

**Fine-scale Disease Mapping by Exploratory
Pedigree Data Analysis**

February 2013

YUKI SUGAYA

主 論 文 要 旨

報告番号	甲乙 第 号	氏 名	菅谷 勇樹
主 論 文 題 目： Fine-scale Disease Mapping by Exploratory Pedigree Data Analysis (探索的家系データ解析による疾患原因遺伝子座位の高精度な推測)			
(内容の要旨) 本論文は、家系データの尤度を注意深く調べ組み換え確率の最尤推定値を精査することで、これまでより高精度に疾患原因遺伝子座位を推測できることを示すものである。またその推測の前提となる疾患発症確率 ϕ を、疾患情報を観測値とする家系データから推定する方法を開発し、その結果を用いれば座位の推測精度が更に増すことも示している。 第一章は、これまでの家系データ解析では、父親由来、母親由来の組み換え確率 θ_0, θ_1 は同一であると仮定した上で、確率 $1/2$ の場合と比較した尤度比にもとづいてマーカーと疾患遺伝子の近さを推測してきたが、これだけでは高精度の座位の推測は困難であることを過去の研究をふり返りながら解説する。 第二章は、 θ_0, θ_1 を同一であるとせず解析するためには、尤度を θ_0, θ_1 のさまざまな値の組に対して数値的に求めるのでは計算量が膨大になるため、尤度が θ_0, θ_1 の多項式となることに着目し、家系のグラフ表現を利用することにより確率継承アルゴリズムと称する多項式の係数だけを再帰的に求めるアルゴリズムを考案した結果について述べている。 第三章は、確率継承アルゴリズムを実装した実験的な R プログラムを開発し、2種類の疾患家系データをとりあげその原因遺伝子座位を推測した結果について述べている。POAG データ解析の場合には、 θ_0, θ_1 の $[0,1] \times [0,1]$ 上の等高線図から尤度の様子を視覚的に把握することによって、挙動のおかしな推定値を検出することができることを示している。また、Fisher 情報量を計算することで信頼性のある推定値を選別することができ、従来手法では導くことができなかった真の原因遺伝子座位の推測に成功したことも報告している。また FJHN データについても同様の精査法を適用することで、疾患原因遺伝子座位の精密な推測を行なうことができることが判明した。この精査法ではメッシュの幅を細かく区切った尤度と微分の数値計算が必要となるが、得られた尤度多項式からそれらの値を計算することで、すべての計算を一から繰り返し行なう必要がある既存のプログラムに比べて早く実行可能となり、計算時間の点で優位であることもシミュレーションにより示している。 第四章では、疾患情報のみを観測値とする家系データが ϕ の多項式になることを示すとともに、その尤度多項式を確率継承アルゴリズムにより評価し ϕ の推測を容易にするために開発した R のパッケージ MLEP について述べている。尤度関数は3つのパラメータの関数となるため注意深く推定値を検討する必要があるが、パッケージにより尤度多項式が求まったためパラメータの意味を考慮した制約を取り入れるなど、探索的な ϕ の推測が可能となった。また、鳥瞰図による尤度の理解も容易になった。MLEP パッケージにより得られた ϕ の推定値を用いて疾患原因座位の推測を行なうことによって、座位の推測が正しく行われることをシミュレーションにより明らかにした。 本論文は、探索的家系データ解析の手法についてまとめ、 ϕ 及び θ_0, θ_1 の推定値を精査することによって疾患原因遺伝子座位が高精度に推測できることを明らかにしたものである。			

SUMMARY OF Ph.D. DISSERTATION

School of Fundamental Science and Technology	Student Identification Number	SURNAME, First name SUGAYA, Yuki
Title Fine-scale Disease Mapping by Exploratory Pedigree Data Analysis		
Abstract <p>In this thesis careful exploration of pedigree likelihood and investigation of maximum likelihood estimates of recombination fraction leads us to more accurate disease mapping. It is also shown that estimation of penetrance probabilities ϕ from affected status on a pedigree increases the accuracy of recombination fraction.</p> <p>In Chapter 1, a brief history of gene mapping is given. A problem of conventional procedure of disease mapping becomes clear. Although it is common that likelihood ratio test or maximum likelihood estimation procedure is applied under the assumption that paternal θ_0 and maternal recombination fractions θ_1 are equal and reside in $[0, 1/2]$, but such an assumption may mislead us as to the disease locus.</p> <p>In Chapter 2, so-called probability inheritance algorithm is explained which has been developed along with a graphical representation of pedigree to treat efficiently θ_0 and θ_1 as different parameter. In the algorithm, likelihood is evaluated by a single recursive call inheriting coefficient with considering the likelihood becomes polynomial; therefore computational complexity of grid evaluation over lots pairs of θ_0 and θ_1 is reduced.</p> <p>In Chapter 3, an experimental R program implemented probability inheritance algorithm is developed and disease loci of two real pedigree datasets are estimated by the program. In POAG pedigree data analysis, contour plot of the likelihood function on the whole region of θ_0 and θ_1 enables deep understanding the likelihood, resulting in ruling out unreliable estimates. Superimposed Fisher information matrix to the contour plot provides reliable estimates alone, so that POAG disease locus is mapped correctly unlike the original result. For FJHN, the disease locus is also localized clearly. Efficacy of functional evaluation in iterative computations needed for the precise grid evaluation search giving refined estimates of θ_0 and θ_1 is reported by comparing run times with existing software program, in which whole evaluation is necessary for each input.</p> <p>In Chapter 4, MLEP package is developed to support exploring the estimates of ϕ. Likelihood of affected status is explicitly expressed as a polynomial of ϕ, so that the package evaluates its polynomial using probability inheritance, resulting in more flexible exploration such as by introducing feasible parameter constraints. The package also provides a function to draw a perspective plot for understanding the likelihood and validating its convergence. Disease mapping using the MLEP estimates enables feasible localisation of disease locus superior to the analysis using intuitive penetrance values is reported through simulation study.</p> <p>In this thesis, a way of fine-scale disease mapping is described together with its powerful result by a careful exploration of ϕ with θ_0 and θ_1.</p>		