

学薬品・薬剤の機能」の3つで構成されている。「化学薬品・薬剤機能」は、「薬理物質」「化学修飾」「臨床試験要素」「生理学的調整因子」「色素沈着」「食物・食品」「薬物乱用」「試薬」「産業生成物」「競争阻害薬」「構造類似体」に分類されていた。また、「薬剤・化学薬品構造」は、大きく「有機化合物」「非有機化合物」に分かれており、「有機化合物」は、「アミノ酸蛋白質」「芳香環化合物」「脂質」「有機金属・隣・ホウ素誘導体」「炭化水素」「炭水化物」「核酸・ヌクレオチド・ヌクレオシド」「マレイン酸」「メシル酸」「有機オキソ酸」「アルカロイド」「アルコール化合物」「ラクトン化合物」「メチルエーテル」「有機塩素化合物」と is-a 関係で構成されていた。非有機化合物質としては、「元素・同位元素」「フリーラジカル」「イオン」「塩」「鉱物」「電解質」「臭化シアン」で構成されていた。例えば「クルクリン」や「フラボノイド」は、「バイオフィラボノイド」と is-a 関係にあり「バイオフィラボノイド」は「フラボノイド」と is-a 関係にあり、「フラボノイド」は「栄養製剤」と is-a 関係にあり、「栄養製剤」は「食品構成要素」と is-a 関係にある。この関係を基に検索やデータ情報の依存性を加味する機能を付け加えた。

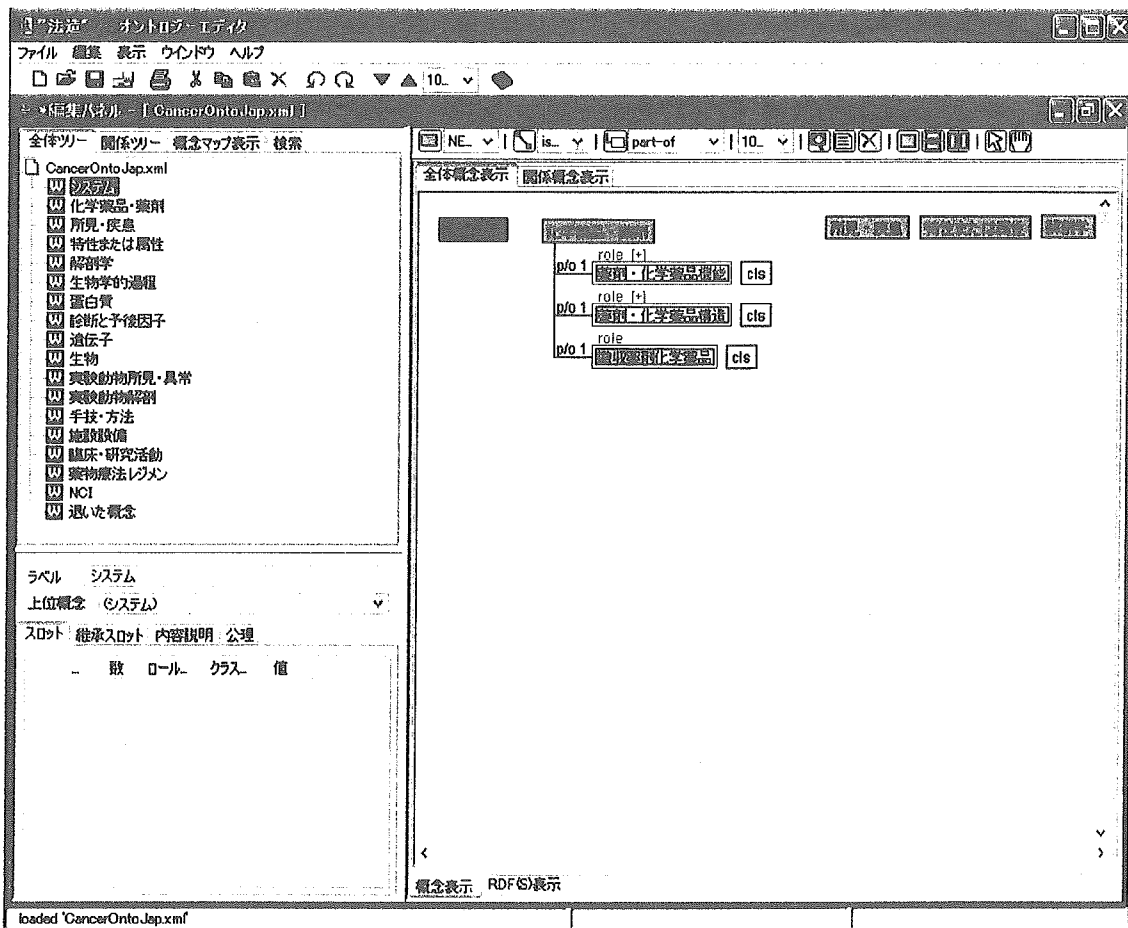


図 III-1.3. オントロジーエディタ「法造」を用いたがん化学薬品・薬剤オントロジーの日本語化。

「法道」 - オントロギーエディタ

ファイル 編集 表示 ウィンドウ ヘルプ

検索

全体マッピング 関係マッピング

CancerOntoJep.xml

- 薬学
- 化学薬品・薬剤
- 所見・疾患
- 特性または属性
- 解剖学
- 生物学的過程
- 蛋白質
- 診断と予後因子
- 遺伝子
- 生物
- 実験動物所見・異常
- 実験動物解剖
- 手技・方法
- 施設設備
- 臨床・研究活動
- 薬物療法レジメン
- NCI
- 選いた概念

ロール概念
クラス制約
ロールホルダー
全体ロール 内容説明 公理
部分ロール 基本ロール

全体概念表示 関係概念表示

システム 化学薬品・薬剤 所見・疾患 特性または属性 解剖学 生物学的過程 遺伝

role

p/o 1 薬剤・化学薬品精造 c/s

role [+]

p/o 1 薬理物質 c/s

role

p/o 1 化学傷筋 c/s

role

p/o 1 毒物 c/s

role

p/o 1 アルキル化剤 c/s

role

p/o 1 腫瘍原性物質 c/s

role

p/o 1 腫瘍プロモーター c/s

role

p/o 1 腫瘍形成因子 c/s

role

a/o 1 癌原性物質 c/s

role

p/o 1 診断試験薬素 c/s

role

p/o 1 生理学的調節因子 c/s

role

p/o 1 色素沈着 c/s

role [+]

p/o 1 食物・食品 c/s

role

p/o 1 薬物利用 c/s

role

p/o 1 医薬品 c/s

role

p/o 1 薬生成物 c/s

role

p/o 1 薬物相互作用 c/s

role

p/o 1 遺伝子発現 c/s

role [+]

p/o 1 薬剤・化学薬品精造 c/s

role

p/o 1 添加剤化学薬品 c/s

概念表示 RDF(S)表示

loaded 'CancerOntoJep.xml'

図 III-1.4. 化学薬品・薬剤の機能についての概念構成.

```

<?xml version="1.0" encoding="Shift_JIS"?>
<rdf:RDF xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#">

<rdfs:Class rdf:ID="WholenessConcept">
<rdfs:label>WholenessConcept</rdfs:label>
</rdfs:Class>
<rdfs:Class rdf:ID="RelationalConcept">
<rdfs:label>RelationalConcept</rdfs:label>
</rdfs:Class>
<rdfs:Property rdf:ID="hasPart">
<rdfs:label>hasPart</rdfs:label>
</rdfs:Property>
<rdfs:Property rdf:ID="hasAttribute">
<rdfs:label>hasPart</rdfs:label>
</rdfs:Property>
<rdfs:Class rdf:ID="システム">
<rdfs:label>システム</rdfs:label>
<rdfs:subClassOf rdf:resource="#WholenessConcept" />
<rdfs:comment>
Def:
Axiom:
</rdfs:comment>
</rdfs:Class>
<rdfs:Class rdf:ID="化学薬品・薬剤">
<rdfs:label>化学薬品・薬剤</rdfs:label>
<rdfs:subClassOf rdf:resource="#WholenessConcept" />
<rdfs:comment>
Def:
Axiom:
</rdfs:comment>
</rdfs:Class>
<rdf:Property rdf:about="化学薬品・薬剤_hasPart_as_role">
<rdfs:label>化学薬品・薬剤_hasPart_as_role</rdfs:label>
<rdfs:subPropertyOf rdf:resource="#hasPart"/>
<rdfs:domain rdf:resource="#化学薬品・薬剤"/>
<rdfs:range rdf:resource="#cls"/>
</rdf:Property>
<rdf:Property rdf:about="化学薬品・薬剤_hasPart_as_role">
<rdfs:label>化学薬品・薬剤_hasPart_as_role</rdfs:label>
<rdfs:subPropertyOf rdf:resource="#hasPart"/>
<rdfs:domain rdf:resource="#化学薬品・薬剤"/>
<rdfs:range rdf:resource="#cls"/>
</rdf:Property>
<rdf:Property rdf:about="化学薬品・薬剤_hasPart_as_role">
<rdfs:label>化学薬品・薬剤_hasPart_as_role</rdfs:label>
<rdfs:subPropertyOf rdf:resource="#hasPart"/>
<rdfs:domain rdf:resource="#化学薬品・薬剤"/>
<rdfs:range rdf:resource="#cls"/>
</rdf:Property>
(中略)
Def:
Axiom:
</rdfs:comment>
</rdfs:Class>
</rdf:RDF>

```

図 III-1.5. RDF 形式で出力した場合.

D. 考察

がん予防ドメインに関する概念抽出を試み、NCI で作成されている **Cancer Ontology** 中の「化学薬品・薬剤」についてオントロジーエディタ「法造」を用いて日本語化を行った。

また、本研究では、主にごがん予防の行為に関するオントロジーについて構造化を試みた。個人のごがん予防の行為には、①発がんの環境の改善に関する行為、②生活習慣の改善に関する行為、③がん予防関連治療、④体質の改善に関する行為に大きく分けることができる。この行為をそれぞれさらに粒度を細かくしていった。がん予防に関する概念の特徴として食品や禁煙や節酒のような

生活習慣に関わる概念など NCI のオントロジーにない概念まで追加する必要があることがわかる。NCI では図 III-1.6. のようながんに関する情報を計算機可読とするための国家プロジェクトを進めている。本研究では、がん予防の中でも「行為」に関するドメインのみでのオントロジーの開発を行ったが、分担研究において後述されるように、英語と日本語の概念がことなることも少なくなく、日本語におけるがん予防に関するオントロジーの開発は、本研究を契機にさらに強力に進めていく必要があると思えた。

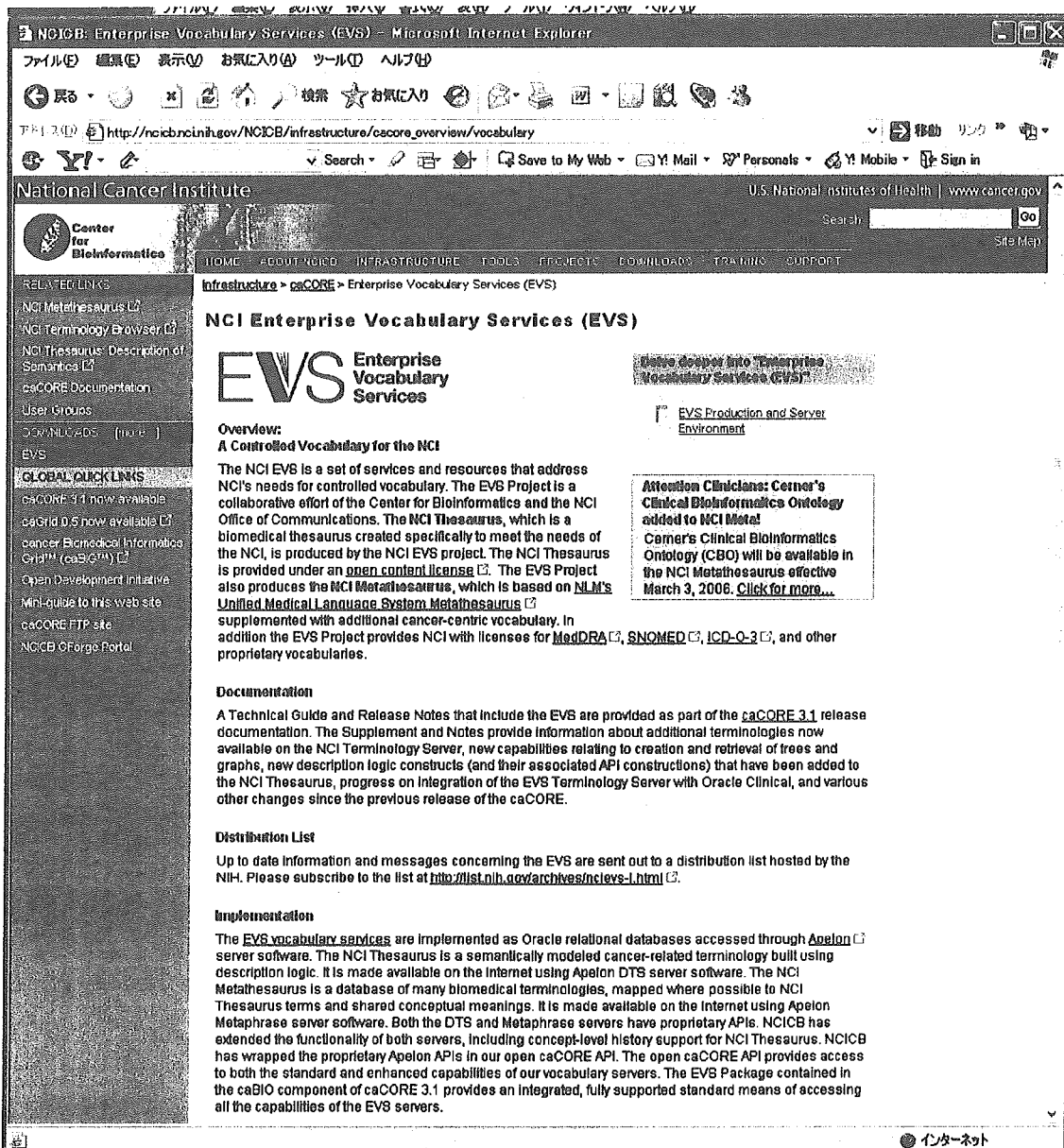


図 III-1.6. NCI の計算機可読な語彙開発に関するサイト。

参考文献:

[1] <http://www.cancer.gov/prevention/index.html>

- [2] 米国対がん協会 . 「がん」になってからの食事療法—米国対がん協会の最新ガイド. 法研. (2002/08) .
- [3] ウェンディ ウェイガー . がんの代替療法—有効性と安全性がわかる本 ハーバード大学の研究グループによる最新報告. 法研. (2004/03) .
- [4] 坪野 吉孝. 「がん」は予防できる. 講談社 (2004/01) .
- [5] 津金 昌一郎 . がんになる人 ならない人. 講談社 (2004/3/21) .
- [6] 溝口理一郎. オントロジー工学. オーム社 (2005 年) .

III-2. がん予防情報データベースの設計

A. 目的と背景

がん発生のメカニズムそのものがゲノム情報と深く関連しているため、がん予防研究におけるゲノム情報解析基盤の整備は必要不可欠のものとなっている。特に、がんの化学予防やワクチン開発、発がん物質の解毒に関与するp-450などの代謝酵素の遺伝子多型情報はがん予防と深く関連しているといわれている。今後、化学発癌情報や抗がん剤、治療情報からがん予防や薬剤治療効果に関連する酵素や受容体、イオンチャンネルとその関連遺伝子を探索を可能とするデータベースシステムはがん予防研究における情報基盤として極めて重要となってくると思われる。

近年ゲノム関連データベースの構築には莫大な投資が行われ、DDBJやSwissProt、JSNPのみならず PharmGKBのような薬理遺伝学情報も含め物名、代謝酵素名、遺伝子多型情報、有害事象、病名から構成された薬物動態関連情報データベースが整備され、インターネット上に公開されている。しかし、これらの公開データベースの多くは、内部的には関連データベース化されている可能性が高いが、外部ファイルとして公開されているものはフラットな主にXML形式である。つまり、がん予防に関連するデータを探索するためには、これらのおのおのインターネットサイトからファイルをダウンロードして、相互解析が可能なデータベースを設計しなければならない。

そこで本研究では、がん予防研究の支援を目的とした関連性解析等のデータマイニングを臨床情報と統合解析可能なデータベースの設計を試みた。

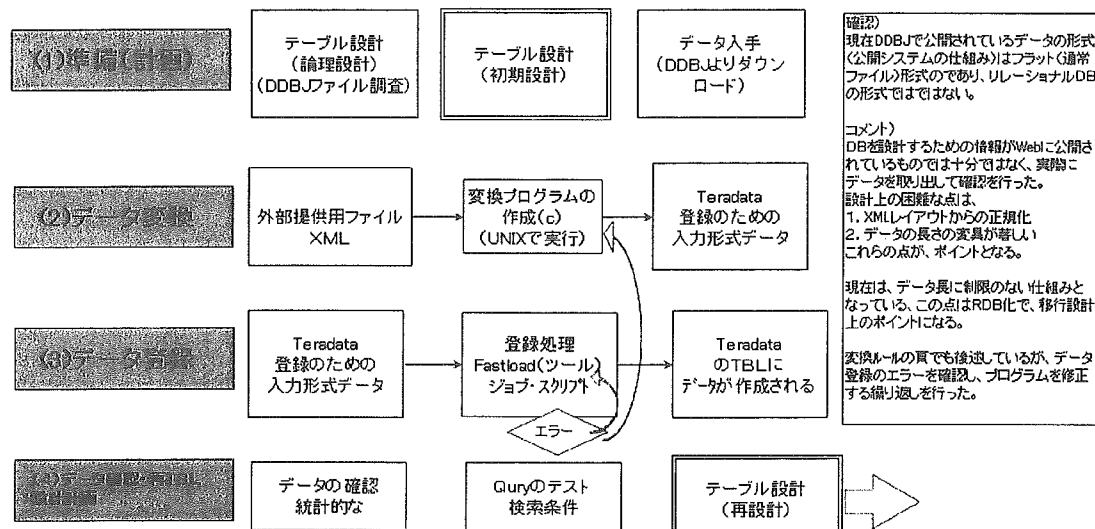
B. 方法

ゲノム関連情報データベースとして遺伝子塩基配列データベースとしてDDBJ(DNA Data Bank of Japan)、遺伝子多型情報データベースとしてJSNP、アミノ酸配列データベースとしてSwissProt、臨床情報との統合解析を可能とすることを目的として国立がんセンターで公開されている院内がん登録システムの臨床情報項目とのデータマイニングが可能なデータベースの論理設計を行い、TeraData(日本NCR社)に実装をおこない検証することとした。

C. 結果

1. DDBJ (<http://www.ddbj.nig.ac.jp/Welcome-j.html>)

- 開発フローを下記に示す。



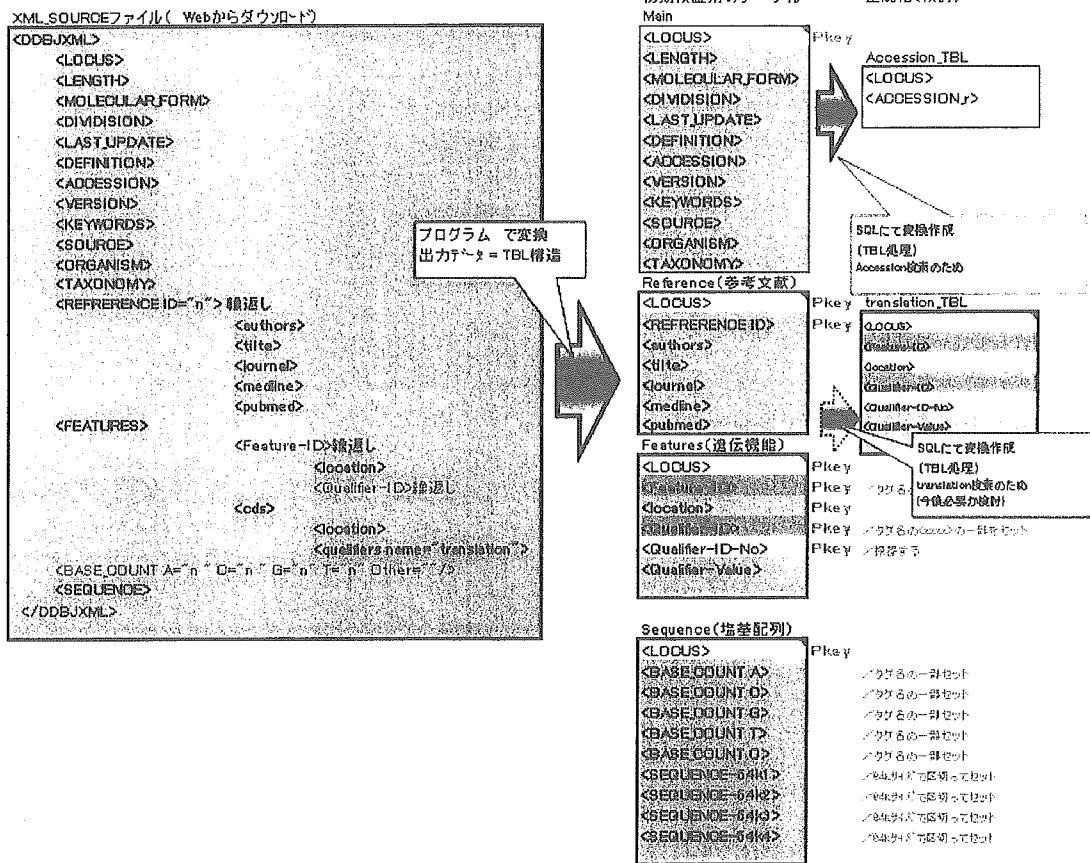
- 正規化前のテーブル構造を下記に示す。

正規化前	正規化後		
テーブル名称	テーブル名称	テーブルの説明	備考
DDBJ	Main_humT	遺伝子ID(Locus)に対する基本情報(長さ、キーワード、由来組織など) 1対1	
Fast形式など複数のフォームが 選択できるが、XML形式を検索	Reference_humT	参考文献に関する情報(文献著者、タイトル、ジャーナル名など) 1対1~3(ぐらい)	
XMLのタグ構造を解析し、プログラムを作成しCSV形式のテーブル 対応に変換した。	Features_humT	遺伝の機能に関する情報(source, site, peptide, cdsなど多数) 1対 n (通常20個程度、大量ケースありばらつきがある)	
生物種は人間(Human)のみを 登録している(レイアウトは他の 生物種も同じ。)	Sequence_humT	遺伝子の塩基配列(シーケンス)、64Kを越えて長いものも含まれる。 1対1	

- XML形式での外部提供ファイル構造を下記に示す。

外部提供ファイル XML 構造	
<?xml version="1.0" standalone="no" ?>	
<!DOCTYPE DDBJXML SYSTEM "DDBJXML.dtd">	
<DDBJXML>	
<LOCUS>	※ XMLファイルのタグに注目した範囲、一部、終りのタグを省略しています。
<LENGTH>	
<MOLECULAR_FORM>	
<DIVISION>	
<LAST_UPDATE>	
<DEFINITION>	
<ACCESSION>	
<VERSION>	
<KEYWORDS>	
<SOURCE>	
<ORGANISM>	
<TAXONOMY>	
<REFERENCE ID="1">注1 (1~3) * 繰り返し(出現)	
<authors>	
<title>	
<journal>	
<medline>	
<pubmed>	
</REFERENCE ID>	
<REFERENCE ID="2">	
</REFERENCE ID>	
<FEATURES>	
<xxxxxx>注2 Feature-ID(Feature 複数タイプあり)	
</xxxxxx>	
<source>	
<location>	</location>
<qualifiers name="xxxxxxx">注3 Qualifier-ID(Feature 複数タイプあり)	
</source>	
<sig_peptide>	
<location>	</location>
<cds>	
<location>	</location>
<qualifiers name="translation">注4 (この場合、ある程度長い文字列である。)	
</cds>	
</FEATURES>	
<BASE_COUNT A=" 340 " C=" 635 " G=" 565 " T=" 463 " Other="" />注5(4種の塩基の数が示されている。)	
<SEQUENCE>注5(塩基の配列、非常に長い場合がある、64k以上)	
</SEQUENCE>	
</DDBJXML>	

- データ変換に関する設計プロセスを下記に示す。



- TBL の定義内容を下記に示す。

Main	LOCUS	Index	CHAR	20	
	LENGTH		CHAR	80	
	MOLECULAR_FORM		CHAR	80	
	DIVISION		CHAR	900	
	LAST_UPDATE		CHAR	80	
	DEFINITION		CHAR	900	
	ACCESSION		CHAR	500	
	VERSION		CHAR	80	
	KEYWORDS		VARCHAR	700	
	SOURCE		CHAR	250	
	ORGANISM		CHAR	80	
	TAXONOMY		CHAR	200	
Reference	LOCUS	Index	CHAR	20	
	Reference ID	Index	CHAR	5	
	Authors_c		VARCHAR	3000	
	Title_c		VARCHAR	500	
	Journal_c		VARCHAR	3000	
	Medline		CHAR	80	
	Pubmed		CHAR	80	
Featuers	LOCUS	Index	CHAR	20	
	Feature_id	Index	CHAR	20	
	Location	Index	VARCHAR	4600	CHARACTER SET Unicode
	Qualifier_id	Index	CHAR	100	
	Qualifier_id_NO	Index	CHAR	20	
	Qualifier_value		VARCHAR	12000	CHARACTER SET Unicode
Translation_value		VARCHAR	12000	CHARACTER SET Unicode	
Sequence	LOCUS	Index	CHAR	20	
	Seq_No	Index	VARCHAR	6	
	Base_count		CHAR	80	
	Sequence_64k1		VARCHAR	64000	CHARACTER SET LATIN
	Sequence_64k2		VARCHAR	64000	CHARACTER SET LATIN
	Sequence_64k3		VARCHAR	64000	CHARACTER SET LATIN
	Sequence_64k4		VARCHAR	64000	CHARACTER SET LATIN

- Teradata へのデータの変換ルールについて下記に示す。

データ変換																							
TeradataへのTBL登録 用にファイルに変換する XMLソースを絞んで、タグ囲まれたデータ単位にまとめ処理を行い、ファイル出力する。																							
<p>(1) ファイル出力について</p> <p>1つのXMLのソースデータを、4つの データ・ファイルに分割する。</p> <table border="1"> <tr> <th>ソースファイル</th> <th>変換されたファイル</th> </tr> <tr> <td>XML (source)</td> <td>Main.txt</td> </tr> <tr> <td></td> <td>Reference.txt</td> </tr> <tr> <td></td> <td>Features.txt</td> </tr> <tr> <td></td> <td>Seaquence.txt</td> </tr> </table> <p>入力に対する、レコード数の対応は、次の通り</p> <table border="1"> <tr> <th>XML・source</th> <th>Main</th> <th>1対1</th> </tr> <tr> <td></td> <th>Reference</th> <td>1対1~3(ぐらい)</td> </tr> <tr> <td></td> <th>Features</th> <td>1対 n (大量ケースあり。ばらつきがある)</td> </tr> <tr> <td></td> <th>Seaquence</th> <td>1対1</td> </tr> </table>		ソースファイル	変換されたファイル	XML (source)	Main.txt		Reference.txt		Features.txt		Seaquence.txt	XML・source	Main	1対1		Reference	1対1~3(ぐらい)		Features	1対 n (大量ケースあり。ばらつきがある)		Seaquence	1対1
ソースファイル	変換されたファイル																						
XML (source)	Main.txt																						
	Reference.txt																						
	Features.txt																						
	Seaquence.txt																						
XML・source	Main	1対1																					
	Reference	1対1~3(ぐらい)																					
	Features	1対 n (大量ケースあり。ばらつきがある)																					
	Seaquence	1対1																					
<p>(2) 区切り文字</p> <p>Teradataのロードツールのためのデータ変換処理。全てのフィールドに区切り文字を挿入する。</p> <p>各フィールドに に設定 (縦棒) フィールドにデータが無い場合も、 を設定する。 の数を揃える。</p>																							
<p>(3) その他</p> <p>TeradataのVersion2R5.1までは、最大カラムサイズが64kのため、Sequence(植替配列)が64kを超えると対応できない。64kを超えるサイズの遺伝子データが含まれる場合があるので、この場合は64で分割した格納した(制限事項) Sequenceと transferのデータの中に改行が入っている場合があるが、これは取り除く処理を追加。</p>																							

- データ例について下記に示す。

(その1)

データ・サンプル

Main Table

LOCUS
LENGTH
MODEL
DIVISION
LAST UPDATE
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
TAXONOMY

MMLAG3
2003
linear
ROD
25-Oct-96
M.musculus mRNA for LAG-3-like cell surface glycoprotein.
X98113
X98113.1
cell surface glycoprotein; Ig superfamily; LAG-3 gene;transmembrane protein.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Crenata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Reference Table

LOCUS
REFERENCE ID
authors
title
journal
medline
pubmed

MMLAG3	MMLAG3	MMLAG3
1	2	3
Mastrandrea R, Micangeli E, and Donini S.	Mastrandrea R	Miyazaki T, Dierich A, Benoist C, and Mathis D.
Cloning of murine LAG-3 by magnetic beads bound homolog Submitted (23-MAY-1996) to the EMBL/Genl Science 272, 405-408(1996)		

(その2)

データ・サンプル

Main Table

LOCUS
LENGTH
MODEL
DIVISION
LAST UPDATE
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
TAXONOMY

MMLAG3
2003
linear
ROD
25-Oct-96
M.musculus mRNA for LAG-3-like cell surface glycoprotein.
X98113
X98113.1
cell surface glycoprotein; Ig superfamily; LAG-3 gene;transmembrane protein.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Crenata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Reference Table

LOCUS
REFERENCE ID
authors
title
journal
medline
pubmed

MMLAG3	MMLAG3	MMLAG3
1	2	3
Mastrandrea R, Micangeli E, and Donini S.	Mastrandrea R	Miyazaki T, Dierich A, Benoist C, and Mathis D.
Cloning of murine LAG-3 by magnetic beads bound homolog Submitted (23-MAY-1996) to the EMBL/Genl Science 272, 405-408(1996)		

• データ量

配布ファイル名	ソースデータサイズ(KB)				配布ファイル名	登録件数/処理時間				(Total)
	2,627	6,475	30,510	180,944		11,031	23,668	347628*er	11031*er	
%	1%	3%	13%	83%		7.37	1215	0		
						1,579	4,586	283652*er	1579*er	
%	0%	0%	0%	0%						
	389	1,520	17,530	205,735		1,804	5,401	274070*er	1804*er	
%	0%	1%	8%	91%		62	128	219	12,555	2,964
	866	3,474	17,532	204,226		4053	10192	267349*er	4053*er	
	12,440	26,434	57,185	135,322						
	11,795	30,414	36,791	141,391						
%	5%	14%	17%	64%						
	1,101	3,257	8,209	13,959		7,850	14,587	84,634	7,851	
%	4%	12%	31%	53%		348	565	1,233	491	2,317

• DDBJ human Teradata Fastload performance

Type	Rec Count	Start	End Time	Processed	Files	size(MB)	KB/Sec
Main	266,946				21	80	316
Reference	531,696				21	155	2,385
Feature	7,952,247	17:19:04	18:27:01		21	615	1,094
Sequence	309,012				21	4,036	5,544
Total	9,059,901			0:26:48		4,886	

DDBJのHumanは、データ量で4.8GB
 データ件数で、Main基本情報(約27万件)、Reference参考文献情報(約53万件)、Feature遺伝子機能情報(795万件)、Sequence(塩基配列情報)(約30万件)。
 全データの登録処理時間は、約27分であった。

Sequence							
Total Records Read		=	309012				
Total Error Table 1		=	0	----	Table has been dropped		
Total Error Table 2		=	0	----	Table has been dropped		
Total Inserts Applied		=	309012				
Total Duplicate Rows		=	0				
Start:	Fri Dec 05 20:23:20 2003						
End :	Fri Dec 05 20:35:28 2003						

2. JSNP (http://snp.ims.u-tokyo.ac.jp/index_ja.html)

- JSNP データの正規化前と正規化後のテーブル構造を下記に示す。

正規化前	正規化後			
テーブル名称	テーブル名称	テーブルの説明	備考	
JSNP Webサイトで公開されているXMLデータをMLデータの構造を解析し、RDB化設計を行った。CSV形式へ変換後、Teradataへローディングした。	contig_account	コンティグ		
	jst_allele_set	JST SNP アレル 情報		
	jst_exon_set	JST 遺伝子エキソン情報		
	jst_gene_set	JST 遺伝子情報		
	jst_snp_set	JST SNP 情報		
	jst_str_set	JST SNP 構造情報		
	ncbi_cds_set	NCBI 遺伝子CDS情報		
	ncbi_exon_set	NCBI 遺伝子エキソン情報		
	ncbi_gene_set	NCBI 遺伝子情報		
	ncbi_snp_set	NCBI SNP 情報		
	ncbi_utr3_set	NCBI 遺伝子UTR3 情報		
	snp_allele_set_screen	SNP アレル 情報		
	snp_screened	SNP スクリーニング 情報		
	snp_screened_gene_set	SNP 遺伝子情報		
	region_gene_set	領域遺伝子情報		
	region_gene_str_set	領域構造情報		
	region_set	領域情報		
method	メソッド情報			

- マッピング情報について下記に示す。

テーブル(日本語)	カラム名(日本語)	テーブル	カラム名
コンテイング		contig_account	contig_account (コンテイング)
JST SNPアレル情報		ist_allele_set	ist_allele_set (JST SNPアレル情報)
JST 遺伝子エキソン情報		ist_exon_set	ist_exon_set (JST 遺伝子エキソン情報)
JST 遺伝子情報		ist_gene_set	ist_gene_set (JST 遺伝子情報)
JST SNP情報		ist_snp_set	ist_snp_set (JST SNP情報)
JST SNP構造情報		ist_str_set	ist_str_set (JST SNP構造情報)
NCBI 遺伝子CDS情報		ncbi_cds_set	ncbi_cds_set (NCBI 遺伝子CDS情報)
NCBI 遺伝子エキソン情報		ncbi_exon_set	ncbi_exon_set (NCBI 遺伝子エキソン情報)
NCBI 遺伝子情報		ncbi_gene_set	ncbi_gene_set (NCBI 遺伝子情報)
NCBI SNP情報		ncbi_snp_set	ncbi_snp_set (NCBI SNP情報)
NCBI 遺伝子UTR3情報		ncbi_utr3_set	ncbi_utr3_set (NCBI 遺伝子UTR3情報)
コンテイング	バージョン	contig_account	contig_version
コンテイング	長さ	contig_account	contig_length
コンテイング	染色体	contig_account	chromosome
コンテイング	コンテイング末端	contig_account	contig_end_in_chr
コンテイング	Contig_Account	contig_account	contig_account
コンテイング	コンテイング先頭	contig_account	contig_start_in_chr
コンテイング	contig_direct_in_chr	contig_account	contig_direct_in_chr
JST SNPアレル情報	snp_allele_na_nuc	ist_allele_set	snp_allele_na_nuc
JST SNPアレル情報	snp_jd	ist_allele_set	snp_jd
JST 遺伝子エキソン情報	exon_end_in_contig	ist_exon_set	exon_end_in_contig
JST 遺伝子エキソン情報	exon_start_in_chr	ist_exon_set	exon_start_in_chr
JST 遺伝子エキソン情報	Gene_Account	ist_exon_set	gene_account
JST 遺伝子エキソン情報	exon_end_in_chr	ist_exon_set	exon_end_in_chr
JST 遺伝子エキソン情報	exon_start_in_contig	ist_exon_set	exon_start_in_contig
JST 遺伝子情報	gene_direct_in_chr	ist_gene_set	gene_direct_in_chr
JST 遺伝子情報	コンテイングバージョン	ist_gene_set	gene_contig_ver
JST 遺伝子情報	Gene_Account	ist_gene_set	gene_account
JST 遺伝子情報	遺伝子最終	ist_gene_set	gene_end_in_chr
JST 遺伝子情報	コンテイング開始	ist_gene_set	gene_start_in_contig
JST 遺伝子情報	Contig_Account	ist_gene_set	gene_contig_acc
JST 遺伝子情報	染色体	ist_gene_set	chromosome
JST 遺伝子情報	コンテイング最終	ist_gene_set	gene_end_in_contig
JST 遺伝子情報	遺伝子開始	ist_gene_set	gene_start_in_chr
JST 遺伝子情報	gene_direct_in_contig	ist_gene_set	gene_direct_in_contig
JST SNP情報	snp_direct_in_chr	ist_snp_set	snp_direct_in_chr
JST SNP情報	snp_primer_backward_seq	ist_snp_set	snp_primer_backward_seq
JST SNP情報	snp_primer_forward_seq	ist_snp_set	snp_primer_forward_seq
JST SNP情報	snp_direct_in_contig	ist_snp_set	snp_direct_in_contig
JST SNP情報	snp_start_in_chr	ist_snp_set	snp_start_in_chr
JST SNP情報	snp_type	ist_snp_set	snp_type
JST SNP情報	snp_end_in_chr	ist_snp_set	snp_end_in_chr
JST SNP情報	chromosome	ist_snp_set	chromosome
JST SNP情報	snp_multi_count	ist_snp_set	snp_multi_count
JST SNP情報	snp_contig_version	ist_snp_set	snp_contig_version
JST SNP情報	snp_jd	ist_snp_set	snp_jd
JST SNP情報	snp_linkout_url	ist_snp_set	snp_linkout_url
JST SNP情報	snp_3flank_seq	ist_snp_set	snp_3flank_seq
JST SNP情報	snp_contig_account	ist_snp_set	snp_contig_account
JST SNP情報	snp_end_in_contig	ist_snp_set	snp_end_in_contig
JST SNP情報	snp_5flank_seq	ist_snp_set	snp_5flank_seq
JST SNP情報	snp_start_in_contig	ist_snp_set	snp_start_in_contig
JST SNP情報	snp_primer_product_size	ist_snp_set	snp_primer_product_size
JST SNP構造情報	snp_str_gene_account	ist_str_set	snp_str_gene_account
JST SNP構造情報	snp_jd	ist_str_set	snp_jd
JST SNP構造情報	snp_str_gene_version	ist_str_set	snp_str_gene_version
JST SNP構造情報	snp_str_type	ist_str_set	snp_str_type
NCBI 遺伝子CDS情報	cds_start_in_chr	ncbi_cds_set	cds_start_in_chr
NCBI 遺伝子CDS情報	gene_account	ncbi_cds_set	gene_account
NCBI 遺伝子CDS情報	cds_start_in_contig	ncbi_cds_set	cds_start_in_contig
NCBI 遺伝子CDS情報	cds_end_in_contig	ncbi_cds_set	cds_end_in_contig
NCBI 遺伝子CDS情報	cds_end_in_chr	ncbi_cds_set	cds_end_in_chr
NCBI 遺伝子エキソン情報	exon_end_in_contig	ncbi_exon_set	exon_end_in_contig
NCBI 遺伝子エキソン情報	exon_end_in_chr	ncbi_exon_set	exon_end_in_chr
NCBI 遺伝子エキソン情報	exon_start_in_contig	ncbi_exon_set	exon_start_in_contig
NCBI 遺伝子エキソン情報	exon_start_in_chr	ncbi_exon_set	exon_start_in_chr
NCBI 遺伝子エキソン情報	gene_account	ncbi_exon_set	gene_account
NCBI 遺伝子情報	cds_na_seq	ncbi_gene_set	cds_na_seq
NCBI 遺伝子情報	gene_version	ncbi_gene_set	gene_version
NCBI 遺伝子情報	gene_direct_in_contig	ncbi_gene_set	gene_direct_in_contig
NCBI 遺伝子情報	gene_account	ncbi_gene_set	gene_account
NCBI 遺伝子情報	gene_symbol	ncbi_gene_set	gene_symbol
NCBI 遺伝子情報	gene_contig_acc	ncbi_gene_set	gene_contig_acc
NCBI 遺伝子情報	gene_start_in_contig	ncbi_gene_set	gene_start_in_contig

NCBI遺伝子情報	cds_aa_seq	ncbi_gene_set	cds_aa_seq
NCBI遺伝子情報	gene_direct_in_chr	ncbi_gene_set	gene_direct_in_chr
NCBI遺伝子情報	gene_end_in_chr	ncbi_gene_set	gene_end_in_chr
NCBI遺伝子情報	gene_end_in_contig	ncbi_gene_set	gene_end_in_contig
NCBI遺伝子情報	chromosome	ncbi_gene_set	chromosome
NCBI遺伝子情報	gene_location	ncbi_gene_set	gene_location
NCBI遺伝子情報	gene_start_in_chr	ncbi_gene_set	gene_start_in_chr
NCBI遺伝子情報	gene_contig_ver	ncbi_gene_set	gene_contig_ver
NCBI SNP情報	chromosome	ncbi_snp_set	chromosome
NCBI SNP情報	snp_end_in_chr	ncbi_snp_set	snp_end_in_chr
NCBI SNP情報	snp_start_in_contig	ncbi_snp_set	snp_start_in_contig
NCBI SNP情報	snp_id	ncbi_snp_set	snp_id
NCBI SNP情報	snp_contig_version	ncbi_snp_set	snp_contig_version
NCBI SNP情報	snp_start_in_chr	ncbi_snp_set	snp_start_in_chr
NCBI SNP情報	snp_contig_account	ncbi_snp_set	snp_contig_account
NCBI SNP情報	snp_end_in_contig	ncbi_snp_set	snp_end_in_contig
NCBI SNP情報	snp_linkout_url	ncbi_snp_set	snp_linkout_url
NCBI遺伝子UTR3情報	utr_3_start_in_chr	ncbi utr3_set	utr_3_start_in_chr
NCBI遺伝子UTR3情報	utr_3_end_in_contig	ncbi utr3_set	utr_3_end_in_contig
NCBI遺伝子UTR3情報	utr_3_end_in_chr	ncbi utr3_set	utr_3_end_in_chr
NCBI遺伝子UTR3情報	utr_3_start_in_contig	ncbi utr3_set	utr_3_start_in_contig
NCBI遺伝子UTR3情報	gene_account	ncbi utr3_set	gene_account
NCBI遺伝子UTR5情報	utr_5_start_in_contig	ncbi utr5_set	utr_5_start_in_contig
NCBI遺伝子UTR5情報	utr_5_end_in_contig	ncbi utr5_set	utr_5_end_in_contig
NCBI遺伝子UTR5情報	gene_account	ncbi utr5_set	gene_account
NCBI遺伝子UTR5情報	utr_5_end_in_chr	ncbi utr5_set	utr_5_end_in_chr
NCBI遺伝子UTR5情報	utr_5_start_in_chr	ncbi utr5_set	utr_5_start_in_chr

- SNP 探索時情報について下記に示す。

テーブル(日本語)	カラム名(日本語)	テーブル	カラム名
SNPアレル情報		snp_allele_set_screen	snp_allele_set_screen (SNPアレル情報)
SNPスクリーニング情報		snp_screened	snp_screened (SNPアレル情報)
SNP遺伝子情報		snp_screened_gene_set	snp_screened_gene_set (SNP遺伝子情報)
SNPアレル情報	snp_allele_na_nuc	snp_allele_set_screen	snp_allele_na_nuc
SNPアレル情報	snp_id	snp_allele_set_screen	snp_id
SNPスクリーニング情報	snp_last_update_date	snp_screened	snp_last_update_date
SNPスクリーニング情報	snp_sample_size	snp_screened	snp_sample_size
SNPスクリーニング情報	db_id_jn_consensus_end_pos	snp_screened	db_id_jn_consensus_end_pos
SNPスクリーニング情報	snp_release_date	snp_screened	snp_release_date
SNPスクリーニング情報	db_snp_rs_id	snp_screened	db_snp_rs_id
SNPスクリーニング情報	genome_cytogenetic	snp_screened	genome_cytogenetic
SNPスクリーニング情報	db_snp_ss_id	snp_screened	db_snp_ss_id
SNPスクリーニング情報	snp_organism	snp_screened	snp_organism
SNPスクリーニング情報	db_id_vc_pcr_reg_id	snp_screened	db_id_vc_pcr_reg_id
SNPスクリーニング情報	db_id_vc_contig	snp_screened	db_id_vc_contig
SNPスクリーニング情報	snp_end_jn_genome	snp_screened	snp_end_jn_genome
SNPスクリーニング情報	snp_start_jn_genome	snp_screened	snp_start_jn_genome
SNPスクリーニング情報	db_id_jn_consensus_start_pos	snp_screened	db_id_jn_consensus_start_pos
SNPスクリーニング情報	snp_jinkout_url	snp_screened	snp_jinkout_url
SNPスクリーニング情報	snp_5flank_seq	snp_screened	snp_5flank_seq
SNPスクリーニング情報	genome_definition	snp_screened	genome_definition
SNPスクリーニング情報	genome_ver	snp_screened	genome_ver
SNPスクリーニング情報	genome_chromosome	snp_screened	genome_chromosome
SNPスクリーニング情報	snp_type	snp_screened	snp_type
SNPスクリーニング情報	snp_id	snp_screened	snp_id
SNPスクリーニング情報	chromosome	snp_screened	chromosome
SNPスクリーニング情報	snp_comment	snp_screened	snp_comment
SNPスクリーニング情報	genome_acc	snp_screened	genome_acc
SNPスクリーニング情報	snp_3flank_seq	snp_screened	snp_3flank_seq
SNP遺伝子情報	snp_id	snp_screened_gene_set	snp_id
SNP遺伝子情報	snp_gene_evidence	snp_screened_gene_set	snp_gene_evidence
SNP遺伝子情報	snp_gene_account	snp_screened_gene_set	snp_gene_account
SNP遺伝子情報	snp_gene_str_type	snp_screened_gene_set	snp_gene_str_type
SNP遺伝子情報	snp_gene_name	snp_screened_gene_set	snp_gene_name
SNP遺伝子情報	snp_gene_molecular_type	snp_screened_gene_set	snp_gene_molecular_type
SNP遺伝子情報	snp_gene_account_release	snp_screened_gene_set	snp_gene_account_release

- SNP 探索時 PCR 領域情報について下記に示す。

テーブル(日本語)	カラム名(日本語)	テーブル	カラム名
領域遺伝子情報		region_gene_set	region_gene_set (領域遺伝子情報)
領域構造情報		region_gene_str_set	region_gene_str_set (領域構造情報)
領域情報		region_set	region_set (領域情報)
領域遺伝子情報	gene_str_start_in_genome	region_gene_set	gene_str_start_in_genome
領域遺伝子情報	gene_str_type	region_gene_set	gene_str_type
領域遺伝子情報	gene_name	region_gene_set	gene_name
領域遺伝子情報	gene_account	region_gene_set	gene_account
領域遺伝子情報	gene_account_release	region_gene_set	gene_account_release
領域遺伝子情報	gene_str_end_in_genome	region_gene_set	gene_str_end_in_genome
領域遺伝子情報	gene_molecular_type	region_gene_set	gene_molecular_type
領域遺伝子情報	gene_evidence	region_gene_set	gene_evidence
領域遺伝子情報	region_id	region_gene_set	region_id
領域構造情報	gene_id	region_gene_str_set	gene_id
領域構造情報	snp_id	region_gene_str_set	snp_id
領域構造情報	region_gene_account	region_gene_str_set	region_gene_account
領域情報	region_id	region_set	region_id
領域情報	pcr_primer_backward_seq	region_set	pcr_primer_backward_seq
領域情報	genome_acc	region_set	genome_acc
領域情報	pcr_primer_forward_seq	region_set	pcr_primer_forward_seq
領域情報	region_experiment_code	region_set	region_experiment_code
領域情報	chromosome	region_set	chromosome
領域情報	seq_primer_backward_pos	region_set	seq_primer_backward_pos
領域情報	seq_primer_forward_seq	region_set	seq_primer_forward_seq
領域情報	region_electrophoresis_pc	region_set	region_electrophoresis_pc
領域情報	genome_ver	region_set	genome_ver
領域情報	pcr_primer_forward_pos	region_set	pcr_primer_forward_pos
領域情報	region_sequence	region_set	region_sequence
領域情報	region_size	region_set	region_size
領域情報	region_end_in_genome	region_set	region_end_in_genome
領域情報	pcr_primer_backward_pos	region_set	pcr_primer_backward_pos
領域情報	seq_primer_backward_seq	region_set	seq_primer_backward_seq
領域情報	seq_primer_forward_pos	region_set	seq_primer_forward_pos
領域情報	region_start_in_genome	region_set	region_start_in_genome

- SNP 探索時実験情報について下記に示す。

テーブル(日本語)	カラム名(日本語)	テーブル	カラム名
メソッド情報		method	
メソッド情報	pcr_method_post_extention	method	pcr_method_post_extention
メソッド情報	additional_cycle_pcr_cycles	method	additional_cycle_pcr_cycles
メソッド情報	add_cycle_polymerization	method	add_cycle_polymerization
メソッド情報	reaction_10PCR_buffer	method	reaction_10PCR_buffer
メソッド情報	reaction_Primer	method	reaction_Primer
メソッド情報	buffer_EDTA_PH8.8	method	buffer_EDTA_PH8.8
メソッド情報	reaction_DMSO	method	reaction_DMSO
メソッド情報	buffer_MgCl2	method	buffer_MgCl2
メソッド情報	pcr_method_denaturation	method	pcr_method_denaturation
メソッド情報	buffer_Ammonium_sulfate	method	buffer_Ammonium_sulfate
メソッド情報	method_experiment_cd	method	method_experiment_cd
メソッド情報	additional_cycle_denaturation	method	additional_cycle_denaturation
メソッド情報	reaction_dNTP	method	reaction_dNTP
メソッド情報	reaction_MgCl2	method	reaction_MgCl2
メソッド情報	pcr_method_cycles	method	pcr_method_cycles
メソッド情報	reaction_Taq_DNA_Polymerase	method	reaction_Taq_DNA_Polymerase
メソッド情報	pcr_method_polymerization	method	pcr_method_polymerization
メソッド情報	seq_method_kit	method	seq_method_kit
メソッド情報	buffer_Trис_HCl_PH8.8	method	buffer_Trис_HCl_PH8.8
メソッド情報	buffer_b_mercaptoethanol	method	buffer_b_mercaptoethanol
メソッド情報	pcr_method_annealing	method	pcr_method_annealing
メソッド情報	additional_cycle_annealing	method	additional_cycle_annealing
メソッド情報	reaction_template	method	reaction_template
メソッド情報	seq_method_dye_strategy	method	seq_method_dye_strategy
メソッド情報	seq_method_sequencing_enzyme	method	seq_method_sequencing_enzyme
メソッド情報	seq_method_sequencing_machine	method	seq_method_sequencing_machine
メソッド情報	pcr_method_pre_denaturation	method	pcr_method_pre_denaturation

3. Swiss-Prot (<http://www.ebi.ac.uk/swissprot/>)

- 正規化前と正規化後のテーブル構造について下記に示す。

正規化前	正規化後		
テーブル名称	テーブル名称	テーブルの説明	備考
SWISSPROT Webサイトで公開されているXMLデータをMLデータの構造を解析し、RDB化設計を行った。CSV形式へ変換後、Teradataへローディングした。	PRT Accession	アセション情報	
	PRT Citation	引用情報	
	PRT Citation Author	引用著者	
	PRT Citation dbRef	リファレンス情報	
	PRT Citation Editor		
	PRT citingCitation		
	PRT citingCitation Author		
	PRT citingCitation Editor		
	PRT Comment		
	PRT Comment Link		
	PRT Entry		
	PRT Entry dbRef		
	PRT Entry dbRef Property		
	PRT Entry New		
	PRT Feature		
	PRT Gene		
	PRT geneLocation		
	PRT Keyword		
	prt_master		
	PRT Organism		
	PRT Organism Name		
	PRT Organism Taxon		
	PRT Protein	プロテイン	
	PRT Protein Component		
	PRT Protein Domain		
	PRT Ref Source Transposon		
	PRT Reference Scope		
PRT Reference Source Plasmid			
PRT Reference Source Species			
PRT Reference Source Strain			
PRT Reference Source Tissue			