

# SNP および臨床データベースを対象としたハプロタイプ解析による知識発見方式

An Implementation Method of Knowledge Discovery for SNP and Clinical Databases by Haplotype Analysis

吉田 尚史<sup>†</sup> 清木 康<sup>‡</sup>  
藤島 清太郎<sup>¶</sup> 相磯 貞和<sup>¶</sup>

Naofumi YOSHIDA Yasushi KIYOKI  
Seitaro FUJISHIMA Sadakazu AISO

本稿では、SNP データベースおよび臨床データベースを対象としたハプロタイプ解析による知識発見方式について示す。本方式は、個人差を規定する因子として着目されている遺伝子上の多型、特に SNP (Single Nucleotide Polymorphism: 一塩基多型) のデータベースと臨床データベースとの組み合わせを対象とし、全 SNP より遺伝子上に近接して存在する複数の SNP を遺伝子の機能単位で一括抽出するハプロタイプ解析を用いて、SNP と臨床情報間の関連を知識として効率的に抽出する方式である。本方式は、さらに、大量のデータから短い時間で知識発見を行う必要のある場合、および、精密な知識発見を行う必要のある場合のどちらの場合においても効率的な知識発見を可能とする。本稿では、実験により本方式の実現可能性および有効性を検証する。

In this paper we present an implementation method of knowledge extraction for SNP (Single Nucleotide Polymorphism) and clinical databases with haplotype analysis. SNPs are genetic individual variation and will accelerate the identification of disease genes. Our method makes it possible to extract associations between SNP data and clinical data for SNP and clinical databases with haplotype analysis, which extracts sets of SNPs for each genetic function from genetic bases. Our method also enables us to perform both efficient knowledge extraction for large databases and precise data mining. In this paper we clarify the feasibility and effectiveness of our method by experimental results.

## 1. はじめに

現在、ゲノム医学が急激な発展を遂げている。人の持つゲノムの全塩基配列の決定が報告され、主要な研究の対象は、ゲノムの塩基配列の決定から、塩基配列が人に与える影響へと移りつつある。

一般に人の持つ約30億塩基対のうち99.9%は共通であり、0.1%は個人により差異が存在すると言われている。この差異のう

ち個人間において1%以上の頻度で差異が存在し、かつ、1塩基のみ差異がSNP (Single Nucleotide Polymorphism: 一塩基多型) と呼ばれており、個人差を決定付ける因子として重要な研究対象となっている。現在では、ポスト・ゲノム・プロジェクトの一つとして、ヒトゲノム上のSNPバンク構築を目指すSNP Consortium[11]やJSNP[5]において、SNPのデータベースが構築されつつある。

本稿では、SNPデータベースおよび臨床データベースを対象としたハプロタイプ解析による知識発見システムの実現方式について示す。本方式は、個人差を規定する因子として着目されている遺伝子上の多型、特にSNPのデータベースと臨床データベースとの組み合わせを対象とし、全SNPより遺伝子上に近接して存在する複数のSNPを遺伝子の機能単位で一括抽出するハプロタイプ解析を用いて、SNPと臨床情報間の関連を効率的に抽出する方式である。

本方式の特徴は、次の2点に集約される。第1に、SNPおよび臨床データベースの組み合わせを対象として、SNPの膨大な組み合わせからハプロタイプ解析により組み合わせの数を削減し、SNPと臨床情報間の関連を効率的に抽出することを可能とする。第2に、本方式は、大量のデータから短い時間で知識発見を行う必要のある場合、および、精密な知識発見を行う必要のある場合のどちらの場合においても効率的な知識発見を可能とする。

SNPデータベースおよび臨床データベースを対象としてSNPおよび臨床情報間の関連を抽出する場合、相関ルール抽出アルゴリズム[1]の適用が有効である。しかし、単純に相関ルール抽出アルゴリズムを適用する場合、SNPの組み合わせ数が膨大な数となり、現在の計算機で実際に許容できる時間内に相関ルールを抽出することは困難である。SNP - 臨床情報間の関連を抽出することは医学的に緊急の課題である。計算量を削減し、臨床情報と関連する可能性が高い組み合わせのみに着目して相関ルールを抽出する方法が有効である。

本方式は、全SNPより遺伝子上に近接して存在する複数のSNPを遺伝子の機能単位で抽出するハプロタイプ解析を用いて、効率的に相関ルールを抽出する方式である。本方式では、特にSNPを対象とするが、他の遺伝子多型を対象とした場合も一般性を失わない。

従来の遺伝子を解析する方法では、主に機能的アプローチおよび統計的アプローチの2つのアプローチが採用されていた[6]。前者は、遺伝子に含まれるSNPを分子生物学的に分析することによりSNPの機能を解析するアプローチである。後者は、SNPを臨床情報と組み合わせ、統計的に関連を抽出するアプローチであり、遺伝統計学と呼ばれる。遺伝統計学の分野において、遺伝子データベースおよび臨床データベースを対象として、両データベースに潜在する関連を統計的に抽出することが研究されている。遺伝統計学は、両データベースに潜在する関連の傾向を統計的に抽出することを可能としている。これらの従来のアプローチと比較して、本方式の特徴は、SNPデータベースおよび臨床データベースの組み合わせを対象として相関ルールを適用することにより、SNP と臨床情報間の有効なルールを分析的かつ網羅的に抽出可能な点である。

## 2. 遺伝子多型, SNP, および, ハプロタイプ解析

人の遺伝子の塩基の配列において、固体間に1%以上存在する変異の事を遺伝子多型(genetic polymorphism)と呼ぶ。特に、1塩基について存在する多型を一塩基多型(Single Nucleotide Polymorphism, SNP)と呼び、遺伝子多型の中でも特に頻度が高いことから特に着目されている。人における個人差は、このSNPの違いにより大部分が決定されることが推測されている。SNPの遺

<sup>†</sup> 正会員 慶應義塾大学大学院政策・メディア研究科  
[naofumi@sfc.keio.ac.jp](mailto:naofumi@sfc.keio.ac.jp)

<sup>‡</sup> 正会員 慶應義塾大学環境情報学部 [kiyoki@sfc.keio.ac.jp](mailto:kiyoki@sfc.keio.ac.jp)

<sup>¶</sup> 慶應義塾大学医学部 [fujishim@sc.itc.keio.ac.jp](mailto:fujishim@sc.itc.keio.ac.jp)

<sup>¶</sup> 慶應義塾大学医学部 [aizo@sc.itc.keio.ac.jp](mailto:aizo@sc.itc.keio.ac.jp)

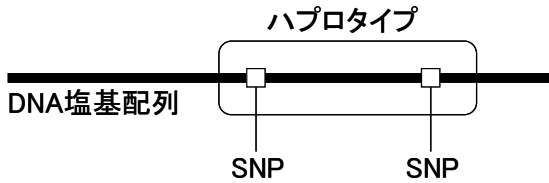


図1 DNA 塩基配列, SNP, および, ハプロタイプ

Fig.1 Base Arrays in DNA, SNPs, and Haplotypes

伝子機能や遺伝子発現に与える影響を解明することにより、個人の体質に合わせたより良い治療法、薬剤の選択や医薬品の開発が可能になると考えられている[7][8][9]。本方式は、このSNPに着目し、臨床情報と組み合わせ有効な関連を抽出することを目的としている。本方式では、全SNPより遺伝子上に近接して存在する複数のSNPを、遺伝子の機能単位で一括抽出した組をハプロタイプ(haplotype)と呼ぶ。図1に、DNA塩基配列、およびSNPとハプロタイプの関係を示す。

### 3. 実現方式

#### 3.1 SNP および臨床データベースを対象とした知識発見システム的方式

本方式は、SNP データベースと臨床データベースを対象として知識発見を行い、SNP と臨床情報間に存在する有用なルールの抽出を目的とする。知識発見においてハプロタイプ解析を用いて、すべての組み合わせを対象として網羅的に解析を行った場合と比較して計算量を削減する方式を設定する。これにより、SNP の組み合わせが多い場合においても実際に許容できる時間内での関連ルール抽出が可能となる。

$X, Y$  を塩基とすると、生物は一般に両親由来の塩基群を有するため、SNPは1塩基部位に対して  $X/X, X/Y, Y/Y$  の3パターンを取る。ただし、 $X, Y$ はA(アデニン), T(チミン), G(グアニン), C(シトシン)いずれかの塩基を現す。

臨床データベースの属性  $i$  が持つ属性値の種類数を  $a_i$ 、臨床データベースの属性数(カラム数)を  $n$  とし、SNP データの属性数(カラム数)を  $m$  とした場合、SNP および臨床情報間の関連ルール抽出において、SNP 全ての組み合わせを反映した関連ルールを抽出する計算量は式(1)となる。ここで  $\bar{O}$  は、関連ルールを抽出するための計算量を示す。ただし、 ${}_x C_y$  は  $x$  個から  $y$  個を選択する組み合わせの数を示す。

$$\bar{O} \left( \prod_{i=1}^n a_i \sum_{j=1}^m {}_m C_j 3^j \right) \quad (1)$$

式(1)は、 $m$  が大きい場合には実際的に許容できる時間内での計算が困難であることを示している。本方式では、SNP の組み合わせを削減するハプロタイプ解析を用いて、関連ルール抽出の計算量を削減する方式を示す。

SNP データベースおよび臨床データベースを対象とした、関連ルール抽出アルゴリズムによる知識発見において、次の4種類の性質を用いて、計算量を削減する。

##### (性質1) 疾患と特に関連が疑われる遺伝子

特定の疾患について、遺伝子の機能的解析によって特に関連が疑われる遺伝子が存在する。その特定の遺伝子上に存在するSNPを対象に、優先的に解析を行う。

##### (性質2) SNP の存在する塩基配列の構成

遺伝子であるDNA塩基配列の構成は次のように定められている。プロモーター領域とは mRNA への転写を制御する領域、またコード領域(エクソン)とはタンパク質をコードする領域である。エクソン間に存在する非コード領域を、イントロンと呼ぶ。本方式で

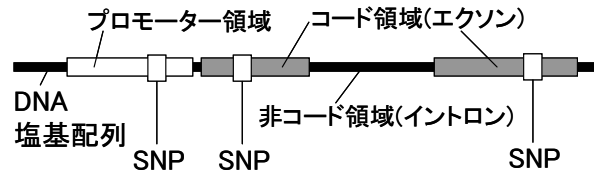


図2 SNP の存在する塩基配列の構成

Fig.2 Structure of Base Arrays with SNPs

は、プロモーター領域を DNA 塩基配列において、コード領域(エクソン)の約 1.5kb 前までと便宜的に定義する。ただし、kb とは千塩基対(kilo base pair)であり、 $p$  kb とは、 $p$  千塩基対の距離を示す。一般的にプロモーター領域およびコード領域に存在するSNPは、それ以外の領域にあるSNPより遺伝子発現量や遺伝子機能に影響する可能性が高いと言える。よって、プロモーター領域とコード領域に存在するSNPを優先して関連ルール抽出を行うことにより、疾患などの臨床情報との関連性を有効に解析することが可能である。図2に、SNPの存在する塩基配列の構成を示す。

##### (性質3) 近傍の遺伝子群と臨床情報との相関

一般に、1つのタンパク質を規定するエクソンおよびプロモーター領域は、同一染色体上に互いに近接して存在する(gene)。よって、各 gene 上に存在し、geneの発現量や機能に影響する複数のSNPを一括しハプロタイプとして分析対象とすることで、臨床情報との関連を有効に解析することが可能である。

##### (性質4) common disease common variant hypothesis

遺伝子が親から子へ遺伝する際、塩基の配列は、近傍にある塩基ほど関連して伝わる。この現象を連鎖不平衡と呼ぶ。ハプロタイプは、連鎖不平衡により生じる塩基の変異の組である。連鎖不平衡により、それぞれのハプロタイプに頻度の差が生じる。一般に高頻度のハプロタイプは、特定の臨床情報との相関が高いと言われている。この現象は、「common disease common variant hypothesis」と呼ばれる[6]。

#### 3.2 本方式の構成

2節における(性質1)~(性質4)までの性質を利用した、SNP データを対象とした関連ルール抽出アルゴリズムを用いた知識発見の方法を、下記のように設定する。関連ルールを抽出するための組み合わせを、Method-1 から Method-5 まで段階的に設定し、関連ルール抽出による知識発見の実行時間と精度を分析者が自由に設定できる環境を実現する。

また本方式ではMethod-1からMethod-5までの各段階において、さらにハプロタイプの性質を2分類し、それぞれの分類に対応する分析方法を次のOption-1およびOption-2として設定する。

#### 3.3 Method-1: 疾患との関連が特に疑われる遺伝子上に存在するSNPとそのハプロタイプを対象とした関連ルール抽出

Method-1では、特定のSNPデータの属性を対象としているため、計算量は少ないが、重要なルールを発見できない可能性がある。臨床データベースの属性  $i$  が持つ属性値の種類数を  $a_i$ 、属性数(カラム数)を  $n$  とし、SNP データの属性数(カラム数)を  $m$  とする。

臨床データベース中の疾患との関連が特に疑われる特定の遺伝子上に存在するSNPを対象とする。疾患との関連性が疑われる特定の遺伝子上にSNPが  $s$  個存在する場合、それらを対象として関連ルールを抽出するための計算量は、式(2)となる。

$$\bar{O} \left( \prod_{i=1}^n a_i \sum_{j=1}^s {}_s C_j 3^j \right) \quad (2)$$

### 3.4 Method-2: 特定のタンパク質をコードするエクソン, およびその前 1.5kb 塩基のプロモーターをハプロタイプと見なしその領域 $k$ 個の gene 内の SNP を対象とした相関ルール抽出

Method-2 では, gene 内のみの SNP の組み合わせを対象としているため, 計算量は少ないが, gene 間にまたがった重要なルールを発見できない可能性がある. Method-2 では, それぞれの gene 内の SNP 群を独立と見なし, SNP の gene をまたがった組み合わせは対象としない.

gene 内の平均塩基数を  $l$  ( $l < m$ ) とすると相関ルールを抽出するための計算量は, 式(3)となる.

$$\bar{O}(\prod_{i=1}^n a_i k \sum_{j=1}^l ({}_i C_j 3^j)) \quad (3)$$

### 3.5 Method-3: Method-2 において $k$ が 2 以上の場合の一定距離内 gene 間の SNP の組み合わせを対象とした相関ルール抽出

Method-3 では, 一定の距離内の gene 間の組み合わせのみを対象としているため, 距離の離れた gene 間に存在するルールを発見できない可能性がある. Method-3 では, それぞれの gene 間において一定の距離以内の組み合わせのみを対象とする.

Method-3 では, Method-2 における  $n$  が 2 以上の場合の gene 間の組み合わせをハプロタイプと見なし, 分析対象とする. この場合, gene 間の距離が一定個  $w$  までの gene をハプロタイプとすると, 相関ルールを抽出するための計算量は, 式(4)となる.

$$\bar{O}(\prod_{i=1}^n a_i (k - (w - 1)) \sum_{j=1}^{i \cdot w} ({}_{i \cdot w} C_j 3^j)) \quad (4)$$

### 3.6 Method-4: Method-2 において $k$ が 2 以上の場合の gene 間の SNP の組み合わせを対象とした相関ルール抽出

Method-4 では, コード領域およびプロモーター領域に存在する全ての SNP の組み合わせを対象とするため, コード領域およびプロモーター領域以外に存在する SNP 間に存在するルールを発見できない可能性がある. Method-4 では, 複数の gene をまたがった SNP のハプロタイプも対象とする.

相関ルールを抽出するための計算量は, 式(5)となる. ただし式(5)は, Method-3 の式(4)において  $w=k$  とした場合と同値である.

$$\bar{O}(\prod_{i=1}^n a_i \sum_{j=1}^{i \cdot k} ({}_{i \cdot k} C_j 3^j)) \quad (5)$$

### 3.7 Method-5: 全 SNP の全ての組み合わせを対象とした相関ルール抽出

Method-5 では, 全 SNP に存在するルールを全て抽出することが可能であるが, 他の Method と比較して計算量が多い.

臨床データベース中の属性と SNP の全ての組み合わせを対象とした相関ルールを抽出するための計算量は, 先に示した式(1)となる.

### 3.8 Option-1: ハプロタイプのうち, 前処理として行われたハプロタイプ解析により得られた高頻度の組を, 分析対象ハプロタイプとして設定した相関ルール抽出

分析対象ハプロタイプの数を  $h$ , そのハプロタイプを構成する SNP データの属性数を  $\#h$ , Method-1 ~ Method-5 の各段階において, 対象 SNP データの属性数を  $x$ , 式(2) ~ 式(6)の

$\sum$  の係数を  $d$  とすると, 相関ルールを抽出するための計算量は式(7)となる.

Option-1 では, あらかじめ高頻度のハプロタイプの抽出が行われていることを前提とする.

$$\bar{O}(\prod_{i=1}^n a_i d \sum_{j=1}^{x-\#h} ({}_x C_j 3^j)^2) \quad (6)$$

### 3.9 Option-2: ハプロタイプの全ての組を分析対象ハプロタイプと設定した場合の相関ルール抽出

この場合の計算量は, Method-1 から Method-5 の各方法における計算量である式(2)~(4), および式(1)と同値となる.

Option-2 では, 2.2 節で示したハプロタイプの抽出が行われていることを前提とする.

## 4. 実験

### 4.1 本方式の検証

本方式の実現可能性および有効性を示すことを目的として, システム実現を行い, 文献[12]を参照して生成した実験データを対象として実験を行った. 具体的には, 本方式による相関ルールを抽出するのに要した時間を計測した. さらに, 文献[12]により正解となる相関ルールを設定し, その正解ルールと比較した再現率および適合率により, 本方式抽出されたルールの妥当性について検証した.

### 4.2 実験結果と考察

全てのハプロタイプを対象とした場合, および頻度の高いハプロタイプのみを対象とした場合の実行時間について, 次のように考察できる. Option-2 では, gene 数の増加に伴い, 急激な実行時間の増加が見られる. それに対して, Option-1 では gene 数に伴って, 実際的に許容できる実行時間の増加に留まる(図3). また, 抽出されたルールの再現率および適合率に関して, 次のように考察できる. Option-1 は Option-2 に比べて, 再現率および適合率ともに低い値を示しているが, gene 数の増加に伴う実行時間の増加が小さい. さらに, Option-1 では, Option-2 と比較して, 抽出されている相関ルールにノイズが少ないことを示している(図4). 以上 2 点は, 組み合わせ数が多く実行時間の点において分析が困難な場合, 本方式の提供する Option-1 により, 困難であった分析が可能となり, さらに抽出されている相関ルールにノイズが少なくなることを示している.

SNP の全ての組み合わせを対象として相関ルール抽出アルゴリズムを適用した場合, 現在の計算機で実際的に許容できる時間内に相関ルールを抽出することは困難である. そのような場合において, Option-1 に示す方法を適用することにより, 臨床情報および SNP データとの実際的に許容可能な時間内での知識発見が可能であると伴に, ノイズの少ない相関ルールの効率的な抽出が可能であることを実証した.

本方式で示している知識発見の方法を用いることにより, 具体的には次に示すような知識発見が有効であると考えられる. 分析対象の SNP データの属性数および gene 数が大きい場合は, まず実行時間が分析対象データの属性数に比例して, ほぼ線形に増加する Option-1 を適用した知識発見を行う. 分析対象の SNP データの属性数および gene 数が比較的小さい場合は, Option-2 を適用して知識発見を行う. 分析対象の SNP データの属性数および gene 数が少量の場合, あるいは網羅的な相関ルール抽出を行う必要がある場合には, Option-2 を適用した知識発見を行うことが適切である.

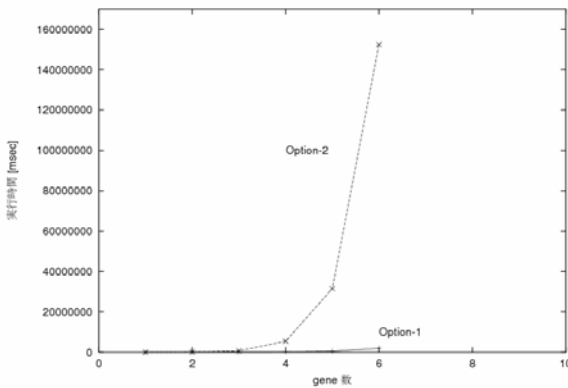


図3 Option-1およびOption-2について gene 数を変化させた場合の相関ルール抽出の実行時間

Fig. 3 Processing Time for Association Rule Extraction by Option-1 and Option-2 for a Variation of Number of Gene

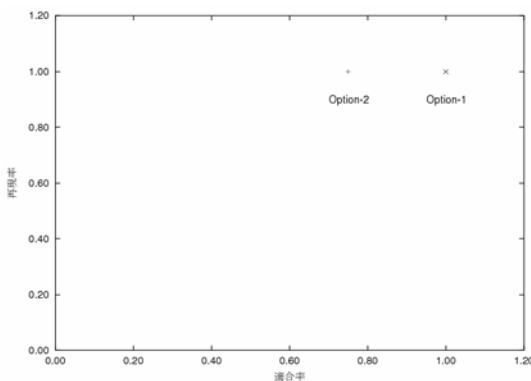


図4 Option-1およびOption-2について gene 数を変化させた場合の再現率と適合率

Fig. 4 Recall and Precision Rate for Association Rule Extraction by Option-1 and Option-2 for a Variation of the Number of Gene

5. まとめと今後の課題

本稿では、SNP および臨床データベースを対象としたハプロタイプ解析による知識発見システムの実現方式について示した。本方式は、個人差を規定する因子として着目されている遺伝子上の SNP のデータベースと臨床データベースとの組み合わせを対象として、相関ルール抽出アルゴリズムを適用することにより、SNP と臨床情報間の相関ルールを効率的に抽出する方式である。本方式は、全 SNP より遺伝子上に近接して存在する複数の SNP を遺伝子の機能単位で一括抽出するハプロタイプ解析を用いることにより、相関ルールを効率的に抽出することを可能とする。

今後の課題としては、本方式の実際の SNP および臨床データベースを対象とした実験が挙げられる。現在は、文献[12]を参照して作成したデータを対象として実験を行っている。次の段階として、実際の医療データである SNP および臨床データベースを対象とした知識発見を行い、その実現可能性および有効性を確認することが必要であると考えられる。さらに、ハプロタイプの頻度やハプロタイプの組み合わせであるディプロタイプに関する研究(例えば[4])を反映した方式の実現、SNP データや臨床情報などの実際の医療データを対象とした場合のプライバシー保護機能を有する SNP データ・臨床情報等の医療データベースの構築、および知識発見のためのシステムの構築が重要であると考えられる。また、現在、国際共同プロジェクト HapMap[3]により、ハプロタイプに関する情報が蓄積されており、これらのデータを活用す

る上でも、本方式が有効に機能すると考えられる。

【謝辞】

システム実現の構築についてご協力頂いた河本穰氏(慶應義塾大学大学院政策・メディア研究科)小川健二氏(慶應義塾大学、現在、京都大学大学院情報学研究所)に感謝致します。

【文献】

[1] R. Agrawal and R. Srikant: "Fast Algorithms for Mining Association Rules," Proc. of the 20th International Conference on Very Large Data Bases, pp.487-489, 1994.  
 [2] Jiawei Han, Micheline Kamber: "Data mining : concepts and techniques", Morgan Kaufmann Publishers, 2001.  
 [3] The International HapMap Consortium: "The International HapMap Project," Nature 426, pp. 789-796, 18 December 2003.  
 [4] Toshikazu Ito, Eisuke Inoue and Naoyuki Kamatani: "Association Test Algorithm between a Qualitative Phenotype and a Haplotype or Haplotype Set Using Simultaneous Estimation of Haplotype Frequencies, Diploptype Configurations and Diploptype-Based Penetrances," Genetics, Vol. 168, pp. 2339-2348, December 2004.  
 [5] JSNP Database: <http://snp.ims.u-tokyo.ac.jp/>.  
 [6] 鎌谷直之(編):"ポストゲノム時代の遺伝統計学", 羊土社, 2001.  
 [7] 中村裕輔(編):"SNP 遺伝子多型の戦略", 中山書店, 2000.  
 [8] 中村裕輔:"先端のゲノム医学を知る", 羊土社, 2000.  
 [9] 中村裕輔(編):"ゲノム解析実験法", 羊土社, 2002.  
 [10] 小川健二, 吉田尚史, 清木康, 藤島清太郎, 相磯貞和:"SNP および臨床データベースを対象としたハプロタイプ解析による知識発見方式とその実現,"電子情報通信学会 第 13 回データ工学ワークショップ (DEWS2002), 2002.  
 [11] The SNP Consortium: <http://snp.cshl.org/>.  
 [12] Matthew Stephens, Nicholas J. Smith, and Peter Donnelly:"A New Statistical Method for Haplotype Reconstruction from Population Data," Am. J. Hum. Genet., Vol. 68, pp.978-989, 2001.  
 [13] 吉田尚史, 清木康, 藤島清太郎, 相磯貞和,"SNP および臨床データベースを対象としたハプロタイプ解析による知識発見システムの実現方式,"情報処理学会研究報告, 2002-DBS-128, 情報処理学会データベースシステム研究会, pp.443-448, 2002.

吉田 尚史 Naofumi YOSHIDA

1996 年筑波大学第三学群情報学類卒業.1998 年同大学院修士課程理工学研究科修了.2001 年同大学院博士課程工学研究科修了.博士(工学).2001 年より慶應義塾大学大学院政策・メディア研究科講師.データベースシステム,マルチメディアシステム,医学情報システムに関する研究に従事.日本データベース学会,ACM,IEEE,情報処理学会各会員.

清木 康 Yasushi KIYOKI

1983 年慶應義塾大学工学研究科博士課程修了,工学博士.同年,日本電信電話公社武蔵野電気通信研究所入所.1984~1996 年筑波大学電子・情報工学系講師,助教授を経て,1996 年慶應義塾大学環境情報学部助教授,1998 年同教授.データベースシステム,知識ベースシステム,マルチメディアシステムの研究に従事.日本データベース学会,ACM,IEEE,電子情報通信学会,情報処理学会各会員.

藤島 清太郎 Seitaro FUJISHIMA

1982 年慶應義塾大学医学部卒業,同大学内科研.1988 年~1991 年米国 Stanford 大学留学.1992 年博士(医学)取得.1997 年より同大救急部講師.同大学病院で患者診療に従事する傍ら,炎症性肺疾患,生体侵襲などの研究に従事.日本救急医学会,日本内科学会,日本呼吸器学会などの各会員.

相磯 貞和 Sadakazu AISO

慶應義塾大学医学部教授.1976 年慶應義塾大学医学部卒業,1980 年慶應義塾大学大学院医学研究科博士課程修了,医学博士.慶應義塾大学医学部助手,専任講師,Stanford 大学医学部微生物学免疫学 Post-doctoral Fellow を経て,1993 年より現職,形態形成学の研究に従事.日本解剖学会,日本内科学会,日本消化器病学会,各会員.