

「テーラーメイド医療を目指したゲノム情報活用基盤技術」
平成 15 年度採択研究代表者

井ノ上 逸朗

東海大学医学部・教授

sub-common disease の感受性遺伝子同定と個人型易罹患性診断への応用

1. 研究実施の概要

2007 年は疾患遺伝子研究にとって特筆すべき年であった。冠動脈疾患、糖尿病などの多因子疾患の感受性遺伝子多型が続々と一流誌に報告され、*Human Genetic Variation* が *Science* 誌の *breakthrough of the year* に選ばれた。それらは HapMap 計画による SNP データベースの充実と、全ゲノム SNP タイピングを可能にしたプラットフォームの寄与が大きい。疾患感受性、薬剤感受性に関与する SNP は個人差医療の重要な鍵を握ると期待される。しかしながら感受性遺伝子を統計遺伝学的に同定できても、寄与度が小さいため、その生物学的機能が捉えにくく、病態との関連を明らかにすることは困難である。また個々人の将来予測に応用することは関連遺伝子の寄与度が小さいため現時点では非常に困難である。本研究は、比較的遺伝背景の強い、脳動脈瘤、男性不妊症、家族性甲状腺癌といった sub-common disease を対象とし、疾患感受性遺伝子を同定する。そして遺伝要因、生活習慣、環境要因の関与を統合した量的形質モデルを構築し、集団から個人への諸因子の関与を明らかにする。そして個人型易罹患性診断への応用を目指し、テーラーメイド医療の実現を目指す。

本年度は脳動脈瘤において Illumina HumanHap300 BeadChip によるゲノム全域アソシエーション・スタディをおこなった。家族歴を有する 300 例の脳動脈瘤患者、200 例の脳動脈瘤を有しない対照で 30 万 SNP を用いて解析した。有意差を有する上位 2000SNP について GoldenGate assay による 2 次スクリーニングを開始したところである（九州大学山本研との共同研究）。また脳動脈瘤に関連する遺伝子多型がヨーロッパ集団で報告されたが ($P = 2 \times 10^{-6}$)、日本人患者 800 例でも同様の結果であった ($P = 2.6 \times 10^{-5}$)。

家族性甲状腺癌について、3 番染色体、6 番染色体に連鎖領域（それぞれ Z score 3.25, 3.13）を認めている。連鎖領域内の遺伝子について遺伝子機能、発現情報などから候補遺伝子を選択し、144 家系サンプルを用いて変異スクリーニングを行ったところ、受容体型

チロシンキナーゼ-RASシグナル伝達系上の候補遺伝子A内に家族性甲状腺癌家系でミスセンス変異が集積していることを見出している。

SNPでの遺伝疫学研究は集団での易罹患性(リスク)を示しており、個人がどの程度病気に罹りやすいかにはまったくアプローチされていない。SNPデータ、生活習慣、臨床検査データを組み合わせ、相互作用を検討する新規モデルを構築することにより個人の疾患予測をおこなう未来型診断システムを提案する。特に個人の易罹患性は連続的な変数を示すという量的形質とそれがある閾値を越えたとき発症するという閾値モデルの考えを基礎におき、そこにvariance component modelを展開することで遺伝情報、生活習慣、臨床データなどを組み込んだ易罹患性モデルを構築し、将来予測へ繋げる。

2. 研究実施内容

患者数の増大が強く懸念されている、アレルギーや自己免疫疾患、さらには糖尿病などの生活習慣病は、単一因子のみによって支配されるものではなく、遺伝的要因(いわゆる「体質」)の他、生活習慣や環境的要素が、相互かつ複雑に影響しながらその易罹患性を決定しており、その複雑さが、これらの疾患の発症メカニズムの解明を困難なものにしている。現在でも、複数の因子による疾患感受性への影響をモデル化した解析は着手されつつある。しかしその大半は、ロジスティック重回帰分析など、従来の統計学的手法を用いたものであり、その特性上、多数の因子を取り上げることや、それら因子を取り巻く複雑な関係のモデリングはほぼ不可能と言ってよい。結果として、根本的な解決策はほとんど講じられていないのが現状である。本研究では脳動脈瘤、男性不妊症、家族性甲状腺癌の網羅的感受性遺伝子同定を目指し、ゲノム情報を基盤に個人の罹患性を予測するプログラムを開発する。

脳動脈瘤の感受性遺伝子同定

本年度はゲノム全域アソシエーション・スタディにより感受性遺伝子同定を目指した。30万SNPをタイピングできるIllumina HumanHap300 duoを用いた。家族歴を有する脳動脈瘤患者300例、非罹患対照200例でゲノム全域アソシエーション・スタディをおこなった。1次スクリーニングの解析プロセスにおいて、Cockran-Armitage検定で最小の p 値を示した21個のSNPに対してTaqManアッセイを用い、脳動脈瘤患者と対照各196例に対して追加タイピングを行った。これは、Parkinson病のGWASを一例として、一般の多因子性疾患では数100サンプル規模の1次スクリーニングでは「トップ10」戦略がうまく機能しない可能性が高いことが示唆されているが、脳動脈瘤の遺伝様式が不明確であるため検証実験を試みたものである。結果、有意差の増強は観察されなかった。対照サンプルをRIKEN JSNP550のサンプル(934例)と拡大して得られた有意性の上位21個のSNPに対しても同様の結果

が得られた。したがって、脳動脈瘤感受性遺伝子同定のためには、包括的な 2 次スクリーニングが必要不可欠であることが明確になった。包括的な 2 次スクリーニングのための SNP 選抜には、単一座位関連検定の結果を用いた。まず、 p -値が 0.008 以下の SNP を抽出し、この集合から SNP 対の遺伝子型の相関係数を用いたクラスター分析（最長距離法）により tagSNP を選定し、GoldenGate アッセイ可能な SNP 2,304 個を 2 次スクリーニングに供した。

日本人検体のみでなく、フィンランド 1500 例、オランダ 1200 例(ともに患者と対照の計)でのゲノム全域アソシエーション・スタディデータの解析を現在担当しており(Yale 大学との共同研究)、日本人と共通するもしくは異なる遺伝要因を明らかにできる。

家族性甲状腺癌の関連遺伝子解析

ゲノム全域罹患同胞対連鎖解析を完了している。SNP タイピング用 DNA マイクロアレイ (GeneChip Human Mapping Xba 50K Array (Affymetrix 社製)) 及び GeneSpring GT (Agilent 社製) (Multipoint nonparametric, Genetic Map; deCODE) を使用し、3 番染色体、6 番染色体に連鎖領域 (それぞれ Z score 3.25, 3.13) を認めている。罹患同胞対の収集を継続しており、未解析の罹患同胞対検体の解析を進めている。連鎖領域内の遺伝子について遺伝子機能、発現情報などから候補遺伝子を選択し、144 家系サンプルを用いて PCR 産物のダイレクトシーケンシング法によるコード領域の変異スクリーニングを行ったところ、受容体型チロシンキナーゼ-RAS シグナル伝達系上の候補遺伝子 A 内に家族性甲状腺癌家系でミスセンス変異が集積していることを見出している。現在、患者、健常者の解析サンプルを追加して変異スクリーニングを継続中である。同時に、連鎖領域を中心に DNA マイクロアレイを用いて体系的な SNP スクリーニングを行なっている。

また、DNA マイクロアレイのデータを用いて、近年その重要性が指摘されているヒトゲノムのコピー数多型解析を行うため、理研ゲノムセンター ゲノム情報グループ 小長谷明彦博士、長谷川亜樹氏と共同で独自アルゴリズムによるゲノム変異部位解析システム VarSearch を開発した。VarSearch は大規模データベースとグラフィックユーザインタフェースを備えおり、数十億のスポット情報を同時に比較することでゲノム上の変異候補を解析することが可能である。また、クラスタリング機能や実験データのクオリティチェック機能を備えた、DNA マイクロアレイデータ統合解析システムである。この VarSearch を用いて腫瘍組織 70 サンプルのゲノム構造異常解析、遺伝子発現解析を行っており、非髄様甲状腺癌発症の分子機構の解明も目指している。

グラフィカルモデリングによる高次元因子解析システムの開発

本研究チームでは様々な理論や手法を取り入れ、それらの特徴を踏まえつつ、多因子疾患を支配する発症メカニズムの解明および個人レベルでの発症予測に向けた、新たな統合解析システム作りを進めている。グラフィカルモデリングは、多変量の関連構造をネット

ワークグラフによって表す手法である。今年度は、そのグラフィカルモデリングの一種である PC (path consistency) アルゴリズムを実装、ソフトウェア化し、同時に、本手法の有効性を検証するため、新潟大学桑野教授との共同研究により遅発性アルツハイマー病 (LOAD) のデータ解析を行なった。その結果、SNP 間の連鎖不平衡のみならず、アポリポタンパク質 E 遺伝子 (*APOE*) と LOAD との関連や、女性だけに認められる、第 10 染色体上の数個の遺伝子多型と LOAD との関連までも再現することができた (投稿中)。

このネットワークグラフが示唆する生物学的な意味については、別途理論的および実験的検証が必要であるが、本手法により、疾患と SNP の相対的な関係性がより理解しやすい形で視覚化され、様々な形で疾患に関与する SNP の網羅的な抽出が可能となる。PC アルゴリズムはまた、大規模なデータにおいても、その効力を十二分に発揮し得るという特長を有している。現在、ゲノム全域に存在する数十万単位の SNP をマーカーとした大規模な関連解析が広く行なわれており、今後、そのような大規模なデータを対象として、多因子疾患を支配する複雑なメカニズムの解明が進められていく中で、非常に有用な手法として、その可能性が期待される。

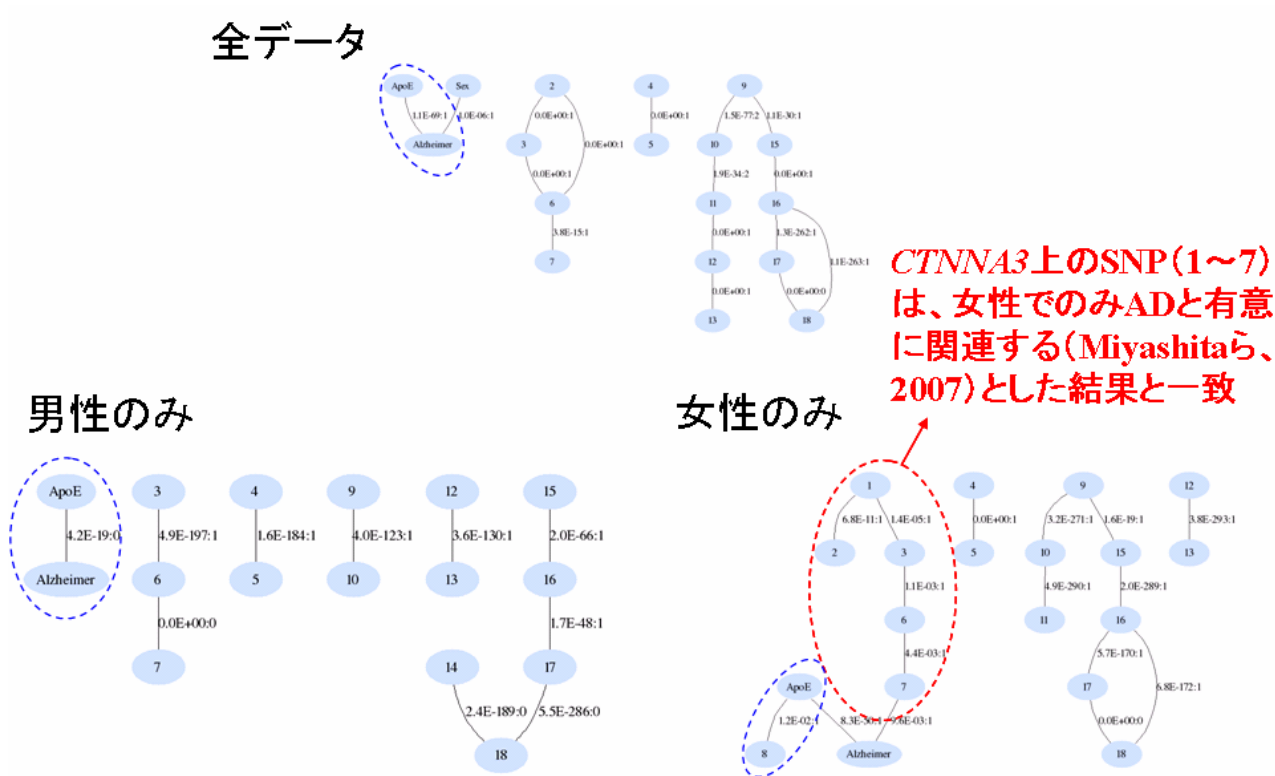


図 1. PC (path consistency) アルゴリズムにより作成された、遅発性アルツハイマー病 (LOAD) と SNP の相対的な関係性を示すネットワークグラフ

SNP 間の連鎖不平衡のみならず、アポリポタンパク質 E 遺伝子 (*APOE*) と LOAD との関連や、女性のみ認められる、第 10 染色体上の *CTNNA3* 遺伝子多型と LOAD との関連まで再現することができた。

臨床データを用いた将来予測プログラム開発

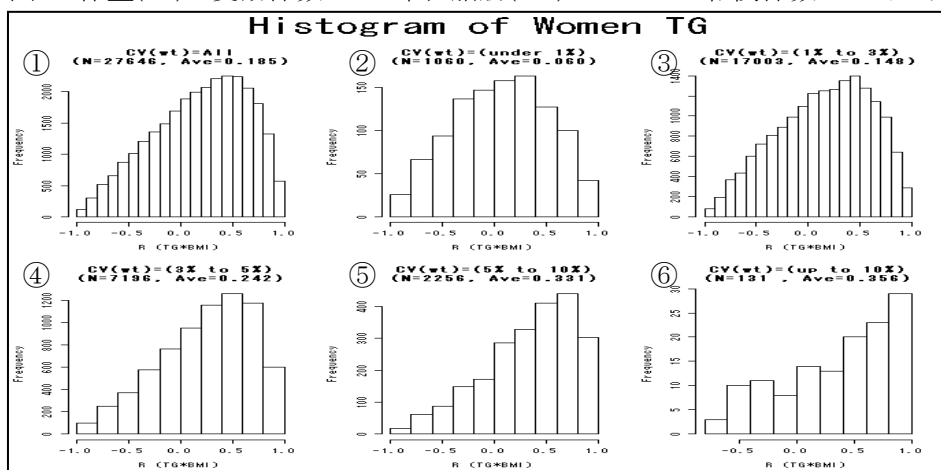
健康診断の経時的な変化から将来の罹患リスク予測や健康状態の判別が可能となれば、健康診断の有用性はさらに高まる。本研究では、匿名化された7万人の健診データを利用し、高血圧、高脂血症、糖尿病等の将来予測プログラムを作成することを目的とし、検証を行っている。

まず全健診データに対して厚生労働省により策定および公表されている『標準的な健診・保健指導プログラム』中の『データ範囲のチェック』に従ってフィルタリングを行い、このうち 6 年以上のデータを有する約 35,000 名を母集団として統計解析を実施した。

これまでの研究で、体重と中性脂肪などの関係を集団で見た場合では相関関係が認められず (0.1~0.2 程度)、また、他の検査項目と合わせたロジスティック回帰分析なども実施したが予測は困難であったが、個人別に 10 年間程度のデータで相関を求めると 32%の人が 0.5 以上の正の相関関係を示していた。そこで、中性脂肪や血圧、HbA1c の推移に対して相関関係をもつ検査項目の選別を行った。その結果、6 年間の体重変動が一定以上となる人については、BMI などとの間に正の高い相関を持つ群が存在することがわかった(図 2)。この群については、大きな体質の変化等が無い条件の基で BMI という個人で計測できる項目から検査項目の変動を把握できる可能性が示唆された。一方、6 年間のうち体重変動があまりない、あるいは正の相関関係を持たないような予測不能な群も存在する。これに加えて過去のデータ数が少ないサンプルに対しても適切な推定結果を表示することが望まれると考え、約 35,000 名のデータに対し年齢や性別、BMI で集団の層別化を行い、予測対象サンプルが該当する集団の分布を予測に適用する方法を検討した(図 3)。このうち、対象サンプルのデータ数が 2 年以上かつ規定年数に満たない対象については、該当する集団の分布を事前分布として Markov Chain Monte Carlo (MCMC) 法を適用し、確率分布として予測式を求めることが効果的と示唆された。

現在これらのアルゴリズムを採用したソフトウェアの開発と予測精度の評価をすすめている。

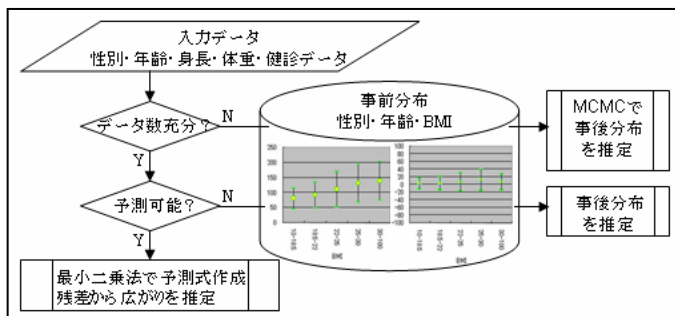
図 2. 体重(wt)の変動係数ごとの中性脂肪(TG)と BMI との相関係数のヒストグラム



①は女性全集団の相関分布、②は体重の変動係数が 1%未満の群、③は変動係数が 1%以上 3%未満の

群、④は変動係数が3%以上5%未満の群、⑤は変動係数が5%以上10%未満の群、⑥は変動係数が10%以上の群についてヒストグラムを表示。変動係数の上昇とともに、分布の最頻値は1に近づく。

図 3. フローチャート



予測不能な群に対しては性別、年齢、BMIにより層別化した集団データ分布を利用して推定結果を表示するためのアルゴリズム。

3. 研究実施体制

(1)「井ノ上」グループ

- ①研究者名:井ノ上 逸朗(東海大学)
- ②研究項目
 - ・脳動脈瘤感受性遺伝子同定
 - ・男性不妊症感受性遺伝子同定
 - ・将来予測プログラム開発

(2)「羽田」グループ

- ①研究者名:羽田 明(千葉大学)
- ②研究項目
 - ・検診臨床データの収集および解析
 - ・将来予測プログラムの開発

(3)「小島」グループ

- ①研究者名:小島 俊男(理化学研究所)
- ②研究項目
 - ・家族性甲状腺癌家系収集、マイクロサテライトタイピングと連鎖解析、関連遺伝子同定

4. 研究成果の発表等

(1) 論文発表(原著論文)

1. Uno Y, Suzuki Y, Wakaguri H, Sakamoto Y, Sano H, Osada N, Hashimoto K, Sugano S, Inoue I. Analysis of expressed sequence tags from liver in cynomolgus monkey (*Macaca fascicularis*): A systematic identification of drug-metabolizing enzyme genes. *FEBS Let* in press.
2. Osada N, Hashimoto K, Kameoka Y, Hirata M, Tamura R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. *BMC Genomics* 9,90, 2008.
3. Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, Asano T, Kenmochi T, Inoue I. Integration of hepatitis B virus DNA into the MPP (mixed lineage leukemia) 4 gene and rearrangements of MLL4 in human hepatocellular carcinoma cells. *Hum Mutat* in press.
4. Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I. Genome-wide expression analyses of testes of non-obstructive azoospermia patients demonstrate a specific gene expression profile and implicate ART3 in genetic susceptibility. *PLoS Genet* 4e26, 2008.
5. Kohara K, Tabara Y, Nakura J, Imai Y, Ohkubo T, Hata A, Soma M, Nakayama T, Umemura S, Hirawa N, Ueshima H, Kita Y, Ogihara T, Katsuya T, Takahashi N, Tokunaga K, Miki T. Systemic multiple candidate gene approach for identification of susceptible genes and susceptible pathways for hypertension: a millennium genome project for hypertension. *Hypertens Res* 31, 203-212, 2008.
6. Nagashima T, Suzuki T, Kondo S, Takahashi K, Ide K, Yumoto N, Hasegawa A, Kojima T, Konagaya A, Suzuki H, Hayashizaki Y, Sakaki Y, Hatakeyama M. Integrative genome-wide expression analysis bears evidence of estrogen receptor-independent transcription in heregulin-stimulated MCF-7 cells. *PLoS ONE*, 3:e1803, 2008.
7. Tomoyasu Y, Yamaguchi T, Tajima A, Nakajima T, Inoue I, Maki K. Further evidence for an association between mandibular height and the growth hormone receptor gene in the Japanese population. *Am J Ortho and Dento Ortho* in press.
8. Saheki T, Iijima M, Li MX, Kobayashi K, Horiuchi M, Ushikai M, Okumura F, Meng XJ, Inoue I, Tajima A, Moriyama M, Eto K, Kadowaki T, Sinasac DS, Tsui LC, Tsuji M, Okano A, Kobayashi T. Citrin/mitochondrial glycerol 3-phosphate dehydrogenase double-knockout mice recapitulate features of human citrin

- deficiency. *J Biol Chem* 282, 25041-254052, 2007.
9. Sasahara A, Kasuya H, Akagawa H, Ujiie H, Kubo O, Sasaki T, Onda H, Sakamoto Y, Krischek B, Hori T, Inoue I. Increased expression of ephrin A1 in brain arteriovenous malformation. *Neurosurg Rev* 30, 299-305, 2007.
 10. Ikeda R, Yoshida K, Inoue I. Identification of FAZF as a novel BMP2-induced transcription factor during osteoblastic differentiation. *J Cell Biochem* 101,147-154 2007.
 11. Onouchi Y, Tamari M, Takahashi A, Tsunoda T, Yashiro M, Nakamura Y, Yanagawa H, Wakui K, Fukushima Y, Kawasaki T, Nakamura Y, Hata A. A genomewide linkage analysis of Kawasaki disease: evidence for linkage to chromosome 12. *J Hum Genet* 52, 179-190, 2007.
 12. Yatsu K, Mizuki N, Hirawa N, Oka A, Itoh N, Yamane T, Ogawa M, Shiwa T, Tabara Y, Ohno S, Soma M, Hata A, Nakao K, Ueshima H, Ogihara T, Tomoike H, Miki T, Kimura A, Mano S, Kulski JK, Umemura S, Inoko H. High-resolution mapping for essential hypertension using microsatellite markers. *Hypertension* 49, 446-52, 2007.
 13. Campos E, Shimojo N, Inoue Y, Arima T, Suzuki S, Tomiita M, Matsuura T, Hata A, Suzuki Y, Aoyagi M, Kohno Y. No Association of Polymorphisms in the 5' Region of the CD14 Gene and Food Allergy in a Japanese Population. *Allergol Int* 56, 23-27, 2007.
 14. Matsuzawa D, Hashimoto K, Miyatake R, Shirayama Y, Shimizu E, Maeda K, Suzuki Y, Mashimo Y, Sekine Y, Inada T, Ozaki N, Nakao Iwata N, Harano M, Komiyama T, Yamada M, Sora I, Ujike H, Hata A, Sawa A, Iyo M. Identification of functional polymorphisms in the promoter region of the human PICK1 gene and their association with methamphetamine psychosis. *Am J Psychiatry* 164, 1105-1114, 2007.
 15. Yamazaki K, Onouchi Y, Takazoe M, Kubo M, Nakamura Y, Hata A. Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. *J Hum Genet.* 52, 575-583, 2007.
 16. Inoue Y., Shimojo N., Suzuki Y., Campos Alberto E.J., Yamaide A., Suzuki S., Arima T., Matsuura T., Tomiita M., Aoyagi M., Hoshioka A., Honda A., Hata A. Kohno, Y. CD14 -550 C/T, Which Is Related to the Serum Level of Soluble CD14, Is Associated with the Development of Respiratory Syncytial Virus Bronchiolitis in the Japanese Population. *J Infect Dis* 195, 1618-24, 2007.

(2) 特許出願

平成 19 年度 国内特許出願件数：1 件（CREST 研究期間累積件数：2 件）