"HAEMORRHAGE"や"ISCHAEMIA"等ラテン表記用語に関しては、文献情報における使用頻度がより高い英語表記のレコードを加えた。

初期システムにおいては、3813 の医薬品名（drug name）、3824 の適応（indication）を DN-INDI list に、9712 の有害反応名（reaction name）を RN list に収載した。AERS データベースの拡張とトレーニングよる最適化に伴う増減を経て、カレントシステムにおいては医薬品名 6870、適応 5930、有害反応名 12875 がそれぞれ収載されている。

4．TSADR システムの構築

質的なばらつきが大きい大量のテキストソースを網羅的に検索して医薬品の安全性に特化した情報を選別・抽出する作業は通常のキーワード検索では事実上不可能であり、これを実現するために医薬品の副作用情報の収集・解析に最適化した DN-INDI list および RN list を実装した Perl スクリプトを新たに開発した。

初期システムを用いて、日本時間 06 年 5 月 29 日に取得した PubMed アブストラクト 500 件（11,924 センテンス）より 74 件（138 センテンス）が抽出された（抽出率 14.8%）。74 件中、有害反応名が正しく表示されていたものが 26 件（正解率 35.1%）、32 件では医薬品名と直接には無関係な反応がヒットし、16 件は適応症が有害反応として誤って表示されていた。誤った選別のパターンとしては、"glucose"、"oxygen"、金属イオン等の生体成分が医薬品名として拾われて起こる事例や"alchol"、"antibiotic"、"chemotherapy"等、医薬品分類名に関して誤った選別が起こる例が多く認められた。前者のパターンに関しては、隣接する単語との関連から医薬品名としての取捨を判断するフィルターをスクリプトに加えて対応し、後者のパターンは DN-INDI list の適応エントリー数を増やす手段で対応した。上記対応を施したシステムを、新たに（日本時間 06 年 6 月 14 日）に取得したシステムトレーニング用テキスト（STTXT）に適用し、その結果を基に辞書ファイルを修正する作業を繰り返した。また、誤って選別されたアブストラクトに癌・腫瘍関係の雑誌のものが多かったことから、癌・腫瘍関係語彙用テキスト（CTBTXT: PubMed から Subsets の Limit に Cancer を設定して取得した）を用いたトレーニングも行った。これらのトレーニングによって最適化された語彙リストを用いて、両トレーニングテキスト自身を検索した最終的な成績（正解件数／抽出件数）は STTXT が 35/62（正解率 56.5%）、CTBTXT が 42/64（正解率 65.6%）であった。必要な場合にはスクリプトの修正も行った。

TSADR の実用化に向けた、上記トレーニングの有効性を検討する目的で、新たに取得したテキスト（日本時間 06 年 8 月 1 日）をトレーニング前のシステム TSADR-original とトレーニング後のシステム TSADR-trained を用いて解析した。TSADR-original、TSADR-trained それぞれの成績（正解件数／抽出件数）は 14/44（正解率 31.8%）および 22/54（正解率 40.7%）と算出され、抽出率、正解率ともにトレーニングの有効性が認められた。

一方、選別されるべきセンテンスの拾い漏れがどの程度起こっているかの予備的検討として、医薬品文献情報の有力サイトである英国の National electronic Library for Medicines にピックアップされた文献（発行期日 06 年 8 月 1 日ー18 日）のうち副作用情報に分類される 15 レコードを対象に TSADR による重要文献の抽出漏れを検討した。15 件中 3 件にはアブストラクトが無かった。残り 12 件のアブストラクトを PubMed より取得し TSADR-trained により解析すると、3 件が抽出されなかった。抽出漏れの原因は、いずれも DN-INDI list に医薬品名が収載されていなかったためで、1 件は第一層臨床試験における開発コード、2 件は治療方法の一般呼称で記載されていた。

Mozilla Firefox

Drug Names Reactions Indications Journal Information

Perphenazine-treated patients had a higher incidence of extrapyramidal symptom-related adverse events, mean increases (i.e., worsening) in extrapyramidal symptom rating scale scores, and a higher rate of elevated prolactin levels than aripiprazole (57.7% vs. 4.4%, p < .001).
PMID: 17335319 = J Clin Psychiatry. 2007 Feb;68(2):213-23.

We also determined whether the strength of antipsychotic or combination trials was associated with age, the duration of the current depressive episode, medical burden, cognitive status, or the severity of depressive or psychotic symptoms.
PMID: 17335316 = J Clin Psychiatry. 2007 Feb;68(2):194-200.

Bloodstream infections among patients treated with intravenous epoprostenol or intravenous treprostinil for pulmonary arterial hypertension--seven sites, United States, 2003-2006.
In September 2006, CDC received a report from a PAH specialist of a suspected increase in the number of gram-negative bloodstream infections (BSIs) among PAH patients treated with IV treprostinil.
The results do not suggest intrinsic contamination of IV treprostinil as a cause of the infections; the difference in rates might have been caused by differences in preparation and storage of the two agents, differences in catheter care practices, or differences in the anti-inflammatory activity of the agents.
PMID: 17332729 = MMWR Morb Mortal Wkly Rep. 2007 Mar 2;56(8):170-2.

Fig. 3 TSADR による解析結果の一例（部分）

システムのトータルな実効性に関する検討の一例として、（日本時間 07 年 3 月 22 日）に取得した PubMed アブストラクトを TSADR のカレントシステムで解析して得られた副作用シグナル「Perphenazine 投与に関連した錐体外路症候群」（Fig. 3 参照）を AERS データベースで検索にかけた。「錐体外路症候群」は MedDRA 用語ではあるが、AERS で使用を認められている PT（主要語）の下の階層の用語であるため、この組み合わせではヒットが無かったが、有害反応のキーワードを「錐体外路」とすると、部分一致により「錐体外路障害」を含む 4 件の報告がヒットした。さらに、「錐体外路障害」に連想される「振戦」、「筋固縮」、「ジスキネジー」、「筋骨格硬直」等、運動機能系統の有害事象に広げると合計 34 件の報告がヒットした。

D. 考察

Fig. 4 は、ヒトに関する PubMed 英文アイテムの年間エントリー数の推移を示したものである。年を追ってエントリー数の増加が認められるが、2005 年のデータを参考にすると、このコーパスから網羅性を重視して必要なアイテムを選別していくには、一日平均 800 ないし 1000 件を処理する必要があり、その実行には大きな組織、または自動化による補助が必須であると考えられる。

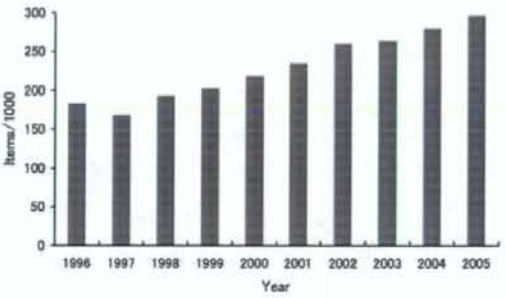


Fig. 3 PubMed 臨床医学系アイテムの年間エントリー数

本研究において医薬品安全性に関連する語彙のソース、および TSADR による副作用シグナルの実効性の検討に用いた AERS データベースは、米国 FDA が収集・管理している、製薬企業からの義務報告と医療従事者・患者およびその家族からの自発報告を総合した巨大な副作用データベースであり、四半期分（収載報告件数 8~9 万）ごとに半年遅れの生データが ASCII または SGML ファイルとして FDA のサイトから入手できる。

TSADR システムにおいては、医薬品の副作用が記載されているセンテンス中には医薬品名と有害反応・症状名が共起するという基本原則に従ってセンテンスの抽出が行われる。従って、いかに網羅的なデータ抽出が実行できるか（抽出率）は、実装された DN-INDI list および RN list の充実度に依存する。一方、この原則に合致するセンテンスには、医薬品副作用以外にも、医薬品の適応をはじめ、生体成分でもある医薬品（ホルモン等）

に関しては副作用のリスクファクターおよび臨床検査成績等の記載されたものもあり、原則的にはこれらも同時に抽出されてくる。従って、システム完成度の指標となる、いかに正しい抽出が行われたか（正解率）は医薬品副作用のシグナルをいかにしてこれらのノイズと分離するかに依存する。システムの構造上、語彙リストの医薬品名、有害反応名のレコード数を増やせば抽出率は上昇し、検索の網羅性に関しては有利に働く一方で、副作用以外の医薬品名と有害反応名の組み合わせを拾う可能性も高くなり、正解率が下がれば人間による最終的な選別操作の負担が大きくなる。これに対して、医薬品を投与する原因となる病態などの名称、すなわち適応名の語彙を増やすことは、誤った選別を抑制し、検索の精度を上昇させる。本研究において語彙リストの基本ソースとして用いた AERS データベースは、リレーショナルデータベースの構造を持つため、医薬品名と適応名を対応させて取得できる大きなメリットがある一方で、MedDRA に準拠した用語標準化によるある種の方言的な偏りが文献情報中のより多様な用語に対するヒット率を下げている可能性は否定できない。また、本研究において再構築した AERS データベースは、TSADR システムによるテキストマイニングの成績の検証ツールとしてもある程度の有用性は示せたが、MedDRA の階層構造を活用した柔軟な検索システムの開発が不可欠であることも明らかになった。

今後の研究においては、本検索システムの網羅性および選別性をさらに向上させるために、LSD シソーラスの実装を含めて、語彙リストの補強・改良を自動化する方法を検討する予定である。

E. 結論

本研究において開発された医薬品安全性監視支援システムは、副作用シグナルの早期検知あるいは副作用の予測用の情報抽出に大きな寄与をしうるものと期待される。

F. 研究発表

1．論文発表

1. 天野博夫，金子周司，医薬品安全性に関する文献情報自動抽出システムの考案，医療情報学，26 （Suppl.），1193-1194 （2006）

2．学会発表

1. 天野博夫，金子周司．医薬品安全性に関する文献情報自動抽出システムの考案．第 26 回医療情報学連合大会 （札幌，2006 年 11 月）

G. 知的財産権の出願・登録状況（予定も含む）

1．特許取得

なし

2．実用新案登録

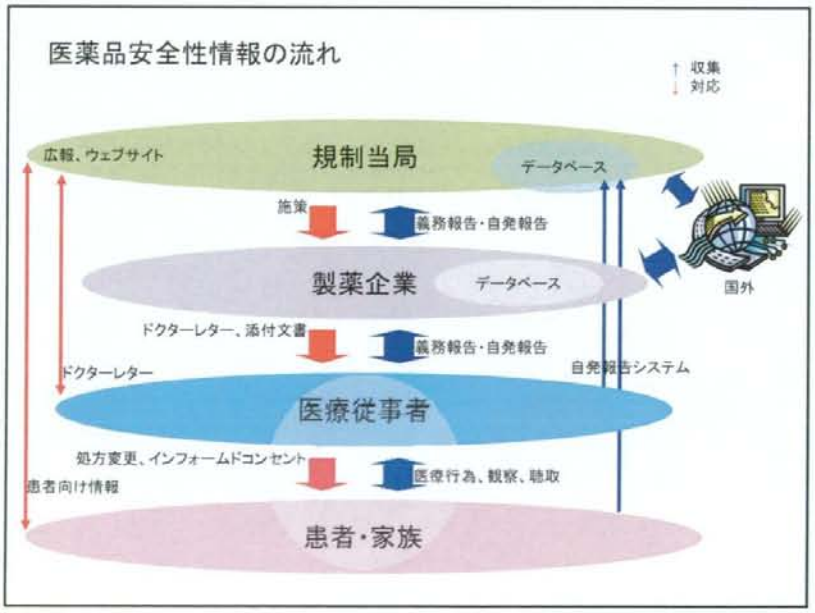
なし

3．その他

なし

医薬品安全性に関する文献情報自動抽出システムの考案

天野 博夫　　金子 周司
京都大学大学院薬学研究科生体機能解析学分野
Powered by LSDプロジェクト



医薬品安全性に関する文献情報

・原則として情報の独立性、中立性が高い
・自発報告システムと並んでファーマコビジランスの重要なチェックポイント
・副作用データベースの重要なソースでもある
・種々雑多な大量のテキストデータ中に点在する
・出版バイアス

研究方法

研究素材:

PubMed検索サイトよりダウンロードしたアブストラクト(Humansを対象とするアブストラクト付き英語文献)500件分のテキストファイルを抽出サンプルの1単位とした。

研究方法

研究素材:

PubMed検索サイトよりダウンロードしたアブストラクト(Humansを対象とするアブストラクト付き英語文献)500件分のテキストファイルを抽出サンプルの1単位とした。

照合用語彙リスト:

AERSデータベース(米国FDA)ASCIIデータベースファイルから作成。

1. DN-INDI list(医薬品名と適応)
2. RN list(有害反応名)

DN-INDI list

ISR	DRUG_SEQ	INDI_PT	DRUG05: DRUGNAME
4092310	1006117480	BREAST CANCER METASTATIC	HERCEPTIN
4092310	1006144815	BREAST CANCER METASTATIC	VINFLUNINE (VINFLUNINE)
4092311	1006117423	CHEMOTHERAPY	ETHYOL
4092311	1006117423	RADIOTHERAPY	ETHYOL
4092312	1006117422	ACUTE PROMYELOCYTIC	TRISENOX
4092313	1006117425	CONJUNCTIVITIS ALLERGIC	CLARITIN
4092314	1006117424	COMPUTERISED TOMOGRAM	OPTIRAY 160ML SYR
4092315	1006117426	LYMPHOMA	MABTHERA(RITUXIMAB) CONC FOR
4092315	1006146198	LYMPHOMA	ALDESLEUKIN (ALDESLEUKIN)
4092316	1006117427	GAUCHER'S DISEASE	CEREZYME
4092317	1006117429	HEPATITIS C	PEG INTRON
4092317	1006114286	HEPATITIS C	REBETOL
4092317	1006114280	HEPATITIS C	PEGASYS
4092319	1006117431	ACUTE LEUKAEMIA	CYTARABINE
4092319	1006142371	INFECTION	AMPHOTERICIN
4092319	1006142375	ACUTE MYELOID LEUKAEMIA	FLUDARABINE PHOSPHATE
4092319	1006142378	FUNGAL INFECTION	VORICONAZOLE
4092319	1006142378	ILL-DEFINED DISORDER	VORICONAZOLE
4092319	1006142378	LOWER RESPIRATORY TRACT	VORICONAZOLE
4092320	1006117432	DEPRESSION	CYMBALTA
4092321	1006117433	ACUTE LYMPHOCYTIC LEUKAEMIA	METHOTREXATE
4092321	1006117433	NERVOUS SYSTEM DISORDER	METHOTREXATE
4092321	1006117433	PROPHYLAXIS	METHOTREXATE
4092321	1006142304	ACUTE LYMPHOCYTIC LEUKAEMIA	CYTARABINE
4092321	1006142304	NERVOUS SYSTEM DISORDER	CYTARABINE
4092321	1006142304	PROPHYLAXIS	CYTARABINE

RN list

ISR	PT
4649555	DYSPNOEA
4649555	EYE IRRITATION
4649555	HYPERSENSITIVITY
4649555	NAUSEA
4649555	PALPITATIONS
4649555	PARAESTHESIA
4649555	URTICARIA
4649618	LEUKOPENIA
4649618	MEAN CELL VOLUME INCREASED
4649618	MONOCYTE PERCENTAGE
4649618	RED BLOOD CELL COUNT
4649618	TRACHEOBRONCHITIS
4649621	BALANCE DISORDER
4649621	CONFUSIONAL STATE
4649621	HYPOTONIA
4649622	PANCREATIC CARCINOMA
4649733	HEPATIC ENZYME INCREASED
4649733	HEPATIC FUNCTION ABNORMAL
4649733	JAUNDICE
4649733	MULTIPLE SCLEROSIS RELAPSE
4649812	CONVERSION DISORDER
4649812	CRYING
4649812	DIFFICULTY IN WALKING
4649812	HALLUCINATION, AUDITORY
4649812	MUSCLE SPASMS
4649812	NAUSEA

研究方法

研究素材:

PubMed検索サイトよりダウンロードしたアブストラクト(Humansを対象とするアブストラクト付き英語文献)500件分のテキストファイルを抽出サンプルの1単位とした。

照合用語彙リスト:

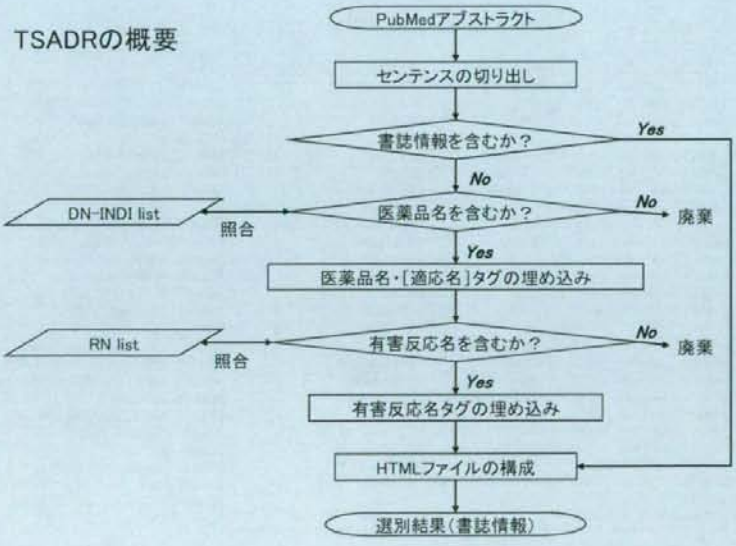
AERSデータベース(米国FDA)ASCIIデータベースファイルから作成。

1. DN-INDI list(医薬品名と適応)
2. RN list(有害反応名)

抽出システム:

TSADR(Text-Search system for Adverse Drug Reaction)
上記語彙リストを照合用ファイルとするPerlスクリプト。

TSADRの概要



Drug Names Reactions Indications Journal Information

There was no significant difference in the rates of death from any cause or total stroke according to group assignment, but raloxifene was associated with an increased risk of fatal stroke (59 vs. 39 events; hazard ratio, 1.49; 95 percent confidence interval, 1.00 to 2.24; absolute risk increase, 0.7 per 1000 woman-years) and venous thromboembolism (103 vs. 71 events; hazard ratio, 1.44; 95 percent confidence interval, 1.06 to 1.95; absolute risk increase, 1.2 per 1000 woman-years).
The benefits of raloxifene in reducing the risks of invasive breast cancer and vertebral fracture should be weighed against the increased risks of venous thromboembolism and fatal stroke.
PMID: 16837676 [N Engl J Med. 2006 Jul 13;355(2):125-37.]

研究方法

研究素材:
PubMed検索サイトよりダウンロードしたアブストラクト(Humansを対象とするアブストラクト付き英語文献)500件分のテキストファイルを抽出サンプルの1単位とした。

照合用語彙リスト:
AERSデータベース(米国FDA)ASCIIデータベースファイルから作成。
1. DN-INDI list(医薬品名と適応)
2. RN list(有害反応名)

抽出システム:
TSADR(Text-Search system for Adverse Drug Reaction)
上記語彙リストを照合用ファイルとするPerlスクリプト。

システムトレーニング:
抽出されたセンテンスにおいて医薬品の副作用が正しく抽出・表示されているかを検証し、問題点を改善。

誤った選別のパターンと対策（システムトレーニング）

パターン	対策
適応を有害反応として拾う	適応名をリストに追加
生体成分、検査項目等を医薬品名として拾う（Calcitonin, Insulin, Glucose, Oxygen 等）	取捨を判断するフィルター
転移癌、検査・治療行為を有害反応として拾う（CARCINOMA METASTATIC, BIOPSY, TRANSPLANT 等）	有害反応のリストから削除
ラテン名の英語表記が見落とされる（ANAEMIA, ISCHAEMIA, OEDEMA 等）	英語表記を有害反応のリストに追加（ANEMIA, ISCHEMIA, EDEMA 等）

Drug Names Reactions Indications Journal Information

Eligible patients were discharged with an ICD-9-CM diagnosis consistent with community-acquired pneumonia and divided into 2 groups: 1) a "not observed" cohort, in which patients were discharged on the same day as the switch from IV to oral antibiotics and 2) an "observed for 1 day" cohort, in which patients remained hospitalized for 1 day after the switch from IV to oral antibiotics.
PMID: 16750965 [Am J Med. 2006 Jun;119(6):512.e1-7.]

Opioids were more effective than placebo for both pain and functional outcomes in patients with nociceptive or neuropathic pain or fibromyalgia.
Among the side effects of opioids, only constipation and nausea were clinically and statistically significant.
PMID: 16717269 [CMAJ. 2006 May 23;174(11):1589-94.]

Calcitonin measurements for early detection of medullary thyroid carcinoma or its premalignant conditions in Hashimoto's thyroiditis.
The measurement of basal serum calcitonin (CT) in patients with evidence of Hashimoto's thyroiditis (HT) has been proposed in a recent study demonstrating an increased prevalence of elevated basal and stimulated CT.
PMID: 16739344 [Anticancer Res. 2006 Jan-Feb;26(1B):723-7.]

It regulates a number of cellular processes including mitosis, apoptosis, secretion and signal transduction as well as critical events in physiological processes as diverse as insulin release, T cell cytokine production, wound healing, vision and neurotransmission.
PMID: 16719771 [Curr Drug Targets. 2006 May;7(5):607-627.]

Failure to complete the TT6 was due to oxygen toxicity (4) and claustrophobia (3).
PMID: 16716057 [Undersea Hyperb Med. 2006 Mar-Apr;33(2):85-8.]

Convincing data now exist that show an association of salmeterol with an increase in asthma-related deaths and life-threatening experiences, while formoterol is associated with more frequent serious asthma exacerbations.
PMID: 16715643 [J Fam Pract. 2006 Apr;Suppl:1-6.]

研究方法

研究素材:

PubMed検索サイトよりダウンロードしたアブストラクト(Humansを対象とするアブストラクト付き英語文献)500件分のテキストファイルを抽出サンプルの1単位とした。

照合用語彙リスト:

AERSデータベース(米国FDA)ASCIIデータベースファイルから作成。

1. DN-INDI list(医薬品名と適応)
2. RN list(有害反応名)

抽出システム:

TSADR(Text-Search system for Adverse Drug Reaction)
上記語彙リストを照合用ファイルとするPerlスクリプト。

システムトレーニング:

抽出されたセンテンスにおいて医薬品の副作用が正しく抽出・表示されているかを検証し、問題点を改善。

評価指標:

- 1単位のサンプルから抽出されたアブストラクト数(抽出率)。
- 抽出されたアブストラクトのうち医薬品安全性に関する記載のあったものの%数(正解率)。

システムトレーニングの効果

	抽出率	正解率
初期システム	8.8% (44/500)	35.1% (14/44)
修正後システム	10.8% (54/500)	40.7% (22/54)

＊一般に、抽出率が高ければ情報の取りこぼしは少なくなるが最終選別者の負担は増加する。

抽出漏れの検証

サンプル:

National electronic Library for Medicines収載の副作用情報
120レコード

TSADRを用いて抽出処理

抽出成績:

抽出されなかったレコード数14件(抽出率88.3%)
抽出されなかった原因

- DN-INDI listに医薬品名がなかったもの11件
- RN listに有害反応名がなかったもの5件
- 2件は医薬品名、有害反応名ともリストに収載されていたが別個のセンテンス中に記載

厚生労働科学研究費補助金（医療安全・医療技術評価総合研究事業）
分担研究報告書

医療情報解析のためのテキストマイニングエンジンの開発

分担研究者：奥野恭史（京都大学大学院薬学研究科・システム創薬科学講座）

[研究要旨]
本研究は、医薬品の副作用（有害事象）のレポートや医療情報の解析・評価に、テキストマイニング技術を適用し、薬物有害事象の情報解析システムを開発することにより、IT時代を迎える医療における効率良く確かな安全体制の実現を情報技術的に支援することを目的とする。
本研究では、「XMLデータベースの構築とテキストマイニングエンジンの開発」、「薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発」、「医療情報とゲノム情報の連携システム（GEM-TREND）の開発」の３点に分けて研究開発を行った。
ここで、「XMLデータベースの構築とテキストマイニングエンジンの開発」としては、JAPIC添付文書記載病名集に加え、米国FDAが提供する副作用情報データAERS (Adverse Event Reporting System)をXMLデータベースとして構築し、全文検索ツールHyper Estraierをテキストマイニングエンジンとして実装したシステムの開発に至った。また、「薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発」としては、医薬品の薬効や副作用の総合的な解析のインフォマティクス基盤として開発した薬物とタンパク質との相互作用データベースGLIDA (GPCR-Ligand interaction database)を開発し、Web公開を行った。さらに、「医療情報とゲノム情報の連携システム（GEM-TREND）の開発」としては、これまで開発してきた医療情報テキストマイニングシステムとゲノム情報の連携を図るために、米国NIH/NCBIが提供する遺伝子発現データベース（GEO: Gene Expression Omnibus）への独自検索システムGEM-TREND（Gene Expression Data Mining Toward Relevant Network Discovery）を開発し、Web公開を行った。

A. 研究目的

本研究は、医薬品の副作用（有害事象）のレポートや医療情報の解析・評価に、テキストマイニング技術を適用し、薬物有害事象の情報解析システムを開発することにより、ＩＴ時代を迎える医療における効率良く確かな安全体制の実現を情報技術的に支援することを目的とする。

B. 研究方法

１．XMLデータベースの構築とテキストマイニングエンジンの開発

薬物有害事象の自動抽出を目的としたテキストマイニングエンジンの開発素材として、（財）日本医薬情報センター（JAPIC）の添付文書記載病名集を用いた。JAPIC添付文書記載病名集は、医薬品の薬効や副作用情報など本研究対象に必要

な情報が記載されており、さらに XML 形式での電子データが供給されている。また、JAPIC の添付文書に加え、米国 FDA が提供する副作用情報データ AERS (Adverse Event Reporting System)もテキストマイニングコンテンツとして用いた。

テキストマイニングエンジンには、開発当初は有料の NeoCore を用いたが、最終的には無償ツールの全文検索エンジン Hyper Estraier への実装を行った。さらに、検索のユーザビリティ向上を目指し、研究代表者が開発するライフサイエンス辞書（http://www.pharm.kyoto-u.ac.jp/channel/ja/dictionary/index.html）および医学用語辞書 MedDRA (Medical Dictionary for Regulatory Activities)、UMLS (Unified Medical Language System)との連携を図り、検索した単語の関連語・共起表現機能を実装した。

2．薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発

医薬品による副作用の総合的な解析には、薬物の作用点となる標的タンパク質との相互作用様式を情報学的に処理する基盤技術の整備が必須となる。本研究では、薬物の主作用点データベースとして GPCR－リガンド相互作用データベース（GLIDA データベース）を構築し、http://pharminfo.pharm.kyoto-u.ac.jp/services/glida より公開している。市販医薬品の大半は、G タンパク質共役型受容体(GPCR)を薬物作用点にしていることから、本データベースがプロトタイプになり得る。GLIDA は、GPCR のバイオ情報、リガンドのケミカル情報、および GPCR とリガンドの相互作用情報の 3 種類の情報より構成される。GPCR のエントリはヒト、ラット、マウスに限定し、バイオ情報は GPCRDB から取得した。また、GPCR と結合するリガンドのエントリとそのケミカルデータ（化学名、構造式、分子量, MDL Mol ファイルなど）は IUPHAR Receptor Database, PubMed, PubChem および MDL ISIS などの公共または商用のデータベースから取得した。

3．医療情報とゲノム情報の連携システム（GEM-TREND）の開発

ヒトゲノムが解読された今日、ゲノム情報と医療や医薬に用いる試みが精力的に研究展開されている。そこで、これまで開発してきた医療情報データベースおよびテキストマイニングシステムとゲノム情報の連携を図るために、米国 NIH/NCBI が提供する遺伝子発現データベース（GEO: Gene Expression Omnibus）への独自検索システム GEM-TREND (Gene Expression Data Mining Toward Relevant Network Discovery) を開発した。本ツールは Linux server 上に、MySQL、PHP および R を用いて実装し、遺伝子ネットワークの可視化には Java Applet graphical user interface を用いた。

なお、本研究は計算機によるシステム開発であり、倫理面に関する問題は一切無い。

C. 研究結果

1．XMLデータベースの構築とテキストマイニングエンジンの開発

複雑・多岐に渡るデータの構成に柔軟に対応し、かつ大量のデータを高パフォーマンスで処理できる仕組みとして、全文検索エンジン Hyper Estraier と XML データベースとの連携を行った。XML データベースとして構築する文書データには、JAPIC 医薬品添付文書記載病名集に加え、米国 FDA が公開している大規模副作用症例データベース Adverse Event Reporting System (AERS)を新たに実装した。JAPIC 医薬品添付文書記載病名集は、日本における医薬品の効能効果、禁忌情報、副作用情報などが記載されているのに対し、米国 FDA の AERS は世界各国における副作用症例が集積されたデータベースであり、薬物有害事象のテ

キストマイニングの対象素材としても最適なコンテンツである。
また、開発したテキストマイニングシステムに、研究代表者が開発するライフサイエンス辞書や医学用語辞書 MedDRA (Medical Dictionary for Regulatory Activities)、UMLS (Unified Medical Language System)を組み込み、検索した単語の関連語・共起表現を日本語・英語で示すことにより、検索のユーザビリティ向上を図るとともに、PubMed および Google などの他の公共データベースへの同時検索を可能にした。

2．薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発

本研究では、薬物と GPCR の相互作用データベースとして開発、公開している GLIDA の医薬品情報の大幅増加を行った。具体的には、24,077 件ものリガンドエントリー、および 39,140 件ものリガンドーGPCR の相互作用エントリーの登録に至っており、公共の医薬品データベースとしては世界最大級のエントリーを誇っている。
また、リガンドエントリーの大幅増加に伴い、リガンド検索ツールの大幅改良も行った。以前のリガンド分類は階層型クラスタリングに基づくものであったが、計算量の問題から、数万エントリーにはこのクラスタリングは適さない。そこで、本研究では、主成分分析（PCA）に基づいて、全リガンドエントリーの分類を行った。さらに、リガンドの部分構造に基づく類似化合物検索エンジンの開発と実装を行い、一般ユーザーからのデータベース検索を可能にしている。
なお、本データベース GLIDA は
http://pharminfo.pharm.kyoto-u.ac.jp/services/glida より公開している。

3．医療情報とゲノム情報の連携システム（GEM-TREND）の開発

近年のマイクロアレイ技術の発展に伴い、大量の遺伝子発現データが GEO（gene expression omnibus）に代表される公共データベースに蓄積されてきている。我々は、キーワードや独自の遺伝子発現データを問合せ情報として、これら遺伝子発現データベースから、遺伝子発現データを検索するツール GEM-TREND（Gene Expression Data Mining Toward Relevant Network Discovery）を開発した。
GEM-TREND では、任意のキーワードに加え、gene expression signature をベースにしたマイクロアレイデータの検索および得られたデータのネットワークを可視化する機能を実装している。gene expression signature をベースにした検索では、nonparametric、rank-based pattern matching approach を用いて、query とデータベース内に蓄積された GEO 発現データに対応した sample ごとの遺伝子ランクデータの signature を比較し、得られた similarity によって検索結果を決定する評価系を用いている。また、描画画面の遺伝子ネットワークの構築には Pearson の相関係数と K-means clustering を用いて行った。開発したシステムは、以下の URL で Web より公開している。
http://cgs.pharm.kyoto-u.ac.jp/services/network/

D. 考察

1．XMLデータベースの構築とテキストマイニングエンジンの開発

今回改良導入を行った新規のテキストマイニングエンジン Hyper Estraier は、高速全文検索を可能にする特徴は言うまでもなく、無償ツールであることからシステムの公開も可能であり、汎用性の高いシステムとして大きな利点を有する。
また、JAPIC 添付文書に加え、AERS の導入により、国内に止まらず、海外の副作用事例を加味した情報抽出、検索が可能となり、医薬品の安全性に大きな寄与を示すものと期待される。今後は、

本研究で実装した全文検索エンジンを活用して、医薬品添付文書内に収録されている副作用情報の検索と分析を展開する。

2．薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発

GLIDA データベースは市販の医薬品の半分以上の標的分子となっているGPCRとそれに作用する薬物の相互作用に関する知識データベースであるとともに、その相互作用メカニズムの解明に関する知識を提供し得るケミカルゲノミクスのためのデータベースである。医薬品の薬効や副作用の総合的な解析には、薬物の作用基点となる遺伝子との相互作用様式を情報学的に処理する基盤技術の整備は必須であり、本データベース構築によりその基盤は確立された。

3．医療情報とゲノム情報の連携システム（GEM-TREND）の開発

今回開発したGEM-TRENDを用いて、薬物名によるデータベース検索を実行し、その検証を行ったところ、高い精度で薬物投与遺伝子発現データの取得が確認できた。今後は、GEM-TRENDデータベースと、上述、テキストマイニングデータベースとの直接的な連携を図る。

E. 結論

1．XMLデータベースの構築とテキストマイニングエンジンの開発

JAPIC 添付文書記載病名集データベースに加え、米国 FDA の大規模副作用症例データベース Adverse Event Reporting System (AERS)を新たに実装したテキストマイニングツールの開発を行った。全文検索アルゴリズムとして、無償ツールである Hyper Estraier の実装に成功した。

2．薬物の主作用点データベース（GPCR－リガンド相互作用データベース）の開発

医薬品の薬効や副作用の総合的な解析のインフォマティクス基盤として、薬物とタンパク質との相互作用データベース GLIDA の大規模データベースの構築を行った。これらは、http://pharminfo.pharm.kyoto-u.ac.jp/services/glida より公開している。

2．医療情報とゲノム情報の連携システム（GEM-TREND）の開発

これまで開発してきた医療情報テキストマイニングシステムとゲノム情報の連携を図るために、米国 NIH/NCBI が提供する遺伝子発現データベース（GEO: Gene Expression Omnibus）への独自検索システム GEM-TREND（Gene Expression Data Mining Toward Relevant Network Discovery）を開発した。これらは、http://cgs.pharm.kyoto-u.ac.jp/services/network/ より公開している。

F. 研究発表

1．論文発表

1. van der Horst, E., Okuno, Y., Bender, A. and Ijzerman, A.P., "Substructure mining of GPCR ligands reveals activity-class specific functional groups in an unbiased manner", J. Chem. Inf. Model., 49, 348-60, 2009

2. Tsuchiya, S., Tachida, Y., Segi-Nishida, E., Okuno, Y., Tamba, S., Tsujimoto, G., Tanaka, S. and Sugimoto, Y., "Characterization of gene expression profiles for different types of mast cells pooled from mouse stomach subregions by an RNA amplification method", BMC Genomics, 10: 35, 2009

3. ケミカルゲノミクスに基づくインシリコ創薬
新島　聡, 奥野 恭史, 日薬理誌, 133, 173, 2009

4. Ruike, Y., Ichimura, A., Tsuchiya, S., Shimizu, K., Kunimoto, R., Okuno, Y., and Tsujimoto,

G., "Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines", J. Hum. Genet., 53, 515-23, 2008

5. Kawanishi, H., Matsui, Y., Ito, M., Watanabe, J., Ta.kahashi, T., Nishizawa, K., Nishiyama, H., Kamoto, T., Mikami, Y., Tanaka, Y., Jung, G., Akiyama, H., Nobumasa, H., Guilford, P., Reeve, A., Okuno, Y., Tsujimoto, G., Nakamura, E. and Ogawa, O., "Secreted CXCL1 is a potential mediator and marker of the tumor invasion of bladder cancer", Clin. Cancer Res., 14, 2579-87, 2008

6. Takano, H., Nakazawa, S., Okuno, Y., Shirata, N.,Tsuchiya, S., Kainoh, T., Takamatsu, S., Furuta, K., Taketomi, Y., Naito, Y., Takematsu, H., Kozutsumi, Y., Tsujimoto, G., Murakami, M., Kudo, I., Ichikawa, A., Nakayama, K., Sugimoto, Y. and Tanaka, S., "Establishment of the culture model system that reflects the process of terminal differentiation of connective tissue-type mast cells", FEBS Lett., 582, 1444-50, 2008

7. Okuno, Y., "In silico drug discovery based on the integration of bioinformatics and chemoinformatics", YAKUGAKU ZASSHI, 128 (11), 1645-51, 2008

8. ケミカルゲノミクス情報を用いた新規リガンド探索手法
藪内 弘昭, 奥野 恭史, SAR News, 14, 2-6, 2008

9. Inoue, T., Adachi, H., Murakami, S., Takano, K., Matsumura, H., Mori, Y., Fukunishi, Y., Nakamura, H., Kinoshita, T., Nakanishi, I., Okuno, Y., Minakata, S., Shimojo, S., Sakata, T. "New progress in crystallization technology of membrane protein and introduction of pharamaceutical innovation value chain" *YAKUGAKU ZASSHI,* **128 (4)**, 497-505, 2008

10. Okuno, Y., Tamon, A., Yabuuchi, H., Niijima, S., Minowa, Y., Tonomura, K., Kunimto, R. and Feng, C. "GLIDA: GPCR-Ligand database for chemical genomics drug discovery - Database and tools update." *Nucleic Acids Res*, **36**, D907-12, 2008

11. Niijima, S. and Okuno, Y. "Laplacian Linear Discriminant Analysis Approach to Unsupervised feature selection." *IEEE/ACM Trans.Comput.Biol. Bioinform,* IEEE computer Society Digital Library., http://doi.ieeecomputersociety.org/10.1109/TCBB.2007.70257

12. Kitajima, M., Minowa, Y., Matsuda, H. and Okuno, Y. "Compound-transporter interaction studies using canonical correlation analysis." *Chem-Bio Inform J.*, **7**, 24-34, 2007

13. Yamamoto, H., Takematsu, H., Fujinawa, R., Naito, Y., Okuno, Y., Tsujimoto, G., Suzuki, A. and Kozutsumi, Y. "Correlation index-based responsible-enzyme gene screening (CIRES), a novel DNA microarray-based method for glycan biosynthesis enzyme gene." *PLoS ONE*, **2**, e1232, 2007

14. Ikeda, A., Miyazaki, T., Kakizawa, S., Okuno, Y., Tsuchiya, S., Myomoto, A., Saito, SY., Yamamoto, T., Yamazaki, T., Iino, M., Tsujimoto, G., Watanabe, M. and Takeshima, H. "Abnormal features in mutant cerebellar Purkinje cells lacking junctophilins." *Biochem. Biophys. Res. Commun.*, **363**, 835-9, 2007

15. Yamazaki, T., Sasaki, N., Nishi, M., Yamazaki, D., Ikeda, A., Okuno, Y., Komazaki, S., and Takeshima, H. "Augmentation of drug-induced cell death by ER protein BRI3BP." *Biochem. Biophys. Res. Commun.*, **362**, 971-5, 2007

16. ケミカル・バイオ情報に基づく創薬インフォマティクス研究
奥野 恭史, Pharma VISION NEWS, 9, 13-16, 2007

17. Naito, Y., Takematsu, H., Koyama, S., Miyake, S., Yamamoto, H., Fujinawa, R., Sugai, M., Okuno, Y., Tsujimoto, G., Yamaji, T., Hashimoto, Y., Itohara, S., Kawasaki, T., Suzuki, A. and Kozutsumi, Y., "Germinal center marker GL7 probes activation-dependent repression of N-glycolylneuraminic acid, a sialic acid species involved in the negative modulation of B cell activation", *Mol. Cell Biol.*, **27**, 3008-22, 2007

18. Zhu, S., Okuno, Y., Tsujimoto, G. and Mamitsuka, H., "Application of a new probabilistic model for mining implicit

associated cancer genes from OMIM and Medline", *Cancer Inform.*, **2**, 361-71, 2006

19. Osada, S., Naganawa, A., Misonou, M., Tsuchiya, S., Tamba, S., Okuno, Y., Nishikawa, J., Satoh, K., Imagawa, I., Tsujimoto, G., Sugimoto, Y. and Nishihara, T., "Altered gene expression of transcriptional regulatory factors in tumor marker-positive cells during chemically induced hepatocarcinogenesis". *Toxicol. Lett.*, **167**, 106-13, 2006

20. Tsuchiya, S., Okuno, Y. and Tsujimoto, G., "MicroRNA: biogenetic and functional mechanisms and involvements in cell differentiation and cancer", *J. Pharmacol. Sci.*, **101**, 267-70, 2006

21. Okuno, Y., Yang, J., Taneishi, K., Yabuuchi, H. and Tsujimoto, G., "GLIDA: GPCR-Ligand database for chemical genomic drug discovery", *Nucleic Acids Res.*, **34**, D673-7, 2006

2．学会発表

1. 関西バイオネットワーク「創薬バリューチェーンの構築に向けて」発表交流会
「非結晶性標的タンパク質に対する化合物探索」2008年12月8日

2. 応用トキシコロジーリカレント講座
「ケモゲノミクスとトキシコゲノミクスの融合」2008年9月12日

3. 日本たばこ産業株式会社　医薬総合研究所　社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」
2008年9月2日

4. キッセイ薬品工業株式会社　中央研究所
社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」
2008年8月20日

5. 生化学工業株式会社　中央研究所
社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」
2008年8月19日

6. アスビオファーマ株式会社　生物医学研究所セミナー「ケミカルゲノミクス情報を用いた生物活性に富んだケミカル空間の合理的探索」2008年6月5日

7. 日本薬学会128年会　日本薬学会奨励賞受賞講演「バイオ空間とケミカル空間の包括的相関解析とそのインシリコ創薬への研究展開」2008年3月28日

8. 第2回ClassAシステムバイロジーセミナー in 東京「ケミカル・スペースでの化合物探索」2008年2月28日

9. 第2回ClassAシステムバイロジーセミナー in 関西「ケミカル・スペースでの化合物探索」2008年2月25日

10. 第3回三重ゲノム創薬フォーラム「薬学研究におけるアレイインフォマティクス」
2008年2月15日

11. 平成19年度 第2回産業情報交流会「ケミカルゲノミクスに基づくインシリコ化合物探索他」2007年10月22日

12. 新産業を創る先端科学技術フォーラム2007「ポストゲノム創薬のための新技術」セッション
「ケミカルゲノミクスに基づく創薬インフォマティクス」2007年10月18日

13. 第66回日本癌学会学術総会 International Session - Chemical Genomics for Cancer

Research -「Knowledge Discovery and Data Mining in Chemical Genomics」
2007年10月3日

14. バイオビジネスステーション卒業生交流会「先端シーズビジネス化の実際・画期的創薬への期待」2007年7月21日

15. 日本薬物動態学会 第21回年会「ケミカルゲノミクスからの創薬インフォマティクス」2006年12月1日

16. 第32回情報処理技術検討交換会「ケミカルゲノミクス情報のデータマイニング」
2006年11月30日

17. モレキュラーライブラリー研究会「ケミカルゲノミクス情報を用いた化合物ライブラリーの合理的設計」2006年11月16日

18. 第34回構造活性相関シンポジウム奨励講演「ケミカルゲノミクス情報に基づく化合物探索」2006年11月14日

19. 第269回ＣＢＩ学会「ケミカルゲノム情報に基づくGPCR創薬」2006年11月1日

20. 第21回21世紀の薬学を探る京都シンポジウム - 薬学教育フロンティア -「インフォーマティクスと創薬」2006年10月14日

21. 第２回バイオメディカル研究会「ケミカルゲノミクスのための創薬インフォマティクス」2006年9月12日

22. 第４回先端医療セミナー「ケミカルゲノミクスからのIn silico創薬」2006年8月25日

23. 第45回バイオグリッドビジネスサロン「創薬におけるバイオインフォマティクスの可能性」2006年8月18日

24. 第16回近畿バイオインダストリー振興会議技術シーズ公開会「創薬リード化合物自動合成装置の研究開発」2006年7月28日

25. コンピュータ化学部会 平成18年度例会「ケミカルゲノム情報に基づくIn silico創薬」2006年6月13日

26. バイオグリッド研究会 - ITプログラム成果報告と今後の展開 -「化合物空間を利用した化合物検索」2006年5月27日

27.（財）サントリー生物有機科学研究所　コロキウム「ケミカルゲノミクスのためのインフォマティクス」2006年2月21日

G. 知的財産権の出願・登録状況（予定も含む）

1．特許出願

1. 特願2007-53322、「マイクロRNA標的遺伝子予測装置」、平成19年3月2日出願、出願人 東レ株式会社、発明者 奥野恭史、辻本豪三、国本亮、寺澤和哉、土屋創健、秋山英雄、妙本明

2. 公開番号 WO2007/004479A; 特開2007-11752、「データ処理装置、データ処理プログラム、それを格納したコンピュータ読み取り可能な記録媒体、およびデータ処理方法」、平成19年1月11日公開、出願人 京都大学、発明者 奥野恭史、辻本豪三、梁智允、種石慶

3. 公開番号 WO2007/139037A1;
PCT/JP2007/060736 特願2006-147433「ケミカ

ルゲノム情報に基づく、タンパク質－化合物相互作用の予測と化合物ライブラリーの合理的設計」、平成 18 年 5 月 26 日（国内）平成 19 年 5 月 25 日（国際）出願、出願人　京都大学、発明者　奥野恭史、種石慶、辻本豪三

2．実用新案登録

無し

3．その他

無し

研究成果の刊行に関する一覧表

書籍

著者氏名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年
金子周司ほか	総編集　伊藤正男，井村裕夫，高久史麿	医学大辞典 第2版	医学書院	日本	2008
金子周司ほか	ライフサイエンス辞書プロジェクト監修	ライフサイエンス英語表現使い分け辞典	羊土社	日本	2007
金子周司ほか	ライフサイエンス辞書プロジェクト監修	ライフサイエンス論文作成のための英文法	羊土社	日本	2007
金子周司ほか	今堀和友・山川民夫監修	生化学辞典第4版	東京化学同人	日本	2007
金子周司 奥野恭史ほか	藤井信孝・辻本豪三・奥野恭史　編集	インシリコ創薬科学—ゲノム情報から創薬へ—	㈱京都廣川書店	日本	2008
奥野恭史ほか	石渡信一・桂勲・桐野豊・美宅成樹 編	生物物理学ハンドブック	㈱朝倉書店	日本	2007
奥野恭史ほか	日本バイオインフォマティクス学会	バイオインフォマティクス事典	共立出版㈱	日本	2006

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
金子周司，鵜川義弘，大武博，河本健，竹内浩昭，竹腰正隆，天野博夫，藤田信之	医学用語シソーラスに基づく効率的医療情報検索システムの開発	**医療情報学**	28 (suppl)	639-642	2008
伊藤悦子，金子周司	分子薬理学的知識を記述する新たな三項関係データベースの開発	**医療情報学**	27 (suppl)	299-300	2007
天野博夫，金子周司	医薬品安全性に関する文献情報自動抽出システムの考案	**医療情報学**	26 (suppl)	1193-1194	2006
金子周司	ライフサイエンス辞書とは	**情報管理**	49 (1)	24-35	2006
金子周司	無料ライフサイエンス辞書の活用と効能	**ファルマシア**	42 (5)	463-467	2006