



Title	疾患関連遺伝子探索のためのMeSHによる遺伝子・疾患情報解析手法に関する研究
Author(s)	仲里, 猛留
Citation	大阪大学, 2008, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/49680
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

【11】	
氏 名	なか ざと たけ も 仲 里 猛 留
博士の専攻分野の名称	博 士（情報科学）
学 位 記 番 号	第 2 2 5 1 5 号
学 位 授 与 年 月 日	平 成 20 年 9 月 25 日
学 位 授 与 の 要 件	学位規則第4条第1項該当 情報科学研究科バイオ情報工学専攻
学 位 論 文 名	疾患関連遺伝子探索のための MeSH による遺伝子・疾患情報解析手法に 関する研究
論 文 審 査 委 員	(主査) 教 授 松田 秀雄 (副査) 教 授 清水 浩 教 授 四方 哲也 教 授 前田 太郎

論 文 内 容 の 要 旨

ライフサイエンス分野において、疾患関連遺伝子や新薬候補化合物の探索は、非常に注目されるテーマの一つである。ゲノム配列が次々に解読される現在、各遺伝子の生物学的な機能は、Gene Ontology (GO) の語を付与することで表現されているが、GOは各遺伝子の機能についての用語集であり、疾患や薬剤についての語は用意されていない。そこで、これらの面から遺伝子群についての生物学的な解釈を行うため、疾患や薬剤の領域をカバーしている用語集であるMeSHを用いて、遺伝子群を解釈する手法を提案した。MeSHは、遺伝子に直接付与された用語でなく、文献に付与された用語集である。つまり、遺伝子とMeSHとを関連づけるためには遺伝子と文献とを対応づけねばならない。しかしながら、文献は自然言語で書かれた情報であるため、遺伝子名称について、表現の揺らぎや別名、名前の衝突を適切に処理する必要がある。本研究では、MEDLINE収載の全文献データを対象に遺伝子と文献の対応づけを行い、各文献に付与されたMeSH termを抽出した。また、得られた遺伝子/関連MeSHペアについて情報利得と*p*値にて関連度のスコアリングを行った。

翻って、疾患情報について見てみると、遺伝子の情報はEntrez Geneなどでアノテーション情報を簡単に得ることができる一方、疾患については、OMIM (Online Mendelian Inheritance in Man) が情報源として広く用いられている。しかしながら、OMIMは、文献情報と同様に、自然言語で記載されており、疾患がサブタイプごとにグループ化されていたりするわけでもない。さらに、疾患の態様と疾患原因遺伝子のエントリが混じっており、各疾患について、その特徴を概観したり、比較したりすることは非常に困難である。また、既存手法では、臨床所見の面でしか解釈が行われてこなかった。そこで、本研究では、MeSHを用いて分子生物学の立場から各OMIMエントリを特徴づける手法を提案した。遺伝子/関連MeSHペア作成と同様に、OMIMエントリ/関連MeSHペアを作成した。また、各OMIMエントリについて、その

特徴を各MeSHとの情報利得を成分とするベクトルと見なし、コサイン係数を計算することで、その類似度を計算し、OMIMエントリどうしの比較を行えるようにした。

また、本提案手法をウェブアプリケーションとして実装し (<http://gendoo.dbcls.jp/>)、実際のマイクロアレイデータに本提案手法を適用することで、本提案手法の評価を行った。このことにより遺伝子と疾患情報とを合わせて解釈するための基盤が構築できた。

論文審査の結果の要旨

ライフサイエンス分野において、疾患関連遺伝子や新薬候補化合物の探索は、非常に注目されるテーマの一つである。ゲノム配列が次々に解読される現在、各遺伝子の生物学的な機能は、Gene Ontology (GO) の語を付与することで表現されているが、GOは各遺伝子の機能についての用語集であり、疾患や薬剤についての語は用意されていない。そこで、これらの面から遺伝子群についての生物学的な解釈を行うため、疾患や薬剤の領域をカバーしている用語集であるMeSHを用いて、遺伝子群を解釈する手法を提案している。MeSHは、遺伝子に直接付与された語でなく、文献に付与された用語集である。つまり、遺伝子とMeSHを関連づけるためには遺伝子と文献を対応づけなければならない。しかしながら、文献は自然言語で書かれた情報であるため、遺伝子名称について、表現の揺らぎや別名、名前の衝突を適切に処理する必要がある。本研究では、MEDLINE収載の全文献データを対象に遺伝子と文献の対応づけを行い、各文献に付与されたMeSH termを抽出している。また、得られた遺伝子/関連MeSHペアについて情報利得と p 値にて関連度のスコアリングを行っている。

次に、疾患情報については、OMIM (Online Mendelian Inheritance in Man) が情報源として広く用いられている。しかしながら、OMIMは、文献情報と同様に、自然言語で記載されており、疾患がサブタイプごとにグループ化されていたりするわけではなく、疾患の態様と疾患原因遺伝子のエントリが混じっており、各疾患について、その特徴を概観したり、比較したりすることは非常に困難である。このため、既存手法では、臨床所見の面でしか解釈が行われてこなかった。そこで、本研究では、MeSHを用いて分子生物学の立場から各OMIMエントリを特徴づける手法を提案し、遺伝子／関連MeSHペア作成と同様に、OMIMエントリ／関連MeSHペアを作成している。各OMIMエントリについて、その特徴を各MeSHとの情報利得を成分とするベクトルと見なし、コサイン係数を計算することで、その類似度を計算し、OMIMエントリどうしの比較を行えるようにしている。さらに、本提案手法をウェブアプリケーションとして実装し (<http://gendoo.dbcls.jp/>)、実際のマイクロアレイデータに本提案手法を適用することで評価を行った結果、本手法により遺伝子と疾患情報を合わせた解釈が可能となることが示されている。

以上により、本論文の成果は計算機による遺伝子機能および疾患情報の解析に関する研究の発展に貢献するものと考えられる。よって博士（情報科学）の学位論文として価値あるものとして認める。