

200401341B

厚生労働科学研究研究費補助金

小児疾患臨床研究事業

小児科診療における効果的薬剤使用のための遺伝子多型  
スクリーニングシステムの構築に関する研究  
(若手医師・協力者活用等に要する研究チーム)

平成14年度～16年度 総合研究報告書

主任研究者：小崎健次郎

平成17(2005)年 4月

## 目 次

### I. 総合研究報告書

小児科診療における効果的薬剤使用のための遺伝子多型スクリーニングシステムの構築に関する研究（若手医師・協力者活用等に要する研究チーム）	小崎健次郎... 1
（資料）薬理遺伝学におけるコンピューティング技術の整備	小崎健次郎... 8
（資料）薬理遺伝学検査の運用	山岸 敬幸... 30
（資料）ジアゾキシドの母集団薬物動態試験法の検討	谷川原祐介... 37
（資料）薬理遺伝学的検査の実施における個人情報匿名化および臨床データ利用の効率化に関する検討	高橋 孝雄... 47
II. 研究成果の刊行に関する一覧表.....	57
III. 研究構成員名簿.....	59

厚生労働科学研究費補助金

(小児疾患臨床研究:小児疾患臨床研究分野 若手医師・協力者活用等に要する研究チーム)

小児科診療における効果的薬剤使用のための遺伝子多型スクリーニングシステムの構築に関する研究 (若手医師・協力者活用等に要する研究チーム)

主任研究者 小崎 健次郎

研究要旨

薬理遺伝学におけるコンピューティング技術の整備

薬理遺伝学的研究においては、遺伝学的データを組み込んだ統計モデル分析(多重回帰、ロジスティック解析、生存分析)、ハプロタイプ解析、母集団薬物動態解析を含むいくつかの応用統計学的手法の習得が不可欠である。また大量のデータを様々な角度から分析するために、1つのプログラムの出力を、他のプログラムの入力として容易に利用できる環境を整備する必要がある。本研究においては、これら解析用コンピューティングの基盤を整備し、若手医師の教育を行った。

薬理遺伝学検査の運用

本研究班では、抗けいれん薬クロバザムについて「CYP2C19の遺伝子多型による薬剤反応性の予測に関する研究」および抗凝固薬ワーファリンについて「CYP2C9の遺伝子多型による薬剤反応性の予測に関する研究」を行った。インフォームドコンセントを実施、検体の匿名化、検体管理の方法、検査等における問題点を検討した。その上で、日本人に認められるTPMT\*3C多型の患者情報を臨床で利用可能にする為、迅速なTPMT\*3C多型スクリーニングシステムを確立した。

ジアゾキシドの母集団薬物動態試験法の検討

ジアゾキシドは、高インスリン血性低血糖症の治療薬として使用されている薬剤であるが、我が国においては未承認である。本薬剤の適正使用を図るためには、血中薬物濃度を測定し薬物動態を検討することが必要である。本研究では、実施可能な例数でPPK解析が行えるかどうか、シミュレーションの手法を用いて検討、ジアゾキシドの血中濃度測定法を確立、臨床試験プロトコルを策定した。

薬理遺伝学的検査の実施における個人情報匿名化および臨床データ利用の効率化に関する検討  
薬理遺伝学的研究を円滑に進めるためには、患者のゲノム情報を的確に解析するとともに、臨床情報を確実に集積することが必要である。European Medicines Evaluation Agency(EMA)の薬理遺伝学における匿名化の定義についてのposition paperを参照して遺伝学研究における匿名化の問題を薬理遺伝学的立場から検討した。個人情報を保護しつつ、臨床情報を集積するシステム開発を推進した。

## 臨床研究指導者

高橋孝雄 慶應義塾大学医学部小児科 教授  
谷川原祐介 慶應義塾大学医学部薬剤部 部長  
小崎健次郎 慶應義塾大学医学部小児科 助教授  
山岸敬幸 慶應義塾大学医学部小児科 専任講師

### A. 研究目的

従来、薬物の副作用や治療効果の予見は困難であると考えられてきたが、近年の薬理ゲノム学の進歩により、薬物代謝酵素の遺伝子多型のうち、低代謝能のアレルを有する患者が、副作用を発症するリスクが高いことが明らかにされてきた。小児においては、副作用がおきた場合には重篤化する可能性が高いため、薬剤投与前に遺伝子多型スクリーニングにより副作用のリスクを予測する、すなわち薬理遺伝学のアプローチの臨床的意義は大きい。

若手研究者が薬理遺伝学的研究を遂行するには、ヒト分子生物学・遺伝子検査診断学の知識に加えて、以下の2分野の研究能力の獲得が必須である。本チームでは、遺伝学・統計学の全般的な知識の獲得とともに下記の分野の教育・研究を行った。とくに薬理遺伝学・統計遺伝学・母集団薬物動態学的解析の2領域について若手医師に対する重点的な実地教育を行った。

薬理遺伝学的解析において、特殊な統計解析が必須である。既報告の海外成人データのメタ解析、遺伝子多型以外の臨床データを検討対象に含めるための多変量解析の手法、複数の遺伝子多型を同時に解析するためのハプロタイプ解析、薬剤反応性遺伝子を網羅するためのDNAマイクロアレイ解析の何れにおいても複雑な統計解析を必要とする。ここ数年、生物統計学者や遺伝統計学者が、新しい統計手法を発表する際、数学理論と同時に実行可能なプログラムを公開することが多くなった。遺伝薬理学では、最近開発された重要な統計的手法がプログラミング言語「S」を用いて記述されている。これを受けて、本臨床治験チーム整備計画においては統計処理用コンピュータ言語としてS言語を選択した。

遺伝薬理学・統計遺伝学においては一つのプログラムからの出力を他のプログラムを用いて処理し、さらに他のプログラムを用いて処理するというように、連続的に一連の処理をする必要がある。このため一連の統計解析に使用される対話型コンピュータ・プログラムではなく、バッチ処理が可能な言語を使用して各プロ

ラムの入出力をインターフェースする必要があった。初年度よりS言語を用いて必要なインターフェースプログラムを作成し、統計処理のためのコンピュータ処理環境を整備している。本年度は電子診療録から、抗癌剤の副作用発症に関する時系列データを系統的に抽出するプログラム・モジュールを開発した。本プログラムはとくに化学療法後の特定のイベントの発生の有無を評価する生存分析のための前処理に有用である。

小児の適切な薬物治療のためには、小児における安全で有効な薬用量について、科学的なエビデンスを得ることが望ましい。本研究では、積極的に母集団薬物動態学的解析法を取り入れ、新薬あるいは既承認薬を小児に適用するにあたり安全で有効な使用法を開発するためのモデルケースを作ること目標としている。母集団薬物動態学的解析法の重要な利点は、複数の要因の影響を同時に検討することができること、被験者の一人あたりの採血回数が少数であっても、多数の被験者からデータを得ることによって解析が可能となる点である。また既承認薬を小児に適用するにあたり、明らかとなっている成人のデータから外挿するためのブリッジングツールそしても期待できる。さらに遺伝的要因を組み込むことにより、個人間変動を説明できればTDMによる1点の測定点から血中濃度を予測でき、年齢層間で特性の異なる小児患者の初期投与設計に役立てることができる。

### B. 研究方法

整備した研究・教育環境を利用して、実施したプロジェクトの方法・成果について以下に要約する。

小児神経病学・小児血液腫瘍学・小児内分泌学・小児循環器病学の各領域をプロトタイプとし、代表的な薬物代謝酵素の遺伝子多型と、臨床情報(年齢・体重・用量・血中濃度、薬効、副作用の発症等)と薬物代謝酵素の遺伝子多型との関連を多変量解析により評価した。評価するアウトカムにより二元配置分散分析(定状態の血中濃度の比較)、混合効果モデル(母集団薬物動態解析)、多重回帰分析(至適用量予測)、ロジスティック回帰分析(単回投与時の副作用発症の有無)、一般化推定方程式(悪性腫瘍の化学療法等、複数回投与時の副作用発症の有無)などの統計手法を利用した。

各遺伝子の翻訳領域を中心として、遺伝子多型を同定し、患者群を対象に熱変性高速液体クロマトグラフィー法を用いてジェノタイピングをおこなった。研究実施期間中に解析した主た

る薬剤とその代謝酵素の組み合わせは、①p450 酵素 CYP2C19-抗てんかん薬クロバザム、②チオプリンメチル転移酵素(TPMT)-抗白血病薬 6-メルカプトプリン、③還元型葉酸キャリアタンパク(RFC1)-抗白血病薬メトレキセートである。加えて、ジアゾキシドの薬物動態試験を推進した。

#### (倫理面への配慮)

文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」、厚生労働省「疫学的手法を用いた研究等に関する倫理指針」に準拠した研究計画に基づき参加施設における倫理委員会の審査を経て実施した。対象者への詳細な説明と同意に充分配慮した。ヒト検体を採取する際に、試料等提供者のプライバシーの保護、検体提供の任意性、提供を受けた検体の取り扱い方、得られる研究成果の医学的貢献度等について、試料等提供者ないしはその保護者に充分説明したうえで、文書により同意を得た。個人情報の外部への持ち出し禁止、試料等の匿名化など個人情報の保護に努めた。必要に応じて遺伝カウンセリングを提供した。

### C. 研究結果

#### 1) p450 酵素 CYP2C19 多型にもとづく抗てんかん薬クロバザムの至適用量の予測

クロバザム(CLB)の活性代謝物である N-デスマチルクロバザム(N-CLB)は、血中半減期が長く、定常状態血中濃度が親化合物に比べ極めて高いことから、CLB 治療における効果や副作用の発症に重要な影響を与える。しかし、血中 N-CLB/CLB 濃度比には大きな個人差が認められるため、治療域の設定が困難である。本研究では CYP2C19 遺伝子多型が N-CLB・CLB 代謝に及ぼす影響について検討した。

1.5~33 歳の 22 名のてんかん患者を対象とし、CLB・N-CLB の定常状態血中濃度を測定した。CYP2C19 の遺伝子型により、対象患者を第 1 群(\*1/\*1: 7 名)、第 2 群(\*1/\*2 または \*1/\*3: 10 名)、第 3 群(\*2/\*2 または \*2/\*3: 5 名)に分け、二元配置分散分析法にて検討した。血中 N-CLB/CLB 濃度比は、第 1 群: 5.4、第 2 群 10.9、第 3 群 32.6 となり、CYP2C19 遺伝子型の影響が顕著であった ( $P < 0.0001$ )。N-CLB 血中濃度/CLB 投与量比 [(ng/mL)/(mg/kg/day)] は各群でそれぞれ 2111、5396、12836 となり、やはり変異アレル数に依存していた ( $P = 0.0001$ )。N-CLB 血中濃度/CLB 投与量比および血中 N-CLB/CLB 濃度比に存在する大きな個人

差は CYP2C19 の遺伝多型により説明できることを示した。

この研究成績にもとづき、クロバザム治療前に CYP2C19 多型検査を実施し、第 3 群における初期投与量を減量して、重篤な意識障害等の副作用の回避に成功した。小児科領域におけるオーダーメイド医療の最初の事例の 1 つと考えられた。

#### 2) チオプリンメチル転移酵素遺伝子変異スクリーニングによる造血障害発症リスクの予測

チオプリンメチル転移酵素(TPMT)は、急性リンパ芽球性白血病治療薬である 6-メルカプトプリン(6-MP)を不活性化する酵素であり、白血病患者が遺伝子変異を有していると、6-MP 治療中に強い造血障害を来すことが知られている。治療前に TPMT 変異陽性者をスクリーニングすることができれば、造血障害の発症を予見し、用量調節が可能になると期待されてきた。しかしこれまで、高効率・高感度のスクリーニング法が開発されていなかった。

本研究では、熱変性高速液体クロマトグラフィー法を利用して、チオプリンメチル転移酵素の全翻訳領域内の遺伝子変異を同時に検出するシステムを開発した。既知の遺伝子変異(13 種類)を全て検出することが可能であることを示した。白人・黒人・アジア人由来の 288 サンプルを解析し、新規の遺伝子変異(非同義アミノ酸置換)を 2 種類、同定した。日本人における遺伝子変異(\*3 変異)頻度は約 1/20 であり、\*3 変異を持つ患者では、造血障害が強い傾向を示した。

本研究により開発した変異解析システムを利用して、関東地区の 4 カ所の主要な白血病治療施設において、新規発症の急性リンパ芽球性白血病症例を対象とした TPMT 遺伝子の治療前遺伝子変異スクリーニングを開始した。

#### 3) 葉酸代謝酵素多型とメトレキセート投与後の副作用の発症

メトレキセート(以下 MTX)は急性リンパ芽球性白血病(以下 ALL)やリンパ腫など小児リンパ系腫瘍の治療に用いられる葉酸拮抗剤である。MTX による毒性の標的臓器は口内炎・嘔吐などの消化器から、皮膚、中枢神経、肝、腎、および造血器(骨髄)まで多様である。各症例における毒性発症の予測は困難である。小児 ALL あるいはリンパ芽球性リンパ腫症例において、葉酸を細胞内にとりこむ輸送蛋白 Reduced folate carrier 1 (RFC1)80G/A 多型と大量 MTX 療法に伴う毒性発症の関連を後方視的に解析した。

1992 年から 2003 年の間に慶應義塾大学病

院で治療が行われた ALL・リンパ芽球性リンパ腫例で、他の細胞障害性薬剤の全身投与による併用なく大量 MTX (3g/m<sup>2</sup>) の投与が行われた症例を研究対象とし、毒性を米国癌研究所診断基準によりスコア化した。15 例における 43 回の MTX 投与を解析対象とした。1) MTX 血中濃度と各種毒性の発症、RFC1 80G/A における変異アレル数と各種毒性の発症について前者を説明変数、後者を目的変数として解析した。A アレル数の増加と NCI-CTC Grade 2 以上の嘔吐発症の減少に有意な関連を認め (Odds Ratio=0.319, P=0.034)。化学療法における副作用の重症度が患者の遺伝的背景に規定される可能性が示唆された。

#### 4) ジアゾキシドの薬物動態試験

ジアゾキシドは、高インスリン血性低血糖症の治療薬として使用されている薬剤であるが、我が国においては未承認である。本薬剤の適正使用を図るためには、血中薬物濃度を測定し薬物動態を検討することが必要である。本研究では、実施可能な例数で PPK 解析が行えるかどうか、シミュレーションの手法を用いて検討、ジアゾキシドの血中濃度測定法を確立、臨床試験プロトコルを策定した。

#### D. 考察

本研究プロジェクト開始時に、若手医師の教育について以下の4点の具体的な目標を設定した。各目標についての達成度は以下の通りである。

① 効率的 (安価かつ迅速) な遺伝子多型のルチーンの検査体制の整備→熱変性高速液体クロマトグラフィー法による遺伝子検査法を実用化し、CYP2C19・チオプリンメチル転移酵素 (TPMT)・SRD5A2 の各遺伝子の変異スクリーニングを、国内外より受託する体制を確立しえた。若手医師は、検査の最適化を通じて、薬理遺伝学の知識と技術を習得した。これらの知識・技術は他の薬剤代謝酵素の基礎研究に直接に応用可能である。

② 薬剤使用禁忌の者 (薬剤代謝酵素の機能喪失変異ホモ接合体) や用量を少なくする必要のある者 (薬剤代謝酵素の機能喪失変異ヘテロ接合体) を治療前に見出し、副作用による被害を未然に防止→関東地区の主要な白血病治療施設において、6-メルカプトプリン治療前 TPMT 変異スクリーニングの運用を開始しえた。若手医師は、検査の社会システム化を経験した。この経験は他の薬剤代謝酵素の応用研究 (トランスレーショナルリサー

チ) に直接に応用可能である。

③ 遺伝子多型を薬物治療に生かすためのガイドラインを策定→CYP2C19 遺伝子多型の変異患者においては、クロバザムの初期投与量を通常量より減量 (1/5-1/10) してスタートすることにより、意識障害等の副作用を回避しうることを示した。これらの解析を通じて、若手医師は、臨床統計学の知識と技術を体得した。これらの知識・技術は他の広汎な分野の臨床研究に応用可能である。

「テーラーメイド医療」という概念が提唱されて数年になるが、これまで小児科臨床の現場で、治療前に薬物代謝酵素の遺伝子検査を実施し、その結果を実地診療に反映させた事例はなかった。上に示した成果は、遺伝子多型を考慮した薬物療法のプロトタイプとして学術的、社会的意義を有する。特に、また本研究で検討した遺伝子変異の一部 (例: CYP2C19) は、欧米においては低頻度だが日本人を含むアジア人においては比較的高頻度 (15%程度) であることから、国際的にも重要な意義を持つと考えられる。

本研究においては、小児神経学領域で抗てんかん薬クロバザム投与前の CYP2C19 検査の、小児血液学領域で抗白血病薬 6-メルカプトプリン投与前の TPMT 検査の検査の意義を示し、効率的に検査を実施するシステムを確立した。今後、同様の手法が、他の小児科領域の薬剤に適用できると期待される。

#### E. 結論

従来の遺伝子多型と薬剤の効果に関する臨床研究は主に欧米人成人に限られていたが、成長につれて代謝経路の遺伝子発現が変化すること、遺伝子多型には大きな民族差があることから日本人小児におけるデータを収集し効率的な遺伝子多型スクリーニング検査体制を整備した。抗てんかん薬クロバザム、抗白血病薬 6-メルカプトプリン、抗白血病薬メトトレキサート、について、各薬物の代謝酵素の遺伝子多型と副作用の発症または薬効との関連を示し、遺伝子多型スクリーニングの利用が、小児科診療における効果的薬剤使用に有用であることを示した。これらの研究を通じて、若手医師は、薬理遺伝学・統計遺伝学・母集団薬物動態学的解析について知識・技術を学び、他の薬剤に関する臨床試験を実施しうる基礎的なスキルを習得した。

#### F. 研究発表

##### 1. 論文発表

Udaka T, Udaka T, Torii C, Takahashi D, Mori T, Aramaki M, Kosaki R, Tanigawara Y, Takahashi T, Kosaki K. Comprehensive screening of the thiopurine methyltransferase polymorphisms by using denaturing high-performance liquid chromatography. *Genetic Testing*, in press.

Kosaki K, Tamura K, Sato R, Samejima H, Tanigawara Y, Takahashi T. A major influence of CYP2C19 genotype on the steady-state concentration of N-desmethylclobazam. *Brain Dev* 26(8) 2004, 530-4.

Soneda S, Fukami M, Fujimoto M, Hasegawa T, Koitabashi Y, Ogata T. Association of Micropenis with Pro185Ala Polymorphism of the Gene for Aryl Hydrocarbon Receptor Repressor Involved in Dioxin Signaling. *Endocr J* 52(1) 2005, 83-8

Sasaki G, Ogata T, Ishii T, Kosaki K, Sato S, Homma K, Takahashi T, Hasegawa T, Matsuo N. Micropenis and the 5alpha-reductase-2 (SRD5A2) gene: mutation and V89L polymorphism analysis in 81 Japanese patients. *J Clin Endocrinol Metab* 88(7)2003, 3431-6

Ishii T, Sasaki G, Hasegawa T, Sato S, Matsuo N, Ogata T. Testosterone enanthate therapy is effective and independent of SRD5A2 and AR gene polymorphisms in boys with micropenis. *J Urol* 172(1) 2004 319-24

Tatami S, Sarashina A, Yamamura N, Igarashi T, Tanigawara Y. Population Pharmacokinetics of an Angiotensin II Receptor Antagonist, Telmisartan, in Healthy Volunteers and Hypertensive Patients. *Drug Metab Pharmacokinet* 18(3) 2003 203-11

Kosaki K, Tamura K, Sato R, Samejima H, Tanigawara Y, Takahashi T. A major influence of CYP2C19 genotype on the steady-state concentration of N-desmethylclobazam. *Brain Dev* 26(8) 2004, 530-4.

Soneda S, Fukami M, Fujimoto M, Hasegawa T, Koitabashi Y, Ogata T. Association of Micropenis with Pro185Ala Polymorphism of the Gene for Aryl Hydrocarbon Receptor Repressor Involved in Dioxin Signaling. *Endocr J* 52(1) 2005, 83-8

Sasaki G, Ogata T, Ishii T, Kosaki K, Sato S, Homma K, Takahashi T, Hasegawa T, Matsuo N. Micropenis and the 5alpha-reductase-2 (SRD5A2) gene: mutation and V89L polymorphism analysis in 81 Japanese patients. *J Clin Endocrinol Metab* 88(7)2003, 3431-6

Ishii T, Sasaki G, Hasegawa T, Sato S, Matsuo N, Ogata T. Testosterone enanthate therapy is effective and independent of SRD5A2 and AR gene polymorphisms in boys with micropenis. *J Urol* 172(1) 2004 319-24

Tatami S, Sarashina A, Yamamura N, Igarashi T, Tanigawara Y. Relationship between pharmacokinetic parameters and occurrence of adverse events in clinical trials performed in Europe and United States for an angiotensin II receptor antagonist, telmisartan. *Drug Metab Pharmacokinet* 19(1) 2004 24-32

Tatami S, Yamamura N, Sarashina A, Yong CL, Igarashi T, Tanigawara Y. Pharmacokinetic comparison of an angiotensin II receptor antagonist, telmisartan, in Japanese and western hypertensive patients using population pharmacokinetic method. *Drug Metab Pharmacokinet* 19(1) 15-23

Sasaki G, Ishii T, Sato S, Hoshino K, Morikawa Y, Kodama H, Matsuo N, Takahashi T, Hasegawa T. Multiple polypoid masses in the gastrointestinal tract in patient with Menkes disease on copper-histidinate therapy *Eur J Pediatr* 163(12) 2004

Udaka T, Udaka T, Torii C, Takahashi D, Mori T, Aramaki M, Kosaki R, Tanigawara Y, Takahashi T, Kosaki K. Comprehensive screening of the thiopurine methyltransferase polymorphisms by using denaturing high-performance liquid chromatography. Genetic Testing, in press.

Kosaki K, Tamura K, Sato R, Samejima H, Tanigawara Y, Takahashi T. A major influence of CYP2C19 genotype on the steady-state concentration of N-desmethylclobazam. Brain Dev 26(8) 2004, 530-4.

Sasaki G, Ogata T, Ishii T, Kosaki K, Sato S, Homma K, Takahashi T, Hasegawa T, Matsuo N. Micropenis and the 5alpha-reductase-2 (SRD5A2) gene: mutation and V89L polymorphism analysis in 81 Japanese patients. J Clin Endocrinol Metab 88(7)2003, 3431-6

Tatami S, Sarashina A, Yamamura N, Igarashi T, Tanigawara Y. Population Pharmacokinetics of an Angiotensin II Receptor Antagonist, Telmisartan, in Healthy Volunteers and Hypertensive Patients. Drug Metab Pharmacokinet 18(3) 2003 203-11

Kosaki K, Tamura K, Sato R, Samejima H, Tanigawara Y, Takahashi T. A major influence of CYP2C19 genotype on the steady-state concentration of N-desmethylclobazam. Brain Dev 26(8) 2004, 530-4.

Sasaki G, Ogata T, Ishii T, Kosaki K, Sato S, Homma K, Takahashi T, Hasegawa T, Matsuo N. Micropenis and the 5alpha-reductase-2 (SRD5A2) gene: mutation and V89L polymorphism analysis in 81 Japanese patients. J Clin Endocrinol Metab 88(7)2003, 3431-6

Tatami S, Sarashina A, Yamamura N, Igarashi T, Tanigawara Y. Relationship between pharmacokinetic parameters and occurrence of adverse

events in clinical trials performed in Europe and United States for an angiotensin II receptor antagonist, telmisartan. Drug Metab Pharmacokinet 19(1) 2004 24-32

Tatami S, Yamamura N, Sarashina A, Yong CL, Igarashi T, Tanigawara Y. Pharmacokinetic comparison of an angiotensin II receptor antagonist, telmisartan, in Japanese and western hypertensive patients using population pharmacokinetic method. Drug Metab Pharmacokinet 19(1) 15-23

Sasaki G, Ishii T, Sato S, Hoshino K, Morikawa Y, Kodama H, Matsuo N, Takahashi T, Hasegawa T. Multiple polypoid masses in the gastrointestinal tract in patient with Menkes disease on copper-histidinate therapy Eur J Pediatr 163(12) 2004

Homma K, Hasegawa T, Masumoto M, Takeshita E, Watanabe K, Chiba H, Kurosawa T, Takahashi T, Matsuo N. Reference values for urinary steroids in Japanese newborn infants: gas chromatography/mass spectrometry in selected ion monitoring. Endocr J 50(6) 2003

小崎健次郎

【遺伝子診療の現状と方向性】各遺伝子診療部の取り組み 慶応義塾大学における取り組み SRL宝函 28巻2号 2004

小崎健次郎

【遺伝学はゲノム情報でどう変わるか ポストゲノム時代を展望する】ゲノム情報と医療 ゲノム情報で医療はどう変わるのか 生物の科学 遺伝 別冊15号 2002

小崎健次郎, 佐藤玲子, 谷川原祐介

【小児薬物療法に関する最近の話題】安全で効果的な小児薬物療法の展望 小児科 44巻9号 2003

島崎紀子, 森鉄也, 吉原宏樹, 嶋田博之, 小崎健次郎, 高橋孝雄  
6MP, MTXによるALL維持療法における休



薬期間と TPMT, MTHFR 多型の関連  
日本小児血液学会雑誌 18 巻 4 号 2004

小崎健次郎, 田村和代, 佐藤玲子, 谷川原  
祐介, 高橋孝雄 CYP2C19 変異によるク  
ロバザム活性中間体(N-デスメチルクロバ  
ザム)定常状態血中濃度の変動 脳と発達  
36 巻 2004

島崎紀子, 森鉄也, 嶋田博之, 小崎健次郎,  
高橋孝雄 メトトレキサート関連肝障害と  
MTHFR C677T 多型 日本小児科学会雑誌  
108 巻 2 号 2004

小崎健次郎, 前山克博, 菅谷明則, 百々秀  
心, 山岸敬幸, 高橋孝雄  
CYP2C9\*3 ヘテロ接合体患者におけるワー  
ファリン投与量のメタ解析  
日本小児科学会雑誌 108 巻 2 号 2004

島崎紀子, 森鉄也, 嶋田博之, 木下明俊, 小  
崎健次郎, 高橋孝雄  
メトトレキサート関連急性毒性と薬物血中  
濃度・個体差の関連 日本小児血液学会雑誌  
17 巻 4 号 2003

佐藤玲子, 谷川原祐介 【臨床に活かす  
PK/PD 薬物の体内動態と薬効・毒性】  
抗菌薬の PK/PD  
医薬ジャーナル 41 巻 1 号 2005

古道一樹, 林拓也, 仲澤麻紀, 土橋隆俊, 福  
島裕之, 山岸敬幸  
ワーファリン内服中, 止血困難な鼻出血を  
呈した CYP2C9\*3 ヘテロ保因者  
日本小児循環器学会雑誌 20 巻 3 号 2004

曾根田瞬, 藤本昌敏, 佐々木理恵, 長谷川  
奉延, 松尾宣武, 深見真紀, 緒方勤  
男児外陰部異常症とダイオキシン アリー  
ル hidrocarbon 受容体(AHR)遺伝子と  
AHR 抑制因子遺伝子多型の相関解析  
日本小児科学会雑誌 108 巻 4 号 2004

古山順一, 黒木良和, 千代豪昭, 藤田潤, 福  
嶋義光, 左合治彦, 松原洋一, 奥山虎之  
遺伝子医療の基盤整備に関する研究(II) 日  
本遺伝カウンセリング学会誌 25 巻 1 号  
2004

三原喜美恵, 田村智英子, 斉藤理恵子, 奥  
山虎之 国立成育医療センターにおける遺  
伝カウンセリング 医療 57 巻増刊 2003

奥山虎之, 小崎里華, 福原康之  
わが国における遺伝子医療の現状  
日本臨床 63 2005

## 2. 学会発表

Kosaki K, Udaka T, Samejima H, Fujita H,  
Yahagi N, Takahashi T. The COPPER plate  
system: Effective use of DHPLC for clinical  
genetics and pharmacogenetics. American  
College of Medical Genetics Annual Meeting,  
March 2004, Orlando, Florida.

Kosaki K, Tamura K, Sato R, Samejima H,  
Tanigawara Y, Takahashi T. A major influence  
of cytochrome 2C19 genotype on the  
steady-state concentration of  
Ndesmethylclobazam. The 9th Conference of  
the European Society of Developmental,  
Perinatal & Paediatric Pharmacology. June  
2004, Marburg, Germany.

## G. 知的財産権の出願・登録状況

特願平 11-357701 変異型 DNA を迅速に検  
出するキット。

(資料)

厚生労働科学研究費補助金

(小児疾患臨床研究：小児疾患臨床研究分野 若手医師・協力者活用等に要する研究)

総合研究報告書

薬理遺伝学におけるコンピューティング技術の整備

小崎 健次郎 (慶應義塾大学医学部小児科 助教授)

研究要旨 薬理遺伝学的研究においては、遺伝学的データを組み込んだ統計モデル分析（多重回帰、ロジスティック解析、生存分析）、ハプロタイプ解析、母集団薬物動態解析を含むいくつかの応用統計学的手法の習得が不可欠である。また大量のデータを様々な角度から分析するために、1つのプログラムの出力を、他のプログラムの入力として容易に利用できる環境を整備する必要がある。本研究においては、これら解析用コンピューティングの基盤を整備する意味で取り組んだ以下の2点について報告する。

①情報解析用コンピュータ言語(S言語)習得のための電子マニュアル作成

②Web公開SNPsデータ(国際Hapmap計画)利用によるSNPs解析効率化

薬理遺伝学的研究を進める上でプログラミング技術は必要不可欠である。同時に、ゲノム情報を含めた、基礎情報の効率的利用を促進するソフトウェア環境を継続的に整備してゆく必要がある。

## A. 研究目的

薬理遺伝学的研究においては、遺伝学的情報と体重等の臨床情報を組み合わせて統計的な評価を行う必要がある。ところが一般的に使用されるSASやSPSSといった市販の統計解析ソフトウェア・パッケージには遺伝学的情報について統計処理するためのルーチンは含まれていないために、研究者が、独自の開発を行わなければならないのが現状である。また大量のデータを様々な角度から分析するために、1つのプログラムの出力を、他のプログラムの入力として容易に利用できる環境を整備する必要がある。対話型コンピュータ・プログラムではなく、バッチ処理が可能なプログラムを利用する必要がある。

本研究では、薬理遺伝学解析の基盤をなすコンピュータ・プログラム環境を整備し、若手医師がR言語の基礎を習得するために必要なトレーニングシステムを明文化し、これらの効果的な利用法について検討

した。

次に、2004年に一般に公表された国際Hapmap計画 (<http://hapmap.jst.go.jp/>) のデータを有効活用し、迅速なSNPs解析に役立てた手法を報告する。Web上における公共ゲノムデータベースの重要性は非常に大きい。これらを有効に活用する手法が確立されることで、薬物代謝関連遺伝子の迅速なスクリーニングが可能になることを期待する。

### 1. R言語コンピューティング

使用するコンピュータ言語として、S言語を選択した。同言語を選択した理由は、インタプリタ型言語で、多変量解析を含む統計計算用の多くの関数があらかじめ含まれている、計算結果のグラフィックス表示が容易である、・探索的なデータ解析に適する、・UNIXのみならずマイクロソフト・ウィンドウズ、マッキントシュにも移植されており、ほとんどのコンピュータで利用可能である、オープンソースのS言語のバージョンであ

る“R”が開発され、各国の統計学者により、新しい統計解析のためのプログラム群が追加されている、などの理由が挙げられる。

本研究においては、慶應義塾大学における臨床研究チームに参加する若手医師の薬理遺伝学的研究能力の向上を目的として、R言語の習得のための電子マニュアルを作成した。R言語の包括的な記述を目的とせず、薬理遺伝学的研究のためにR言語を使用する場合に必要な基礎知識を要約し、実際に慶應義塾大学医学部小児科の臨床治験チームに参加している若手医師の教育目的に使用した。

この電子マニュアルの抜粋版を巻末に記す。以下、項目の概要である。

- ① ゲノム解析データのフォーマット変換
- ② 統計モデル(スプライン関数を用いた多重回帰分析、ロジステック回帰、生存分析)
- ③ ゲノムデータの統計モデルへの組み込み
- ④ ハプロタイプ解析
- ⑤ ゲノム関連研究のメタ解析
- ⑥ マイクロアレーを用いた薬剤投与前後の発現レベルの解析

薬理ゲノム研究においては、遺伝学的データを組み込んだ統計モデル分析(多重回帰、ロジステック解析、生存分析)、ハプロタイプ解析、母集団薬物動態解析を含むいくつかの応用統計学的手法の習得が不可欠である。

S言語(実装はR言語)は薬理ゲノム研究における統計処理において、きわめて有効である。電子トレーニングマニュアルの使用により、大学院レベルの若手医師が、S言語を用いたハプロタイプ解析や統計モデリングなどの高度な処理を短期間で習得可能である。

## 2. Bioconductor

異なる条件下にある遺伝子発現パターンの比較法としてDNAマイクロアレー解析が注目されているが、マイクロアレーのデータを有効に解析するためには、高度な統計処理が必要とされる。最近になり、汎用統計

解析用プログラミング言語「R」を用いたBioconductorというマイクロアレー解析に特化したプログラム群が開発された(ハーバード大学とシアトルのFred Hutchinson Cancer Research Centerのグループによる。[www.bioconductor.org](http://www.bioconductor.org))。

R言語は、上記1で整備したS言語のオープンソースバージョンである。

現在、Bioconductorパッケージの使用法については、英文の断片的な解説文書が発表されているのみである。本研究では臨床治験チームの教育のために日本語テキストを作成した。

Bioconductor パッケージは 3 群から構成される。

- ① affymetrix 社の DNA マイクロアレー・データに対して、バックグラウンドの消去や正規化・各種形式のデータへの変換を行う前処理プログラム群「affy」
- ② 2群における遺伝子発現パターンの差を解析するプログラム群「genefilter」
- ③ マイクロアレーの各プローブに対応する遺伝子についての情報をデータベースから入手し、表示するためのプログラム群「annotate」である。

以下に、それぞれのプログラム群を使用する際に留意すべき点について解説を加えた。

### <Affy パッケージ>

Affy パッケージは affymetrix 社の DNA マイクロアレー・データに対して バックグラウンドの消去や正規化・各種形式のデータへの変換を行う前処理プログラム群である。Affymetrix 社の マイクロアレー上には 1 個の遺伝子に対して複数個のオリゴヌクレオチド・プローブ(25 塩基長)が付加され、全体で数千から数万個の遺伝子に対するプローブが付加されている。各プローブに対してハイブリダイズした cDNA の量をシグナル強度として表現したファイル形式を CEL フォーマットと称する。CEL ファイルを affybatch 形式データに変換する関数が read.affybatch() であり、指定したディレクトリ内に存在する全ての CEL ファイルを変換する。次に、読み

込んだプローブレベルのデータに対して、バックグラウンド処理を行う。

平行して、分析に用いたRNAの質を評価する。先述した通り、Affymetrix社のマイクロアレイでは、各遺伝子について複数のprobeが対応しているために、特定遺伝子の各プローブについて5'端からの距離と、ハイブリダイゼーションによる信号強度の関係を検討することが可能である。5'側のプローブの信号強度が弱ければ、ハイブリダイゼーションに用いたRNAの分解の程度が大きいと判断できる。図1に例を示した。

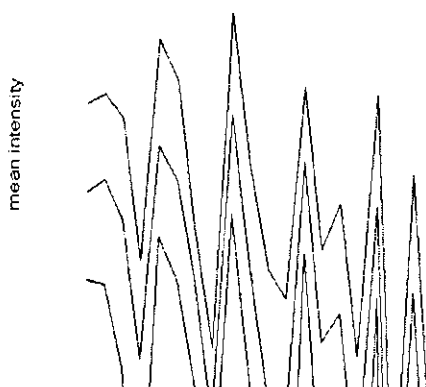


図1

各遺伝子に対応する複数のプローブからの信号データに対して統計学的処理(平均値算定等)を加えて個々の遺伝子の発現量を評価する。個々の遺伝子の発現量は、サンプル調整過程、アレイ作成過程、ハイブリダイゼーション過程におけるばらつきに起因する攪乱因子“obscuring variation”を含んでいる。これらのばらつきによる影響を最小限とするために全遺伝子の発現量の総和を一定値とする正規化処理(normalization)を行う。

得られたデータの分布を確認するために関数 `hist()` を利用する。

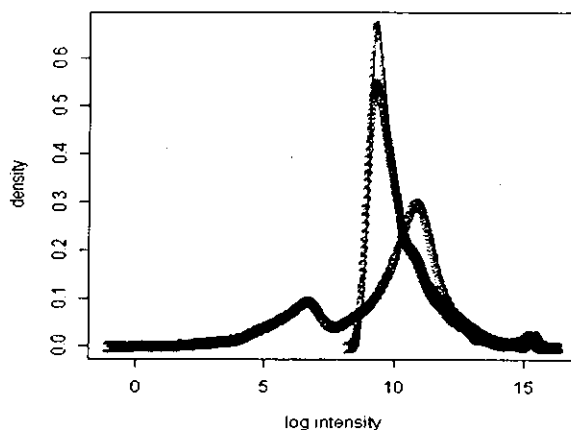


図2. 正規化処理した後の画像を確認するために関数 `image()` を利用する。

Affy パッケージでは、複数のデータ形式が用意されている。最も基本となるデータ形式は `Affybatch` 形式であり、各プローブの信号のみを含んでいる。関数が `write.exprs()` を用いるとデーテキストファイル形式として書き出すことが出来る。プローブデータ以外に臨床症状・重症度・検査値などのサンプルに関する情報の付加された形式を `phenodata` 形式と称し、関数 `read.Phenodata()` により変換することが出来る。

### <genefilter パッケージ>

`genefilter` パッケージは2群における遺伝子発現パターンの差を解析するプログラム群である。

マイクロアレイ・データの解析において、2群における発現の差の検討が頻繁に行われる。2群の比較にさいして、`genefilter()` という関数を使用する。以下のフィルターが利用できる。

`kOverA()` - 発現レベルが値 A 以上の遺伝子を K 個選択する関数

`gapFilter()` - サンプル間の発現レベルの差が特定値以上の遺伝子を選択する関数

`ttest` - t検定によるp値が特定値以下の遺伝子を選択する関数

`Anova` - ANOVA 検定によるp値が特定値

以下の遺伝子を選択する関数  
coxfilter – Cox 検定によるp値が特定値以下の遺伝子を選択する関数

以下にフィルタ関数の利用例を示す。

```
f1 <- kOverA(5, 200) #フィルタ設定を定義
ffun <- filterfun(f1)
which <- genefilter(exprs(eset), ffun) #データにフィルタを適用する
sample <- eset[which,] #条件に合う"True"のものだけを抽出
```

### <annotate パッケージ>

Affymetrix 社のマイクロアレイ上の個々のプローブは AffyID という識別番号で管理されている。この識別番号に対応する NCBI の Entrez (遺伝子データベース), PubMed (文献データベース) などのデータエントリを取得するための関数群が annotate である。

ベクトル mouse1 に調べたい遺伝子の AffyID のリストを格納しておく場合を例示する。NIH GenBank の遺伝子識別番号 LocusID を取得するには、

```
mouse.name <- multiget(mouse1, env = mgu74aLOCUSID)
```

と入力する。同様に染色体上における位置 map position および遺伝子省略名を取得するには

```
map<-multiget(mouse1,mgu74aMAP)
symb<-multiget(mouse1,env = mgu74aSYMBOL)と入力する。
```

上記の方法で得られる、遺伝子識別番号や、染色体上における位置 map position および遺伝子省略名その他必要なデータ項目を HTML 形式の一覧表として出力することが可能である。作成された一覧表内の各プローブデータをクリックすると、関連する Entrez (遺伝子データベース), PubMed (文献データベース) などのデータエントリにジャンプすることが出来る。


```
res<-data.frame(mouse.name,cbind(unlist
```

```
t(symb),unlist(map)))
nres<-data.frame(res,
mouseIP[c(3,5,7,9,11,13,15,17)])
names(nres)<-c("Affy ID","Gene symbol",
"Chromosomal location",
"exposed.1","exposed.2","exposed.3","exposed.4","exposed.5",
"exposed.6","exposed.7","exposed.8")
ll.htmlpage(mouse.name,
filename="Locus1.html",
othernames=nres,
table.head=c("LocusID",names(nres)),
table.center = TRUE)
```

Bioconductor パッケージは R 言語環境上で作動するマイクロアレイ解析用プログラム群である。R 言語には多くの統計解析用プログラム群が多く関数が用意されており、ユーザーがマイクロアレイデータの詳細な解析を行うことができる。しかも若手医師レベルの研究者がアレイを使った薬理遺伝学的研究を遂行することが充分可能であることが示された。R 言語を遺伝薬理学研究の基本ツールとして、採用することの妥当性が確認された。

### 3. 国際 Hapmap 計画について

HapMap プロジェクトは、多型解析・ハプロタイプ解析を通じてヒトの疾患に関わる遺伝子を同定するための基盤整備プロジェクトである。染色体上のある領域について、遺伝子多型に関するアレル頻度・遺伝子型等のデータが簡便に閲覧できる。また、Hapmap 計画が独自に多型頻度を算出した一部 SNPs に関しては、実際の解析データをダウンロードすることも可能である。したがって公開されているデータのみからハプロタイプブロックを予測することもできる(注)。



International  
**HapMap**  
Project

## 国際 HapMap 計画

Home | プロジェクトについて | データ | 文献 | Conference

中文 | English | Français | 日本語 | Yoruba

国際 HapMap 計画は、ヒトの病気や薬に対する反応性に関わる遺伝子を発見するための基盤を整備するプロジェクトです。この計画は、カナダ、中国、日本、ナイジェリア、英国、米国の科学者と各国政府、財団などの協力により行われています。詳細は [国際 HapMap 計画](#) についてをご覧ください。

### プロジェクトの情報

- プロジェクトについて
- HapMap 文献
- HapMap Conference
- HapMap News メールングリスト
- 参加施設, 参加者
- HapMap 日本ミラーサイト

### データ

- Genetic Genome Browser
- データ全体のダウンロード
- ENCODE Project
- データ使用に関するガイドライン

### リンク

### News

- 2005年02月18日: 第 15 回のデータリリースで Han Chinese, Japanese, Yoruba のサンプルについての情報はここから。ジェノタイプデータの一括ダウンロードはここから。データのサマリーは。

Populations	CEU	HCB	JPT	YRI
Genotyped SNPs	1,050,307	1,015,490	1,015,226	1,003,169

- 2005年2月7日: 国際 HapMap プロジェクトはハプロタイプ地図をより高密度に作成することとしました。330万ドルの追加予算により、従来想定していたものより高密度に地図を作成できるようになりました。このマップによって、喘息、癌、糖尿病や心臓病といった疾患に関連する遺伝子をより迅速に同定することができます。
- 2004年12月10日: genotype データにアクセスするための clickwrap agreement は必要なくなりました。ユーザーのあなたは [data usage guidelines](#) を既読するようお願いいたします。参考までに、以前の

図 3. 国際 Hapmap プロジェクト Web サイト

(<http://www.hapmap.org/> 日本ミラーサイト: <http://hapmap.jst.go.jp/>)

(注・これら hapmap 作成のためのサンプルを提供した母集団に関しては、以下の 4 グループが定義されている。

ナイジェリア、イバダンのヨルバ族 (YRI)

日本、東京の日本人 (JPT)

中国、北京の漢民族系中国人 (HCB)

CEPH(北ヨーロッパや西ヨーロッパから来た祖先を持つ米国ユタ州の住民) (CEU)

ヨルバ族については、ドナーには祖父母四人が全員

ヨルバ族であることが必要とされた。漢民族系中国人については、ドナーは四人の祖父母のうち少なくとも三人が漢民族系中国人であることが必要とされた。日本人については、ドナーには祖先が日本人である人からサンプルを収集することが目的である旨が伝えられた。CEPH 母集団における構成員としての基準については、ドナーが全員ユタ州の住民であるという以外には明記されていない。

また、この集団は表現型の情報を収集されていないため、正常対照と見なすことはできない。)

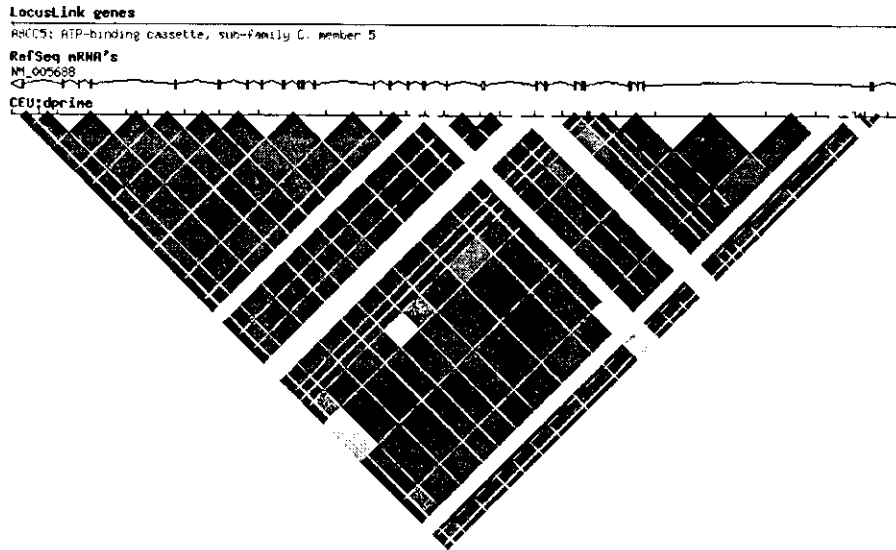


図 4.Hapmap 計画による ABCC5 遺伝子連鎖不平衡地図 (LD Map)

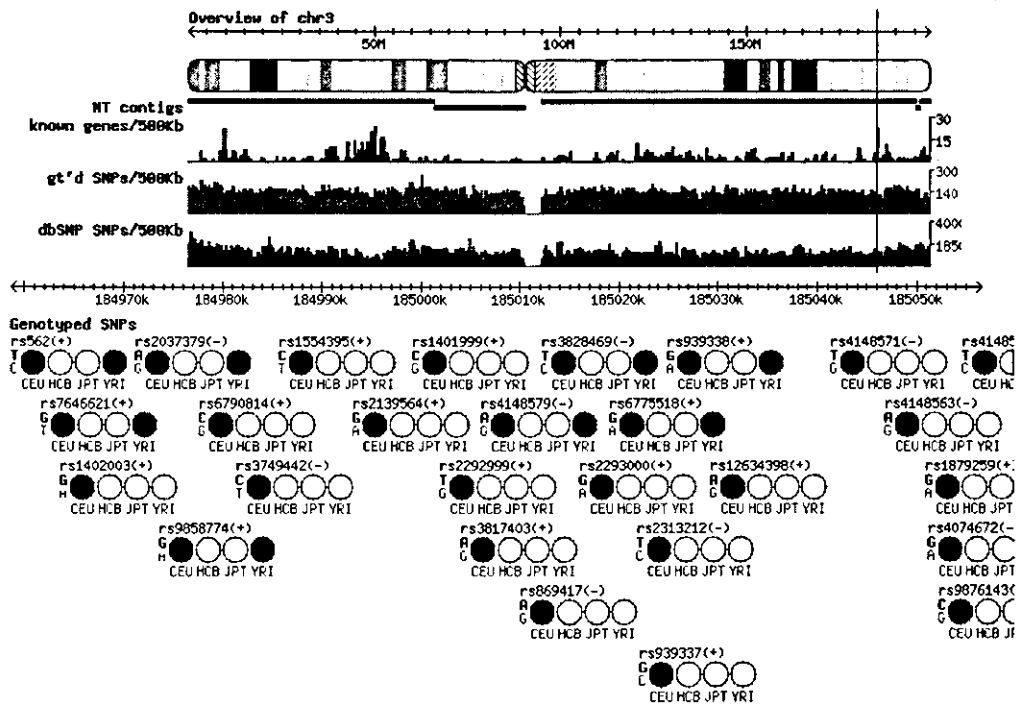


図 5.各 SNPs頻度の独自解析結果

rs#	SNPalleles	chrom	pos	strand	genome_build	center	NA18940	NA18942	NA18943	NA18944	NA18945	NA18947
rs3883674	C/T	Chr10	32360	-	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs10751931	C/T	Chr10	49949	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC
rs12775203	C/T	Chr10	52277	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs2313116	C/T	Chr10	57161	-	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs12218303	A/C	Chr10	58178	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC
rs12781585	C/T	Chr10	64658	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC
rs7909677	A/G	Chr10	65955	+	ncbi_b34	sanger	AA	AA	AA	AA	AA	AA
rs11253113	C/T	Chr10	67224	+	ncbi_b34	sanger	CT	CT	TT	TT	TT	CT
rs11253178	C/T	Chr10	68164	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC
rs4881493	G/T	Chr10	73893	+	ncbi_b34	sanger	GT	GT	GG	GG	GG	GT
rs12413667	A/G	Chr10	73959	+	ncbi_b34	sanger	GG	GG	GG	GG	GG	GG
rs9419537	A/G	Chr10	74660	+	ncbi_b34	sanger	GG	GG	GG	GG	GG	GG
rs12773042	C/G	Chr10	81636	+	ncbi_b34	sanger	GG	GG	GG	GG	GG	GG
rs4242802	C/G	Chr10	87998	+	ncbi_b34	sanger	GG	GG	GG	GG	GG	GG
rs12768206	A/G	Chr10	88767	+	ncbi_b34	sanger	AG	AA	AA	AA	AG	AG
rs7475011	C/G	Chr10	97076	+	ncbi_b34	sanger	CC	CG	GG	GG	CG	CC
rs2379077	C/T	Chr10	99698	+	ncbi_b34	sanger	CT	CT	CC	CC	CC	CT
rs12268912	C/T	Chr10	103173	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC
rs7081047	C/T	Chr10	103465	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs9419561	G/T	Chr10	106552	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs11253553	A/T	Chr10	111441	+	ncbi_b34	sanger	AT	AT	TT	TT	TT	AT
rs4881551	A/G	Chr10	113076	+	ncbi_b34	sanger	AG	AG	GG	GG	GG	AG
rs2379050	A/G	Chr10	116271	+	ncbi_b34	sanger	GG	GG	GG	GG	GG	GG
rs9419418	C/T	Chr10	117536	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs9645540	A/C	Chr10	118417	+	ncbi_b34	sanger	AA	AA	AA	AA	AA	AA
rs4372370	C/T	Chr10	122814	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs7081782	G/T	Chr10	123560	+	ncbi_b34	sanger	TT	TT	TT	TT	TT	TT
rs4469807	A/G	Chr10	125126	+	ncbi_b34	sanger	AA	AA	AA	AA	AA	AA
rs4880781	C/T	Chr10	129653	+	ncbi_b34	sanger	CT	CT	CC	CC	CT	CC
rs7089889	G/T	Chr10	132434	+	ncbi_b34	sanger	GT	GG	GG	GG	GT	GT
rs11250357	C/G	Chr10	136783	+	ncbi_b34	sanger	CC	CC	CC	CC	CC	CC

図 6.各 SNPs頻度の独自解析結果

(データはテキスト形式、gz 圧縮データとしてダウンロード可能。上図はテキストデータをエクセルで表形式に表示している)

従来、薬物代謝関連遺伝子の多型を解析する上では、検討するSNPsを限定するためには、独自に連鎖不平衡地図を作り、多型頻度を調べていく必要があった。しかし Hapmap 計画のデータを効果的に用いることにより、この課程は短縮され、実験を非常に効率的に行うことができる(実際に実験に応用した結果については、平成 16 年度チーム整備報告書「Hapmap 計画データの有効活用による多型解析の効率化」を参照されたい)。

多くの遺伝子多型解析にこの手法を用いることで、SNPs タイピング作業は高速化される。解析の効率化において、Hapmap 計画を利用することはソフトウェア的側面から大きな意義がある。

## B. 参考文献

Bioconductor 公式ウェブサイト

[www.bioconductor.org](http://www.bioconductor.org)

R 言語公式ウェブサイト [www.r-project.org](http://www.r-project.org)

## C. 学会発表

Fujita H, Kosaki R, Suzuki T, Takahashi T, Kosaki K. Bioinformatic screening of xenobiotic responsive elements in search of downstream targets of aryl hydrocarbon receptor: identification of DMRT1 as a candidate. American Society of Human Genetics, October 2002, Baltimore, Maryland. American Journal of Human Genetics 2002;71: A967.



## 補遺

### <薬理遺伝学研究のための R 言語使用法要約>

#### I. 注意すべき点

R 言語を使用する上で、注意すべき点。

##### 1. ソフトインストール時の注意

R インストール

(フリーソフトウェア、対応機種 windows95, 98, 2000)

以下のアドレスにアクセスし、「rwl062.exe」をダウンロード

<ftp://ftp.u-aizu.ac.jp/pub/lang/R/CRAN/bin/windows/base>

指示に従い、セットアップを進める。なおシステムをインストールする場所(フォルダ)は任意だが、その際フォルダ名に日本語を用いないこと。

##### 2. ライブラリの読み込み

既成の関数群がセットになっているものをライブラリーと呼び、必要に応じてデフォルトのライブラリー以外のものを DL し、インストール→ロードする。ネットワークにつながっていれば、DL サイト「CRAN」より R 上で直接呼び出すことが可能。

「ツールバー」の「Packege」より「InstallPackagefrom CRAN」を選択 → 「ツールバー」の「Packege」より「Road Package」を選択

ライブラリがインストールされていない場合にロードのエラー例

```
> library(ISwR)
> Error in library(ISwR) : There is no package called `ISwR`
```

##### 3. 前に実行したコマンドの再利用

↑や↓(方向キー)を使うことで、使用したコマンドをたどることが可能。なおこのコマンド履歴はソフト終了時に保存することができる。

#### 4. 一般的なコマンド関数

合計

```
> sum (ベクトル名)
```

データ数

```
> length (ベクトル名)
```

平均

```
> mean (ベクトル名)
```

平方根

```
> sqrt (4)
```

#### 5. 変数設定

ベクトルなどの変数をユーザが設定する場合任意の名前を付けることができる。基本コマンドは

```
>aaaaa <-10
```

ただし変数名にアンダーバーは使用禁止(「\_」は「<-」の代わりに解釈されてしまう)。dot はいくつでもつけることができる。

```
>height. age. sex. NO1<- c (1, 2, 3, 4)
```

#### 6. ヘルプ

関数に関する説明ヘルプは、コマンド

```
>help (関数名)
```

キーワードをサーチするコマンドは

```
help. search ("keyword")
```

オフラインのヘルプファイルをブラウザから開くには(マニュアル的ヘルプ)

```
help. start ()
```

#### 7. 参考ページ

<http://web.sfc.keio.ac.jp/~watanabe/r.htm>

8. コメントの処理	1	5.1	3.5	1.4
プログラム中にコメント行、処理しない説明文などを入れる場合は、行頭に次のよう	2	4.9	3.0	1.4
#コメントの内容	3	4.7	3.2	1.3
表記すると、その行はプログラム処理されない。	4	4.6	3.1	1.5
	5	5.0	3.6	1.4

## II. データの処理の流れ

### 1. 元データの読みとり

原型となる数値データを読み出す場合、そのデータは CSV 形式でなければならない（※CSV 形式=数値・表などを完全にテキストデータとして簡易的な形式で保存するもの）。またこのデータ（例・ファイル名「iris.csv」）は R のプライマリフォルダにあるものとする。

```
> iris <- read.csv("iris.csv")
> data(iris)
> iris
      Sepal.Length Sepal.Width Petal.Length
1           5.1         3.5         1.4
2           4.9         3.0         1.4
3           4.7         3.2         1.3
```

### 2. データの修正

既に作成され、読み込んだデータを修正する場合

```
> edit(iris)
あるいは
> fix(iris)
```

どちらも data editor が別ウィンドウで立ち上がり、修正可能になる。

### 3. データ操作

データフレームにおける基本操作

```
> data(iris)
> iris
      Sepal.Length Sepal.Width Petal.Length
```

#特定の列を取り出す指定

```
> iris$Sepal.Length
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4
 4.8 4.8 4.3 5.8 5.7 5.4 5.1
 [19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2
 4.7 4.8 5.4 5.2 5.5 4.9 5.0
 [37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1
 4.6 5.3 5.0 7.0 6.4 6.9 5.5
```

固定のデータフレームを操作している場合、以下のコマンドで一度読み込めば、上と同じ操作が短縮できる。

```
> attach(iris)
> Sepal.Length
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4
 4.8 4.8 4.3 5.8 5.7 5.4 5.1
 [19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2
 4.7 4.8 5.4 5.2 5.5 4.9 5.0
 [37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1
 4.6 5.3 5.0 7.0 6.4 6.9 5.5
 [55] 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6
 6.7 5.6 5.8 6.2 5.6 5.9 6.1
```

attach で読み込んだデータフレームを外す場合

```
> detach(iris)
```

#一行一列を抽出

```
> iris[1,1]
 [1] 5.1
```

#一列目全て抽出

```
> iris[,1]
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4
 4.8 4.8 4.3 5.8 5.7 5.4 5.1
```

```
[19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2
4.7 4.8 5.4 5.2 5.5 4.9 5.0
[37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1
4.6 5.3 5.0 7.0 6.4 6.9 5.5
[55] 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6
6.7 5.6 5.8 6.2 5.6 5.9 6.1
[73] 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8
6.0 5.4 6.0 6.7 6.3 5.6 5.5
[91] 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3
5.8 7.1 6.3 6.5 7.6 4.9 7.3
[109] 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7
6.0 6.9 5.6 7.7 6.3 6.7 7.2
[127] 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3
6.4 6.0 6.9 6.7 6.9 5.8 6.8
[145] 6.7 6.7 6.3 6.5 6.2 5.9
```

#一行目全て抽出

```
> iris[1,]
 Sepal.Length Sepal.Width Petal.Length
1           5.1           3.5           1.4
```

#### 4. データの保存

作成したオブジェクト（関数オブジェクト、データオブジェクト）は、自動的にカレントディレクトリに「.RData」として保存されている。

### III. ISWR

基本的な動作とそのおおまかな紹介をする。

まず教科書で用いるライブラリデータ集を取り込む。

```
> library(ISWR)
```

#### 1-1

##### 1.1.1

基本演算——通常の四則演算はそのまま行うことができる

```
> 2+2
```

```
[1] 4
```

```
> 6/2
```

```
[1] 3
```

```
> 3*3
```

```
[1] 9
```

関数を呼び出して行う計算は、引数を () 内に記述する

#指数関数

```
> exp(-2)
```

```
[1] 0.1353353
```

#サイン

```
> sin(5)
```

```
[1] -0.9589243
```

#### 1.1.2

変数の取り扱い——x に変数 2 を指定するコマンド。代入後 x を入力すると代入された変数を確認できる

```
> x <- 2
```

```
> x
```

```
[1] 2
```

```
> x+x
```

```
[1] 4
```

#### 1.1.3

ベクトル（データオブジェクト）の作成。ベクトル概念については 1.2.3 も参照のこと

作成例 1

```
> weight <- c(60, 72, 57, 90, 95, 72)
```

```
> weight
```

```
[1] 60 72 57 90 95 72
```

作成例 2

```
> height <- c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
```

```
> height
```

```
[1] 1.75 1.80 1.65 1.90 1.74 1.91
```

例 1 と例 2 の複合定義

```
> bmi <- weight/height^2
```

```
> bmi
[1] 19.59184 22.22222 20.93664 24.93075 31.37799
19.73630
```

```
19.73630
> t.test(bmi, mu=22.5)
```

## 関数による定義数の操作

### 平均

```
> sum(weight)
[1] 446
> length(weight)
[1] 6
> sum(weight)/length(weight)
[1] 74.33333
> mean(weight)
[1] 74.33333
```

## One Sample t-test

```
data: bmi
t = 0.3449, df = 5, p-value = 0.7442
alternative hypothesis: true mean is not equal to
22.5
95 percent confidence interval:
 18.41734 27.84791
sample estimates:
mean of x
 23.13262
```

### #計算例

```
> xbar<- sum(weight)/length(weight)
> xbar
[1] 74.33333
> weight-xbar
[1] -14.333333 -2.333333 -17.333333 15.666667
20.666667 -2.333333
> (weight-xbar)^2
[1] 205.444444 5.444444 300.444444 245.444444
427.111111 5.444444
> sum((weight-xbar)^2)
[1] 1189.333
> sqrt(sum((weight-xbar)^2)/(length(weight)-1))
[1] 15.42293
```

1.1.5

### グラフィック機能

#### 散布図の作成

```
> plot(height, weight)
```

表示されるポイントの種類は変えることができる。

pchの値はデフォルトで1、2以上を指定

```
> plot(height, weight, pch=2)
```

傾き: 表示された散布図に、任意の一次直線を加える。

```
> hh<- c(1.65, 1.70, 1.75, 1.80, 1.85, 1.90)
```

```
> hh
```

```
[1] 1.65 1.70 1.75 1.80 1.85 1.90
```

```
> plot(height, weight)
```

```
> lines(hh, 22.5*hh^2)
```

#### ヒストグラムの作成

```
> hist(weight)
```

## 1-2

R言語の基本的なアウトライン

あるいは標準偏差は

```
> sd(weight)
[1] 15.42293
```

### 1.1.4

t-検定 (student-t) — bmi は weight/height<sup>2</sup> で定義済みのものをそのまま使用

```
> bmi
[1] 19.59184 22.22222 20.93664 24.93075 31.37799
```