力されていることを確認した上で業務完了とする。

#### 9. その他

##### (1) 応札条件

本業務の受け入れは、ヒト遺伝学・ゲノム科学の基礎知識を有するスタッフを擁し、上記作業内容、特にヒトの遺伝子変異・多型の記載に関する国際学会での推奨ルールをよく理解でき、支障なくデータ入力可能な業者に限定することとする。それを証明するために、以下の業務実績を提出すること。

- ・ バイオインフォマティクスの業務実績
- ・ 遺伝子疾患データベースデータ入力業務実績
- ・ 遺伝子疾患に関連する学会（日本人類遺伝学会、日本遺伝子診療学会、日本遺伝子治療学会、日本分子生物学会等）における遺伝子疾患データベースに関する発表実績
- ・ 業務担当者は医学・薬学・生物学等の大学院修士以上の学歴を持ち遺伝学・分子生物

学の基礎知識を持つこと。

補記：業務実績、に関しては、請負業務の中のものであっても差し支えない。発表実績、に関しては、当該発表に、主たる著者または共著者として記されている必要がある。

項目（１）内の前２項目は、データ処理の実績を問うものであるが、後２項目は、キュレーターから抽出・転記されたデータを理解・評価する際に必要な学問的バックグラウンドを有するかを間接的に問うためのものであるため、本項に示した判断とする

## （２）特記

- ・ 入力したデータの著作権は当塾に帰属することとする。
- ・ 提供したデータの、他の目的での使用、無断公開、第三者への提供を行わないこと。
- ・ 提供したデータ及び入力したデータはセキュリティ措置を講じ、情報漏えい、不正アクセスから保護すること。

以上