

G-010

ベイジアンネットワークモデルを用いた胃癌関連遺伝子調節ネットワークの構築

Gastric-cancer-related gene regulatory network using Bayesian network model

柳 祐貴†
Yuuki Yanagi

知念 佑奈‡
Yuna Chinen

石井 寛之†
Hiroyuki Ishii

山崎 敏正†
Toshimasa Yamasaki

1. はじめに

胃癌は未分化型と分化型の2つのサブタイプに分類され、これにより成長様式や転移先、予後が異なる[1]。そのため、適切な治療を行うためにはサブタイプの識別が重要とされる。

本研究の目標は胃癌関連遺伝子を利用して胃癌のサブタイプを識別することである。ここでは遺伝子単体からサブタイプを識別するのではなく、ベイジアンネットワークモデル[2]を利用した遺伝子間相互作用ネットワークを用いることで識別率を向上させる。このモデルは事象間の関係を確率で表現し、因果の方向を簡明に表現する。このモデルは非循環有向グラフ(DAG:Directed Acyclic Graph)であり、実際の相互作用ネットワークに適合しない部分がある。本研究ではこの問題を克服し、胃癌関連遺伝子調節ネットワークを構築する。

2. 方法

2.1 重要遺伝子の抽出

胃癌のサブタイプの分類を示唆する遺伝子を抽出する。今回は t 検定を用いて分化型と未分化型に特異的な遺伝子を抽出した。この際に、Benjamini-Hochberg 法[3]を利用して、有意水準を設定している。これによって得られた遺伝子を DAVID[4]に入力し、遺伝子機能による分類を行った。この中の1つの機能に注目し、関連遺伝子を抽出した。

2.2 遺伝子調節ネットワークの構築[5]

(1) トリプレットの作成

抽出された遺伝子をネットワークのノード、ノード間の関係を表す矢印をアークとし、ノード3個で1組とするトリプレットの全ての組み合わせを作成する。このトリプレットを用いて小規模のベイジアンネットワークを推定する。そのため、このトリプレット内では循環型となるパターンを除いた 25 パターンのアークの組み合わせを想定する。

(2) ノード値

ノードに付与する値は(遺伝子発現サンプル数)/(全サンプル数)とする。遺伝子発現の有無を判断する基準値は分化型と未分化型の各平均値の中間値に設定する。ここで、未分化型と分化型のノード値を、表1、表2のように付与する。

表1 未分化型の遺伝子調節ネットワークのノード値

	未分化型発現確率	未分化型非発現確率
未分化型特異遺伝子の場合	未分化型サンプルの未分化型特異遺伝子を発現した確率	分化型サンプルの未分化型特異遺伝子を発現した確率
分化型特異遺伝子の場合	未分化型サンプルの分化型特異遺伝子を発現した確率	未分化型サンプルの未分化型特異遺伝子を発現した確率

表2 分化型の遺伝子調節ネットワークのノード値

	分化型発現確率	分化型非発現確率
分化型特異遺伝子の場合	分化型サンプルの分化型特異遺伝子を発現した確率	未分化型サンプルの分化型特異遺伝子を発現した確率
未分化型特異遺伝子の場合	分化型サンプルの非分化型特異遺伝子を発現した確率	未分化型サンプルの未分化型特異遺伝子を発現した確率

(3) ベイズの定理を用いたトリプレットの評価

トリプレット内の3つの事象間のアークの評価値の和をトリプレットの評価値とする。このアークの評価値はベイズの定理から次のように計算する。

◆ A→B

$$P(B|A)=P(A,B)/P(A) \quad P(B|A):アークの評価値$$

$$P(A,B):A \text{ と } B \text{ の同時確率}$$

$$P(A):A \text{ の起こる確率}$$

◆ AB間アーク無

$$\{P(B^c|A)+P(A^c|B)\}/2$$

◆ A→B→C(線形経路)

$$P(C|A,B)=P(A,B,C)/P(A,B) \quad P(C|A,B):線形経路での評価値$$

$$P(A,B,C):A,B,C \text{ の同時確率}$$

◆ A→B←C(合流経路)

$$P(B|A,C)=P(A,B,C)/P(A)P(C) \quad P(B|A,C):合流経路での評価値$$

◆ A←B→C(分岐経路)

$$P(A,C|B)=P(A,B,C)/P(B) \quad P(A,C|B):分岐経路での評価値$$

†九州工業大学,KIT

‡日立ソリューションズ

(4) ネットワーク全体の推定

評価値の高いトリプレットを使って遺伝子調節ネットワーク全体を推定する。この評価値の和をネットワーク全体の評価値とする。ここで、トリプレット間でアークの食い違いがある場合はネットワーク全体の評価値が高くなるようにトリプレットを選択する。

3. 結果

今回、6 遺伝子を重要遺伝子としてネットワークを描いた。この6 遺伝子は分化型サンプルにおいて多く発現していたため、分化型胃癌における遺伝子調節ネットワークを描いた。6 遺伝子を遺伝子 A,B,C,D,E,F とし、分化型の遺伝子調節ネットワークを推定した(図1)。

また、図1と KEGG_Pathway[6]による調節ネットワークを比較した。遺伝子 C と遺伝子 D のノード関係が反対となっていた(図1点線)が、その他のノードに関しては同様の位置関係、方向を示していた(図1実線)。

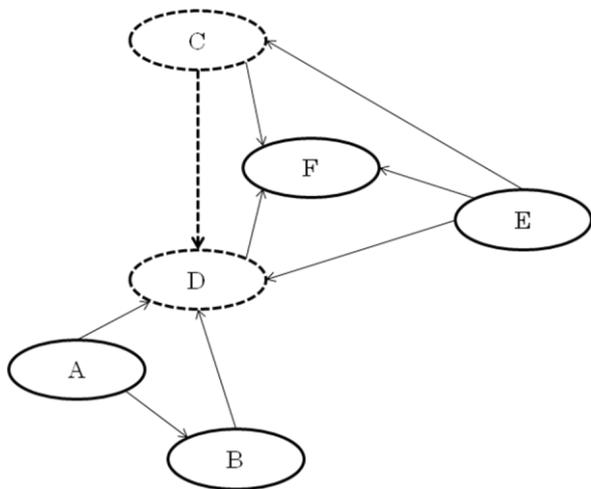


図1 分化型胃癌における遺伝子調節ネットワーク

4. 考察

図1と KEGG_Pathwayのネットワークにより、遺伝子Aの作用から開始されていることが分かる。次いで、遺伝子B,C,Dという順に遺伝子が作用していくと考えられる。推定したネットワークを KEGG_Pathwayのネットワークと比較して、大まかな部分が重なったことから、今回の手法の信憑性が示された。

また、遺伝子Eがアークの開始点となっている。これにより遺伝子Eが遺伝子A,B,C,Dと異なる部分から作用していると考えられる。KEGG_Pathwayのネットワークでも同様に遺伝子Eは他の遺伝子とは異なる経路から発現していた。

今回、推定した図1のネットワークでは遺伝子Cの他に遺伝子Aと遺伝子Bが作用することで遺伝子Dが発現するという結果となったが、KEGG_Pathwayでは遺伝子Dの発現に遺伝子Aは関与していなかった。これは遺伝子

A,C間で今回推定に用いなかった遺伝子が存在し、それが表現できなかったことも考えられる。

また、ネットワークが一致しなかった原因として、今回用いたデータのサンプル数が少なかったことも挙げられる。その理由は本研究では事象について確率を用いて表現するため、推定するネットワークの信憑性を高めるためにはサンプル数が非常に重要となるためである。

5. おわりに

今回、生物学的プロセスを表す遺伝子調節ネットワークを事象の確率から計算するベイジアンネットワークモデルを用いて表現した。ベイジアンネットワークモデルの欠点である非循環型モデルという課題は完全には克服できなかったが、これから手法中の評価値の計算方法や小規模ネットワーク(トリプレット)の組み合わせ方の見直しを行い、この課題を克服することを試みようとする。

また、今回よりも近接した遺伝子のネットワークを推定し、より細かな因果関係を推定できるのか試みたい。そして、遺伝子相互作用の生物学的プロセスと比較しながら、より信憑性のある手法を目指したい。

最後に、このネットワーク推定法を利用して胃癌関連遺伝子から遺伝子調節ネットワークを構築し、サブタイプの識別に有効な指標となるよう試みていきたい。

6. 謝辞

貴重な遺伝子発現データを提供していただいた国立がん研究センター研究所の佐々木博己ユニット長に深謝する。

7. 参考文献

- [1]佐々木博己、青柳一彦、深川剛生、片井均(2009)『マイクロアレイ解析から見てきた未分化型胃癌の発生母地とびまん性浸潤・増殖の分子機構』最新医学社 64 巻 9 月増刊号, pp.111-124.
- [2] 繁榊算男、植野真臣、本村陽一(2006)『ベイジアンネットワーク概説』培風館
- [3]Y.Benjamini and D.Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann Stat, 2001, 29,pp. 1165-1188.
- [4] DAVID Functional Annotation Bioinformatics Microarray Analysis, <<http://david.abcc.ncifcrf.gov/>>(2013年6月アクセス)
- [5]Y.Watanabe et al., An estimation method for inference of gene regulatory network using Bayesian network with uniting of partial problems, BMC Genomics 2012, 13(Supple 1):S12.
- [6] KEGG: Kyoto Encyclopedia of Genes and Genomes, <<http://www.genome.jp/kegg/>>(2013年6月アクセス)