International Symposium on Applied Genomics 2006

主催：文部科学省特定領域研究「応用ゲノム」（平成16－21年度）
「ゲノム」4領域URL：http://genome-sci.jp/
領域3「応用ゲノム」URL：https://www.genome-sci.jp/modules/contents2/rewrite/tc_25.html

Supported by Grant-in-Aid for Scientific Research on Priority Areas,
"New research initiatives in medicine and microbiology based on genome"
from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

平成18年12月14日（木）～15日（金）

グランドアーク半蔵門
「光」（14日）、「富士」（15日）

東京都千代田区隼町1番1号
アクセス：http://www.grandarc.com/access/access.htm

Satellite Symposium
Human Genome, Evolution, and Disease

平成18年12月16日（土）

東京大学理学部新1号館2階　小柴ホール（文京区本郷7-3-1）

「応用ゲノム」
代表：辻 省次（東京大学医学部附属病院神経内科）

［事務局］
東京大学医学部附属病院神経内科
〒113-8655 東京都文京区本郷7-3-1
Tel：03-3815-5411 内線33810
Fax：03-5800-9255
E-mail：genome3-office@umin.ac.jp

［お問い合わせ］
株式会社クバプロ
〒102-0072 千代田区飯田橋3-11-15 UEDAビル6階
Tel：03-3238-1689
Fax：03-3238-1837
E-mail：genome@kuba.jp

予稿集
Proceedings

International Symposium on Applied Genomics 2006

Satellite Symposium

Human Genome, Evolution, and Disease

December 14 (Thursday)

13 :00 ～13 :05 **Opening Remarks** **Shoji Tsuji** (The University of Tokyo, Japan)

1 Integration of Clinical and Genome Informatics (Japanese Session)

司会：春日雅人（神戸大学）／Masato Kasuga (Kobe University, Japan)

15 :05 ～15 :20 **Break**

2 Prospective Cohort Study and Genome Analysis

Chair: Akira Hata (Chiba University, Japan)

December 15 (Friday)

3. Structural Variation in the Human Genome

Chair: Ituro Inoue (Tokai University, Japan)

9:00～9:40 Discovery of Structural Variation in the Human Genome: New Insights for Disease Study. ……20
Stephen W. Scherer, Ph.D. (The Hospital for Sick Children, Toronto, Canada)

9:40～10:20 Copy number variation in the human genome: Implication to diagnostics and cancer research ……21
Charles Lee, Ph.D. (Harvard Medical School, USA)

10:20～11:00 Gene copy number variation and susceptibility to common diseases ……22
Tim Aitman (Medical Research Council and Imperial College Faculty of Medicine, UK)

11:00～11:10 **Break**

Chair: Tatsushi Toda (Osaka University, Japan)

11:10～11:40 Detecting copy-number variation in the human genome using BAC-array based comparative genomic hybridization. ……23
Issei Imoto (Tokyo Medical and Dental University, Japan)

11:40～12:10 Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays ……24
Hiroyuki Aburatani (The University of Tokyo, Japan)

12:10～13:00 **Lunch**

4. Strategy for Genomics Research of Mental Disorders

Chair: Tadafumi Kato (RIKEN Brain Science Institute, Japan)

13:00～13:40 Theoretical basis of genome-wide association studies and its application ……26
Jurg Ott, Ph.D. (Rockefeller University, USA)

13:40～14:20 A potential role for copy number polymorphisms in the etiology of bipolar disorder ……27
Francis J. McMahon, MD (Genetic Basis of Mood & Anxiety Disorders, NIH/NIMH, USA)

14:20～14:30 **Break**

Chair: Yuji Okazaki (Mie University, Japan)

14:30～15:10 Approaches to map genetic underpinnings of schizophrenia and mood disorders ……28
Takeo Yoshikawa (RIKEN Brain Science Institute, Japan)

15:10～15:50 Genomic studies of bipolar disorder: from gene expression to genetics and epigenetics ……29
Tadafumi Kato (RIKEN Brain Science Institute, Japan)

15:50～16:00 **Closing Remarks** Yuji Okazaki (Mie University, Japan)

December 16 (Saturday)

Kosiba Hall,
Faculty of Science,
The University of Tokyo

Satellite Symposium

"Human Genome, Evolution, and Disease"

10 : 00 ～10 : 05 **Opening Remark** Shoji Tsuji (The University of Tokyo, Japan)

Session I Chair: Naruya Saitou (National Institute of Genetics, Japan)

10 : 05 ～10 : 40 Genomic evolution of human nervous system-specific genes and its implication to methods for hunting neuropathological disease-sensitive genes
Takashi Gojobori (National Institute of Genetics, Japan) …… 32

10 : 40 ～11 : 15 Evolution and linkage disequilibrium: implications for common diseases
Lynn B. Jorde (University of Utah School of Medicine, USA) …… 33

Session II Chair: Takashi Gojobori (National Institute of Genetics, Japan)

11 : 15 ～11 : 40 Utility of chimpanzee and other non-human primate genomes for understanding evolution of modern human
Naruya Saitou (National Institute of Genetics, Japan) …… 34

11 : 40 ～12 : 15 Issues and Approaches in the Population Genomics Paradigm of Studying Complex Diseases
Ranajit Chakraborty (University of Cincinnati College of Medicine, USA) …… 35

12 : 15 ～13 : 30 **Lunch**

Session III Chair: Ituro Inoue (Tokai University, Japan)

13 : 30 ～14 : 05 Molecular Genetics of Autism
Stephen W. Scherer, Ph.D. (The Hospital for Sick Children, Toronto, Canada) …… 36

14 : 05 ～14 : 40 Genome-wide scan of disease genes by association analysis using microsatellites
Hidetoshi Inoko (Tokai University School of Medicine, Japan) …… 37

14 : 40 ～14 : 55 **Coffee break**

Session IV Chair: Hidetioshi Inoko (Tokai University School of Medicine, Japan)

14 : 55 ～15 : 30 Hotspots for copy number variation in humans and chimpanzee
Charles Lee, Ph.D. (Harvard Medical School, USA) …… 38

15 : 30 ～15 : 55 Natural selection and genes for human common disease
Toshiaki Nakajima (Tokyo Medical and Dental University, Japan) …… 39

15 : 55 ～16 : 00 **Closing Remark**
Takashi Gojobori (National Institute of Genetics, Japan)

Integration of Clinical and Genome Informatics
—Japanese Session

1

December 14
Thursday

司会 : **春日雅人**（神戸大学）
Chair : Masato Kasuga（Kobe University, Japan）

永井 良三
Ryozo Nagai
●東京大学大学院医学系研究科
The University of Tokyo Graduate School of Medicine, Japan

統計解析可能な電子カルテとゲノム研究への応用

Development of the electronic medical record-linked database system for clinical practice and genome research

医療の効率化や安全な医療を提供するために、数年前から医療の電子化が強く求められるようになっている。既に多くの病院やクリニックでは、様々な血液検査や画像検査のオーダーと結果の参照が電子化されており、また電子カルテの導入も急速に進んでいる。その結果として、医療機関受診のたびにコンピュータで処理する無数のデータが生み出されるようになっている。ところが、大多数のデータは、その都度利用されるだけで、後に解析できるような形では蓄積されていない。例えば、血液検査の結果は、検査のオーダーに対して結果を返す形でしか用いられず、電子カルテ上の病名や病気の状態のデータと組み合わせて解析することができないのが現状である。現在の医療で最大の問題である生活習慣病は、遺伝的背景に様々な環境要因が組み合わさることによって発症し、また異なった経過をたどる。生活習慣の改善や各種薬物の効果も、個々人によって異なる。したがって、個々人に最適で安全な医療を提供するためには、多様な臨床情報を詳細に検討することが必須であり、このような臨床情報の解析をゲノム解析と組み合わせることによって、多数の遺伝子の相互作用で発症すると考えられる生活習慣病のメカニズムや治療にも迫れると考えられる。我々は、正確な医療情報を収集し、全ての情報をリアルタイムに解析可能なシステムを確立し、運用してきた。このシステムは、心臓カテーテル検査などの医療現場での詳細な情報収集、正確な病名や患者背景の登録、HL7 標準プロトコールを介した病院情報システムから検査や薬物情報の取得、詳細な統計解析と外来などの医療現場で簡便に活用できる情報提供システムから構成されており、日常臨床に用いることも、詳細な臨床情報の解析やゲノム解析に用いることもできるようになっている。また、明確な診断クライテリアと標準病名を設定し、医療情報の構造化・可視化を進めている。今後、糖尿病など他の生活習慣病へと展開するとともに複数の医療機関を結ぶことにより、安全で有効な個別医療を推進することができると考える。

Professor／Cardiovascular Medicine／The University of Tokyo Hospital／http://plaza.umin.ac.jp/nagai/

鎌谷 直之
Naoyuki Kamatani

●東京女子医科大学膠原病リウマチ痛風センター
Institute of Rheumatology, Tokyo Women's Medical University, Japan

ゲノム情報に基づいた個別化医療の展望

Prospect of personalized medicine based on genome information

はじめに　関節リウマチの治療は疾患修飾性抗リウマチ薬（DMARDs）と生物学的製剤の導入により大幅に改善した。しかし、治療無効例、副作用例が少なからず存在することが大きな問題となっている。個別化医療（オーダーメイド医療）の実現可能性について関節リウマチ患者を対象に研究を行った。

個人のゲノム情報を基礎に医療介入を行うためのロードマップ　個人間のゲノム配列の違いに関する情報が蓄積され、1人当たり50万のSNP情報が明らかになる時代を迎えた。そのような技術的進歩に基づき、個人間のゲノム情報の差異（ゲノム多型）と疾患や薬物反応性との関連の証拠が次々に発表されている。しかし、これらの情報を実際に診療に応用するために何が必要かについての合意はできていない。患者のゲノム情報と薬物反応性、合併症の関係に基づいて、ゲノム情報を基礎に医療介入を行うために次の条件をクリアする必要がある。即ち、(1) 正当な遺伝統計学を用いた解析により個人のゲノム情報と薬物反応性の間に関連があること（仮説検定）、(2) 独立の集団を用いた疫学研究で同じ結果が出ること（再現）、(3) ゲノム情報に基づく医療介入のアルゴリズムが構築でき、それが患者に有利であると予想されること（アルゴリズム）、である。

条件をクリアした4つの項目　次の4項目が上の3つの条件をクリアした。即ち、(a) メトトレキサートの副作用を予測する（MTHFR遺伝子のC677T多型を用いる）、(b) メトトレキサートの必要用量を予測する（MTHFR遺伝子のA1298C多型を用いる）、(c) スルファサラジンの副作用を予測する（NAT2遺伝子のハプロタイプを用いる）、(d) 合併症アミロイドーシスの発症を予測する（SAA1遺伝子の−13C/T多型を用いる）である。それに基づいて、個別化医療を始めた。

個別化医療の実績　これまでに約240名の関節リウマチ患者について、上記の4項目の個別化医療を約1年間行っている。この間に3名のスルファサラジンの重症副作用患者が入院し、そのうち2名はNAT2遺伝子のリスク型であった。リスク型にはスルファサラジンの投与を控えるというアルゴリズムであったので、重症副作用の2/3はオーダーメイド医療の実施により予防できたと考えられる。オーダーメイド医療を行った患者の中からはスルファサラジンの重症副作用は起きていない。

おわりに　関節リウマチの個別化医療は重症副作用の予防に極めて有効である事が示された。ゲノム情報に基づいた個別化医療の実現のためには、信頼できる遺伝統計学を基礎にしたロードマップに基づく必要がある。

Professor, Director／Institute or Rheumatology／Tokyo Women's Medical University

大江 和彦
Kazuhiko Ohe

●東京大学大学院医学系研究科医療情報経済学分野
Department of Medical Informatics and Economics, Graduate School of Medicine, The University of Tokyo, Japan

電子カルテデータベースの臨床ゲノム研究への活用と課題

Implications for applying EMR-database to clinical genome research

大規模病院を中心に電子カルテ（EMR）の導入が進みつつあり、日常診療における検査データ、症状所見、画像診断レポートなどに加えて診療経過のテキストが電子的処理可能な形で蓄積されつつある。さらに、こうした状況のもとで従来の検査データや画像に加えて診療上の意思決定に有用な遺伝子型が検査されるようになり、これらのデータも同時に管理し診療に活用する情報システムが必要になってきた。たとえば東大病院ではpharmacogenomicsworkingグループでの検討を経て、2006年8月からCYP2C19のオーダリングシステムと結果データの管理が開始している。このシステムでは、特定の診療科の胃枝のみがオーダーを仮登録することができ、その情報にもとづいて薬剤部で患者に説明を行い同意をとったことが入力される。システムからは患者IDをシステム管理者だけが把握する方法で匿名番号化したラベルが出力され、このラベルを貼布した採血管のままで検査部技師により検査され結果がシステムに入力される。このように検査スタッフは匿名化された状況で検査データを取り扱う。時間により変動する従来からの検査値と異なり遺伝子多型情報の組み合わせは時間経過に影響を受けない個人識別情報としての性格を有すると同時に、未知の健康状態の予測と高い相関を有する可能性があるので、そのデータの取扱いにはこのように特段の留意を要すると考えられるが、システム運用面でも情報管理技術面でも試行錯誤の域を出ていない。

一方、既存の電子カルテデータベースからデータマインニングなどの手法を用いて使用薬剤ごとの検査データの推移を解析したり、特異的な薬剤反応性を来たす患者を自動的に発見することによって、新たな遺伝子型を持つ患者集団を抽出することは、今後の臨床ゲノム研究にとって非常に重要であると考えられる。こうしたアプローチを可能とするには、日常診療で蓄積される電子カルテデータをデータマインニング可能な形態のデータベース構造に変換することが必要である。発表者はこれを実現するため2006年4月から日常診療で発生するすべての検査データをリアルタイムで臨床研究解析が容易なデータベース構造に変換し蓄積を開始している。本発表ではこのデータベース構造の概略とその解析可能性についても論じる。今後の課題として、注射や服薬データと関連づけた検査データの時系列データ解析手法、副作用など特定の患者に特定的と考えられる臨床的反応事象の検出をいかに高感度で実現するか、そのために最低限どの程度の入力をどのタイミングでできるようなシステムが現実的かなどについても研究が必要である。

Professor／Department of Medical Informatics and Economics／The University of Tokyo／
http://www.m.u-tokyo.ac.jp/medinfo/

清水 俊郎
Toshiro Shimizu

●株式会社エスビーエス情報システム
SBS Information Systems, Japan

電子カルテからの情報抽出の現状

Extraction of data from EMR

臨床情報の統計解析のためには、大量の臨床データを格納し分析するための強力なデータベースが必要であることは言うまでもありません。しかし、これ以上に重要なのは、このデータベースに情報を取り込むためのお膳立て―すなわち臨床情報を容易に収集し、収集された情報を柔軟にデータベース化するための仕組みです。

病院情報システムはベンダ毎、もしくは施設毎に異なる仕様に基づいたシステムが導入されているのが現状です。これは、日本における電子カルテをはじめとする医療情報システム導入・発展の歴史的経緯に起因するものですが、異なる言語を使用している民族間ではコミュニケーションと情報収集が困難であるのと同様の状態であり、臨床情報の統計解析の進展を阻んできた大きな原因の一つであると考えられます。さまざまな理由から施設毎に異なる医療情報システムを用いなければならない事情はあるとしても、複数の施設からのアウトプットを統一的に解釈できるような機能実現が急務であり、このためには異ベンダ・異システム間で標準語にてコミュニケーションが取れるようにする、すなわち医療情報の標準化に対応するための機能が必要不可欠です。

電子カルテの機能と範囲を如何に定義するかは諸説があるところでありますが、ここでは大まかにオーダエントリと診療記録管理に大別して臨床情報の標準化とデータベース化、情報の抽出の現状を説明します。

オーダエントリは医療従事者による診療行為に関する各種の指示入力を行うとともに、調剤・検査・RIS 等のさまざまな部門システムとの連携を司り、診療行為の履歴や各種検査の結果を管理するものです。ここでは、患者基本情報・病歴・処方歴・臨床検査結果、等に関する診療情報がHL7 にて標準化され、このアウトプットをデータベースに取り込むことにより、柔軟且つ迅速なデータ抽出を行う様を、静岡県版電子カルテプロジェクトにおける成果を例に実演します。

Chief Officer／Healthcare System Division／SBS Information Systems Co.,Ltd／http://www.sys.sbs-np.co.jp

Prospective Cohort Study and Genome Analysis

2

December 14
Thursday

Chair : Akira Hata (Chiba University, Japan)

Hilary Burton ●Public Health Genetics Unit, Cambridge, UK

Public Health Genomics: translation in action. The experience of the Public Health Genetics Unit, Cambridge UK

Public health genomics is defined internationally as the 'responsible and effective translation of genomebased knowledge and technologies for the benefit of human health'. In this presentation I will provide background information on the concept of Public Health Genomics, describe the work of the Public Health Genetics Unit in Cambridge, and show how this contributes to new international networks concerned with this area.

Research in genetics and molecular biology, and, in particular the information emerging from the Human Genome Project, offers new opportunities for the promotion of population health. Such benefits may arise through more accurate diagnosis, disease treatments with better specificity, innovative drug therapies and screening programmes. In the future the possibility is presented of personalised preventive care and lifestyle advice.

The volume and complexity of new genomic information presents an enormous challenge in collecting, analysing and evaluation to identify potential beneficial interventions. This will require an integrated multidisciplinary effort including epidemiological studies in multiple populations and the findings from other disciplines such as social sciences, psychology, ethics, political science and law. Not only must the science be robust, but it must also be used in a way that is acceptable to society, including its legal and regulatory frameworks.

Public Health Genomics has knowledge integration at its heart accompanied by a process through which this knowledge is used to change practice. Important aspects include informing public policy, developing and evaluating health services (both preventive and clinical), communication and stakeholder involvement and education of health professionals and generally within society. This is undertaken through a cycle of analysis, strategy, action (implementation) and evaluation, representing a widely recognised approach to public health practice.

The Public Health Genetics Unit in Cambridge, which is the core facility of the Cambridge Genetics Knowledge Park (funded by Department of Health and Department of Trade and Industry) has nine years experience in this area. It includes a wide number of disciplines under one roof and also has close links with university, clinical, research, educational and commercial communities as well as the public. Its work programme includes:

- International leadership in Public Health Genomics
- Epidemiology: contributing to a sound evidence base on epidemiological associations among genes, the environment and health and on genetic tests
- Knowledge integration: integration of epidemiological and other knowledge with comprehensive analysis

· Service development: applying knowledge to analyse, develop and evaluate existing and new health programmes
· Education: developing a specialist workforce and enabling public health professionals and other specialists to incorporate genetics into their own practice
· Ethical, legal and social policy analysis

The presentation will include examples of this work and information about how to access it. The Unit is also involved in a number of international networks in public health genomics whose work will also be described in the presentation.

Consultant in Public Health Medicine／Public Health Genetics Unit, Cambridge／phgu.org.uk

Michiaki Kubo ●Laboratory for Genotyping, SNP Research Center, RIKEN, Department of Environmental Medicine, Graduate School of Medical Sciences, Kyushu University, Japan

Yusuke Nakamura ●Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan

Yutaka Kiyohara ●Department of Environmental Medicine, Graduate School of Medical Sciences, Kyushu University, Japan

Search for susceptible genes of cerebral infarction in the Hisayama study

The development of common disease is thought to be result from complex interactions between numerous environmental factors and variations of many genes. One of the main aims for applying epidemiological aspects to genetic analysis is to clarify this complex interaction. When we understand the mechanisms which environmental factors increase the risk of the development of common disease in persons with certain genetic variation, we will put forward to a new preventive strategy by controlling environmental factors for the individuals with specific genetic background (personalized prevention). To achieve this goal, we first need to know which genetic variations affect the development and progression of common diseases. Unfortunately, our current understanding of susceptible genetic variations for common diseases is limited. After a large volume of SNP information and their large-scale genotyping methods became available, genome-wide association studies have been applied and successfully identified genes susceptible to common diseases.

Stroke is the major leading cause of death in Japan as well as in many developed countries. In Japan, the incidence of cerebral infarction, the most common type of stroke, has been significantly decreased for the last three decades, but still remained high even in recent years. To identify genes associated with susceptibility to cerebral infarction, we performed a large-scale genome-wide case-control study using 1,112 Japanese cases with cerebral infarction and 1,112 age- and sex-matched control subjects. First, we carried out genotyping of 188 cases with cerebral infarction and 188 age- and sex-matched controls using 52,608 gene-based tag-SNPs selected from the JSNP database. We compared allele frequencies of 48,083 successfully genotyped SNPs (overall success rate of 91.4％) between the two groups and selected 1,098 SNPs showing p-values of <0.01. We subsequently genotyped the remaining cases and controls for these SNPs as the second screening. Through this screening we identified 12 SNPs with a strong association with cerebral infarction or its subtype. Further analysis of these candidate loci is ongoing. Among these loci, we further examined the impact of candidate SNP on the development of cerebral infarction using the ongoing cohort of the Hisayama study, and found that the candidate SNP significantly increased the risk of the development of cerebral infarction in the Japanese population. To combine epidemiological study and genetic analysis will be an important tool to clarify the mechanisms of common disease with complex environmental and genetic factors.

Michiaki Kubo／Laboratory Head／Laboratory for Genotyping, SNP Research Center／The Institute of Physical and Chemical Research（RIKEN）／http://www.riken.jp/r-world/research/lab/idenn/idenn-ta/index.html http://www.src.riken.go.jp/jpn/group/genome/index.html

Takeo Nakayama ●Department of Health Informatics, Kyoto University School of Public Health, Japan

Prospective cohort study at Nagahama

Public health practice is defined as the science and art of disease prevention, prolonging life, and promoting health and well-being through organized community effort (Winslow). Epidemiology is the study of the distributions and determinants of health-related states or events in specified populations (Last). Final evidence for the involvement of candidate genes in human diseases must come from extensive epidemiological studies, preferably in different populations (Peltonen and McKusick). The US CDC stresses that importance of the epidemiologic approach to genetic information as follows: (1) prevalence (2) associations (3) interactions (gene-environment).

One of the community-based cohort studies in Japan that involve genomic analysis, we are planning to launch Nagahama "zeroji" cohort in Shiga prefecture, Japan. Among 84,000 residents in Nagahama city, approximately 10,000 residents who are aged 30- 74 are expected to participate in this cohort. Being different from case-control studies, both initial survey at entry and long-term follow are necessary in cohort studies. Therefore, development of longstanding relationship of researchers to residents (not limited to participants), community organizations and authority is critical. It is also necessary to consider diverse values of stakeholders in a community when conducting and maintaining a cohort study.

I will point out some issues arising from community-based genome cohort studies, that is, the crossroad of "life science" and "public health" and introduce a part of our plan in the Nagahama project.

Professor／Department of Health Informatics, Kyoto University School of Public Health／
http://square.umin.ac.jp/healthim/

Cashell E. Jaquish, Ph.D.
●National Heart, Lung, and Blood Institute,
National Institutes of Health, USA

Genomics Research in the Framingham Heart Study: Opportunities and Challenges.

The Framingham Heart Study (FHS) is the National Heart, Lung, and Blood Institute's (NHLBI) longest running prospective cohort study. The study began in 1948 with recruitment of adults in households within the town of Framingham. Since then, the FHS has been a leader in cardiovascular epidemiology, helping to define the important role of many major risk factors that have been measured such as hypertension and hyperlipidemia. In the early 1970s the children of the original cohort were recruited and more recently the third generation joined the study. Because of this recruitment strategy and the availability of many well-measured phenotypes, the FHS is ideal for implementing genetics research.

Estimates of heritabilities for traditional risk factors and subclinical disease measures revealed moderate to strong genetic effects. The NHLBI Mammalian Genotyping Service was used to type approximately 400 microsatellite markers in 7,144 members of the three generations. Initial linkage analyses revealed substantial LOD scores for hypertension, body mass index, and lipid measures. The extensive longitudinal, well measured phenotypes provide many opportunities for investigating changes in genetic and environmental risk with age as well as genetic contributions to disease progression.

With the availability of high density SNP typing, 100,000 SNPs were typed on approximately 1,400 members of the FHS. This led to the discovery of an association between the INSIG2 gene and BMI which was replicated in 5 other studies. The Framingham 100K SNP data were also used to confirm associations involving the CAPON and Factor 7 variants. Currently, genotyping of 550,000 SNPs is underway on the entire FHS cohort, offspring and third generation (N ～10,000). The program FHS SHARe (SNP Health Association Resource) is among the largest genome wide association efforts at the National Institutes of Health. FHS SHARe will provide an unparalleled resource for gene discovery. The NHLBI is working with the National Center for Biotechnology information (NCBI) to construct a database containing all of the genotype information and 50+ years of phenotypic information. This database will be accessible to members of the scientific community in 2007.

The opportunities for research and collaboration in the FHS go beyond gene discovery to replication, validation, methods development, assessment of gene by environment interaction, epistasis, etc. However, there are also challenges. Open access to the vast amount of genetic and medical information in the FHS could compromise the privacy and confidentiality of participating individuals and families. The FHS investigators and NHLBI have put in place an application process which places the participant's privacy as its highest priority. This process includes local Institutional Review Board approval and signature of a data distribution agreement pledging protection of confidentiality and privacy protection. Through such a process the Framingham Heart Study will continue to be an international resource leading the way in cardiovascular epidemiology.

Cashell Elizabeth Jaquish／Program Director／Division of Prevention and Population Science／National Heart, Lung, and Blood Institute, National Institutes of Health

Isabel Fortier ●P³G Consortium, Canada

The Public Population Project in Genomics (P³G)

Over the last decade, genetic epidemiology has experienced an important shift from family-based studies of genetic linkage to individual-based studies of genetic association and there is an increasing investment worldwide in large population-based biobanks. This is primarily because they provide an ideal infrastructure for the study of the joint effects of genes and environment in causing common chronic diseases of great impact on public health. But samples size calculations that take realistic account of bio-clinical complexity indicate that in order to comprehensively study the etiological determinants of a complex disease one ideally needs 10,000 cases, and if interest focuses on a gene:gene or gene:life-style interaction, 20,000 cases would be preferable. Currently, at least 20 population-based biobanks around the world include more that 100,000 participants and 7 of them have recruited of plan to recruit more than 500,000 participants. However, even these largest population-based biobanks cannot, on their own, guarantee enough cases of most diseases of interest to ensure that many of the important scientific questions can be answered with complete clarity. In order to optimize the return from the global investment in biobanking, it is essential to focus on the quality of study design, conduct and analysis, but also on biobank harmonization. Harmonization can be defined as *a set of procedures that promote, both now and in the future, the effective interchange of valid information and samples between a number of studies or biobanks, accepting that there may be important differences between those studies.* Biobank scientists around the world have demonstrated a clear desire to work together on these issues, but success demands an active international organization to promote collaboration.

In order to maximize international synergy, the Public Population Project in Genomics (P³G) coordinates and fosters an international harmonized vision for research in population genomics. P³G is **not** a central repository of data. The organization is designed to encourage new collaborations and to aid P³G members and partners in the development of their own programs. The main targets of P³G are to: 1) **foster collaboration** between biobanks; 2) **optimize the design**, set up and research activities of population-based biobanks; 3) **promote harmonization**; and, 4) facilitate **transfer of knowledge.** The P³G Observatory is one the principal P³G tool. It is a central Internet repository (www.p3gobservatory.org) aimed at facilitating the development, realization and harmonization of population genomics research projects. It gives access to information on biobanks around the world and to interactive tools for description of studies and harmonization activities. All of the activity that takes place within the P³G consortium is ultimately driven by leading scientists and other professionals in the field of population-based biobanking and by partner institutions sharing common interests with P³G.

Director／P³G Observatory／P³G Consortium

Structural Variation in the Human Genome

3

December 15
Friday

Chair : Ituro Inoue (Tokai University, Japan)

Chair : Tatsushi Toda (Osaka University)

Stephen W. Scherer, Ph.D.

●The Centre for Applied Genomics and
Program in Genetics and Genomic Biology,
The Hospital for Sick Children;
Department of Molecular and Medical
Genetics, University of Toronto, Canada

Discovery of Structural Variation in the Human Genome: New Insights for Disease Study.

The advent of genome-scanning technologies and comparative DNA assembly analyses has uncovered a significant extent of 'structural variation' in the human genome. Structural variants can include microscopic and more commonly submicroscopic deletions, duplications, and large-scale copy number variants - collectively termed copy number variants (CNVs) or polymorphisms - as well as insertions, inversions and translocations. A growing body of literature indicates that structural variants can comprise millions of nucleotides of heterogeneity within every genome, having an important contribution to human diversity and disease susceptibility. To generate a first-generation CNV map of the human genome our Consortium (with N. Carter, M. Hurles, C. Tyler-Smith: Wellcome Trust Sanger Institute; C. Lee: Brigham Women's Hospital/Harvard; K Jones: Affymetrix; H. Aburatani: Univ. of Tokyo) screened the 270 individuals from the four HapMap populations for copy number variation using two complementary technologies: the Affymetrix 500k-SNP platform and comparative genome hybridization (CGH) on a microarray containing ~27,000 clones representing the genome tile-path. We rigorously assessed the rate of false positive CNV calls by several independent means, including the validation of hundreds of loci by quantitative PCR and FISH. We identified 1447 CNVs in these four populations, many of which have not previously been identified, and which contain hundreds of genes and non-coding functional sequences (samples from the CEPH-Human Diversity panel and Phase II Hapmap are now also being assessed using a similar experimental design). In a second approach we have performed comparative DNA analysis of human vs. human and human vs. chimp assemblies followed by experimentation to find hundreds of other structural variants in the human genome. Our collective data, integrated with all other available information, is released in the 'Database of Genomic Variants' (http://projects.tcag.ca/variation/). The database serves as a resource to assist numerous clinical research studies. Our latest data assessing CNV content for involvement in the study of the complex disease autism will also be presented.

Stephen Wayne Scherer／Senior Scientist and Professor／Genetics and Genomic Biology／
Hospital for Sick Children／http://www.tcag.ca/scherer/

Charles Lee, Ph.D. ●Department of Pathology, Harvard Medical School, USA

Copy number variation in the human genome: Implication to diagnostics and cancer research

Array-based comparative genomic hybridization (aCGH) can be used for genome-wide assessment of chromosomal imbalances in a robust fashion and at high resolution. Indeed, earlier studies using a 1-Mb resolution aCGH platform on clinically-indicated, postnatal cases (with normal GTG-banded karyotype results) found abnormal aCGH results in as many as 20% of the referred cases. Genome-wide aCGH testing in higher risk prenatal cases (e.g. cases with abnormal ultrasound but normal GTG-banded karyotype results), may prove to be similarly useful. However, the recent appreciation of the widespread existence of copy number variation (CNV) in the genomes of healthy, normal individuals needs to be carefully considered when interpreting aCGH results in both the clinical and research settings. Indeed, recent observations suggest that CNVs may play a more important role in cancer research, than previously appreciated. Hence, comprehensive identification and characterization of CNVs in different human populations need to be made a priority, and in context of the limitations with aCGH technology.

Director of Cytogenetics and Assistant Professor／Pathology／Harvard Medical School／
http://www.chromosome.bwh.harvard.edu

Tim Aitman ●MRC Clinical Sciences Centre, Imperial College London, UK

Gene copy number variation and susceptibility to common diseases

Identification of the genes underlying complex phenotypes and the definition of the evolutionary forces that have shaped eukaryotic genomes are among the current challenges in molecular genetics. Variation in gene copy number is increasingly recognized as a source of inter-individual differences in genome sequence and has been proposed as a driving force for genome evolution and phenotypic variation. Several years ago, we showed that copy number variation amongst rat strains in the *Cd36* gene, caused by a chromosomal deletion at the *Cd36* locus, is a cause of defects in insulin action and fatty acid metabolism in hypertensive rats. More recently we showed by positional cloning that loss of the newly described, rat-specific *Fcgr3* paralogue, Fcgr3-related sequence (*Fcgr3-rs*), is a determinant of macrophage overactivity and glomerulonephritis in Wistar Kyoto rats. In humans, low copy number of *FCGR3B*, an orthologue of rat *Fcgr3*, was associated with glomerulonephritis in the autoimmune disease systemic lupus erythematosus. These findings show that copy number variation of the orthologous rat and human *Fcgr3* genes is a determinant of susceptibility to immunologically mediated glomerulonephritis. Since copy number variation is known to be highly prevalent in mammalian genomes, our data suggest that this type of transmissible polymorphism is a likely driver of evolutionary selection and susceptibility to common diseases.

Timothy John Aitman／Professor of Clinical & Molecular Genetics／Physiological Genomics & Medicine／Medical Research Council and Imperial College Faculty of Medicine／http://www.csc.mrc.ac.uk/ResearchGroups/PhysiologicalGenomicsAndMedicine/PhysiologicalGenomicsAndMedicineResearch.html

Issei Imoto
Shin Hayashi
Shozo Honda
Johji Inazawa

●Department of Molecular Cytogenetics, Medical Research Institute,
Tokyo Medical and Dental University and
CREST, JST, Japan

Detecting copy-number variation in the human genome using BAC-array based comparative genomic hybridization.

The near completeness of human chromosome sequences is facilitating accurate characterization and assessment of all classes of genomic variation. Although single nucleotide polymorphisms (SNPs) has been thought to be the main source of those genetic and phenotypic human variation, the advent of genome-scanning technologies has now uncovered an unexpectedly large extent of what we term‘ structural variation ’in the human genome. Array-based comparative genomic hybridization (CGH) is one of the most reliable technologies that have recently revealed a newly appreciated type of genetic variation: sub-microscopic copy-number variation (CNV), in which many regions of the human genome are now known to be variable in number between individuals.

Last 5 years, we have constructed various types of in-house bacterial artificial chromosome (BAC)-based arrays (designated as MCG arrays) to explore genomic copy-number alterations in cancers as well as genetic diseases in a genome-wide manner, and identified a number of genes affected with those human diseases. Through those studies, we have also detected many CNVs: some of them are unique regions/sequences, which have never been shown in databases probably due to clones used for our array or population we analyzed. Since more comprehensive cataloging and characterization of CNVs will provide the basis to distinguish directly “causative” genomic alterations from variations, especially that in a specific ethnicity, we are continuing more analysis in Japanese healthy subjects and patients with congenital abnormalities and/or mental retardation with their parents or sibling by array-CGH using arrays for scanning whole genome or X-chromosome as well as newly constructed array for CNV analysis (MCG Genome Variation Array Ver.1) followed by fluorescence in situ hybridization (FISH) to collect data of the position, context of genome, and frequency. Since CNVs may account for a significant proportion of normal phenotypic variation predispose to certain diseases, those data with the list of genes included within CNV will have a significant impact on how each CNV shows pathological and/or biological significance. To unravel the structural complexity and/or variation among CNVs, we performed FISH with each CNV, and found that BACs within CNVs delineate variegated signal pattens, such as a simple copy-number change and intra- and inter-chromosomal multiple hybridizations, suggesting that CNVs detected by array-based analysis may be more complicated as considered.

Issei Imoto／Associate Professor／Department of Molecular Cytogenetics／Medical Research Institute, Tokyo Medical and Dental University／http://www.tmd.ac.jp/mri/cgen/framepage.htm

Hiroyuki Aburatani
Shumpei Ishikawa
Daisuke Komura
Keith W. Jones

●Genome Science division, RCAST, The University of Tokyo, Japan

Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays

Recent reports indicate that copy number variations (CNVs) within the human genome contribute to nucleotide diversity to a larger extent than single nucleotide polymorphisms (SNPs). In addition, the contribution of CNVs to human disease susceptibility may be greater than previously expected, although a complete understanding of the phenotypic consequences of CNVs is incomplete. We have recently reported a comprehensive view of CNVs among 270 HapMap samples using high-density SNP genotyping arrays and BAC array CGH. In this report, we describe a novel algorithm using Affymetrix GeneChip Human Mapping 500K Early Access (500K EA) arrays that identified 1203 CNVs ranging in size from 960 bp to 3.4 Mb. The algorithm consists of three steps: (1) *Intensity pre-processing* to improve the resolution between pairwise comparisons by directly estimating the allele-specific affinity as well as to reduce signal noise by incorporating probe and target sequence characteristics via an improved version of the Genomic Imbalance Map (GIM) algorithm; (2) *CNV extraction* using an adapted SW-ARRAY procedure to automatically and robustly detect candidate CNV regions; and (3) *copy number inference* in which all pairwise comparisons are summarized to more precisely define CNV boundaries and accurately estimate CNV copy number. Independent testing of a subset of CNVs by quantitative PCR and mass spectrometry demonstrated a >90 % verification rate. The use of high-resolution oligonucleotide arrays relative to other methods may allow more precise boundary information to be extracted, thereby enabling a more accurate analysis of the relationship between CNVs and other genomic features. Software called GEMCA (Genotyping Microarray based CNV Analysis), which implements this CNV calling algorithm, can be freely downloaded from http://www2.genome.rcast.u-tokyo.ac.jp/CNV/.

Although the 500K EA platform has good resolving power to identify CNVs <100 kb in size, CNVs spanning segmental duplications are underrepresented because of the difficulty in developing robust SNP genotyping assays in these regions. To minimize this discrepancy and have appropriate representation for segmentally duplicated regions, we are currently testing a new high-density array that contains multiple nonpolymorphic probes for every predicted NspI fragment, which can be used in conjunction with whole-genome sampling analysis (WGSA). This array covers >1.3 million fragments with a median intermarker distance of just less than 800 bp. Preliminary results will be presented.

References:

1) Komura D, Shen F, Ishikawa S, et al. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res.* **16**(12):1575-1584, 2006
2) Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. **444**(7118):444-454. 2006
3) Komura D, Nishimura K, Ishikawa S, et al. Noise Reduction from genotyping microarrays using probe level information. *In Silico Biol*. **11**;6(1-2):79-92, 2006
4) Ishikawa S, Komura D, Tsuji S, et al. Allelic dosage analysis with genotyping microarrays. *Biochem Biophys Res Commun*. **333**(4):1309-1314, 2005

Hiroyuki Aburatani／Professor／Research Center for Advanced Science and Technology／The University of Tokyo／http://www.genome.rcast.u-tokyo.ac.jp/

Strategy for Genomics Research of Mental Disorders

4

December 15
Friday

Chair : Tadafumi Kato（RIKEN Brain Science Institute, Japan）

Chair : Yuji Okazaki（Mie University, Japan）

Jurg Ott, Ph.D. ●Rockefeller University, New York, USA

Theoretical basis of genome-wide association studies and its application

"Disease gene mapping by genome-wide association analysis"

For many so-called complex traits that commonly occur in the population (diabetes, schizophrenia, etc.), there is good evidence that genes play a role in disease susceptibility but nothing is known about these genes. As a first step towards elucidating disease etiology, we want to localize such genes on the human gene map, that is, the human chromosome. This can be achieved with (1) genetic linkage or (2) case-control association studies. The former has been very successful for many mendelian diseases (following a mendelian mode of inheritance) such as Huntington disease (dominant) or cystic fibrosis (recessive). Now interest is shifting towards the much more difficult task of finding genes as risk factors for common, non-mendelian traits that nonetheless have a genetic basis, for example, because they are more common among relatives of affected individuals than in the general population. For such traits, it has been shown that genetic association (linkage disequilibrium) mapping is more powerful than linkage analysis. In recent years, large numbers of genetic marker loci have been developed throughout the genome. In particular, 100,000s of single-nucleotide polymorphisms (SNPs) are beginning to be used in case-control association studies. The first large-scale such study was carried out by Ozaki *et al* (2002) with >90,000 SNPs and furnished clear evidence for functional SNPs in the lymphotoxin-α gene as genetic risk factors for myocardial infarction. I will discuss various current approaches and potential pitfalls in genetic association studies. In particular, I will review a study, published 2005 in Science magazine, leading to the discovery of a gene partly responsible for age-related macular degeneration (AMD). The idea behind all these studies is that discovery of risk genes will make it possible to elucidate the etiology of diseases and eventually find a cure for such traits. In some situations, however, one needs to carefully consider the potential benefits of genetic studies for public health. More direct intervention by public health measures often promise to be more beneficial, for example, to curb the growing obesity epidemic in many countries even though there seems to be some degree of genetic liability for obesity.

Professor／Laboratory of Statistical Genetics／Rockefeller University／http://www.genemapping.cn

Francis J. McMahon, MD

●Chief, Genetic Basis of Mood & Anxiety Disorders, National Institute of Mental Health, National Institutes of Health, Bethesda MD, USA

A potential role for copy number polymorphisms in the etiology of bipolar disorder

Copy number polymorphisms (CNPs), which involve deletion or duplication of chromosomal segments, are coming to be recognized as a common and important form of variation within the human genome. CNPs can range in size from a few hundred to a few hundred million basepairs and are widely distributed throughout the genome. Since CNPs frequently affect coding regions, they are likely to have functional significance through their impact on gene dosage. Some of the larger CNPs, such as those involved in Velo-cardio-facial Syndrome and Williams Syndrome lead to prominent neuropsychiatric symptoms. This raises the question of whether CNPs may also play an etiologic role in more common neuropsychiatric disorders, such as bipolar disorder. This presentation will review the rapidly-evolving field of CNPs, focusing on their epidemiology, methods of detection, and known functional consequences, along with a discussion of study designs aimed optimizing the detection of pathogenic CNPs in bipolar disorder.

Chief, Genetic Basis of Mood & Anxiety Disorders／National Institute of Mental Health Intramural Research Program／National Institutes of Health／http://neuroscience.nih.gov/Lab.asp?Org_ID=495

Takeo Yoshikawa ●RIKEN Brain Science Institute, Japan

Approaches to map genetic underpinnings of schizophrenia and mood disorders

Strategies to identify the genetic elements underlying complex psychiatric phenotypes

Psychiatric illnesses such as substance abuse, schizophrenia and mood disorders are complex traits controlled by multiple genes. This makes the identification of genes that increase the risk of disease a formidable task. Another compounding factor is that compared to the physical diseases that are also multi-factorial in origin, the diagnosis of mental disorders is largely subjective, since there are few unique or consistent biological markers. This reduces the obvious phenotype-genotype correlation, further complicating the identification of risk genes. I will introduce our experiences in investigating schizophrenia and mood disorders, by covering candidate gene analysis based on the hypotheses of disease mechanisms, combination of gene expression and genetic association analyses, metabolic pathway-based genetic analysis, and extensive genome-wide association scans. The feasibility of the latter two approaches has increased with improvements in large-scale genotyping methodology and the genetic information on human genome polymorphisms, although further improvements in genotype costing, SNP (single nucleotide polymorphism) density and basic biological knowledge are required. The genetic analysis of experimental animals is another emerging useful tool. Psychiatric traits are 'quantitative' by nature. Animals that show traits similar to human mental disorders can be genetically dissected using quantitative trait loci (QTL) analysis. I will also introduce our attempts to detect depression/schizophrenia-related genes by performing QTL analyses in mice. The future genetic studies will have to encompass multidisciplinary ideas and techniques.

Laboratory Head／Laboratory for Molecular Psychiatry／RIKEN Brain Science Institute／
http://www.riken.go.jp/engn/r-world/research/lab/nokagaku/age/molecular/index.html

Tadafumi Kato ●Laboratory for Molecular Dynamics of Mental Disorder, RIKEN Brain Science Institute, Japan

Genomic studies of bipolar disorder: from gene expression to genetics and epigenetics

Bipolar disorder is characterized by recurrent manic and depressive episodes, which threatens the life by suicide and severely disturbs the quality of life. Altered calcium signaling has been implicated from the study of peripheral blood cells and impaired neural plasticity and cellular resilience have been suggested by clinical and pharmacological studies. Role of genetic factors in bipolar disorder is well established from twin and family studies. Extensive genetic studies suggested the role of several candidate genes, but none of them has been well established yet. To supplement traditional genetic approach, we have been searching for candidate genes of bipolar disorder by comprehensive gene expression analysis using DNA microarray.

First, we examined the gene expression patterns in lymphoblastoid cells derived from monozygotic twins discordant for bipolar disorder. We found that two genes in endoplasmic reticulum (ER) stress pathway, XBP1 and HSPA5, were commonly downregulated in the twins with bipolar disorder. We found that ER stress response was attenuated in lymphoblastoid cells of patients with bipolar disorder. We found a functional polymorphism of XBP1 reducing ER stress response and reported the association of this polymorphism with bipolar disorder. This association was not replicated in a larger European samples. On the other hand, reduced ER stress response has been reported also in Europeans. We are now searching for epigenetic factors responsible for the discordant phenotypes in the monozygotic twins by methylation-specific representational differential analysis.

We also performed comprehensive gene expression analysis of the postmortem brains of patients with bipolar disorder. We found that calcium signaling related genes were altered. Among them, altered expression of PDLIM5, which encodes a protein linking protein kinase C and Ca^{2+} channel, was also observed in lymphoblastoid cells. Genetic association study suggested association of SNPs of this gene with bipolar disorder. On the other hand, global downregulation of mitochondria related genes reported by the other investigators was replicated but it was affected by sample pH and medication. Subgroup of mitochondria-related genes were rater upregulated. We also searched for the molecular basis of altered expression of LARS2, encoding mitochondrial $tRNA^{Leu}$ synthetase. We found that accumulation of the 3243 mutation of mitochondrial DNA (mtDNA), impairing the aminoacylation of $tRNA^{LeuUUR}$, can cause upregulation of LARS2. We found that the mtDNA 3243 mutation was accumulated in several patients showing upregulation of LARS2 in the brain. We are now searching for DNA methylation differences responsible for altered gene expression in the postmortem brains.

These findings collectively suggest the feasibility of gene expression analysis in identifying the genetic and epigenetic factors responsible for bipolar disorder.

Group Director／Aging and Psychiatric Research Group／RIKEN, Brain Science Institute／
http://www.brain.riken.jp/labs/mdmd/

Satellite Symposium
" Human Genome, Evolution, and Disease "

S

December 16
Saturday

Session Ⅰ Chair : Naruya Saitou (National Institute of Genetics, Japan)

Session Ⅱ Chair : Takashi Gojobori (National Institute of Genetics, Japan)

Session Ⅲ Chair : Ituro Inoue (Tokai University, Japan)

Session Ⅳ Chair : Hidetioshi Inoko (Tokai University School of Medicine, Japan)

Takashi Gojobori ●Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Japan

Genomic evolution of human nervous system-specific genes and its implication to methods for hunting neuropathological disease-sensitive genes

With the aim of elucidating the evolutionary process of the brain and central nervous system (CNS), we took an approach of making genomic comparisons of a protein-coding gene set between human and other species. In practice, we first obtained about 400 protein-coding genes whose level of the mRNA expression is more than 50 % in a human brain or CNS compared with those in other tissues or organs in the H-ANGEL (Human-Anatomical Gene Expression Library) section of the H-Invitational integrated database of human genes. We now call those genes operationally as "human nervous system-specific genes (human NS-specific genes)." We, then, compared these human NS-specific genes with the protein-coding gene sets that were contained in each of the entire genome of the species available, in order to estimate when each of those genes emerged during evolution. As a result, we found that about one thirds of the human NS-specific genes evolutionarily emerged just before the outbreak of vertebrates. Moreover, we found that a major kind of those genes may have had receptor activities. Thus, those findings imply that protein-to-protein interaction network on the basis of ligand and receptor relationships must have played a crucial role in the complex and integrated formation of a brain and CNS. These findings give us a hint for developing novel methods for hunting neuropathological disease-sensitive genes.

Director and Professor／Center for Information Biology and DNA Data Bank of Japan／National Institute of Genetics／
http://www.cib.nig.ac.jp http//www.ddbj.nig.ac.jp

Lynn B. Jorde ●Department of Human Genetics, University of Utah School of Medicine, USA

Evolution and linkage disequilibrium: implications for common diseases

Following two decades of success in identifying genes responsible for Mendelian diseases, we now face the challenge of discovering genes that predispose to complex diseases. Linkage disequilibrium (LD) analysis has gained increasing popularity as a means of detecting genes that underlie predisposition to complex disease. In this presentation, we review the definition and basic applications of LD in disease-gene mapping. Study design issues include sample size, phenotype definition, and statistical techniques. Complex diseases, by definition, are likely to be characterized by extensive locus and allelic heterogeneity. Strategies for minimizing the effects of such heterogeneity (and thus maximizing the chances of identifying disease-causing mutations) include the use of endophenotypes, the identification of clinically defined subsets, and the use of isolated populations. Each of these strategies will be discussed and evaluated, and examples of recent success in isolating genes that underlie complex diseases (e.g., asthma, Crohn disease, age-related macular degeneration, type 2 diabetes) will be highlighted.

Because it reflects the effects of recombination over many past generations, LD can potentially refine the location of a disease-causing mutation. However, LD is also susceptible to the effects of factors like population history and population sampling. To illustrate the effects of population history on LD patterns, we present results of population variation in 145 Africans, 700 Asians, and 120 Europeans in whom 30 RSPs, 60 STRs, 100 *Alu* insertion polymorphisms, and 75 L1 insertion polymorphisms have been assayed. In addition, we report results based on an analysis of 11,000 SNPs, as well as extensive resequencing of the angiotensinogen gene. Our results demonstrate consistently that Africans have substantially higher genotypic and haplotypic diversity than do other populations and that most variation in non-African populations is a subset of variation found within Africa. These patterns have important implications for haplotype block structure in different human populations and thus for the design of LD mapping studies. This is illustrated further by comparisons made between haplotype block structures revealed by HapMap and those revealed by extensive resequencing in more broadly sampled populations. We demonstrate that, in general, patterns revealed by the HapMap can be exported successfully to other populations.

Lynn Burgess Jorde ／ Professor ／ Department of Human Genetics ／ University of Utah School of Medicine ／ http://jorde-lab.genetics.utah.edu/

Naruya Saitou ●National Institute of Genetics, Japan

Utility of chimpanzee and other non-human primate genomes for understanding evolution of modern human

When I was graduate student more than 20 years ago, some teachers never mentioned evolutionary aspects during human genetics lectures. This non-evolutionary attitude, I presume, no longer exists in human genetics studies in present days. Any gene in human genome has long history of evolution, and we have to face this even when considering genetic variations within modern humans. Chimpanzee is phylogenetically closest to human, and chimpanzee genome sequencing already started before determination of human genome was complete. I was involved in determination and analysis of chimpanzee chromosome 22 (Ref. 1), and our group utilized chimpanzee genomic sequences for estimating ancestral states of human SNP loci. We applied H test to human chromosome 21 SNP data and picked up 18 regions as candidates under positive selection during modern human evolution (ref. 1). We then examined chimpanzee SNP for some of these 18 regions, and found that majority of them also showed possibility of positive selection (ref. 2). This suggests that similar selection pressure exists both in human and chimpanzee for these regions.

We also estimated pattern of nucleotide substitution in modern humans as well as modern chimpanzees, and equilibrium frequencies of four nucleotides were somewhat different between human and chimpanzee. While current GC content of human and chimpanzee are both ca. 41 %, that of equilibrium is 42 % and 41 % for human and chimpanzee, respectively (ref. 1). This suggests that some evolutionary changes occurred at least in human lineage after divergence from common ancestor of human and chimpanzee. We further analyzed pattern of nucleotide substitution in various regions of human genome and various evolutionary stages (ref. 3) as well as analysis of including mitochondrial DNA genomes (ref. 4), and I would like to discuss our these recent progress.

References Cited

1. The International Chimpanzee Chromosome 22 Consortium [H.Watanabe, A.Fujiyama, M.Hattori, T. D.Taylor, A.Toyoda, Y.Kuroki, H.Noguchi, A.BenKahla, H.Lehrach, R.Sudbrak, M.Kube,S.Taenzer, P.Galgoczy, M.Platzer, M.Scharfe, G.Nordsiek, H.Blocker, I.Hellmann, P.Khaitovich, S.Paabo, R.Reinhardt, H.-J.Zheng,X.-L.Zhang, G.-F.Zhu, B.-F.Wang, G.Fu, S.-X.Ren9, G.-P.Zhao, Z.Chen, Y.-S.Lee, J.-E.Cheong, S.-H.Choi, K.-M.Wu, T.-T.Liu, K.-J.Hsiao, S.-F.Tsai, C.-G.Kim, S.Oota, T.Kitano, Y.Kohara, N.Saitou, H.-S.Park, S.-Y.Wang, M.-L.Yaspo, and Y.Sakaki] (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature*, vol. **429**, No. 6990, pp. 382-388.
2. Kitano T., OOta S., and Saitou N., unpublished.
3. OOta S., Kawai Y., and Saitou N., unpublished.
4. Kawai Y. and Saitou N., unpublished.

Professor／Division of Population Genetics／National Institute of Genetics／sayer.lab.nig.ac.jp/~saitou/

Ranajit Chakraborty Ph.D.
●University of Cincinnati College of Medicine, USA

Issues and Approaches in the Population Genomics Paradigm of Studying Complex Diseases

Basis and features of the population genomics paradigm of studying the genetic architecture of complex diseases will be described.

Factors governing the extent and pattern of linkage disequilibria are to be built into the interpretation of population genomic data on disease-gene association. These factors dictate the choice of populations and selection of the number and density of genomic markers for such studies. Both whole genome and candidate gene based approaches require adjustment for multiple testing of correlated hypotheses. Theoretical rationale will be given to address these questions, with some suggested solutions for controlling false discovery rates without gross loss of statistical power of the study.

Robert A. Kehoe Professor and Director／Center for Genome Information／Department of Environmental Health／University of Cincinnati College of Medicine

Stephen W. Scherer, Ph.D.

●The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children; Department of Molecular and Medical Genetics, University of Toronto, Canada

Molecular Genetics of Autism

Autism is a neurodevelopmental disorder characterized by impairments in social- communication and by a preference for repetitive activities. Twin and family studies indicate autism is mainly a genetic condition of complex etiology, but none of the major susceptibility genes are known. Our recent studies have shown that ～7.4％(129/1749) of autistic patients have chromosomal anomalies detectable upon karyotyping suggesting gene-dosage effects can contribute to the phenotype. We are now (1) characterizing all samples having cytogenetically detectable rearrangements and (2) assessing the genome in 500 probands (having varying phenotypic complexity) from our collection for sub-microscopic disease-associated DNA variants, using the 500k Affymetrix-SNP array. The combined data should contribute to sample stratification and provide opportunity to find susceptibility loci. As such, we have mapped 67 cytogenetically-visible breakpoints in 36 patients (42 translocation, 14 deletion, 5 inversion, 6 duplication breakpoints; data is in our Autism Chromosome Rearrangement Database at http://projects.tcag.ca/autism/). For aim 2, we have already assessed 200 samples, but so far the preliminary data does not indicate prevalence for sub-microscopic changes in autism compared to controls. However, our combined data (microscopic and sub-microscopic) identifies recurrent apparently phenotype-associated loci on chromosome 7q31.2 (at *FOXP2*), 7q31.3, 7q34, 15q and at other regions. Other non-recurrent, but potentially disease-associated intervals have been found including the chromosome 4p-*GABARG1* locus. Moreover, in addition to two autism patients with deletions at 7q31.2, we also identified 13 cases with Developmental Verbal Dyspraxia (DVD; language disorder, inability to cough/sneeze, or laugh spontaneously), all failing to inherit a paternal *FOXP2* gene. Cases with paternal uniparental disomy do not have DVD, and we show maternally inherited *FOXP2* to be comparatively under-expressed. Our data suggests a role for differential gene expression in the regulation of *FOXP2* in human speech development, and calls further attention to it for having a possible role in autism.

Stephen Wayne Scherer／Senior Scientist and Professor／Genetics and Genomic Biology／Hospital for Sick Children／http://www.tcag.ca/scherer/

Hidetoshi Inoko ●Division of Molecular Medical Science and Molecular Medicine, Tokai University School of Medicine , Japan

Genome-wide scan of disease genes by association analysis using microsatellites

The " human genome diversity project " is conducted by disease gene mapping and association analyses using polymorphic genetic markers. For this purpose, SNP (single nucleotide polymorphism) markers are now being extensively and world-widely collected, and applied to disease mapping in a lot of laboratories. However, SNP is generally bi-allelic, and so polymorphism is not considered to be extensive enough to localize disease genes on the human genome by genome-wide mapping. Instead, we propose to use microsatellite which displays a high degree of polymorphism in repeat number of repetitious unit and so is expected to serve as a more useful genetic marker for genome-wide mapping.

Our mapping studies of diseases in the HLA region on chromosome 6p21.3 demonstrated that the length of linkage disequilibrium with disease locus observed for microsatellite is around 100 kb. This suggests that one microsatellite per 100 kb is enough for genome-wide mapping. Therefore, efficient method of genome-wide mapping is to first use microsatellites markers which enable to narrow down the critical region to 100 kb and thereafter to employ SNP markers within thus determined critical region for fine mapping to identify a causative gene. Based on this strategy, we collected 30,000 polymorphic microsatellite markers (one microsatellite per 100 kb) throughout the human genome (1).

We have applied these 30,000 microsatellites to genome-wide association mapping of more than 10 complex diseases including psoriasis vulgaris and hypertension. We have so far identified more than 30 susceptible loci or candidate loci for them, namely 7 susceptible loci or candidate loci for rheumatoid arthritis (1), 6 for psoriasis vulgaris, 5 for hypertension, 8 for myopia, 1 for endometriosis, 3 for anorexia nervosa, 1 for narcolepsy and 2 for Diabetes mellitus. Hence, microsatellite-based genome-wide association analysis provides a new and useful tool for genetic dissection of multifactorial pathologies, including common diseases, serving as an excellent alternative to the SNP or SNP HapMap-based genome-wide association analysis. .

Reference

1. Tamiya G et al: Whole Genome Association Study of Rheumatoid Arthritis using 27,039 Microsatellites, *Human Molecular Genetics*, **14**: 2305-2321, 2005.

Dean, Professor／School of Medicine, Department of Molecular Life Science, Division of Molecular Medical Science and Molecular Medicine／Tokai University／http://mls.med.u-tokai.ac.jp/

Charles Lee, Ph.D. ●Harvard Medical School, USA

Hotspots for copy number variation in humans and chimpanzee

Copy number variation is surprisingly common among humans and can be involved in phenotypic diversity and variable susceptibility to complex diseases, but little is known of the extent of copy number variation in non-human primates. We have used two array-based comparative genomic hybridization platforms to identify a total of 355 copy number variants (CNVs) in the genomes of 20 wild-born chimpanzees (*Pan troglodytes*) and have compared the chimpanzee CNVs to known human CNVs from previous studies. Many CNVs were observed in the same regions in both chimpanzees and humans; especially those CNVs of higher frequency. Strikingly, these loci are enriched 20-fold for ancestral segmental duplications, which may facilitate CNV formation through non-allelic homologous recombination mechanisms. Therefore, some of these regions may be unstable "hotspots" for the genesis of copy number variation, with recurrent duplications and deletions occurring across and within species.

Director of Cytogenetics and Assistant Professor／Pathology／Harvard Medical School／
http://www.chromosome.bwh.harvard.edu

Toshiaki Nakajima

●Department of Molecular Pathogenesis, Medical Research Institute and Laboratory of Genome Diversity, School of Biomedical Science, Tokyo Medical and Dental University, Japan

Natural selection and genes for human common disease

It is about time to identify the genes responsible for human common disease. Approaches to assessing functional significance of genetic variations promise to enhance reproducibility and plausibility of associations between genotypes and human diseases. To identify sequence variations with functional impacts in human genome seems to be a key issue in further studies for human common diseases. However, our understanding of the functional significance of many sequence variations in human genome is limited.

Several approaches, including experimental and computational methods, are available to assess the functional impacts of sequence variations. A well-knowledge of population genetic structure might also provide much information about the functional aspect of sequence variations in human genome. Especially, identifying the genes under the pressure of natural selection based on population genetics approach might be very useful, because sequence variations under the selective pressure must have functional and phenotypical consequences. Moreover, the evidences, which support the association between natural selection and human common diseases, are accumulating now.

Here, we introduce three evidences to support the association between natural selection and the genes responsible for human common disease. First, HLA genes are the best example to support the association between natural selection and human common disease. It is widely accepted that the patterns of sequence variations in HLA genes had been shaped by natural selection and several human common diseases were linked to sequence variations in HLA genes. Second, T-cell immunoglobulin and mucin domains-containing protein 1 gene (*TIM1*), which is related to human allergic disorder, had been under the pressure of natural selection. We showed unusual patterns of genetic variation in *TIM1*. The extent of polymorphism is extremely high in exon 4 of human *TIM1*. Non-synonymous substitutions and insertion/deletion variants are more frequent than synonymous substitutions in that exon. Moreover, nucleotide diversity in *TIM1*-exon 4 sequences is unusually large among primate species. These patterns are similar to those observed in HLA genes. Finally, we showed that the human angiotensinogen gene (*AGT*), which is one of the genes responsible for essential hypertension, had been under the pressure of natural selection in human evolution. The A(-6)G promoter variant of *AGT* is associated with plasma angiotensinogen level and the risk of essential hypertension. A salt sensitive allele, A(-6) allele, is dominant in Africa, but a salt resistance allele, G(-6) allele, has been favored in Caucasian. Our studies suggested that natural selection has shaped the distributions of A(-6)G variant among human populations.

These evidences might provide the new insight into the strategies to identify the genes for human common disease. Identifying the genes under the pressure of natural selection based on population genetics approach might be a promising approach in studies for human common diseases. HapMap data also provide much information about the genes under the pressure of natural selection.

Associate Professor／Department of Molecular Pathogenesis, Medical Research Institute and Laboratory of Genome Diversity, School of Biomedical Science／Tokyo Medical and Dental University

International Symposium on Applied Genomics 2006