ある.

出生前遺伝カウンセリングにおいて参照すべきガイドライン

出生前診断では遺伝学的検査を用いることが多いことから，出生前遺伝カウンセリングを実施する際には，次の2つのガイドラインを参照すべきである.

1. 厚生労働省「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」＜http://www.mhlw.go.jp/shingi/2004/12/s1224-11.html＞

個人情報保護法が2005年4月に完全施行となることを受けて，厚生労働省は「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」を作成した．個人情報の中でも遺伝情報は特殊であることから，ガイドラインの10番目の項目に「遺伝情報を診療に活用する場合の取扱い」が設けられ，「医療機関等が，遺伝学的検査を行う場合には，臨床遺伝学の専門的知識を持ち，本人及び家族等の心理社会的支援を行うことができる者により，遺伝カウンセリングを実施する必要がある」と記載されている（表1）.

2. 遺伝医学関連10学会「遺伝学的検査に関するガイドライン」（日本遺伝カウンセリング学会，日本遺伝子診療学会，日本産科婦人科学会，日本小児遺伝学会，日本人類遺伝学会，日本先天異常学会，日本先天代謝異常学会，日本マススクリーニング学会，日本臨床検査医学会（以上五十音順），家族性腫瘍研究会）＜http://jshg.jp/＞

1. の厚生労働省のガイドラインでは診療の場で用いられる遺伝情報の扱い方の原則のみを示しているが，具体的には厚生労働省のガイドラインにも正式に引用されている本ガイドラインを参照すべきである.

遺伝学的検査においては，生涯変化しない個人の重要な遺伝学的情報が扱われるため，検査実施時のインフォームド・コンセント，個人の遺伝学的情報の保護，検査に用いた生体試料の取り扱い，検査前後の遺伝カウンセリングなど，慎重に検討すべき問題が存在している．また，個人の遺伝学的情報は血縁者で一部共有されており，その影響が個人に留まらないという際立った特徴も有していることから，新たな生命倫理規範が求められていた．遺伝医学関連学会では代表者が集まり，2001年に「遺伝学的検査に関するガイドライン（案）」を発表し，ある一定の評価を得たが，その後，ガイドライン制定に賛同する2学会および法学専門家，生命倫理専門家も加えて内容を検討し，2003年に公表したのがこのガイドラインである.

まず最初に総論として，遺伝学的検査を実施する場合には，遺伝学的検査の有用性を確認することと，総合的臨床遺伝医療の中で行われる必要性が述べられている．総合的な臨床遺伝医療とは医師による情報提供だけではなく，できるだけ専門の異なる複数の医師，さらには医師以外のコ・メディカルのメンバーを含めたチーム医療として対応することを意味している．特に出生前診断は，「選択的人工妊娠中絶」という大変重い倫理的課題と密接に関連している医療行為なので，検査前後の遺伝カウンセリングの実施，およびチーム医療としての取組みが必要である.

また，総論のあとの各論のうちの一つの項目として「出生前検査と出生前診断」が取り上げられており，表2のように記載されている.

表1 厚生労働省「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」(2004. 12. 24告示)
＜http://www.mhlw.go.jp/shingi/2004/12/s1224-11.html＞

10. 遺伝情報を診療に活用する場合の取扱い
遺伝学的検査等により得られた遺伝情報については，本人の遺伝子・染色体の変化に基づく体質，疾病の発症等に関する情報が含まれるほか，その血縁者に関わる情報でもあり，その情報は生涯変化しないものであることから，これが漏えいした場合には，本人及び血縁者が被る被害及び苦痛は大きなものとなるおそれがある．したがって，遺伝学的検査等により得られた遺伝情報の取扱いについては，UNESCO国際宣言等（別表6参照），別表5に掲げる指針及び関連団体等が定める指針を参考とし，特に留意する必要がある．また，検査の実施に同意している場合においても，その検査結果が示す意味を正確に理解することが困難であったり，疾病の将来予測性に対してどのように対処すればよいかなど，本人及び家族等が大きな不安を持つ場合が多い．したがって，医療機関等が，遺伝学的検査を行う場合には，臨床遺伝学の専門的知識を持つ者により，遺伝カウンセリングを実施するなど，本人及び家族等の心理社会的支援を行う必要がある．

表2 遺伝医学関連10学会「遺伝学的検査に関するガイドライン」に記載されている出生前診断の留意点
＜http://jshg.jp/の中の［参考資料］＞

V. 目的に応じた遺伝学的検査における留意点
5. 出生前検査と出生前診断
(1) 妊娠前半期に行なわれる出生前検査及び診断には，羊水，絨毛，その他の胎児試料などを用いた細胞遺伝学的，遺伝生化学的，分子遺伝学的，細胞・病理学的方法，及び超音波検査などを用いた物理学的方法などがある.
(2) 出生前検査及び診断として遺伝学的検査及び診断を行うにあたっては，倫理的及び社会的問題を包含していることに留意しなければならず，とくに以下の点に注意して実施しなければならない.
(a) 胎児が罹患児である可能性（リスク），検査法の診断限界，母体・胎児に対する危険性，副作用などについて検査前によく説明し，十分な遺伝カウンセリングを行うこと.
(b) 検査の実施は，十分な基礎的研修を行い，安全かつ確実な検査技術を習得した産婦人科医により，またはその指導のもとに行われること.
(3) 絨毛採取，羊水穿刺など，侵襲的な出生前検査・診断は下記のような場合の妊娠について，夫婦からの希望があり，検査の意義について十分な理解が得られた場合に行う.
(a) 夫婦のいずれかが，染色体異常の保因者である場合
(b) 染色体異常症に罹患した児を妊娠，分娩した既往を有する場合
(c) 高齢妊娠の場合
(d) 妊婦が新生児期もしくは小児期に発症する重篤なX連鎖遺伝病のヘテロ接合体の場合
(e) 夫婦のいずれもが，新生児期もしくは小児期に発症する重篤な常染色体劣性遺伝病のヘテロ接合体の場合
(f) 夫婦のいずれかが，新生児期もしくは小児期に発症する重篤な常染色体優性遺伝病のヘテロ接合体の場合
(g) その他，胎児が重篤な疾患に罹患する可能性のある場合
(4) 重篤なX連鎖遺伝病のために検査が行われる場合を除き，胎児の性別を告げてはならない.
(5) 出生前診断技術の精度管理については，常にその向上に務めなければならない.
(6) 母体血清マーカー検査の取り扱いに関しては，厚生科学審議会先端医療技術評価部会出生前診断に関する専門委員会による「母体血清マーカー検査に関する見解」，日本人類遺伝学会倫理審議委員会による「母体血清マーカー検査に関する見解」，及び日本産科婦人科学会周産期委員会による報告「母体血清マーカー検査に関する見解について」を十分に尊重して施行する.
(7) 着床前検査及び診断は，極めて高度な知識・技術を要する未だ研究段階にある遺伝学的検査を用いた医療技術であり，倫理的側面からもより慎重に取り扱わなければならない．実施に際しては，日本産科婦人科学会会告「着床前診断に関する見解」に準拠する.

遺伝カウンセリングの診療体制

一口に遺伝カウンセリングといっても扱われる内容はさまざまであり，要求される診療体制もさまざまである．医療に一次医療，二次医療，三次医療があるように，遺伝カウンセリングもその扱う内容により，一次，二次，三次と分類し，それぞれに要求される診療体制を考えることが必要である．

1．一次遺伝相談

(医療としての位置づけがあいまいなものも含まれるので，ここでは遺伝相談と呼ぶ)

保健所の医療相談の一貫として保健師が対応するものや，医療施設の一般外来で担当医に寄せられる遺伝に関する質問への対応などがこれに含まれる．もっとも重要な役割は，質問の内容を吟味し，本格的な遺伝カウンセリングが必要かどうかを判断することである．近親婚や羊水検査の適応などについての定型的な質問には適切に対応することが求められるが，出生前診断を含む遺伝学的検査が必要な場合は，二次あるいは三次の遺伝カウンセリング施設と連携をとることが望ましい．

2．二次遺伝カウンセリング

二次遺伝カウンセリングにおいては，臨床遺伝学のトレーニングを受けた臨床遺伝専門医が，クライエントから家族歴など必要な情報を収集し，問題となっている状況を遺伝医学的に判断する．必要な場合には医学的検査を施行し，正確な遺伝医学的診断を行う．その結果を，クライエントの心理的背景も考慮したうえで，今後の方針についての選択肢を含め，クライエントに伝える．

厚生労働省のガイドラインに記載されているように，出生前診断を含む遺伝学的検査を行う際には臨床遺伝学的知識をもつ者により遺伝カウンセリングが行われるべきである．出生前診断に携わる産婦人科医は臨床遺伝専門医の資格を得るなど臨床遺伝に関するトレーニングを受けるか，あるいは次項で述べる臨床遺伝専門医または認定遺伝カウンセラーと連携をとることが望ましい．

3．三次遺伝カウンセリング

大学病院の遺伝子診療部などがこれにあたる．出生前，小児期，成人期のあらゆる遺伝的問題に対応できる臨床遺伝専門医が複数勤務しており，看護職や心理職も関与するチーム医療の体制がとられている．種々の倫理的問題に対応するため，スタッフカンファレンスがもたれ，必要な場合には，大学の倫理委員会に諮問する体制がとられている．

出生前遺伝カウンセリングにおいて，出生前診断が考慮される場合は，その倫理的問題を検討しなければならない．遺伝医学関連10学会のガイドラインでは，胎児が「重篤」な疾患に罹患している可能性があり，当事者が希望する場合に実施できるとしているが，「重篤」についての定義が難しい．一般的には生命予後が悪く，知的障害が重い疾患を重篤な疾患と考えられているが，成人期以降に発症する神経疾患や小児期から症状はみられるもののそれほど重度とはいえないような疾患で両親が出生前診断を望む場合，どのように対応すべきかについての結論は出ていない．クライエントが望む出生前診断が技術的には可能であっても，その技術を医療として提供することが社会一般に受け入れられるものであるかを慎重に検討しなければならない．このような問題については一人の医師が単独で結論を出すことは好ましいことではなく，遺伝医療部門の臨床遺伝専門医や産科診療部門の専門医とチームを組み三次遺伝カウンセリングの体制のもとで対応すべきである．

遺伝カウンセリング担当者

わが国では，主治医がさまざまな遺伝に関する情報提供を患者・家族に行っていると考えられるが，遺伝カウンセリングでは単なる情報提供だけではなく，心理的・精神的・社会的サポートを行うことがきわめて重要である．遺伝カウンセリングを行おうとする医師は専門分野だけの知識ではなく幅広い遺伝医学の知識を身に付け，遺伝情報の特殊性と倫理的問題を理解し，心理的・精神的・社会的サポートが可能となるような診療体制を構築したうえで遺伝カウンセリングを行う必要がある．遺伝カウンセリングに関連する2つの研修プログラムを紹介する．

1．臨床遺伝専門医制度＜http://jshg.jp/＞

日本人類遺伝学会と日本遺伝カウンセリング学会は適切な遺伝医療を担う医師を育成するために，2002年に合同で臨床遺伝専門医制度（事務局：東北大学大学院医学系研究科遺伝病学分野）を立ち上げた．2005年8月には厚生労働省からも広告のできる専門医としての認可を受け，2005年度までに認定した臨床遺伝専門医は597名にのぼる．

臨床遺伝専門医はすべての診療科からのコンサルテーションに応じ，適切な遺伝医療を実行するとともに，各医療機関において発生することが予想される遺伝・遺伝子に関係した問題の解決を担う医師であり，1）遺伝医学についての広範な専門知識を持っている，2）遺伝医療関連分野のある特定領域について，専門的検査・診断・治療を行うことができる，3）遺伝カウンセリングを行うことができる，4）遺伝学的検査について十分な知識と経験を有している，5）遺伝医学研究の十分な業績を有しており，遺伝医学教育を行うことができる，などの能力を有する医師である．臨床遺伝専門医は内科，小児科，産婦人科，外科などの基本領域の学会の認定医・専門医となった後，臨床遺伝専門医制度研修施設において3年間の研修を行い，筆記試験と面接試験によって認定されるsub-specialtyの学会の専門医である．従来，臨床遺伝専門医を志す医師は小児科医が多かったが，最近，産科医が急増し，臨床遺伝専門医のうち約30%が産婦人科医である．出生前診断においては遺伝情報の適切な取り扱いが求められることが多いので，出生前診断に関わる多くの産婦人科医に臨床遺伝専門医資格の取得をめざしていただきたい．

2．認定遺伝カウンセラー制度＜http://jshg.jp/＞

わが国には，その必要性は叫ばれてはいるものの「遺伝カウンセラー」という非医師の医療職の制度はまだ確立していない．ヒトゲノム解析研究の進展とともに遺伝・遺伝子情報を適切に医療の場で利用しなければならない機会が増え，遺伝カウンセリングの必要性は広く認識されはじめているが，遺伝カウンセリングは誰がどのように行うべきなのかについては定まっていない．そのような状況下で，厚生労働省科学研究費補助金「遺伝子医療の基盤整備に関する研究」班（班長：古山順一）では，「認定遺伝カウンセラーの養成と資格認定に関する研究」を分担研究課題（責任者：千代豪昭）として，わが国における非医師の遺伝カウンセラーの養成と資格認定について検討を重ね，到達目標（知識レベル，技術レベル，態度レベル）と，標準カリキュラム（学ぶべき科目とその単位数）を定め，平成17年度に認定遺伝カウンセラー制度を開始した．2005年度に第1回目の認定試験を行い，わが国初の認定遺伝カウンセラー5名が誕生した．認定遺伝カウンセラーは臨床遺伝専門医と連携しながら質の高い臨床遺伝医療を提供し，遺伝に関する問題に悩むクライエントを援助するとともに，その権利を守る専門家

であり，その養成は原則として大学院修士課程で行うこととしている．すでに，この制度の教育カリキュラムに則った認定遺伝カウンセラーの養成を目的とした修士コースが，2003年度からは信州大学と北里大学で，2004年度からはお茶の水女子大学で，また，2005年以降には，千葉大学，川崎医療福祉大学，京都大学，および近畿大学で開設されている．

出生前遺伝カウンセリングにおいて，認定遺伝カウンセラーが関与することができれば，非常に質の高い遺伝カウンセリングを提供することが可能になると考えられる．

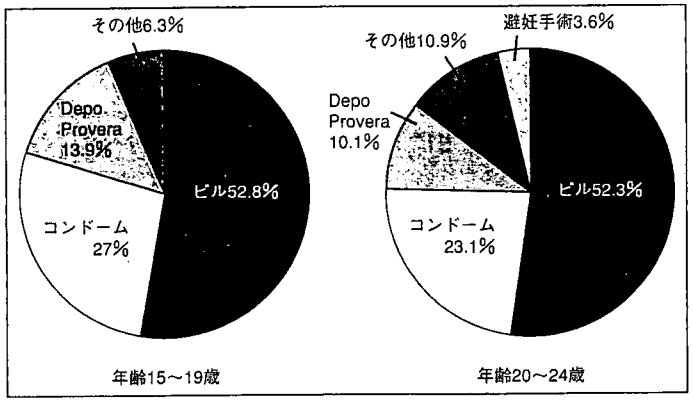
検査前遺伝カウンセリングの重要性

出生前診断でもっとも重要なことは検査を実施する前に遺伝カウンセリングを行うことである．ハイリスクのカップル（遺伝病の家族歴，前児が遺伝病，高齢出産，疑わしい臨床・検査所見をもつ妊婦など）ばかりではなく，とくにリスクが高いとは考えられないカップルに対しても，たとえば，母体血清マーカー検査などを行う場合には，その前に，この検査が中絶の決定の第一歩となるかもしれないことを伝えておくべきである．多くの人は出生前診断やその結果生じ得る人工妊娠中絶の可能性について深くは考えていないので，いざ，そのような状況に直面すると悩み，困惑の度合いは大きくなる．とくにハイリスクのカップルに対しては，あらゆる機会を通じて，妊娠する前から出生前診断を受けるかどうかについて考えておくことを勧めるべきである．

著者連絡先
(〒390-8621)
長野県松本市旭3-1-1
信州大学医学部附属病院
遺伝子診療部
福嶋義光

Quick Uptake

米国女性は避妊に何を使用しているのであろうか？
左図は年齢15～19歳，右図は年齢20～24歳の場合である．



Depo Proveraは1回の筋注で3カ月有効な黄体ホルモン製剤である．
出典は米国疾患コントロールセンター（2002年のデータ）による．　（文責　KY）

日本未熟児新生児学会雑誌　第18巻　第1号　17～22頁（2006年）

●招請講演●

新生児医療に必須の遺伝カウンセリングの基礎知識

信州大学医学部社会予防医学講座遺伝医学分野　信州大学医学部附属病院遺伝子診療部

福嶋義光

Key Words：遺伝カウンセリング，遺伝情報，遺伝学的検査，Dysmorphology，ガイドライン

Ⅰ．はじめに

ヒトゲノム解析研究の進展により，遺伝学（genetics）からゲノム学（genomics）へのパラダイムシフトがおころうとしている。従来の遺伝学は主に単一遺伝子とその影響について研究する学問であったが，新しく生まれつつあるゲノム学は単一遺伝子だけではなく，ゲノム上のすべての遺伝子の機能と相互作用に関する研究であるということができる。遺伝学では主に稀で重篤な単一遺伝子疾患や染色体異常を対象に研究が行われてきたが，ゲノム学の対象はすべての疾患であり，したがってすべての人がその影響を受け，またその恩恵を受けることができる。ヒトゲノム上の生殖細胞系列の遺伝子変異の解析にとどまらず，体細胞における後天的な遺伝子変異やさまざまな時間・空間ごとの遺伝子発現，遺伝子相互間の影響，遺伝子と環境要因との相互作用についてもそれらを研究する手段が生まれ，これらを明らかにすることにより，新しい診断法，治療法，予防法の開発が期待できる。ヒト遺伝子解析により得られる情報は，急速に日々の診療の場で利用されるようになってきており，新生児医学領域においても例外ではない。

本稿では遺伝情報の特殊性に基づき制定されたガイドラインの概要を紹介するとともに，新生児の専門家に有用と考える遺伝カウンセリングに関連した情報提供を行いたい。

別刷請求先：390-8621　長野県松本市旭3-1-1
信州大学医学部社会予防医学講座遺伝医学分野
福嶋義光

Ⅱ．遺伝情報の特殊性

ヒト遺伝情報は，1）個人に関する遺伝的易罹病性を予見しうること，2）世代を超えて，子孫を含めた家族，集団に対して重大な影響を与え得ること，3）試料収集の時点では必ずしも明らかにはされていない情報を含み得ること，4）個人又は集団に対する文化的な重要性を有し得ること（UNESCO：ヒト遺伝情報に関する国際宣言，2003），など通常の医療情報とは異なる側面があるため，慎重な取扱いが求められている。

遺伝子構成は変化しないので，人生のあらゆる時期に診断が可能である。したがって生まれる前に胎児由来の細胞を絨毛穿刺や羊水穿刺によって得ることにより，さまざまな遺伝疾患の出生前診断が可能となる。このことはリスクをかかえるカップルにとっては有用な情報とも考えられるが，この情報により生命の選別が行われてよいのかという倫理的問題が生じる。また，通常臨床検査は病気を発症した人を対象に行われるが，血縁者に遺伝病が発生している場合，本人はその時点で全く健康であっても，リスクのある人を対象として将来その遺伝病が発症するかどうかについての発症前診断も可能となる。治療法・予防法のある疾患では発症の予知は医学的メリットがあるが，治療法・予防法のない疾患の場合には苦痛のみを与えることともなりかねない。さらに遺伝情報は血縁者間で共有されているので，個人の遺伝情報が他の血縁者にも影響を与えることがある点にも注意が必要である。通常の医療は個人を対象に行われるが，遺伝情報を医療の場で用いる際には血縁者も考慮に入れた取組みが必要である。

表1　厚生労働省「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」（2004.12.24告示）

10．遺伝情報を診療に活用する場合の取扱い

遺伝学的検査等により得られた遺伝情報については，遺伝子・染色体の変化に基づく本人の体質，疾病の発症等に関する情報が含まれるほか，生涯変化しない情報であること，またその血縁者に関わる情報でもあることから，これが漏えいした場合には，本人及び血縁者が被る被害及び苦痛は大きなものとなるおそれがある。したがって，検査結果及び血液等の試料の取扱いについては，UNESCO国際宣言，医学研究分野の関連指針及び関連団体等が定める指針を参考とし，特に留意する必要がある。

また，検査の実施に同意している場合においても，その検査結果が示す意味を正確に理解することが困難であったり，疾病の将来予測性に対してどのように対処すればよいかなど，本人及び家族等が大きな不安を持つ場合が多い。したがって，医療機関等が，遺伝学的検査を行う場合には，臨床遺伝学の専門的知識を持ち，本人及び家族等の心理社会的支援を行うことができる者により，遺伝カウンセリングを実施する必要がある。

Ⅲ．ガイドラインの整備と遺伝カウンセリング

我が国では，個人情報保護法の2005年4月全面施行に伴い，医療，研究，産業のそれぞれの分野における適切な個人情報の取扱いに関するガイドラインが作成され，その中にUNESCOの「ヒト遺伝情報に関する国際宣言」の趣旨が盛り込まれている。

診療の場において参照すべきガイドラインには次の2つがあるが，両者とも遺伝情報を扱う際の遺伝カウンセリングの重要性が記載されている。

1．厚生労働省「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」
<http://www.mhlw.go.jp/shingi/2004/12/s1224-11.html>

個人情報保護法が2005年4月に完全施行となることを受けて，厚生労働省は「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」を作成した。個人情報の中でも遺伝情報は特殊であることから，ガイドラインの10番目の項目に「遺伝情報を診療に活用する場合の取扱い」を設け，「医療機関等が，遺伝学的検査を行う場合には，臨床遺伝学の専門的知識を持ち，本人及び家族等の心理社会的支援を行うことができる者により，遺伝カウンセリングを実施する必要がある」と記載している（表1）。

2．遺伝医学関連10学会「遺伝学的検査に関するガイドライン」（日本遺伝カウンセリング学会，日本遺伝子診療学会，日本産科婦人科学会，日本小児遺伝学会，日本人類遺伝学会，日本先天異常学会，日本先天代謝異常学会，日本マススクリーニング学会，日本臨床検査医学会（以上五十音順），家族性腫瘍研究会）<http://jshg.jp/>

1．の厚生労働省のガイドラインでは診療の場で用いられる遺伝情報の扱い方の原則のみを示しているが，具体的には厚生労働省のガイドラインにも正式に引用されている本ガイドラインを参照すべきである。

遺伝学的検査においては，生涯変化しない個人の重要な遺伝学的情報が扱われるため，検査実施時のインフォームド・コンセント，個人の遺伝学的情報の保護，検査に用いた生体試料の取扱い，検査前後の遺伝カウンセリングなど慎重に検討すべき問題が存在している。また個人の遺伝学的情報は血縁者で一部共有されており，その影響が個人に留まらないという際立った特徴も有していることから，新たな生命倫理規範が求められていた。遺伝医学関連学会では代表者が集まり，2001年に「遺伝学的検査に関するガイドライン（案）」を発表し，ある一定の評価を得たが，その後，ガイドライン制定に賛同する2学会および法学専門家，生命倫理専門家も加えて内容を検討し，2003年に公表したのがこのガイドラインである。

Ⅳ．遺伝カウンセリング

1．遺伝カウンセリングの定義

遺伝カウンセリングとは，遺伝性疾患の患者・家族またはその可能性のある人（クライエント）に対して，生活設計上の選択を自らの意思で決定し行動できるよう臨床遺伝学的診断を行い，医学的判断に基づき適切な情報を提供し，支援する医療行為である[1]。遺伝カウンセリングにおいてはクライエントと遺伝カウンセリング担当者との良好な信頼関係に基づき，さまざまなコミュニケーションが行われ，この過程で心理的精神的援助がなされる。遺伝カウンセリングは決して一方的な遺伝医学的情報提供だけではないことに留意すべきである[2][3]。

2．遺伝カウンセリング担当者

我が国では，発端者の診断・治療にあたっている主治医がさまざまな遺伝に関する情報提供を患者・家族

に行っていると考えられるが，遺伝カウンセリングでは単なる情報提供だけではなく心理的・精神的・社会的サポートを行うことが極めて重要である。遺伝カウンセリングを行おうとする医師は専門分野だけの知識ではなく幅広い遺伝医学の知識を身に付け，遺伝情報の特殊性と倫理的問題を理解し，心理的・精神的・社会的サポートが可能となるような診療体制を構築した上で遺伝カウンセリングを行う必要がある。

遺伝カウンセリングに関連する2つの研修プログラムを紹介する。

a）臨床遺伝専門医制度<http://jshg.jp>

日本人類遺伝学会では適切な遺伝医療を担う人材を育成するために，1991年に臨床遺伝学認定医制度を発足させ，2002年からは日本遺伝カウンセリング学会の協力を得て，この認定医制度を臨床遺伝専門医制度（事務局：東北大学大学院医学系研究科遺伝病学分野）としてレベルアップさせている。2005年度までに認定した臨床遺伝専門医は597名にのぼる。

臨床遺伝専門医は，すべての診療科からのコンサルテーションに応じ，適切な遺伝医療を実行するとともに，各医療機関において発生することが予想される遺伝・遺伝子に関係した問題の解決を担う医師であり，1）遺伝医学についての広範な専門知識を持っている，2）遺伝医療関連分野のある特定領域について，専門的検査・診断・治療を行うことができる，3）遺伝カウンセリングを行うことができる，4）遺伝学的検査について十分な知識と経験を有している，5）遺伝医学研究の十分な業績を有しており，遺伝医学教育を行うことができる，などの能力を有する医師であり，3年間の研修の後に筆記試験と面接試験を行って認定する。

b）認定遺伝カウンセラー制度<http://jshg.jp>

我が国には，その必要性は叫ばれてはいるものの「遺伝カウンセラー」という医療職は未だ存在しない。ヒトゲノム解析研究の進展とともに遺伝・遺伝子情報を適切に医療の場で利用しなければならない機会が増え，遺伝カウンセリングの必要性は広く認識されはじめているが，遺伝カウンセリングは誰がどのように行うべきなのかについては定まっていない。そのような状況下で，厚生労働省科学研究費補助金「遺伝子医療の基盤整備に関する研究」班（班長：古山順一）では，「認定遺伝カウンセラーの養成と資格認定に関する研究」を分担研究課題（責任者：千代豪昭）として，我が国における非医師の遺伝カウンセラーの養成と資格認定について検討を重ね，到達目標（知識レベル，技術レベル，態度レベル）と，標準教育カリキュラム（学ぶべき科目とその単位数）を定め，平成17年度に認定遺伝カウンセラー制度を開始した。2005年度に第1回の認定試験を行い，我が国初の認定遺伝カウンセラー5名が誕生した。認定遺伝カウンセラーは臨床遺伝専門医と連携しながら質の高い臨床遺伝医療を提供し，遺伝に関する問題に悩むクライエントを援助するとともに，その権利を守る専門家であり，その養成は原則として大学院修士課程で行うこととしている。すでに，この制度の教育カリキュラムに則った認定遺伝カウンセラーの養成を目的とした修士コースが2003年度からは信州大学医学部と北里大学医学部で，2004年度からはお茶の水女子大学で，また，2005年以降には，千葉大学，川崎医療福祉大学，京都大学，および近畿大学で開設されている。

V．生涯にわたる臨床遺伝医療の必要性

遺伝カウンセリングは確定診断とその告知の場面，あるいは保因者診断や出生前診断など，遺伝学的問題がおきた時だけに必要なものと狭く考えられがちであるが，生涯にわたる臨床遺伝学的フォローアップの過程で継続して行われるべきものである。1）臨床遺伝学的診断，2）臨床遺伝学的治療・フォローアップ，3）保因者診断・出生前診断，などそれぞれの場面で有用と考える事柄を述べておきたい。

1．臨床遺伝学的診断

a）Dysmorphology（異常形態学）

先天異常，とくに先天奇形症候群の診断にDysmorphology（異常形態学）の知識と経験は極めて有用である。我が国では，「Dysmorphologyの夕べ」実行委員会が日本小児遺伝学会の中に設立されており，日本人類遺伝学会，日本小児科学会の前後に研究会を開催している。2006年は日本未熟児新生児学会開催に併せて開催する計画がある。小奇形の見方・考え方，診断がなされていない症例の検討など，Dysmorphologyの基礎を学ぶことができる。

先天奇形症候群の診断には成書（新先天奇形症候群アトラス，Smith's Recognizable Patterns of Human Malformation〈6th ed.〉）の他に，琉球大学の成富研二教授が開発したコンピュータソフトUR-DBMSが極めて有用である。

b）染色体検査

先天異常の診断に染色体検査は欠かすことができない。染色体検査結果の意味について不明な点がある場合には，梶井　正先生（山口大学名誉教授）が染色体の知識のない医師のために公開しているサイト［染色体異常をみつけたら］<http://www16.ocn.ne.jp/~chr.abn/>を参照すべきである。このサイトには次のように記載されている。

染色体検査の報告書を受け取って「染色体異常」だったら，その異常に相当する項目を読めば次のことが分かるはずです。

(1)その染色体異常を理解して報告書を読みこなし，
(2)追加検査をすべきか，
(3)両親の染色体検査をすべきか，
(4)細胞遺伝学の専門家に紹介すべきかを決め，
(5)代表的な染色体異常ではカウンセリングの基礎的方法を知り，
(6)両親（本人）に染色体異常について説明する。

c）遺伝学的検査（遺伝子検査）

遺伝学的検査を行う場合には，遺伝医学関連10学会の「遺伝学的検査に関するガイドライン」を参照すべきである。近年，従来原因が不明だった疾患（先天奇形症候群など）の責任遺伝子が見つかったという論文が次々と発表されているが，これはこの責任遺伝子を個々の症例で検査すれば，必ず見つかる，すなわち確定診断の方法が確立したことを意味していない。多くの場合，典型的な症例であっても，遺伝子の変異が発見されないことも多く，遺伝子解析はほとんど研究レベルで行われていることを理解しておかなければならない。

遺伝学的検査については，慶応大学小児科のサイト［先天異常遺伝子診断<http://www.dhplc.jp/info/info.html>］，京都大学遺伝子診療部の［ヒトGermline遺伝子・染色体検査オンラインデータベース<http://www.kuhp.kyoto-u.ac.jp/idennet/>］，および米国ワシントン大学のGeneTests http://www.geneclinics.org/が有用である。

2．臨床遺伝学的治療・フォローアップ

染色体異常，先天奇形症候群，先天代謝異常など遺伝学的バックグランドに基づく疾患をもって誕生した児に対しては，生涯にわたる医療的ケアが必要である。近年，徐々に稀な疾患の自然歴，長期予後，managementの情報が得られつつあり参考にすべきである。

Management of Genetic Syndromes〈2nd ed.〉(Cassidy SB & Allanson JE, Wiley-Liss)には，先天奇形症候群を中心に55疾患について，頻度，診断基準，原因，病態生理，遺伝学，診断のための検査法，鑑別診断とともに，おこりうる症状とそれぞれに対する詳細な治療法，ケアの方法が記載されている。

インターネットのサイトとしてはワシントン大学のGeneReviews <www.geneclinics.org>と，その日本語訳を許可を得て掲載している信州大学の遺伝子診療部遺伝ネットワークGENETOPIA <genetopia.md.shinshu-u.ac.jp>の中のGeneReview Japanが有用である。

3．新生児領域で特に留意すべき遺伝学的検査と遺伝カウンセリング

a）小児を対象とした遺伝学的検査

小児を対象とした遺伝学的検査について，遺伝医学関連10学会「遺伝学的検査に関するガイドライン」には，以下の記載がある。

Ⅱ4（2）治療法または予防法が確立されていない成人期以後に発症する遺伝性疾患について，小児期に遺伝学的検査を行うことは，基本的に避けるべきである。

Ⅱ4（3）将来の自由意思の保護という観点から，未成年者に対する遺伝学的検査は，検査結果により直ちに治療・予防措置が可能な場合や緊急を要する場合を除き，本人が成人に達するまで保留すべきである。

この記載の趣旨を説明したい。通常，診療の場で行われる検査は，診断を確定し，よりよい医療を行うために行われる。遺伝学的検査には通常の臨床検査と同様，適切な治療の選択を可能とし，今後の予後の推定など医学的メリットがはっきりしているものもあるが，治療法も予防法も確立していない疾患の発症前遺伝子診断など医学的メリットがはっきりしないものもある。たとえ我が子の検査であっても親の考えでこのような医学的メリットがはっきりしない遺伝学的検査を行うことは好ましいことではなく，その小児が成人した後に自分自身の意思により行われるべきであることを述べている。

b）保因者診断

保因者診断についても同様の注意が必要である。一般に患児の遺伝子診断がなされると親の希望により，その同胞の保因者診断が安易になされる傾向がある。保因者とは，1）メンデル遺伝病（主に常染色体劣性病およびX連鎖劣性遺伝）において疾患を決定する遺伝子変異をヘテロ接合の状態で有している人，および2）均衡型染色体異常を有している人，をいう。すなわち，保因者は遺伝子変異を持っているが健康であり，将来にわたって自分自身がその病気を発症することはない；しかし，保因者の子どもは発症する可能性がある（成人期発症の常染色体優性遺伝病で遺伝子変異をもっているがまだ発症していない人は，保因者ではなく未発症者という）。

ここで保因者診断の意義について考えてみたい。たとえ保因者であっても，本人は発症しないので，健康管理に役立てられる一般の臨床検査とは異なる。しかし，X連鎖劣性遺伝病や染色体異常などでは出生前診断が可能なものがあるので，将来の妊娠に備えて保因者かどうかを明らかにしておくことは，生殖行動における選択肢を増やすという意味でメリットのあることである。しかし，これは十分な遺伝カウンセリングを行った上で，本人の自発的希望により行われるものである。

Duchenne 型筋ジストロフィーや血友病などのX連鎖劣性遺伝病および転座型の染色体異常などでは従来，親の希望により患者の同胞の保因者診断が行われることも少なくなかったと思われる。しかし，小児期に保因者であると判明した場合には，さまざまな心理的・精神的軋轢が本人および家族に生じる可能性があることを考え，小児期の保因者診断は安易に行うべきではない。保因者診断においては自己決定権の尊重が最も重要となるので，親の判断ではなく，本人が成人となるまで待ち，十分な遺伝カウンセリングを受けた上で保因者診断を受けるかどうかを本人自身に判断していただくことも考慮すべきである。

c）出生前診断

第1子が遺伝性疾患に罹患している場合，次子の妊娠について出生前診断を求められる場合がある。出生前診断はその結果によっては人工妊娠中絶につながる可能性があることから，さまざまな倫理的諸問題が含まれており，慎重に対応する必要がある。Ⅲ-2の遺伝医学関連10学会「遺伝学的検査に関するガイドライン」では，出生前診断について下記のように記載されている。

Ⅴ5（3）絨毛採取，羊水穿刺など，侵襲的な出生前検査・診断は下記のような場合の妊娠について，夫婦からの希望があり，検査の意義について十分な理解が得られた場合に行う。

(a) 夫婦のいずれかが，染色体異常の保因者である場合
(b) 染色体異常症に罹患した児を妊娠，分娩した既往を有する場合
(c) 高齢妊娠の場合
(d) 妊婦が新生児期もしくは小児期に発症する重篤なX連鎖遺伝病のヘテロ接合体の場合
(e) 夫婦のいずれもが，新生児期もしくは小児期に発症する重篤な常染色体劣性遺伝病のヘテロ接合体の場合
(f) 夫婦のいずれかが，新生児期もしくは小児期に発症する重篤な常染色体優性遺伝病のヘテロ接合体の場合
(g) その他，胎児が重篤な疾患に罹患する可能性のある場合

この中で出生前診断の適応として記載されている「重篤」についての定義が難しい。一般的には生命予後が悪く，知的障害の重い疾患が重篤な疾患と考えられているが，成人期以降に発症する神経疾患や小児期から症状はみられるもののそれほど重度とはいえないような疾患で両親が出生前診断を望む場合，どのように対応すべきかについての結論は出ていない。クライエントが望む出生前診断が技術的には可能であっても，その技術を医療として提供することが社会一般に受け入れられるものであるかを慎重に検討しなければならない。このような問題については一人の医師が単独で結論を出すことは好ましいことではなく，遺伝医療部門の臨床遺伝専門医や産科診療部門の専門医とチームを組み対応すべきである。

文　献

1) 鈴木友和他．わが国における遺伝カウンセリングのあり方について．平成11年度厚生科学研究費補助金（子ども家庭総合研究事業）「遺伝医療システムの構築と運用に関する研究」（主任研究者：古山順一）報告書，2000；27-31.
2) 千代豪昭．遺伝カウンセリング―面接の理論と技術．医学書院，東京，2000.
3) 新川詔夫，福嶋義光．遺伝カウンセリングマニュアル（第2版）．南江堂，東京，2003.

検査室の安全管理・12

生体試料の取り扱いと倫理　その3
遺伝子・染色体検査

涌井敬子[*1]・福嶋義光[*2]

はじめに

ヒトゲノム解析研究の急速な進展により，2003年にはヒトの持つ30億塩基対の塩基配列の一次構造が決定された．さまざまな疾患の責任遺伝子や発症のメカニズムが分子レベルで明らかにされ，新たな診断法の開発や，原因に基づく病態や治療へ向けての研究が進みつつある．このような近年の分子遺伝学の著しい進歩による成果は，遺伝子診断という形で医療現場に導入されつつある．一方，2005年4月より施行された個人情報保護法を受けて，遺伝情報の扱いについて国の倫理ガイドラインが整備された．

本稿では遺伝子・染色体検査を目的別に整理し，それぞれの目的における生体試料の取り扱い上の留意点，特に生殖細胞系列変異を検索する遺伝子・染色体検査の特殊性とその倫理的問題について概説する．

遺伝子・染色体検査の分類

分子遺伝学的技術の進歩により，遺伝子・染色体検査の応用範囲は広がっている．下記にさまざまな検査の分類を示しながらその内容について解説する．

1．検出対象による分類

臨床検査に応用されている遺伝子・染色体検査は，検出しようとする対象により，感染症などにおける病原微生物などの遺伝情報の検出を対象とする検査と，ヒトの身体の遺伝情報を対象とする検査との二つに分類することができる．

1）病原微生物などのヒト以外の遺伝情報の検出を対象とする検査

ヒトに病気を引き起す原因となった微生物やウイルスなどが生体内に感染しているかどうか，感染しているとしたらどのような種類なのかといったことを同定することにより，適切な治療法の選択やその効果の確認などに利用される．ヒトに存在しない外来遺伝子についての検査であり，基本的には遺伝情報に関する倫理的問題の対象とはならない．ただし，社会的影響のある微生物やウイルスなどの検査を行う際には，個人情報保護に留意する必要がある．

2）ヒトの遺伝情報の検出を対象とする検査

ヒトの遺伝情報には，染色体・遺伝子・DNAがある．その遺伝情報にも，ヒトが生命活動を営むために必要な遺伝情報と，ヒトの体質・多様性・個性に関わる遺伝情報とがあり，現在，臨床検査の対象になっているのは，直接疾患の発症に

用語解説

生殖細胞系列変異(germline mutation)と体細胞変異(somatic mutation)：生殖細胞系列の変異とはその個体が形成される基となった精子あるいは卵子(生殖細胞)の段階で既にその変異が存在している，すなわち受精卵の段階で存在している遺伝子変異のことであり，その個体の持つ細胞すべてに共通して存在している．この変異は生涯変化することがなく，また血縁者とも共有している可能性がある．一方，体細胞変異とは体の細胞のごく一部の細胞だけに後天的に生じた変異である．癌細胞にみられる変異がその代表である．変異を持つ細胞(癌細胞など)以外の細胞には，この変異は存在せず，次世代に受け継がれることも，血縁者が共有しているということもない．

[*1] 信州大学医学部社会予防医学講座遺伝医学分野・講師，同大学医学部附属病院遺伝子診療部
〒390-8621 長野県松本市旭3-1-1
[*2] 同大学医学部社会予防医学講座遺伝医学分野・教授，同大学医学部附属病院遺伝子診療部・部長

かかわる前者である．

本稿での解説は，染色体異常と，現在までに責任遺伝子が同定されその解析法が確立し臨床応用されている，いくつかの単一遺伝子疾患やミトコンドリア遺伝病などを想定したものである．将来的には，生活習慣病と総称される高血圧，糖尿病，心筋梗塞などの多因子遺伝疾患においても複数の関連遺伝子などが解明され，個々人の薬物反応性・疾患感受性などのいわゆる体質の違いを考慮に入れた，オーダーメイド医療が導入されることが期待されているが，現在はそれを実現するために必要な研究を推進しているところで，関連すると考えられる複数の遺伝子のうちの一つの遺伝子変異だけをもって，正確に将来の発症予測や予防のための対策がとれる段階ではない．技術は開発されても，積極的に対応すべきものと，慎重な対応が求められるもの，さらには現時点では対応すべきでない検査もあるということを，医療に携わる者は認識する必要がある．

生体試料を扱う臨床検査としての遺伝子・染色体検査が行われる場面としては，先天異常の原因検索，生殖障害の原因検索，遺伝性疾患の確定診断，家系内の遺伝性疾患の保因者診断，胎児の出生前診断，腫瘍細胞の診断などがある．用いる生体試料としては，腫瘍細胞の診断には白血病細胞や癌細胞の含まれる組織・細胞が，出生前診断には羊水細胞・絨毛細胞が，そのほかには主に血液が用いられる．

2．遺伝学的背景による分類

遺伝学的背景からは，個体を構成するすべての細胞に共通して存在する遺伝子変異すなわち生殖細胞系列変異を解析する検査と，腫瘍細胞など体の一部の細胞に後天的に起こった遺伝子変異すなわち体細胞変異を解析する検査とに分類される．臨床検査を進めるうえで，この両者の違いをしっかり認識することは大変重要である．

以下にそれぞれにつき概説する．

1）生殖細胞系列変異を明らかにする遺伝子検査・染色体検査

個体を形成する約60兆個の細胞は原則としてすべて同じ遺伝子構成を有し，細胞のひとつひとつに約30,000の遺伝子を含んでいると予想されている．個体が発生して以降，原則として遺伝子構成は変化しないので，生殖細胞系列変異を検索する目的では，クライエントへの侵襲も比較的少なく用意に採取できる血液細胞を用いるのが一般的である．例えば神経疾患であっても採取困難な神経細胞や脳細胞を採取せずとも，血液中の白血球で検査が可能ということである．また，生殖細胞系列変異は，生涯変化しないので，人生のどの時期でも検査が可能である．例えば成人期以降に発症することの多い遺伝性腫瘍などでは，発症前に検査をして，早期発見・早期治療の医療にのせることができるという利点もある．そして，発端者の遺伝情報は，それを共有する血縁者にもかかわる可能性のある情報であるという特殊性もある．上記のような理由から，一般の臨床検査とは異なる種々の倫理的問題が含まれている．

具体的な適応としては，①先天性の染色体異常症や，責任遺伝子の同定された既知の奇形症候群の診断，②習慣流産や不妊症など，生殖障害の原因検索，③責任遺伝子の同定された遺伝性疾患（奇形症候群，酵素欠損症，骨系統疾患，神経変性疾患，家族性腫瘍など）の確定診断，④責任遺伝子の同定された常染色体優性遺伝性疾患（神経変性疾患，家族性腫瘍など）の発症前診断，⑤責任遺伝子の同定された劣性遺伝性疾患（酵素欠損症など）の保因者診断，⑥親の染色体均衡型構造異常や遺伝子変異が同定された重篤な遺伝性疾患の出生前診断など，が挙げられる．適応の詳細については他書[1]を参照いただきたい．

倫理との関連については，後述の「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」を遵守するためのポイントについて以下に解説する．

生殖細胞系列変異を明らかにする遺伝子検査・染色体検査を実施する際には，検査前にクライエントに検査の目的，検出限界などをわかりやすい言葉で十分に説明し，同意を得ることが必要である．そして，検査前の説明から検査結果の告知，さらに告知後のフォローをも含む，遺伝カウンセリングの一環として実施されなければならない．遺伝カウンセリングは十分な経験を有した臨床遺伝専門医を中心とした総合的臨床遺伝医療として行われるべきである．

特に他施設に検査を依頼する際の注意事項としては，依頼書および検体には連結可能匿名化された匿名化IDを記して，本名や生年月日など個人情報に該当する項目は検査に支障のないかぎり記載しないことが求められるようになってきた．再連結可能な匿名符号を発行しラベル印刷する機能付きのシステムも提案されている〔Secure-Name™（NTTデータ）など〕．

病院内でも，遺伝医療として行われた，遺伝子検査・染色体検査の結果を含めた遺伝カウンセリングに関する診療録は，通常の診療録とは別に，アクセス権限を限定した施錠できる独立したロッカーに保管することが望ましいとされている．

近年，生殖細胞系列変異を明らかにする遺伝子検査のいくつかは検査センターなどでも受託するようになってきたが，保険適応になっている項目はない．一方，染色体検査は保険適応になっている検査で，これまでは通常の臨床検査と同様に扱われてきたが，上記のような対応が求められるようになってきたことを周知願いたい．

2）腫瘍細胞の遺伝子検査・染色体検査

癌化した細胞にどのような遺伝的変化が生じているかどうかを検索しようとするもので，ある組織に分化した体細胞の一部に，正常細胞が腫瘍化するのに関連した突然変異が起こっていないかどうかを検索する検査である．病型分類，治療法の選択，治療効果の判定などのための必須項目となっている．以下のような目的が，腫瘍細胞の遺伝子検査・染色体検査を行うための適応である．①初発時：発生した腫瘍の性質の判断，病型分類，治療法の選択，②経過観察時・寛解期：治療の効果判定，および再発のスクリーニング（初発時に認めた体細胞変異が残存しているか/減少しているか/消失しているかなど），③再発時：腫瘍の性質の判断（初発時と同様の病型か/異なる種類の腫瘍か），その後の治療法の選択，などである．

腫瘍細胞の遺伝子検査・染色体検査は，腫瘍細胞のみに起こった後天的な変異を調べるものなので，基本的には通常の臨床検査と同様に考えることができる．ただし，本人が生まれもった遺伝的背景が基本となるため，生殖細胞系列に生来有している，発生した腫瘍とは関係のない染色体異常や遺伝子変異を有しているクライエントがいることも想定され，解析方法によっては，その生殖細胞系列変異も偶然検出してしまうことがあるので，初回の検査時には生殖細胞系列変異を明らかにする検査としての配慮も必要である．腫瘍細胞の遺伝子・染色体検査の結果，生殖細胞系列変異が疑われた際には，それが生殖細胞系列の変異なのか，腫瘍細胞の変異なのか，そして生殖細胞系列変異と判明した場合には，その変異と発生した腫瘍との関連性について調べる必要がある．

ヒト遺伝情報に関する倫理ガイドライン

UNESCO（United Nations Educational, Scientific and Cultural Organization，国際連合教育科学文化機構）は，2003年「ヒト遺伝情報に関する国際宣言」を採択した〈http://www.mext.go.jp/b=menu/shingi/gijyutu/gijyutu1/shiryo/001/04010701.htm〉．そのなかに，ヒト遺伝情報は，①個人に関する遺伝的易罹病性を予見しうること，②世代を越えて，子孫を含めた家族，集団に対して重大な影響を与えうること，③試料収集の時点では必ずしも明らかにはされていない情報を含みうること，④個人または集団に対する文化的な重要性を有しうること，の理由により特別な地位が与えられるべきであり，したがって，ヒト遺伝情報は，ⓐ医療，ⓑ研究，ⓒ法的措置など，に限って用いられるべきであり，健康に関わる重要な意味を持つ可能性がある遺伝学的検査を行う場合には，当事者が遺伝カウンセリングを適切な方法で受けられるようにすべきである，と謳っている．

わが国においては，2005年4月から実施された「個人情報保護法」に伴って，上記のUNESCO国際宣言の趣旨を盛り込んで策定あるいは改訂された，国が定めた三つのガイドライン，すなわち，①厚生労働省から告示された「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」，②3省（文部科学省，厚生労働省，経済産業省）から告示された「ヒトゲノム・遺伝子解析研究に関する倫理指針」，③経済産業省から告示された「経済産業分野のうち個人遺伝情報を用いた事業分野における個人情報保護ガイドライン」，が現在運用されている．また，④遺伝医学関連10学会が「遺伝学的検査に関す

るガイドライン」を，⑤日本衛生検査所協会が「ヒト遺伝子検査受託に関する倫理指針」のガイドラインを定めている．下記に各ガイドラインについて簡単に解説する．原文は各ウェブサイトを参照されたい．

1．「医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン」(2004年12月28日告示)(厚生労働省)〈http://www.mhlw.go.jp/shingi/2004/12/s1224-11.html〉

医療の現場で診療目的に行われる遺伝情報の取り扱いについて，国が示した初めての指針である．UNESCOの「ヒト遺伝情報に関する国際宣言」や医学研究分野の関連指針および関連団体などが定める指針遺伝(わが国では遺伝医学関連10学会の「遺伝学的検査に関するガイドライン」(後述)が該当する)などを参考とし，「臨床遺伝学の専門的知識を持ち，本人及び家族等の心理社会的支援を行うことができる者により，遺伝カウンセリングを実施する必要がある．」と明記された．

2．「ヒトゲノム・遺伝子解析研究に関する倫理指針」(文部科学省，厚生労働省，経済産業省)(2001年3月29日告示，2004年12月28日全部改正・告示，2005年6月29日一部改正)〈http://www.mext.go.jp/a＝menu/shinkou/seimei/main.htm〉

遺伝子解析研究を実施する際のガイドラインとして定められたもので，2001年に国として定めた最初の遺伝情報の取り扱いに関するガイドラインを，遺伝情報に関する個人情報保護法の趣旨を盛り込むために見直し改訂したものである．診療目的といっても，研究的要素の含まれる場合もあり，ここに記載されている内容は，診療時にも考慮すべき内容を含んでいるので，遺伝医療にかかわる者は理解していなければならない．

3．「経済産業分野のうち個人遺伝情報を用いた事業分野における個人情報保護ガイドライン」(経済産業省)(2004年12月17日告示)〈http://www.meti.go.jp/press/20041217010/041217iden.pdf〉

臨床的意義がまだ確立されていない体質に関する遺伝学的検査(例えば，太りやすいかどうかを調べる目的の遺伝子検査など)がスポーツクラブやエステサロンにおいて行われたり，インターネットを介して親子鑑定などを受託する企業が現れ，社会的混乱をきたすことが憂慮されていたが，経済産業省では，このガイドラインを制定することにより，個人遺伝情報を扱う事業者を網羅的に指導することとした．対象はいわゆる体質検査，DNA鑑定・親子鑑定，遺伝子解析受託(臨床検査を除く)を行う事業者である．遺伝情報の扱いかたについては，「遺伝情報の開示には，臨床遺伝学の専門的知識・経験を有した医師，医療従事者による遺伝カウンセリングを受けられるような体制を整えること」が明記された．

4．遺伝医学関連10学会(日本遺伝カウンセリング学会，日本遺伝子診療学会，日本産科婦人科学会，日本小児遺伝学会，日本人類遺伝学会，日本先天異常学会，日本先天代謝異常学会，日本マススクリーニング学会，日本臨床検査医学会，家族性腫瘍研究会)「遺伝学的検査に関するガイドライン」(2003年8月公表)〈http://jshg.jp〉

診療として生殖細胞系列の変異を解析する遺伝学的検査を行う際のガイドラインとしては，先に述べた厚生労働省のガイドラインが存在するが，そこで謳われている，「参考にすべき～関連団体などが定める指針」の代表的な指針であり，配慮すべき事項について詳細に記載されている．診療を目的として遺伝学的検査を実施する際には，このガイドラインを十分理解したうえで実施することが求められる．

遺伝学的検査を行う際の留意点として，検査を実施する場合には検査の有用性，分析的妥当性(確立された検査法，適切な精度管理の実施)，臨床的妥当性(感度，特異度，陽性的中率の確認)，臨床的有用性が確認できていることが必要であること，遺伝学的検査は，検査前の説明，検査結果告知，フォローアップを含め，遺伝カウンセリングの一環として実施されなければならないこと，専門の異なる複数の医師とコメディカルのメンバーを含めたチーム医療で行う総合的臨床遺伝医療の中で行われる必要があることなどが明記されている．さらに，遺伝学的差別の防止，インフォームドコンセントにおける留意点，遺伝学的検査を行わない場合，小児を対象とした遺伝学的検査，検査試料の取り扱い，個人遺伝学的情報の保護，臨床的有用性の確立していない遺伝学的検査実施の禁止などについて記載されている．

5．日本衛生検査所協会「ヒト遺伝子検査受託に関する倫理指針」のガイドライン(2001年4月10日公表)〈http://www.jrcla.or.jp/〉

検体検査業務を請負っている衛生検査所で組織されている，社団法人日本衛生検査所協会は，ヒト遺伝子検査の一部がさまざまな倫理的・法的・社会的問題を招く可能性が高いとの考えから，独自の取り組みとして「ヒトゲノム・遺伝子解析研究に関する倫理指針」を策定し，受託検査は，検出限界などについての基礎データを把握したうえで，技術的な安全性，精度管理，臨床診断上の有用性の確立したものに限ることなどの遵守事項を定めた．

おわりに

個々人の遺伝情報は究極の個人情報ということもできる．倫理問題に対するあるべき体制についてのガイドラインは作成されたが，これらのガイドラインを遵守するために必須とされている遺伝カウンセリング体制を構築するためにはまだまだ課題が存在する．従来，ヒトゲノムの遺伝子検査は研究でのみ行われていたが，現在は，技術も有用性も妥当性も確立した検査も増えてきた．にもかかわらず保険診療として実施できるものはない．希望するクライエントに提供する遺伝カウンセリングや遺伝学的検査にかかる費用に見合う収入がなんらかの形で保証されなければ，今後，医療として発展できないと考える．さらに，わが国ではまだ遺伝学的検査の精度管理体制が整備されていないのが現状で，医療の質の向上のためにこれらの体制の構築が急がれる．

遺伝学的検査を行う場合には十分な倫理的配慮が必要で，包括的な遺伝子診療システムのなかで取り扱われるべきであることを強調したい．

文　献

1）福嶋義光，涌井敬子，金子安比古：染色体・遺伝子検査．金井正光(編)：臨床検査法提要改訂第32版．金原出版，pp 1161-1231，2005

ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination rates vary systematically around genes and between genes of different function. Finally, we demonstrate increased differentiation at non-synonymous, compared to synonymous, SNPs, resulting from systematic differences in the strength or efficacy of natural selection between populations.

Advances made possible by the Phase I haplotype map

The International HapMap Project was launched in 2002 with the aim of providing a public resource to accelerate medical genetic research. The objective was to genotype at least one common SNP every 5 kilobases (kb) across the euchromatic portion of the genome in 270 individuals from four geographically diverse populations[1,2]: 30 mother–father–adult child trios from the Yoruba in Ibadan, Nigeria (abbreviated YRI); 30 trios of northern and western European ancestry living in Utah from the Centre d'Etude du Polymorphisme Humain (CEPH) collection (CEU); 45 unrelated Han Chinese individuals in Beijing, China (CHB); and 45 unrelated Japanese individuals in Tokyo, Japan (JPT). The YRI samples and the CEU samples each form an analysis panel; the CHB and JPT samples together form an analysis panel. Approximately 1.3 million SNPs were genotyped in Phase I of the project, and a description of this resource was published in 2005 (ref. 3).

The initial HapMap Project data had a central role in the development of methods for the design and analysis of genome-wide association studies. These advances, alongside the release of commercial platforms for performing economically viable genome-wide genotyping, have led to a new phase in human medical genetics. Already, large-scale studies have identified novel loci involved in multiple complex diseases[4,5]. In addition, the HapMap data have led to novel insights into the distribution and causes of recombination hotspots[3,6], the prevalence of structural variation[7,8] and the identity of genes that have experienced recent adaptive evolution[3,9]. Because the HapMap cell lines are publicly available, many groups have been able to integrate their own experimental data with the genome-wide SNP data to gain new insight into copy-number variation[10], the relationship between classical human leukocyte antigen (HLA) types and SNP variation[11], and heritable influences on gene expression[12–14]. The ability to combine genome-wide data on such diverse aspects of genetic variation with molecular phenotypes collected in the same samples provides a powerful framework to study the connection of DNA sequence to function.

*Lists of participants and affiliations appear at the end of the paper.

In Phase II of the HapMap Project, a further 2.1 million SNPs were successfully genotyped on the same individuals. The resulting HapMap has an SNP density of approximately one per kilobase and is estimated to contain approximately 25–35% of all the 9–10 million common SNPs (minor allele frequency (MAF) ≥ 0.05) in the assembled human genome (that is, excluding gaps in the reference sequence alignment; see Supplementary Text 1), although this number shows extensive local variation. This paper describes the Phase II resource, its implications for genome-wide association studies and additional insights into the fine-scale structure of linkage disequilibrium, recombination and natural selection.

Construction of the Phase II HapMap

Most of the additional genotype data for the Phase II HapMap were obtained using the Perlegen amplicon-based platform[15]. Briefly, this platform uses custom oligonucleotide arrays to type SNPs in DNA segmentally amplified via long-range polymerase chain reaction (PCR). Genotyping was attempted at 4,373,926 distinct SNPs, which corresponds, with exceptions (see Methods), to nearly all SNPs in dbSNP release 122 for which an assay could be designed. Additional submissions were included from the Affymetrix GeneChip Mapping Array 500K set, the Illumina HumanHap100 and HumanHap300 SNP assays, a set of ~11,000 non-synonymous SNPs genotyped by Affymetrix (ParAllele) and a set of ~4,500 SNPs within the extended major histocompatibility complex (MHC)[11]. Genotype submissions were subjected to the same quality control (QC) filters as described previously (see Methods) and mapped to NCBI build 35 (University of California at Santa Cruz (UCSC) hg17) of the human genome. The re-mapping of SNPs from Phase I of the project identified 21,177 SNPs that had an ambiguous position or some other feature indicative of low reliability; these are not included in the filtered Phase II data release. All genotype data are available from the HapMap Data Coordination Center (http://www.hapmap.org) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP); analyses described in this paper refer to release 21a. Three data sets are available: 'redundant unfiltered' contains all genotype submissions, 'redundant filtered' contains all submissions that pass QC, and 'non-redundant filtered' contains a single QC+ submission for each SNP in each analysis panel.

The QC filters remove SNPs showing gross errors. However, it is also important to understand the magnitude and structure of more subtle genotyping errors among SNPs that pass QC. We therefore carried out a series of analyses to assess the influence of the long-range PCR amplicon structure on genotyping error, the concordance rates between genotype calls from different genotyping platforms and between those platforms and re-sequencing assays, as well as the rates of false monomorphism and mis-mapping of SNPs (see Supplementary Text 2, Supplementary Figs 1–3 and Supplementary Tables 1–4). We estimate that the average per genotype accuracy is at least 99.5%. However, there are higher rates of missing data and genotype discrepancies at non-reference alleles, with some clustering of errors resulting from the amplicon design and a few incorrectly mapped SNPs.

Table 1 shows the numbers of SNPs attempted and converted to QC+ SNPs in each analysis panel (Supplementary Table 5 shows a breakdown by each major submission). Haplotypes and missing data were estimated for each analysis panel separately using both trio information and statistical methods based on the coalescent model (see Methods). To enable cross-population comparisons, a consensus data set was created consisting of 3,107,620 SNPs that were QC+ in all analysis panels and polymorphic in at least one analysis panel. The equivalent figure from Phase I was 931,340 SNPs. Unless stated otherwise, all analyses have been carried out on the consensus data set. An additional set of haplotypes was created for those SNPs in the consensus where a putative ancestral state could be assigned by comparison of the human alleles to the orthologous position in the chimpanzee and rhesus macaque genomes.

The variation in SNP density within the Phase II HapMap is shown in Fig. 1. On average there are 1.14 genotyped polymorphic SNPs per kilobase (average spacing is 875 base pairs (bp)) and 98.6% of the assembled genome is within 5 kb of the nearest polymorphic SNP. Still, there is heterogeneity in genotyped SNP density at both broad (Fig. 1a) and fine (Fig. 1b) scales. Furthermore, there are systematic changes in genotyped SNP density around genomic features including genes (Fig. 1c).

The Phase II HapMap differs from the Phase I HapMap not only in SNP spacing, but also in minor allele frequency distribution and patterns of linkage disequilibrium (Supplementary Fig. 4). Because the criteria for choosing additional SNPs did not include consideration of SNP spacing or preferential selection for high MAF, the SNPs added in Phase II are, on average, more clustered and have lower MAF than the Phase I SNPs. Because MAF predictably influences the distribution of linkage disequilibrium statistics, the average r^2 at a given physical distance is typically lower in Phase II than in Phase I; conversely, the $|D'|$ statistic is typically higher (data not shown). One notable consequence is that the Phase II HapMap includes a better representation of rare variation than the Phase I HapMap.

The increased resolution provided by Phase II of the project is illustrated in Fig. 2. Broadly, an additional SNP added to a region shows one of three patterns. First, it may be very similar in distribution to SNPs present in Phase I. Second, it may provide detailed resolution of haplotype structure (for example, a group of chromosomes with identical local haplotypes in Phase I can be shown in Phase II to carry

Table 1 | Summary of Phase II HapMap data (release 21)

Phase	SNP categories	Analysis panel		
		YRI	CEU	CHB+JPT
I	Assays submitted	1,304,199	1,344,616	1,306,125
	Passed QC	1,177,312 (90%)	1,217,902 (91%)	1,187,800 (91%)
	Did not pass QC	126,887 (10%)	126,714 (9%)	118,325 (9%)
	>20% missing	82,463 (65%)	95,684 (76%)	78,323 (66%)
	>1 duplicate inconsistent	6,049 (5%)	5,126 (4%)	9,242 (8%)
	>1 mendelian error	18,916 (15%)	11,310 (9%)	N/A
	<0.001 Hardy–Weinberg *P*-value	10,265 (8%)	8,922 (7%)	13,722 (12%)
	Other failures	19,345 (15%)	13,858 (11%)	20,674 (17%)
II	Assays submitted	5,044,989	5,044,996	5,043,775
	Passed QC	3,150,433 (62%)	3,204,709 (64%)	3,244,897 (64%)
	Did not pass QC	1,894,556 (38%)	1,840,287 (36%)	1,798,878 (36%)
	>20% missing	1,419,000 (75%)	1,398,166 (76%)	1,403,543 (78%)
	>1 duplicate inconsistent	0 (0%)	0 (0%)	6,617 (0%)
	>1 mendelian error	172,339 (9%)	127,923 (7%)	N/A
	<0.001 Hardy–Weinberg *P*-value	96,231 (5%)	82,268 (4%)	108,880 (6%)
	Other failures	334,511 (18%)	337,906 (18%)	340,370 (19%)
Overall	Assays submitted	6,349,188	6,389,612	6,349,900
	Passed QC	4,327,745 (68%)	4,422,611 (69%)	4,432,697 (70%)
	Did not pass QC	2,021,443 (32%)	1,967,001 (31%)	1,917,203 (30%)
	>20% missing	1,501,463 (74%)	1,493,850 (76%)	1,481,866 (77%)
	>1 duplicate inconsistent	6,049 (0%)	5,126 (0%)	15,859 (1%)
	>1 mendelian error	191,255 (9%)	139,233 (7%)	N/A
	<0.001 Hardy–Weinberg *P*-value	106,496 (5%)	91,190 (5%)	122,602 (6%)
	Other failures	353,856 (18%)	351,764 (18%)	361,044 (19%)
	Non-redundant (unique) SNPs	3,796,934	3,868,157	3,890,416
	Monomorphic	861,299 (23%)	1,246,183 (32%)	1,410,152 (36%)
	Polymorphic	2,935,635 (77%)	2,621,974 (68%)	2,480,264 (64%)

SNP categories	All analysis panels
Unique QC-passed SNPs	4,000,107
Passed in one analysis panel	88,140 (2%)
Passed in two analysis panels	268,534 (7%)
Passed in three analysis panels (QC+3)	3,643,433 (91%)
QC+3 and monomorphic across three analysis panels	535,813
QC+3 and polymorphic in at least one analysis panel	3,107,620
QC+3 and polymorphic in all three analysis panels	2,006,352
QC+3 and MAF ≥ 0.05 in at least one of three analysis panels	2,819,322

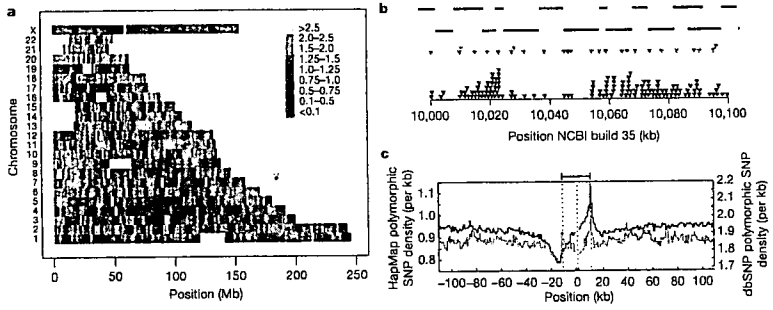


Figure 1 | SNP density in the Phase II HapMap. a, SNP density across the genome. Colours indicate the number of polymorphic SNPs per kb in the consensus data set. Gaps in the assembly are shown as white. b, Example of the fine-scale structure of SNP density for a 100-kb region on chromosome 17 showing Perlegen amplicons (black bars), polymorphic Phase I SNPs in the consensus data set (red triangles) and polymorphic Phase II SNPs in the consensus data set (blue triangles). Note the relatively even spacing of Phase I SNPs. c, The distribution of polymorphic SNPs in the consensus Phase II HapMap data (blue line and left-hand axis) around coding regions. Also shown is the density of SNPs in dbSNP release 125 around genes (red line and right-hand axis). Values were calculated separately 5′ from the coding start site (the left dotted line) and 3′ from the coding end site (right dotted line) and were joined at the median midpoint position of the coding unit (central dotted line).

multiple related haplotypes). Third, the novel SNP (or group of added SNPs) may reveal previously missed recombinant haplotypes. The extent to which each type of event occurs varies among populations and chromosomal regions. The greatest gains in resolution, in terms of identifying new recombinant haplotypes and haplotype groupings, occur in YRI. Consequently, the Phase II HapMap provides increased resolution in the estimated fine-scale genetic map and improved power to detect and localize recombination hotspots (Fig. 2b).

The use of the Phase II HapMap in association studies

The increased SNP density of the Phase II HapMap has already been extensively exploited in genome-wide studies of disease association. In this section, we quantify the gain in resolution and outline how the HapMap data can be used to improve the power of association studies.

Improved coverage of common variation. We previously predicted that the vast majority of common SNPs would be correlated to Phase II HapMap SNPs by extrapolation from the ten HapMap ENCODE regions[3]. Using the actual Phase II marker spacing and frequency distributions (Table 2), we repeated the simulations and estimate that Phase II HapMap marker sets capture the overwhelming majority of all common variants at high r^2. For common variants (MAF ≥ 0.05) the mean maximum r^2 of any SNP to a typed one is 0.90 in YRI, 0.96 in CEU and 0.95 in CHB+JPT. The impact of the increased density of the Phase II HapMap is most notable in YRI (in the Phase I HapMap the mean maximum r^2 was 0.67). Similar results are found if a threshold of $r^2 \geq 0.8$ is used to determine whether an SNP is captured (Table 2). As expected, very common SNPs with MAF > 0.25 are captured extremely well (mean maximum r^2 of 0.93 in YRI to 0.97 in CEU), whereas rarer SNPs with MAF < 0.05 are less well covered (mean maximum r^2 of 0.74 in CHB+JPT to 0.76 in YRI). The latter figure is probably an overestimate because it is based on lower frequency SNPs discovered via re-sequencing 48 HapMap individuals, and does not include a much larger number of very rare SNPs. We also assessed the increase in coverage provided by using two-SNP haplotypes as proxies for SNPs that are poorly captured by single SNPs[16] (Table 2). These two-SNP haplotypes lead to a modest increase in mean maximum r^2 of 0.01 to 0.03 across all allele frequencies. However, in some regions, particularly where marker density is low, gains from multi-marker and imputation approaches in practical situations can be substantial (see below).

Currently, the Phase II HapMap provides the most complete available resource for selecting tag SNPs genome-wide. Using a simple pairwise tagging approach, we find that 1.09 million SNPs are required to capture all common Phase II SNPs with $r^2 \geq 0.8$ in YRI, with slightly more than 500,000 required in CEU and CHB+JPT (Table 3). These numbers are approximately twice those required to capture SNPs in the Phase I HapMap (which has one-third as many SNPs). The number of SNPs required to achieve perfect tagging ($r^2 = 1.0$) in each analysis panel is almost double that required to achieve the $r^2 \geq 0.8$ threshold. It becomes increasingly expensive to improve the coverage afforded by tags from the Phase I and, now, the Phase II HapMap, because additional tag SNPs are unlikely to capture large groups of additional SNPs.

Phase II HapMap and genome-wide association studies. Although the efficient choice of tag SNPs is one use of the Phase II HapMap, for most disease studies the tag SNPs genotyped will be primarily determined by the choice of a commercial platform for the experiment[17,18]. Using Phase II data, we estimated the coverage of several available products on which genome-wide association studies are already underway (Table 4). Similar to earlier estimates[17,18], these products typically perform well in CEU and CHB+JPT, and some also perform well in YRI. For example, arrays of approximately 500,000 SNPs capture 68–88% (depending on selection method) of all HapMap Phase II variation with $r^2 \geq 0.8$ in CEU. SNPs that are not included in the Phase II HapMap will be covered more poorly because most genotyping products were designed using HapMap data.

HapMap data have several additional roles in the analysis of disease-association studies using fixed marker sets. For example, the high-quality haplotype information within the Phase II HapMap can be used to aid the phasing of genotype data from new samples because additional haplotypes are likely to be locally very similar to at least one haplotype in the Phase II data. By a similar argument, missing genotypes can potentially be inferred through comparison to the Phase II haplotypes. Genotypes may be missing either because of genotyping failure or because the SNP was not assayed within the experiment. Therefore, the HapMap haplotypes provide a way of *in silico* genotyping Phase II SNPs that were not included in the experiment.

Although there is no clear consensus yet about the role of SNP imputation in the analysis of genome-wide association studies, high imputation accuracy can be achieved using model-based methods[19–23] and can lead to an increase in power[23,24]. To illustrate the possibilities, in the 500-kb HapMap ENCODE region on 8q24.11 (Supplementary Fig. 5) we evaluated imputation of Phase II SNPs from the Affymetrix GeneChip 500K array. To do this, we used a

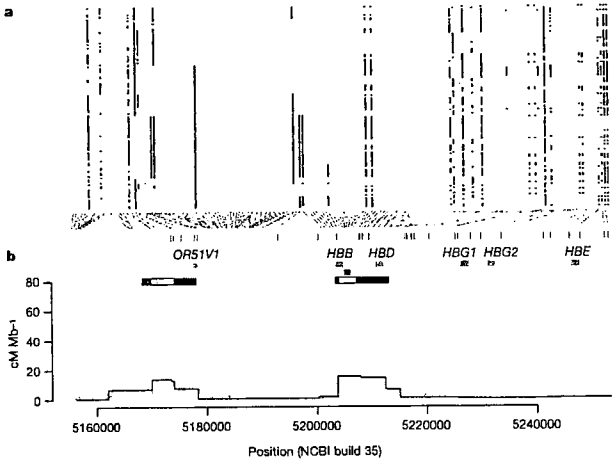


Figure 2 | Haplotype structure and recombination rate estimates from the Phase II HapMap. a, Haplotypes from YRI in a 100 kb region around the β-globin (*HBB*) gene. SNPs typed in Phase I are shown in dark blue. Additional SNPs in the Phase II HapMap are shown in light blue. Only SNPs for which the derived allele can be unambiguously identified by parsimony (by comparison with an outgroup sequence) are shown (89% of SNPs in the region); the derived allele is shown in colour. b, Recombination rates (lines) and the location of hotspots (horizontal blue bars) estimated for the same region from the Phase I (dark blue) and Phase II HapMap (light blue) data. Also shown are the location of genes within the region (grey bars) and the location of the experimentally verified recombination hotspot[37,38] at the 5′ end of the *HBB* gene (black bar).

Table 2 | Estimated coverage of the Phase II HapMap in the ten HapMap ENCODE regions

Panel	MAF bin	Phase I HapMap[3]		Phase II HapMap			
				Pairwise linkage disequilibrium		Additional 2-SNP tests	
		$r^2 \geq 0.8$ (%)	Mean maximum r^2	$r^2 \geq 0.8$ (%)	Mean maximum r^2	$r^2 \geq 0.8$ (%)	Mean maximum r^2
YRI	≥0.05	45	0.67	82	0.90	87	0.93
	<0.05			61	0.76	62	0.78
	0.05–0.10			81	0.89	81	0.89
	0.10–0.25			90	0.94	90	0.95
	0.25–0.50			87	0.93	92	0.96
CEU	≥0.05	74	0.85	93	0.96	95	0.97
	<0.05			70	0.79	72	0.81
	0.05–0.10			87	0.92	88	0.93
	0.10–0.25			94	0.96	95	0.97
	0.25–0.50			95	0.97	97	0.98
CHB+JPT	≥0.05	72	0.83	92	0.95	95	0.97
	<0.05			65	0.74	65	0.74
	0.05–0.10			81	0.89	82	0.89
	0.10–0.25			90	0.94	90	0.95
	0.25–0.50			94	0.96	97	0.98

2-SNP tests, linkage disequilibrium to haplotypes formed from two nearby SNPs.

Table 3 | Number of tag SNPs required to capture common (MAF ≥ 0.05) Phase II SNPs

Threshold	YRI	CEU	CHB+JPT
$r^2 \geq 0.5$	627,458	290,969	277,831
$r^2 \geq 0.8$	1,093,422	552,853	520,111
$r^2 = 1.0$	1,616,739	1,024,665	1,078,959

Table 4 | Estimated coverage of commercially available fixed marker arrays

Platform*	YRI		CEU		CHB+JPT	
	$r^2 \geq 0.8$ (%)	Mean maximum r^2	$r^2 \geq 0.8$ (%)	Mean maximum r^2	$r^2 \geq 0.8$ (%)	Mean maximum r^2
Affymetrix GeneChip 500K	46	0.66	68	0.81	67	0.80
Affymetrix SNP Array 6.0	66	0.80	82	0.90	81	0.89
Illumina HumanHap300	33	0.56	77	0.86	63	0.78
Illumina HumanHap550	55	0.73	88	0.92	83	0.89
Illumina HumanHap650Y	66	0.80	89	0.93	84	0.90
Perlegen 600K	47	0.68	92	0.94	84	0.90

* Assuming all SNPs on the product are informative and pass QC; in practice these numbers are overestimates.

leave-one-out procedure to assess the accuracy of genotype prediction in the YRI. For SNPs with MAF ≥ 0.2, the average maximum r^2 to a typed SNP in the region is 0.59 compared to an average genotype prediction r^2 of 0.86. Furthermore, whereas 44% of such SNPs in the region have no single-marker proxy with $r^2 \geq 0.5$, fewer than 6% of the SNPs have a genotype imputation accuracy of $r^2 < 0.5$, establishing that accurate imputation can be achieved even in the population where linkage disequilibrium is the weakest.

New insights into linkage disequilibrium structure

The paradigm underlying association studies is that linkage disequilibrium can be used to capture associations between markers and nearby untyped SNPs. However, the Phase II HapMap has revealed several properties of linkage disequilibrium that illustrate the full complexity of empirical patterns of genetic variation. Two striking features are the long-range similarity among haplotypes, and SNPs that show almost no linkage disequilibrium with any other SNP.

The extent of recent common ancestry and segmental sharing. A simplified view of linkage disequilibrium is that genetic variation is organized in relatively short stretches of strong linkage disequilibrium (haplotype blocks), each containing only a few common haplotypes and separated by recombination hotspots across which little association remains[25]. Although this view has heuristic value, if chromosomes share a recent common ancestor then similarity between chromosomes can extend over considerable genetic distance and span multiple recombination hotspots[26]. The extent of such recent ancestry in the four populations surveyed here has not been characterized previously. Therefore we identified stretches of identity between pairs of chromosomes, both within and across individuals, reflecting autozygosity and identity-by-descent (IBD) (Fig. 3a). After first checking for stratification within each analysis panel (see Supplementary Text 3; none was found for YRI, CEU and JPT, and only small stratification was found for CHB), we calculated genome-wide probabilities of sharing 0, 1 or 2 chromosomes identical by descent for each pair of individuals (see Supplementary Text 4). In addition to identifying a few close relationships (as reported in HapMap Phase I[5]), we estimate that, on average, any two individuals from the same population share approximately 0.5% of their genome through recent IBD (Table 5). Using a hidden Markov model approach[27] (see Supplementary Text 5), we searched for such shared segments over 1-megabase (Mb) long and containing at least 50 SNPs, after first pruning the list of SNPs to remove local linkage disequilibrium. We find that 10–30% of pairs in each analysis panel share regions of extended identity resulting from sharing a common ancestor within 10–100 generations. These regions typically span hundreds of SNPs and can extend over tens of megabases (Table 5).

Similarly, extended stretches of homozygosity are indicative of recent inbreeding within populations[28,29]. Although short runs of homozygosity are commonplace, covering up to one-third of the genome and showing population differences reflective of ancient linkage disequilibrium patterns (Table 5 and Fig. 3b), very long homozygous runs exist that are clearly distinct from this process. Including two JPT individuals who have unusually high levels of homozygosity (NA18987 and NA18992) and one CEU individual (NA12874), we identified 79 homozygous regions over 3 Mb in 51 individuals, with many segments extending over 10 Mb (Supplementary Tables 7 and 8). Segments intersecting with suspected deletions were first removed from the analysis (Supplementary Text 6).

In studies of rare mendelian diseases, the extended haplotype sharing surrounding recent mutations, usually with a frequency of much less than 1%, has been exploited to great advantage through homozygosity mapping[30,31] and haplotype sharing[32] methods. In studies of common disease, extended haplotype sharing among patients potentially offers a route for identifying rare variants (MAF in the range of 1–5%) of high penetrance[33,34], which tend to be poorly captured through single-marker association with genome-wide arrays. To illustrate the idea, we identified SNPs where only two copies of the minor allele are present (referred to as '2-SNPs'), which have minor allele frequencies of 1–2%. We find that these are enriched approximately sevenfold (Table 5) among regions of IBD identified by the hidden Markov model approach. Notably, identification of IBD regions can be performed with the same genome-wide SNP data being

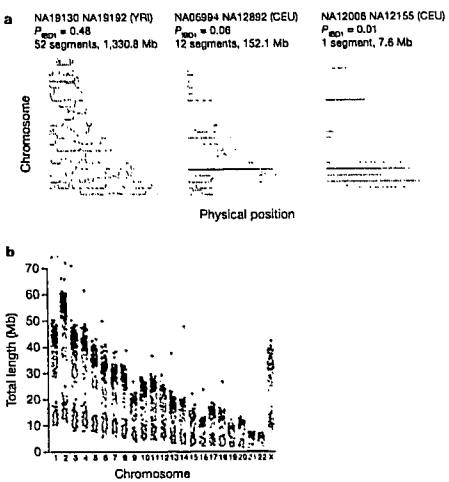


Figure 3 | The extent of recent co-ancestry among HapMap individuals. **a**, Three pairs of individuals with varying levels of identity-by-descent (IBD) sharing illustrate the continuum between very close and very distant relatedness and its relation to segmental sharing. The three pairs are: high sharing (NA19130 and NA19192 from YRI; previously identified as second-degree relatives[5]), moderate sharing (NA06994 and NA12892 from CEU) and low sharing (NA12006 and NA12155 from CEU). Along each chromosome, the probability of sharing at least one chromosome IBD is plotted, based on the HMM method described in Supplementary Text 5. Red sections indicate regions called as segments: in general, the proportion of the genome in segments is similar to each pair's estimated global relatedness. **b**, The extent of homozygosity on each chromosome for each individual in each analysis panel. Excludes segments <106 kb and chromosome X in males. Asterisk, NA12874, length = 107 Mb. YRI, green; CEU, orange; CHB, blue; JPT, magenta.

Table 5 | Relatedness, extended segmental sharing and homozygosity

Property	YRI	CEU	CHB	JPT
Number of pairs included	1,767	1,708	990	861
Mean identity by state (IBS) (%)	81.9	83.7	85.0	85.1
Mean identity by descent (IBD) (%)	0.04	0.34	0.36	0.42
Number of pairs with >1% IBD (%)	8.8	20.4	21.1	29.7
Number of pairs with one or more segment (%)	195 (11.0)	350 (20.5)	135 (13.6)	216 (25.1)
Total number of segments	250	427	146	273
Total distance spanned (Mb)	1,416	2,336	704	1,301
Mean segment length (Mb)	5.7	5.5	4.8	4.8
Maximum segment length (Mb)	51.7	56.2	15.0	25.3
Maximum segment length (Mb) (including close relatives)	141.4	128.5	N/A	N/A
Total number of 2-SNPs	6,219	9,220	8,174	8,750
Number of 2-SNPs in segments	109	162	116	132
2-SNP fold increase	6.7	7.3	7.6	7.0
Number of homozygous segments ($\times 10^3$)*	0.9	2.2	2.6	2.6
SNPs in homozygous segments ($\times 10^5$)	1.6	4.2	5.3	5.4
Total length of homozygous segments (Mb)	160	410	510	520

2-SNP, SNPs where only two copies of the minor allele are present.
* Homozygous segments >106 kb.

collected in large-scale association studies, making haplotype-sharing approaches an attractive and complementary analysis to standard SNP association tests, with the potential to identify rare variants associated with complex disease.

The distribution and causes of untaggable SNPs. Despite the SNP density of the Phase II HapMap, there are high-frequency SNPs for which no tag can be identified. Among high-frequency SNPs (MAF ≥ 0.2), we marked as untaggable SNPs to which no other SNP within 100 kb has an r^2 value of at least 0.2. In Phase II, approximately 0.5–1.0% of all high-frequency SNPs are untaggable and the proportion in YRI is approximately twice as high as in the other panels. Similar proportions are observed across the ten HapMap ENCODE regions.

To identify factors influencing the location of untaggable SNPs we considered their distribution relative to segmental duplications, repeat sequence, CpG dinucleotide density, regions of low SNP density, unusual allele frequency distribution, linkage disequilibrium patterns and recombination hotspots. We find no evidence for an enrichment of untaggable SNPs in segmental duplications or repeat sequence, as would be expected from mis-mapping of SNPs (2% and 35% of common SNPs lie in segmental duplications and repeat sequence, respectively, compared to 1.8% and 29%, respectively, of untaggable SNPs). Untaggable SNPs are slightly enriched in CpG islands (0.37% of common SNPs are in CpG islands compared to 1.4% of untaggable SNPs) and have slightly reduced MAF (Fig. 4). Most notably, untaggable SNPs are strongly enriched in regions of low linkage disequilibrium, particularly in recombination hotspots. To test whether these untaggable SNPs are themselves responsible for the identification of recombination hotspots, we eliminated them from 100 randomly chosen recombination hotspots and reassessed the evidence for a local peak in recombination. In all cases we still find evidence for a considerable increase in local recombination rate.

Over 50% of all untaggable SNPs lie within 1 kb of the centre of a detected recombination hotspot and over 90% are within 5 kb. Because only 3–4% of all SNPs lie within 1 kb from the centre of a detected recombination hotspot (16% are within 5 kb), this constitutes a marked enrichment and implies that at least 10% of all SNPs within 1 kb of hotspots are untaggable. The implication for association mapping is that when a region of interest contains a known hotspot it may be prudent to perform additional sequencing within the hotspot. Many of the variants identified in this manner will be untaggable SNPs that should be genotyped directly in association studies. From a biological perspective, the proximity of untaggable SNPs to the centre of hotspots suggests that they may lie within gene conversion tracts associated with the repair of double-strand breaks. Double-strand breaks are thought to resolve as crossover events only 5–25% of the time[35]. Consequently, SNPs lying near the centre of a hotspot are liable to be included within gene conversion tracts and will experience much higher effective recombination rates than predicted from crossover rates alone.

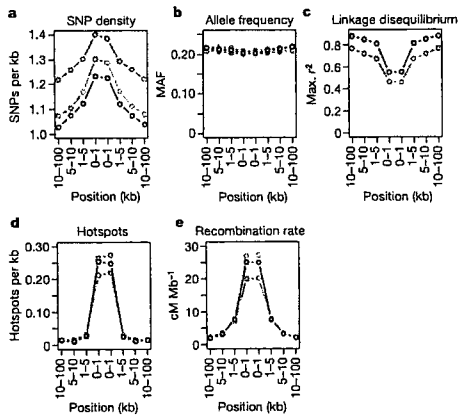


Figure 4 | Properties of untaggable SNPs. a–e, Properties of the genomic regions surrounding untaggable SNPs in terms of: **a**, the density of polymorphic SNPs within the consensus data set; **b**, mean minor allele frequency of polymorphic SNPs; **c**, maximum r^2 of SNPs to any others in the Phase II data; **d**, the density of estimated recombination hotspots (defined from hotspot centres); and **e**, the estimated mean recombination rate. YRI, green; CEU, orange; CHB+JPT, purple.

The distribution of recombination

In the Phase II HapMap we identified 32,996 recombination hotspots[5,6,36] (an increase of over 50% from Phase I) of which 68% localized to a region of ≤5 kb. The median map distance induced by a hotspot is 0.043 cM (or one crossover per 2,300 meioses) and the hottest identified, on chromosome 20, is 1.2 cM (one crossover per 80 meioses). Hotspots account for approximately 60% of recombination in the human genome and about 6% of sequence (Supplementary Fig. 6). We do not find marked differences among chromosomes in the concentration of recombination in hotspots, which implies that obligate differences in recombination among chromosomes of different size result from differences in hotspot density and intensity[6].

The increased number of well-defined hotspots allows us to understand better the influence of genomic features on the distribution of recombination. Previous work identified specific DNA motifs that influence hotspot location[6,37] as well as additional influences of local sequence context including the location of genes[6] and base composition[38]. The Phase II HapMap provides the resolution to separate these influences. Figure 5a shows the distribution of recombination, hotspot motifs and base composition around genes. Within the transcribed region of genes there is a marked decrease in the estimated recombination rate. However, 5′ of the transcription start site is a peak in recombination rate with a corresponding local increase in the density of hotspot motifs. This region also shows a marked increase in G+C content, reflecting the presence of CpG islands in promoter regions. There is also an asymmetry in recombination rate across genes, with recombination rates 3′ of transcribed regions being elevated (as are motif density and G+C content) compared to regions 5′ of genes. Studies in yeast have previously suggested an association between promoter regions and recombination hotspots[39]. Our results suggest a significant, although weak, relationship between promoters and recombination in humans. Nevertheless, the vast majority of hotspots in the human genome are not in gene promoters. The association may reflect a general association between regions of accessible chromatin and crossover activity.

Systematic differences in recombination rate by gene class. Previous work has demonstrated differences in the magnitude of linkage disequilibrium, as measured at a megabase scale, among genes associated with different functions[5,40]. Using the fine-scale genetic map estimated from the Phase II HapMap data we can quantify local increases in recombination rate associated with genes of different function using the Panther gene ontology annotation[41]. Average recombination rates vary more than sixfold among such gene classes (Fig. 5b), with defence and immunity genes showing the highest rates (1.9 cM Mb^{-1}) and chaperones showing the lowest rates (0.3 cM Mb^{-1}). Gene functions associated with cell surfaces and external functions tend to show higher recombination rates (immunity, cell adhesion, extracellular matrix, ion channels, signalling) whereas those with lower recombination rates are typically internal to cells (chaperones, ligase, isomerase, synthase). Controlling for systematic differences between gene classes in base composition and gene clustering, the differences between groups remain significant.

We also find that the density of hotspot-associated DNA motifs varies systematically among gene classes and that variation in motif density explains over 50% of the variance in recombination rate among gene functions (Supplementary Fig. 7).

These results pose interesting evolutionary questions. Because recombination involves DNA damage through double-strand breaks, hotspots may be selected against in some highly conserved parts of the genome. In regions exposed to recurrent selection (for example, from changes in environment or pathogen pressure) it is plausible that recombination may be selected for. However, because the fine-scale structure of recombination seems to evolve rapidly[42,43] it will be important to learn whether patterns of recombination rate heterogeneity among molecular functions are conserved between species.

Natural selection

The Phase I HapMap data have been used to identify genomic regions that show evidence for the influence of adaptive evolution[3,9], primarily through extended haplotype structure indicative of recent positive selection. Using two established approaches[9,44], we identified approximately 200 regions with evidence of recent positive selection from the Phase II HapMap (Supplementary Table 9). These regions include many established cases of selection, such as the genes *HBB* and *LCT*, the HLA region, and an inversion on chromosome 17. Many other regions have been previously identified in HapMap Phase I including *LARGE*, *SYT1* and *SULT1C2* (previously called *SULT1C1*). A detailed description of the findings from the Phase II HapMap is published elsewhere[45].

The Phase II HapMap also provides new insights into the forces acting on SNPs in coding regions. Effort was made to genotype as many known or putative non-synonymous SNPs as possible. Of the 56,789 non-synonymous SNPs identified in dbSNP release 125, attempts were made to genotype 36,777, which resulted in 17,427 that are QC+ in all three analysis panels and polymorphic. We selected only those SNPs for which ancestral allele information was available (approximately 90%). For comparison, we used patterns of variation at synonymous SNPs. As previously reported[46,47], non-synonymous SNPs show an increase in frequency of rare variants and a slight decrease of common variants compared to synonymous SNPs, compatible with widespread purifying selection against non-synonymous mutations (Fig. 6a). In contrast, we find no excess of high-frequency derived non-synonymous mutations, as might be expected if positive selection were widespread.

Natural selection also influences the extent to which allele frequencies differ between populations, not only through local selective pressures that drive alleles to different frequencies[48,49], but also through local variation in the strength of purifying selection. We compared the distribution of population differentiation (as measured by F_{ST}, the proportion of total variation in allele frequency that is due to differences between populations) at non-synonymous SNPs and synonymous SNPs matched for allele frequency (Fig. 6b). We find a systematic bias for non-synonymous SNPs to show stronger differentiation than synonymous SNPs. Among SNPs showing high levels of differentiation there is a strong tendency for the derived allele to be at higher frequency in non-YRI populations. Among SNPs with $F_{ST} > 0.5$ between CEU and YRI, in 79% and 75% of non-synonymous and synonymous variants, respectively, the derived allele is more common in CEU. Although this difference between non-synonymous and synonymous SNPs is not significant, among the eight exonic SNPs with $F_{ST} > 0.95$, all are non-synonymous. We see no such bias towards increased MAF in CEU at high-differentiation SNPs, indicating that SNP ascertainment is unlikely to explain the difference. Rather, this effect can largely be explained by more genetic drift in the non-African populations, as confirmed by simulations (data not shown). In addition, reduced selection against deleterious mutations and local adaptation within non-African populations will both act to increase the frequency of derived variants in non-African populations.

To assess the evidence for widespread local adaptation influencing non-synonymous mutations we considered the distribution of integrated extended haplotype homozygosity (iEHH) statistics[9,44] (Fig. 6c). We find no evidence for systematic differences between non-synonymous and synonymous SNPs, suggesting that local adaptation does not explain their higher differentiation. Although hitch-hiking effects will tend to obscure differences between selected

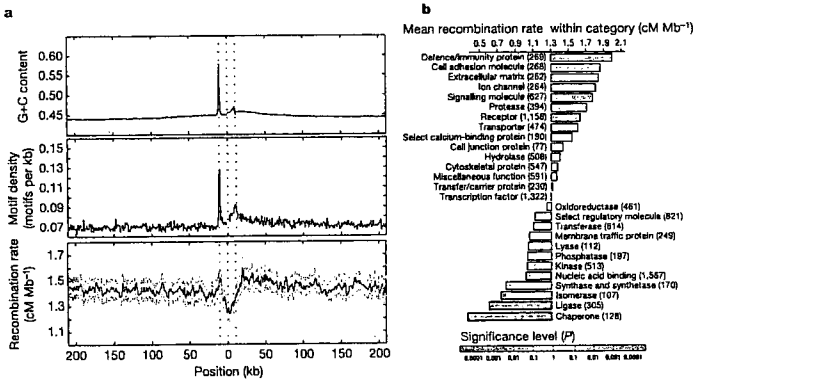


Figure 5 | Recombination rates around genes. a, The recombination rate, density of recombination-hotspot-associated motifs (all motifs with up to 1 bp different from the consensus CCTCCCTNNCCAC) and G+C content around genes. The blue line indicates the mean. For the recombination rate, grey lines indicate the quartiles of the distribution. Values were calculated separately 5′ from the transcription start site (the first dotted line) and 3′ from the transcription end site (third dotted line) and were joined at the median midpoint position of the transcription unit (central dotted line). Note the sharp drop in recombination rate within the transcription unit, the local increase around the transcription start site and the broad decrease away from the 3′ end of genes. These patterns only partly reflect the distribution of G+C content and the hotspot-associated motif, suggesting that additional factors influence recombination rates around genes. **b**, Recombination rates within genes of different molecular function[41]. The chart shows the increase or decrease for each category compared to the genome average. *P* values were estimated by permutation of category; numbers of genes are shown in parentheses.

and neutral SNPs, these results are consistent with a scenario in which the higher differentiation of non-synonymous SNPs is primarily driven by a reduction in the strength or efficacy of purifying selection in non-African populations.

Discussion and prospects

The International HapMap Project has been instrumental in making well-powered, large-scale, genome-wide association studies a reality. It is now clear that the HapMap can be a useful resource for the design and analysis of disease association studies in populations across the world[50–53]. Furthermore, the decreasing costs and increasing SNP density of standard genotyping panels mean that the focus of attention in disease association studies is shifting from candidate gene approaches towards genome-wide analyses. Alongside developments in technology, new statistical methodologies aimed at improving aspects of analysis, such as genotype calling[21,54], the identification of and correction for population stratification and relatedness[55,56], and imputation of untyped variants[21–23], are increasing the accuracy and reliability of genome-wide association studies.

Within this context, it is important to consider the future of the HapMap Project. Currently, additional samples from the populations used to develop the initial HapMap, as well as samples from seven additional populations (Luhya in Webuye, Kenya; Maasai in Kinyawa, Kenya; Tuscans in Italy; Gujarati Indian in Houston, Texas, USA; Denver (Colorado) metropolitan Chinese community; people of Mexican origin in Los Angeles, California, USA; and people with African ancestry in the southwestern United States; http://ccr.coriell.org/Sections/Collections/NHGRI/?SsId=11) will be sequenced and genotyped extensively to extend the HapMap, providing information on rarer variants and helping to enable genome-wide association studies in additional populations. There are also ongoing efforts by many groups to characterize additional forms of genetic variation, such as structural variation, and molecular phenotypes in the HapMap samples. Finally, in the future, whole-genome sequencing will provide a natural convergence of technologies to type both SNP and structural variation. Nevertheless, until that point, and even after, the HapMap Project data will provide an invaluable resource for understanding the structure of human genetic variation and its link to phenotype.

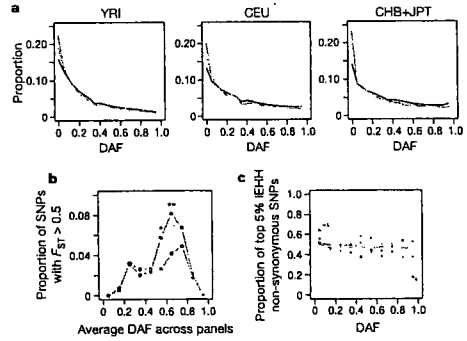


Figure 6 | Properties of non-synonymous and synonymous SNPs. a, The derived allele frequency (DAF) spectrum in each analysis panel for all SNPs (black), synonymous SNPs (green) and non-synonymous SNPs (red). Note the excess of rare variants for coding sequence SNPs but no excess of high-frequency derived variants. **b**, Enrichment of non-synonymous SNPs among genic SNPs showing high differentiation. For each of ten classes of derived allele frequency (averaged across analysis panels) the fraction of non-synonymous (red) and synonymous (green) variants in that class that show $F_{ST} > 0.5$ is shown. Note the strong enrichment of non-synonymous SNPs among SNPs of moderate to high derived-allele frequency (asterisk, $P < 0.05$; double asterisk, $P < 0.01$). **c**, Lack of enrichment of non-synonymous SNPs among those showing long-range haplotype structure. The integrated extended haplotype homozygosity (iEHH) statistic[9] was calculated for non-synonymous and synonymous SNPs in each analysis panel (YRI, green; CEU, orange; CHB+JPT, purple). For each of ten derived allele frequency classes, the proportion of non-synonymous SNPs among those showing the 5% most extreme statistics (within the allele frequency class) is shown (points). Also shown is the proportion of non-synonymous SNPs among SNPs in the coding sequence for each frequency class (dotted lines). Differences between synonymous and non-synonymous SNPs are tested for using a contingency table test.

METHODS SUMMARY

Of approximately 6.9 million SNPs in dbSNP release 122 approximately 4.7 million were selected for genotyping by Perlegen. 2.5 million SNPs were excluded because no assay could be designed and a further 350,000 were excluded for other reasons (see Methods). Perlegen performed genotyping using custom high-density oligonucleotide arrays as previously described[15]. Additional genotype submissions are described in the text. QC filters were applied as previously described[3]. Where multiple submissions met the QC criteria the submission with the lowest missing data rate was chosen for inclusion in the non-redundant filtered data set. Haplotypes were estimated from genotype data as described previously[3]. Ancestral states at SNPs were inferred by parsimony by comparison to orthologous bases in the chimpanzee (panTro2) and rhesus macaque (rheMac2) assemblies. Recombination rates and the location of recombination hotspots were estimated as described previously[3]. Additional details can be found in the Methods section and the Supplementary Information. The data described in this paper are in release 21 of the International HapMap Project.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 April; accepted 18 September 2007.

1. The International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nature Rev. Genet.* 5, 467–475 (2004).
2. The International HapMap Consortium. The International HapMap Project. *Nature* 426, 789–796 (2003).
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 437, 1299–1320 (2005).
4. Bowcock, A. M. Genomics: guilt by association. *Nature* 447, 645–646 (2007).
5. Altshuler, D. & Daly, M. Guilt beyond a reasonable doubt. *Nature Genet.* 39, 813–815 (2007).
6. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310, 321–324 (2005).
7. McCarroll, S. A. *et al.* Common deletion polymorphisms in the human genome. *Nature Genet.* 38, 86–92 (2006).
8. Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genet.* 38, 75–81 (2006).
9. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72 (2006).
10. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* 444, 444–454 (2006).
11. de Bakker, P. I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genet.* 38, 1166–1172 (2006).
12. Pastinen, T. *et al.* Mapping common regulatory variants to human haplotypes. *Hum. Mol. Genet.* 14, 3963–3971 (2005).
13. Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 1, e78 (2005).
14. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437, 1365–1369 (2005).
15. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079 (2005).
16. de Bakker, P. I. *et al.* Efficiency and power in genetic association studies. *Nature Genet.* 37, 1217–1223 (2005).
17. Pe'er, I. *et al.* Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nature Genet.* 38, 663–667 (2006).
18. Barrett, J. C. & Cardon, L. R. Evaluating coverage of genome-wide association studies. *Nature Genet.* 38, 659–662 (2006).
19. Burdick, J. T., Chen, W. M., Abecasis, G. R. & Cheung, V. G. *In silico* method for inferring genotypes in pedigrees. *Nature Genet.* 38, 1002–1004 (2006).
20. Servin, B. R. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* 3, e114 (2007).

21. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–668 (2007).
22. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345 (2007).
23. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies via imputation of genotypes. *Nature Genet.* 39, 906–913 (2007).
24. Chapman, J. M., Cooper, J. D., Todd, J. A. & Clayton, D. G. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum. Hered.* 56, 18–31 (2003).
25. Paabo, S. The mosaic that is our genome. *Nature* 421, 409–412 (2003).
26. McVean, G., Spencer, C. C. & Chaix, R. Perspectives on human genetic variation from the HapMap Project. *PLoS Genet.* 1, e54 (2005).
27. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575 (2007).
28. Broman, K. W. & Weber, J. L. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* 65, 1493–1500 (1999).
29. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795 (2006).
30. Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–1570 (1987).
31. Leutenegger, A. L. *et al.* Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. *Am. J. Hum. Genet.* 79, 62–66 (2006).
32. Te Meerman, G. J., Van der Meulen, M. A. & Sandkuijl, L. A. Perspectives of identity by descent (IBD) mapping in founder populations. *Clin. Exp. Allergy* 25 (Suppl 2), 97–102 (1995).
33. Houwen, R. H. *et al.* Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genet.* 8, 380–386 (1994).
34. Durham, L. K. & Feingold, E. Genome scanning for segments shared identical by descent among distant relatives in isolated populations. *Am. J. Hum. Genet.* 61, 830–842 (1997).
35. Jeffreys, A. J. & May, C. A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genet.* 36, 151–156 (2004).
36. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* 304, 581–584 (2004).
37. Myers, S. *et al.* The distribution and causes of meiotic recombination in the human genome. *Biochem. Soc. Trans.* 34, 526–530 (2006).
38. Spencer, C. C. *et al.* The influence of recombination on human genetic diversity. *PLoS Genet.* 2, e148 (2006).
39. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* 2, 360–369 (2001).
40. Smith, A. V., Thomas, D. J., Munro, H. M. & Abecasis, G. R. Sequence features in regions of weak and strong linkage disequilibrium. *Genome Res.* 15, 1519–1534 (2005).
41. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141 (2003).
42. Winckler, W. *et al.* Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308, 107–111 (2005).
43. Ptak, S. E. *et al.* Fine-scale recombination patterns differ between chimpanzees and humans. *Nature Genet.* 37, 429–434 (2005).
44. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837 (2002).
45. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* doi:10.1038/nature06250 (this issue).
46. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* 437, 1153–1157 (2005).
47. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* 22, 231–238 (1999).
48. Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814 (2002).
49. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* 312, 1614–1620 (2006).
50. de Bakker, P. I. *et al.* Transferability of tag SNPs in genetic association studies in multiple populations. *Nature Genet.* 38, 1298–1303 (2006).
51. Conrad, D. F. *et al.* A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genet.* 38, 1251–1260 (2006).
52. Service, S., Sabatti, C. & Freimer, N. Tag SNPs chosen from HapMap perform well in several population isolates. *Genet. Epidemiol.* 31, 189–194 (2007).
53. Lim, J. *et al.* Comparative study of the linkage disequilibrium of an ENCODE region, chromosome 7p15, in Korean, Japanese, and Han Chinese samples. *Genomics* 87, 392–398 (2006).
54. Rabbee, N. & Speed, T. P. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22, 7–12 (2006).
55. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007).
56. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* 38, 904–909 (2006).
57. Smith, R. A., Ho, P. J., Clegg, J. B., Kidd, J. R. & Thein, S. L. Recombination breakpoints in the human β-globin gene cluster. *Blood* 92, 4415–4421 (1998).
58. Holloway, K., Lawson, V. E. & Jeffreys, A. J. Allelic recombination and *de novo* deletions in sperm in the human β-globin gene region. *Hum. Mol. Genet.* 15, 1099–1111 (2006).
59. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370 (1984).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank many people who contributed to this project: all members of the genotyping laboratory and the sample, primer, bioinformatics, data quality and IT groups at Perlegen Sciences for technical and infrastructural support; J. Beck, C. Beiswanger, D. Coppock, A. Leach, J. Mintzer and L. Toji for transforming the Yoruba, Japanese and Han Chinese samples, distributing the DNA and cell lines, storing the samples for use in future research, and producing the community newsletters and reports; J. Greenberg and R. Anderson for providing funding and support for cell line transformation and storage in the NIGMS Human Genetic Cell Repository at the Coriell Institute; T. Dibling, T. Ishikura, S. Kanazawa, S. Mizusawa and S. Saito for help with genotyping; C. Hind and A. Moghadam for technical support in genotyping and all members of the subcloning and sequencing teams at the Wellcome Trust Sanger Institute; X. Ke for help with data analysis; Oxford E-Science Centre for provision of high-performance computing resources; H. Chen, W. Chen, L. Deng, Y. Dong, C. Fu, L. Gao, H. Geng, J. Geng, M. He, H. Li, H. Li, S. Li, X. Li, B. Liu, Z. Liu, F. Lu, F. Lu, G. Lu, C. Luo, X. Wang, Z. Wang, C. Ye and X. Yu for help with genotyping and sample collection; X. Feng, Y. Li, J. Ren and X. Zhou for help with sample collection; J. Fan, W. Gu, W. Guan, S. Hu, H. Jiang, R. Lei, Y. Lin, Z. Niu, B. Wang, L. Yang, W. Yang, Y. Wang, Z. Wang, S. Xu, W. Yan, H. Yang, W. Yuan, C. Zhang, J. Zhang, K. Zhang and G. Zhao for help with genotyping; P. Fong, C. Lai, C. Lau, T. Leung, L. Luk and W. Tong for help with genotyping; C. Pang for help with genotyping; K. Ding, B. Qiang, J. Zhang, X. Zhang and K. Zhou for help with genotyping; Q. Fu, S. Ghose, X. Lu, D. Nelson, A. Perez, S. Poole, R. Vega and H. Yonath for help with genotyping; C. Bruckner, T. Brundage, S. Chow, O. Iartchouk, M. Jain, M. Moorhead and K. Tran for help with genotyping; N. Addleman, J. Atilano, T. Chan, C. Chu, C. Ha, T. Nguyen, M. Minton and A. Phong for help with genotyping, and D. Lind for help with quality control and experimental design; R. Donaldson and S. Duan for help with genotyping, and J. Rice and N. Saccone for help with experimental design; J. Wigginton for help with implementing and testing QA/QC software; A. Clark, B. Keats, R. Myers, D. Nickerson and A. Williamson for providing advice to NIH; C. Juenger, C. Bennet, C. Bird, J. Melone, P. Nailer, M. Weiss, J. Witonsky and E. DeHaut-Combs for help with project management; M. Gray for organizing phone calls and meetings; D. Leja for help with figures; the Yoruba people of Ibadan, Nigeria, the people of Tokyo, Japan, and the community at Beijing Normal University, who participated in public consultations and community engagements; the people in these communities who donated their blood samples; and the people in the Utah CEPH community who allowed the samples they donated earlier to be used for the Project. This work was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology, the Wellcome Trust, Nuffield Trust, Wolfson Foundation, UK EPSRC, Genome Canada, Génome Québec, the Chinese Academy of Sciences, the Ministry of Science and Technology of the People's Republic of China, the National Natural Science Foundation of China, the Hong Kong Innovation and Technology Commission, the University Grants Committee of Hong Kong, the SNP Consortium, the US National Institutes of Health (FIC, NCI, NCRR, NEI, NHGRI, NIA, NIAAA, NIAID, NIAMS, NIBIB, NIDA, NIDCD, NIDCR, NIDDK, NIEHS, NIGMS, NIMH, NINDS, NLM, OD), the W.M. Keck Foundation, and the Delores Dore Eccles Foundation. All SNPs genotyped within the HapMap Project are available from dbSNP (http://www.ncbi.nlm.nih.gov/SNP); all genotype information is available from dbSNP and the HapMap website (http://www.hapmap.org).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.M. (mcvean@stats.ox.ac.uk) or M.D. (mjdaly@chgr.mgh.harvard.edu).

The International HapMap Consortium (Participants are arranged by institution and then alphabetically within institutions except for Principal Investigators and Project Leaders, as indicated.)

Genotyping centres: Perlegen Sciences Kelly A. Frazer (Principal Investigator)[1], Dennis G. Ballinger[2], David R. Cox[2], David A. Hinds[2], Laura L. Stuve[2]; **Baylor College of Medicine and ParAllele BioScience** Richard A. Gibbs (Principal Investigator)[3], John W. Belmont[3], Andrew Boudreau[4], Paul Hardenbol[5], Suzanne M. Leal[3], Shiran Pasternak[6], David A. Wheeler[3], Thomas D. Willis[4], Fuli Yu[7]; **Beijing Genomics Institute** Huanming Yang (Principal Investigator)[8], Changqing Zeng (Principal Investigator)[8], Yang Gao[8], Haoran Hu[8], Weitao Hu[8], Chaohua Li[8], Wei Lin[8], Siqi Liu[8], Hao Pan[8], Xiaoli Tang[8], Jian Wang[8], Wei Wang[8], Jun Yu[8], Bo Zhang[8], Qingrun Zhang[8], Hongbin Zhao[8], Hui Zhao[8], Jun Zhou[8]; **Broad Institute of Harvard and Massachusetts Institute of Technology** Stacey B. Gabriel (Project Leader)[7], Rachel Barry[7], Brendan Blumenstiel[7], Amy Camargo[7], Matthew Defelice[7], Maura Faggart[7], Mary Goyette[7], Supriya Gupta[7], Jamie Moore[7], Huy Nguyen[7], Robert C. Onofrio[7], Melissa Parkin[7], Jessica Roy[7], Erich Stahl[7], Ellen Winchester[7], Lluda Ziaugra[7], David Altshuler (Principal Investigator)[7,9]; **Chinese National Human Genome Center at Beijing** Yan Shen (Principal Investigator)[10], Zhijian Yao[10]; **Chinese National Human Genome Center at Shanghai** Wei Huang (Principal Investigator)[11], Xun Chu[11], Yungang He[11], Li Jin[12], Yangfan Liu[11], Yayun Shen[11], Weiwei Sun[11], Haifeng Wang[11], Yi Wang[11], Ying Wang[11], Xiaoyan Xiong[11], Liang Xu[11]; **Chinese University of Hong Kong** Mary M. Y. Waye (Principal Investigator)[13], Stephen K. W. Tsui[13]; **Hong Kong University of Science and Technology** Hong Xue (Principal Investigator)[14], J. Tze-Fei Wong[14]; **Illumina** Luana M. Galver (Project Leader)[15], Jian-Bing Fan[15], Kevin Gunderson[15], Sarah S. Murray[1], Arnold R. Oliphant[16], Mark S. Chee (Principal Investigator)[17]; **McGill University and Génome Québec Innovation Centre** Alexandre Montpetit (Project Leader)[18], Fanny Chagnon[18], Vincent Ferretti[18], Martin Leboeuf[18], Jean-François Olivier[4], Michael S. Phillips[18], Stéphanie Roumy[15], Clémentine Sallée[19], Andrei Verner[18], Thomas J. Hudson (Principal Investigator)[20]; **University of California at San Francisco and Washington University** Pui-Yan Kwok (Principal Investigator)[21], Dongmei Cai[21], Daniel C. Koboldt[22], Raymond D. Miller[22], Ludmila Pawlikowska[21], Patricia Taillon-Miller[22], Ming Xiao[21]; **University of Hong Kong** Lap-Chee Tsui (Principal Investigator)[23], William Mak[23], You Qiang Song[23], Paul K. H. Tam[23]; **University of Tokyo and RIKEN** Yusuke Nakamura (Principal Investigator)[24,25], Takahisa Kawaguchi[25], Takuya Kitamoto[25], Takashi Morizono[25], Atsushi Nagashima[25], Yozo Ohnishi[25], Akihiro Sekine[25], Toshihiro Tanaka[25], Tatsuhiko Tsunoda[25]; **Wellcome Trust Sanger Institute** Panos Deloukas (Project Leader)[26], Christine P. Bird[26], Marcos Delgado[26], Emmanouil T. Dermitzakis[26], Rhian Gwilliam[26], Sarah Hunt[26], Jonathan Morrison[27], Don Powell[26], Barbara E. Stranger[26], Pamela Whittaker[26], David R. Bentley (Principal Investigator)[28]

Analysis groups: Broad Institute Mark J. Daly (Project Leader)[7,9], Paul I. W. de Bakker[7,9], Jeff Barrett[7,9], Yves R. Chretien[7], Julian Maller[7,9], Steve McCarroll[7,9], Nick Patterson[7], Itsik Pe'er[29], Alkes Price[7], Shaun Purcell[9], Daniel J. Richter[7], Pardis Sabeti[7], Richa Saxena[7,9], Stephen F. Schaffner[7], Pak C. Sham[23], Patrick Varilly[7], David Altshuler (Principal Investigator)[7,9]; **Cold Spring Harbor Laboratory** Lincoln D. Stein (Principal Investigator)[6], Lalitha Krishnan[6], Albert Vernon Smith[6], Marcela K. Tello-Ruiz[6], Gudmundur A. Thorisson[30]; **Johns Hopkins University School of Medicine** Aravinda Chakravarti (Principal Investigator)[31], Peter E. Chen[31], David J. Cutler[31], Carl S. Kashuk[31], Shin Lin[31]; **University of Michigan** Gonçalo R. Abecasis (Principal Investigator)[32], Weihua Guan[32], Yun Li[32], Heather M. Munro[33], Zhaohui Steve Qin[32], Daryl J. Thomas[34]; **University of Oxford** Gilean McVean (Project Leader)[35], Adam Auton[35], Leonardo Bottolo[35], Niall Cardin[35], Susana Eyheramendy[35], Colin Freeman[35], Jonathan Marchini[35], Simon Myers[35], Chris Spencer[7], Matthew Stephens[36], Peter Donnelly (Principal Investigator)[35]; **University of Oxford, Wellcome Trust Centre for Human Genetics** Lon R. Cardon (Principal Investigator)[37], Geraldine Clarke[38], David M. Evans[38], Andrew P. Morris[38], Bruce S. Weir[39]; **RIKEN** Tatsuhiko Tsunoda (Principal Investigator)[25], Todd A. Johnson[25]; **US National Institutes of Health** James C. Mullikin[40]; **US National Institutes of Health National Center for Biotechnology Information** Stephen T. Sherry[41], Michael Feolo[41], Andrew Skol[42]

Community engagement/public consultation and sample collection groups: Beijing Normal University and Beijing Genomics Institute Houcan Zhang[43], Changqing Zeng[8], Hui Zhao[8]; **Health Sciences University of Hokkaido, Eubios Ethics Institute, and Shinshu University** Ichiro Matsuda (Principal Investigator)[44], Yoshimitsu Fukushima[45], Darryl R. Macer[46], Eiko Suda[47]; **Howard University and University of Ibadan** Charles N. Rotimi (Principal Investigator)[48], Clement A. Adebamowo[49], Ike Ajayi[49], Toyin Aniagwu[49], Patricia A. Marshall[50], Chibuzor Nkwodimmah[49], Charmaine D. M. Royal[48]; **University of Utah** Mark F. Leppert (Principal Investigator)[51], Missy Dixon[51], Andy Peiffer[51]

Ethical, legal and social issues: Chinese Academy of Social Sciences Renzong Qiu[52]; **Genetic Interest Group** Alastair Kent[53]; **Kyoto University** Kazuto Kato[54]; **Nagasaki University** Norio Niikawa[55]; **University of Ibadan School of Medicine** Isaac F. Adewole[49]; **University of Montréal** Bartha M. Knoppers[19]; **University of Oklahoma** Morris W. Foster[56]; **Vanderbilt University** Ellen Wright Clayton[57]; **Wellcome Trust** Jessica Watkin[58]

SNP discovery: Baylor College of Medicine Richard A. Gibbs (Principal Investigator)[3], John W. Belmont[3], Donna Muzny[3], Lynne Nazareth[3], Erica Sodergren[3], George M. Weinstock[3], David A. Wheeler[3], Imtaz Yakub[5]; **Broad Institute of Harvard and Massachusetts Institute of Technology** Stacey B. Gabriel (Project Leader)[7], Robert C. Onofrio[7], Daniel J. Richter[7], Lluda Ziaugra[7], Bruce W. Birren[7], Mark J. Daly[7,9], David Altshuler (Principal Investigator)[7,9]; **Washington University** Richard K. Wilson (Principal Investigator)[59], Lucinda L. Fulton[59]; **Wellcome Trust Sanger Institute** Jane Rogers (Principal Investigator)[26], John Burton[26], Nigel P. Carter[26], Christopher M. Clee[26], Mark Griffiths[26], Matthew C. Jones[26], Kirsten McLay[26], Robert W. Plumb[26], Mark T. Ross[26], Sarah K. Sims[26], David L. Willey[26]

Scientific management: Chinese Academy of Sciences Zhu Chen[60], Hua Han[60], Le Kang[60]; **Genome Canada** Martin Godbout[61], John C. Wallenburg[62]; **Génome Québec** Paul L'Archevêque[63], Guy Bellemare[63]; **Japanese Ministry of Education, Culture, Sports, Science and Technology** Koji Saeki[64]; **Ministry of Science and Technology of the People's Republic of China** Hongguang Wang[65], Daochang An[65], Hongbo Fu[65], Qing Li[65], Zhen Wang[65]; **The Human Genetic Resource Administration of China** Renwu Wang[66]; **The SNP Consortium** Arthur L. Holden[15]; **US National Institutes of Health** Lisa D. Brooks[67], Jean E. McEwen[67], Mark S. Guyer[67], Vivian Ota Wang[67,68], Jane L. Peterson[67], Michael Shi[69], Jack Spiegel[70], Lawrence M. Sung[71], Lynn F. Zacharia[67], Francis S. Collins[72]; **Wellcome Trust Karen Kennedy**[61], Ruth Jamieson[58], John Stewart[58]

[1]The Scripps Research Institute, 10550 North Torrey Pines Road MEM275, La Jolla, California 92037, USA. [2]Perlegen Sciences, Inc., 2021 Stierlin Court, Mountain View, California 94043, USA. [3]Baylor College of Medicine, Human Genome Sequencing Center, Department of Molecular and Human Genetics, 1 Baylor Plaza, Houston, Texas 77030, USA. [4]Affymetrix, Inc., 3420 Central Expressway, Santa Clara, California 95051, USA. [5]Pacific Biosciences, 1505 Adams Drive, Menlo Park, California 94025, USA. [6]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [7]The Broad Institute of Harvard and Massachusetts Institute of Technology, 1 Kendall Square, Cambridge, Massachusetts 02139, USA. [8]Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 100300, China. [9]Massachusetts General Hospital and Harvard Medical School, Simches Research Center, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [10]Chinese National Human Genome Center at Beijing, 3-707 N. Yongchang Road, Beijing Economic-Technological Development Area, Beijing 100176, China. [11]Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China. [12]Fudan University and CAS-MPG Partner Institute for Computational Biology, School of Life Sciences, SIBS, CAS, Shanghai 201203, China. [13]The Chinese University of Hong Kong, Department of Biochemistry, The Croucher Laboratory for Human Genetics, 6/F Mong Man Wai Building, Shatin, Hong Kong. [14]Hong Kong University of Science and Technology, Department of Biochemistry and Applied Genomics Center, Clear Water Bay, Knowloon, Hong Kong. [15]Illumina, 9885 Towne Centre Drive, San Diego, California 92121, USA. [16]Complete Genomics, Inc., 658 North Pastoria Avenue, Sunnyvale, California 94085, USA. [17]Prognosys Biosciences, Inc., 4215 Sorrento Valley Boulevard, Suite 105, San Diego, California 92121, USA. [18]McGill University and Génome Québec Innovation Centre, 740 Dr. Penfield Avenue, Montréal, Québec H3A 1A4, Canada. [19]University of Montréal, The Public Law Research Centre (CRDP), PO Box 6128, Downtown Station, Montréal, Québec H3C 3J7, Canada. [20]Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 500, Toronto, Ontario M5G 1L7, Canada. [21]University of California, San Francisco, Cardiovascular Research Institute, 513 Parnassus Avenue, Box 0793, San Francisco, California 94143, USA. [22]Washington University School of Medicine, Department of Genetics, 660 South Euclid Avenue, Box 8232, St Louis, Missouri 63110, USA. [23]University of Hong Kong, Genome Research Centre, 6/F, Laboratory Block, 21 Sassoon Road, Pokfulam, Hong Kong. [24]University of Tokyo, Institute of Medical Science, 4-6-1 Sirokanedai, Minato-ku, Tokyo 108-8639, Japan. [25]RIKEN SNP Research Center, 1-7-22 Suehiro-cho, Tsurumi-ku Yokohama, Kanagawa 230-0045, Japan. [26]Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [27]University of Cambridge, Department of Oncology, Cambridge CB1 8RN, UK. [28]Solexa Ltd, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK. [29]Columbia University, 500 West 120th Street, New York, New York 10027, USA. [30]University of Leicester, Department of Genetics, Leicester LE1 7RH, UK. [31]Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine, Broadway Research Building, Suite 579, 733 North Broadway, Baltimore, Maryland 21205, USA. [32]University of Michigan, Center for Statistical Genetics, Department of Biostatistics, 1420 Washington Heights, Ann Arbor, Michigan 48109, USA. [33]International Epidemiology Institute, 1455 Research Boulevard, Suite 550, Rockville, Maryland 20850, USA. [34]Center for Biomolecular Science and Engineering, Engineering 2, Suite 501, Mail Stop CBSE/ITI, UC Santa Cruz, Santa Cruz, California 95064, USA. [35]University of Oxford, Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK. [36]University of Chicago, Department of Statistics, 5734 South University Avenue, Eckhart Hall, Room 126, Chicago, Illinois 60637, USA. [37]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. [38]University of Oxford/Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. [39]University of Washington Department of Biostatistics, Box 357232, Seattle, Washington 98195, USA. [40]US National Institutes of Health, National Human Genome Research Institute, 50 South Drive, Bethesda, Maryland 20892, USA. [41]US National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, 8600 Rockville Pike, Bethesda, Maryland 20894, USA. [42]University of Chicago, Department of Medicine, Section of Genetic Medicine, 5801 South Ellis, Chicago, Illinois 60637, USA. [43]Beijing Normal University, 19 Xinjiekouwai Street, Beijing 100875, China. [44]Health Sciences University of Hokkaido, Ishikari Tobetsu Machi 1757, Hokkaido 061-0293, Japan. [45]Shinshu University School of Medicine, Department of Medical Genetics, Matsumoto 390-8621, Japan. [46]United Nations Educational, Scientific and Cultural Organization (UNESCO Bangkok), 920 Sukhumwit Road, Prakanong, Bangkok 10110, Thailand. [47]University of Tsukuba, Eubios Ethics Institute, PO Box 125, Tsukuba Science City 305-8691, Japan. [48]Howard University, National Human Genome Center, 2216 6th Street, NW, Washington DC 20059, USA. [49]University of Ibadan College of Medicine, Ibadan, Oyo State, Nigeria. [50]Case Western Reserve University School of Medicine, Department of Bioethics, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. [51]University of Utah, Eccles Institute of Human Genetics, Department of Human Genetics, 15 North 2030 East, Salt Lake City, Utah 84112, USA. [52]Chinese Academy of Social Sciences, Institute of Philosophy/Center for Applied Ethics, 2121, Building 9, Caoqiao Xinyuan 3 Qu, Beijing 100067, China. [53]Genetic Interest Group, 4D Leroy House, 436 Essex Road, London N130P, UK. [54]Kyoto University, Institute for Research in Humanities and Graduate School of Biostudies, Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan. [55]Nagasaki University Graduate

School of Biomedical Sciences, Department of Human Genetics, Sakamoto 1-12-4, Nagasaki 852-8523, Japan. [56]University of Oklahoma, Department of Anthropology, 455 West Lindsey Street, Norman, Oklahoma 73019, USA. [57]Vanderbilt University, Center for Genetics and Health Policy, 507 Light Hall, Nashville, Tennessee 37232, USA. [58]Wellcome Trust, 215 Euston Road, London NW1 2BE, UK. [59]Washington University School of Medicine, Genome Sequencing Center, Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [60]Chinese Academy of Sciences, 52 Sanlihe Road, Beijing 100864, China. [61]Genome Canada, 150 Metcalfe Street, Suite 2100, Ottawa, Ontario K2P 1P1, Canada. [62]McGill University, Office of Technology Transfer, 3550 University Street, Montréal, Québec H3A 2A7, Canada. [63]Génome Québec, 630, boulevard René-Lévesque Ouest, Montréal, Québec H3B 1S6, Canada. [64]Ministry of Education, Culture, Sports, Science, and Technology, 3-2-2 Kasumigaseki, Chiyodaku, Tokyo 100-8959, Japan. [65]Ministry of Science and Technology of the People's Republic of China, 15 B. Fuxing Road, Beijing 100862, China. [66]The Human Genetic Resource Administration of China, b7, Zaojunmiao, Haidian District, Beijing 100081, China. [67]US National Institutes of Health, National Human Genome Research Institute, 5635 Fishers Lane, Bethesda, Maryland 20892, USA. [68]US National Institutes of Health, Office of Behavioral and Social Science Research, 31 Center Drive, Bethesda, Maryland 20892, USA. [69]Novartis Pharmaceuticals Corporation, Biomarker Development, One Health Plaza, East Hanover, New Jersey 07936, USA. [70]US National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Rockville, Maryland 20852, USA. [71]University of Maryland School of Law, 500 West Baltimore Street, Baltimore, Maryland 21201, USA. [72]US National Institutes of Health, National Human Genome Research Institute, 31 Center Drive, Bethesda, Maryland 20892, USA.

doi:10.1038/nature06258 nature

METHODS

SNP selection and genotyping. All SNPs in dbSNP release 122 were considered for genotyping by Perlegen. Among these the following were excluded: SNPs for which no assay could be designed (primarily through location in repeat-rich regions; approximately 2.5 million); SNPs shown previously in samples from related populations[15] to be most probably in perfect association ($r^2 = 1$) with a Phase I SNP (approximately 122,000); all but one of SNPs shown previously[15] to be most probably in perfect association ($r^2 = 1$) with each other but not with a Phase I SNP (approximately 62,000); and SNPs shown previously[15] to have MAF < 0.05 (approximately 119,000). In addition, a few SNPs were excluded for efficiency (for example, if an amplicon contained a single SNP). Approximately 30,000 SNPs that had been typed in Phase I were deliberately retyped in Phase II to allow detailed comparisons of data quality, and an additional 15,000 SNPs that showed discrepancies between multiple genotyping attempts in Phase I were re-typed in Phase II. A further 2,000 SNPs identified by the Mammalian Gene Collection were also typed.

Perlegen performed genotyping using custom high-density oligonucleotide arrays as previously described[15]. Initially, a pilot phase was carried out on chromosome 2p to optimize experimental workflow and data handling. Details of amplicons used in the experiment and PCR primers can be found at http://genome.perlegen.com/pcr/ and also on the HapMap website. The arrays were tiled with sets of 25-bp probes for each SNP, with either 40 or 24 probes per SNP. These consisted of four sets of features, corresponding to forward and reverse strand tilings of sequences complementary to each of the two SNP alleles. Within a feature set, the position of the SNP within the oligonucleotide varied from position 11 to position 15. Mismatch probes were used to measure background, and by comparison with the perfect match probes, to detect the presence or absence of a specific PCR product. The 40-feature and 24-feature tilings both provided 10 perfect-match features for each SNP allele and differed only in the number of mismatch probes.

Genotypes were scored by clustering intensity measurements as previously described[15]. In addition, quality scores similar to Phred scores were computed for each genotype call, based on a combination of experimental metrics correlated to data quality. Assays with overall call rates less than 80% or with poor average quality scores were flagged as failed. About 38% of the tiled assays failed these basic criteria, and the remainder were processed using the more rigorous HapMap Project data quality control filters. For analysis of the whole genome, probes for 4,373,926 distinct SNPs were tiled onto 32 chip designs, with 32 SNPs tiled in replicate onto each chip design for quality control (QC). Perlegen did not type the samples by plates as had been done for the Phase I genotyping, instead typing large numbers of SNPs one sample at a time. Consequently, blank wells on each plate were not included as a component of QC for this genotyping. In the Phase I HapMap a single JPT sample had been excluded because of technical problems. Perlegen typed a replacement sample (from the original JPT collection) for all new SNPs. This sample was not specifically genotyped on the Phase I SNPs, although a substantial fraction of these was typed in Phase II.

Additional genotype submissions came from the Affymetrix GeneChip Human Mapping 500K array called with the BRLMM algorithm. In release 21a additional genotype submissions were incorporated from the MHC haplotype consortium[11], the Illumina HumanHap300 BeadChip, the Illumina Human-1 Genotyping BeadChip and the 10K non-synonymous SNP set from Affymetrix (ParAllele).

Details of primer design, DNA amplification, DNA labelling and hybridization and signal detection for the Perlegen platform can be found in Supplementary Text 7.

QC analyses. Genotype submissions were assessed for mendelian errors (where possible), missing data rates and Hardy–Weinberg proportions. QC filters were applied as previously described[3]; to achieve QC+ status a SNP had to have fewer than two mendelian errors, less than 20% missing data and $P > 0.001$ for Hardy–Weinberg analysis. The consensus data set consists only of SNPs for which QC+ submissions were available from all analysis panels. Where multiple submissions met the QC criteria the submission with the lowest missing data rate was chosen for inclusion in the non-redundant filtered data set. Comparison of the Phase II HapMap with the Affymetrix 500K genotypes has shown approximately 20 SNPs where the reported minor allele is discrepant (referred to as 'allele-flipping'). Over the entire data set, we expect that 500–2,000 SNPs have this problem and the vast majority will occur in SNPs from Phase I of the project. The Data Coordination Center (DCC) is working to resolve as many of these as possible.

Analyses of data quality. See Supplementary Text 2.

Analyses of population stratification, relatedness and homozygosity. See Supplementary Texts 3–6.

Analysis of recombination rate and gene ontology. We used the Panther Database[41] to obtain details of the gene molecular function and biological process. Genes are grouped into 28 top-level molecular function groups and 30 top-level biological process groups, with each gene allowed to exist in more than one group. We identified 14,979 non-overlapping autosomal genes from the Panther RefSeq Annotation for which we could obtain recombination rates. Of these, 9,735 had at least one assigned molecular function and 9,432 had at least one assigned biological process. Genes without a molecular function or biological process were removed from the corresponding analysis. To control for gene size, we estimated the mean recombination rate over a 20-kb region centred on the mid-point of each gene transcription region.

Genes were grouped based on molecular function and biological process. A mean recombination rate was calculated for each group. The significance of the result from each group was calculated via a permutation test involving 10^5 random groupings of genes. No correction was made for multiple testing. To account for the effect of G+C content on recombination, we performed a linear regression between the G+C content and recombination rate of all genes in each sample. Using the estimated regression parameters, the proportion of recombination explained by G+C content was subtracted from each gene.

Identification of non-synonymous SNPs and tests for natural selection. Using annotations from dbSNP release 125 we identified 17,427 polymorphic non-synonymous SNPs in release 21 and 15,976 polymorphic synonymous SNPs. Of these, 15,583 non-synonymous and 14,324 synonymous SNPs were autosomal and could have ancestral allele status unambiguously assigned by parsimony through comparison to the chimpanzee and macaque genomes. We used the phased haplotypes for analysis in which missing data had been imputed. F_{ST} was calculated using the method of Weir and Cockerham[59].

To detect recent partial selective sweeps we used the long-range haplotype (LRH) test[44,49] and the integrated haplotype score (iHS) test[9]. On simulated data[45], we found that the tests have similar power to detect recent selection but the iHS test has slightly lower power at low haplotype frequency and the LRH test has slightly lower power at high frequency. This can be seen in applications to HapMap Phase I data[3,9], where the iHS test misses the well-known cases of *HBB* and *CD36* and the LRH test misses the *SULT1C2* region. Although both tests are based on the concept of EHH[44], we observed that the false positives produced by the two tests tend not to overlap and thus that signals detected by both tests have a very low false-positive rate.

©2007 Nature Publishing Group

LETTERS

Genome-wide detection and characterization of positive selection in human populations

Pardis C. Sabeti[1]*, Patrick Varilly[1]*, Ben Fry[1], Jason Lohmueller[1], Elizabeth Hostetter[1], Chris Cotsapas[1,2], Xiaohui Xie[1], Elizabeth H. Byrne[1], Steven A. McCarroll[1,2], Rachelle Gaudet[3], Stephen F. Schaffner[1], Eric S. Lander[1,4,5,6] & The International HapMap Consortium†

With the advent of dense maps of human genetic variation, it is now possible to detect positive natural selection across the human genome. Here we report an analysis of over 3 million polymorphisms from the International HapMap Project Phase 2 (HapMap2)[1]. We used 'long-range haplotype' methods, which were developed to identify alleles segregating in a population that have undergone recent selection[2], and we also developed new methods that are based on cross-population comparisons to discover alleles that have swept to near-fixation within a population. The analysis reveals more than 300 strong candidate regions. Focusing on the strongest 22 regions, we develop a heuristic for scrutinizing these regions to identify candidate targets of selection. In a complementary analysis, we identify 26 non-synonymous, coding, single nucleotide polymorphisms showing regional evidence of positive selection. Examination of these candidates highlights three cases in which two genes in a common biological process have apparently undergone positive selection in the same population: *LARGE* and *DMD*, both related to infection by the Lassa virus[3], in West Africa; *SLC24A5* and *SLC45A2*, both involved in skin pigmentation[4,5], in Europe; and *EDAR* and *EDA2R*, both involved in development of hair follicles[6], in Asia.

An increasing amount of information about genetic variation, together with new analytical methods, is making it possible to explore the recent evolutionary history of the human population. The first phase of the International Haplotype Map, including ~1 million single nucleotide polymorphisms (SNPs)[7], allowed preliminary examination of natural selection in humans. Now, with the publication of the Phase 2 map (HapMap2)[1] in a companion paper, over 3 million SNPs have been genotyped in 420 chromosomes from three continents (120 European (CEU), 120 African (YRI) and 180 Asian from Japan and China (JPT + CHB)).

In our analysis of HapMap2, we first implemented two widely used tests that detect recent positive selection by finding common alleles carried on unusually long haplotypes[2]. The two, the Long-Range Haplotype (LRH)[8] and the integrated Haplotype Score (iHS)[9] tests, rely on the principle that, under positive selection, an allele may rise to high frequency rapidly enough that long-range association with nearby polymorphisms—the long-range haplotype[8]—will not have time to be eliminated by recombination. These tests control for local variation in recombination rates by comparing long haplotypes to other alleles at the same locus. As a result, they lose power as selected alleles approach fixation (100% frequency), because there are then few alternative alleles in the population (Supplementary Fig. 2 and Supplementary Tables 1–2).

We next developed, evaluated and applied a new test, Cross Population Extended Haplotype Homozygosity (XP-EHH), to detect selective sweeps in which the selected allele has approached or achieved fixation in one population but remains polymorphic in the human population as a whole (Methods, and Supplementary Fig. 2 and Supplementary Tables 3–6). Related methods have recently also been described[10–12].

Our analysis of recent positive selection, using the three methods, reveals more than 300 candidate regions[1](Supplementary Fig. 3 and Supplementary Table 7), 22 of which are above a threshold such that no similar events were found in 10 Gb of simulated neutrally evolving sequence (Methods). We focused on these 22 strongest signals (Table 1), which include two well-established cases, *SLC24A5* and *LCT*[4,5,13], and 20 other regions with signals of similar strength.

The challenge is to sift through genetic variation in the candidate regions to identify the variants that were the targets of selection. Our candidate regions are large (mean length, 815 kb; maximum length, 3.5 Mb) and often contain multiple genes (median, 4; maximum, 15). A typical region harbours ~400–4,000 common SNPs (minor allele frequency >5%), of which roughly three-quarters are represented in current SNP databases and half were genotyped as part of HapMap2 (Supplementary Table 8).

We developed three criteria to help highlight potential targets of selection (Supplementary Fig. 1): (1) selected alleles detectable by our tests are likely to be derived (newly arisen), because long-haplotype tests have little power to detect selection on standing (pre-existing) variation[14]; we therefore focused on derived alleles, as identified by comparison to primate outgroups; (2) selected alleles are likely to be highly differentiated between populations, because recent selection is probably a local environmental adaptation[2]; we thus looked for alleles common in only the population(s) under selection; (3) selected alleles must have biological effects. On the basis of current knowledge, we therefore focused on non-synonymous coding SNPs and SNPs in evolutionarily conserved sequences. These criteria are intended as heuristics, not absolute requirements. Some targets of selection may not satisfy them, and some will not be in current SNP databases. Nonetheless, with ~50% of common SNPs in these populations genotyped in HapMap2, a search for causal variants is timely.

We applied the criteria to the regions containing *SLC24A5* and *LCT*, each of which already has a strong candidate gene, mutation and trait. At *SLC24A5*, the 600 kb region contains 914 genotyped

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA. [2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. [5]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [6]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.
†Lists of participants and affiliations appear at the end of the paper.

Table 1 | The twenty-two strongest candidates for natural selection

Region	Chr:position (MB, HG17)	Selected population	Long Haplotype Test	Size (Mb)	Total SNPs with Long Haplotype Signal	Subset of SNPs that fulfil criteria 1	Subset of SNPs that fulfil criteria 1 and 2	Subset of SNPs that fulfil criteria 1, 2 and 3	Genes at or near SNPs that fulfil all three criteria
1	chr1:166	CHB + JPT	LRH, iHS	0.4	92	39	30	2	*BLZF1, SLC19A2*
2	chr2:72.6	CHB + JPT	XP-EHH	0.8	732	250	0	0	
3	chr2:108.7	CHB + JPT	LRH, iHS, XP-EHH	1.0	972	265	7	1	*EDAR*
4	chr2:136.1	CEU	LRH, iHS, XP-EHH	2.4	1,213	282	24	3	*RAB3GAP1, R3HDM1, LCT*
5	chr2:177.9	CEU,CHB + JPT	LRH, iHS, XP-EHH	1.2	1,388	399	79	9	*PDE11A*
6	chr4:33.9	CEU,YRI, CHB + JPT	LRH, iHS	1.7	413	161	33	0	
7	chr4:42	CHB + JPT	LRH, iHS, XP-EHH	0.3	249	94	65	6	*SLC30A9*
8	chr4:159	CHB + JPT	LRH, iHS, XP-EHH	0.3	233	67	34	1	
9	chr10:3	CEU	LRH, iHS, XP-EHH	0.3	179	63	16	1	
10	chr10:22.7	CEU, CHB + JPT	XP-EHH	0.3	254	93	0	0	
11	chr10:55.7	CHB+JPT	LRH, iHS, XP-EHH	0.4	735	221	5	2	*PCDH15*
12	chr12:78.3	YRI	LRH, IHS	0.8	151	91	25	0	
13	chr15:46.4	CEU	XP-EHH	0.6	867	233	5	1	*SLC24A5*
14	chr15:61.8	CHB + JPT	XP-EHH	0.2	252	73	40	6	*HERC1*
15	chr16:64.3	CHB + JPT	XP-EHH	0.4	484	137	2	0	
16	chr16:74.3	CHB + JPT, YRI	LRH, iHS	0.6	55	35	28	3	*CHSTS, ADAT1, KARS*
17	chr17:53.3	CHB + JPT	XP-EHH	0.2	143	41	0	0	
18	chr17:56.4	CEU	XP-EHH	0.4	290	98	26	3	*BCAS3*
19	chr19:43.5	YRI	LRH, iHS, XP-EHH	0.3	83	30	0	0	
20	chr22:32.5	YRI	LRH	0.4	318	188	35	3	*LARGE*
21	chr23:35.1	YRI	LRH, iHS	0.6	50	35	25	0	
22	chr23:63.5	YRI	LRH, iHS	3.5	13	3	1	0	
		Total SNPs		16.74	9,166	2,898	480	41	

Twenty-two regions were identified at a high threshold for significance (Methods), based on the LRH, iHS and/or XP-EHH test. Within these regions, we examined SNPs with the best evidence of being the target of selection on the basis of having a long haplotype signal, and by fulfilling three criteria: (1) being a high-frequency derived allele; (2) being differentiated between populations and common only in the selected population; and (3) being identified as functional by current annotation. Several candidate polymorphisms arise from the analysis including well-known *LCT* and *SLC24A5* (ref. 2), as well as intriguing new candidates.

SNPs. Applying filters progressively (Table 1 and Fig. 1a–d), we found that 867 SNPs are associated with the long-haplotype signal, of which 233 are high-frequency derived alleles, of which 12 are highly differentiated between populations, and of which only 5 are common in Europe and rare in Asia and Africa. Among these five SNPs, there is only one implicated as functional by current knowledge; it has the strongest signal of positive selection and encodes the A111T polymorphism associated with pigment differences in

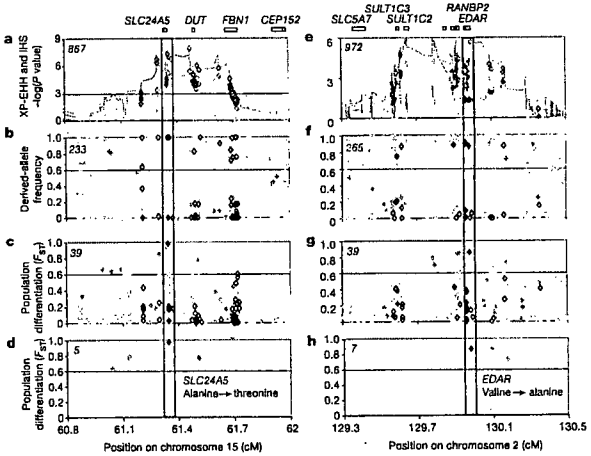


Figure 1 | Localizing *SLC24A5* and *EDAR* signals of selection. a–d, *SLC24A5*. a, Strong evidence for positive selection in CEU samples at a chromosome 15 locus: XP-EHH between CEU and JPT + CHB (blue), CEU and YRI (red), and YRI and JPT + CHB (grey). SNPs are classified as having low probability (bordered diamonds) and high probability (filled diamonds) potential for function. SNPs were filtered to identify likely targets of selection on the basis of the frequency of derived alleles (b), differences between populations (c) and differences between populations for high-frequency derived alleles (less than 20% in non-selected populations) (d). The number of SNPs that passed each filter is given in the top left corner in red. The threonine to alanine candidate polymorphism in *SLC24A5* is the clear outlier. e–h, *EDAR*. e, Similar evidence for positive selection in JPT + CHB at a chromosome 2 locus: XP-EHH between CEU and JPT + CHB (blue), between YRI and JPT + CHB (red), and between CEU and YRI (grey); iHS in JPT + CHB (green). A valine to alanine polymorphism in *EDAR* passes all filters: the frequency of derived alleles (f), differences between populations (g) and differences between populations for high-frequency derived alleles (less than 20% in non-selected populations) (h). Three other functional changes, a D→E change in *SULT1C2* and two SNPs associated with *RANBP2* expression (Methods), have also become common in the selected population.

humans and thought to be the target of positive selection[5]. Our criteria thus uniquely identify the expected allele.

At the *LCT* locus, we found similar degrees of filtration. Within the 2.4 Mb selective sweep, 24 polymorphisms fulfil the first two criteria (Table 1, and Supplementary Fig. 4), with the polymorphism thought to confer adult persistence of lactase among them. However, this SNP was only identified as functional after extensive study of the *LCT* gene[15]. Thus *LCT* shows both the utility and the limits of the heuristics.

Given the encouraging results for *SLC24A5* and *LCT*, we performed a similar analysis on all 22 candidate regions (Table 1). Filtering the 9,166 SNPs associated with the long-haplotype signal, we found that 480 satisfied the first two criteria. We identified 41 out of the 480 SNPs (0.2% of all SNPs genotyped in the regions) as possibly functional on the basis of a newly compiled database of polymorphisms in known coding elements, evolutionarily conserved elements and regulatory elements (Methods; B.F., unpublished), together containing ~ 5.5% of all known SNPs.

Eight of the forty-one SNPs encode non-synonymous changes (Table 1 and Supplementary Table 9). Apart from the well-known case of *SLC24A5*, they are found in *EDAR*, *PCDH15*, *ADAT1*, *KARS*, *HERC1*, *SLC30A9* and *BLFZ1*. The remaining 33 potentially functional SNPs lie within conserved transcription factor motifs, introns, UTRs and other non-coding regions.

To identify additional candidates, we reversed the process by taking non-synonymous coding SNPs with highly differentiated high-frequency derived alleles; these SNPs comprise a tiny fraction of all SNPs and have a higher a priori probability of being targets of selection. Of the 15,816 non-synonymous SNPs in HapMap2, 281 (Supplementary Table 10) have both a high derived-allele frequency (frequency >50%) and clear differentiation between populations (F_{ST} is in the top 0.5 percentile). We examined these 281 SNPs to identify those embedded within long-range haplotypes[16], and identified 26 putative cases of positive selection. These include the eight non-synonymous SNPs identified in the genome-wide analysis above.

Interestingly, analysis of the top regions and the non-synonymous SNPs together revealed three cases of two genes in the same pathway both having strong evidence of selection in a single population.

In the European sample, there is strong evidence for two genes already shown to be associated with skin pigment differences among humans. The first is *SLC24A5*, described above. We further examined the global distribution (Fig. 2) and the predicted effect on protein activity of the *SLC24A5* A111T polymorphism (Supplementary Fig. 5, 6). The second, *SLC45A2*, has an important role in pigmentation in zebrafish, mouse and horse[4]. An L374F substitution in *SLC45A2* is at 100% frequency in the European sample, but absent in the Asian and African samples. A recent association study has shown that the Phe-encoding allele is correlated with fair skin and non-black hair in Europeans[4]. Together, the data support *SLC45A2* as a target of positive selection in Europe[10,17].

In the African sample (Yoruba in Ibadan, Nigeria), there is evidence of selection for two genes with well-documented biological links to the Lassa fever virus. The strongest signal in the genome, on the basis of the LRH test, resides within a 400 kb region that lies entirely within the gene *LARGE*. The LARGE protein is a glycosylase that post-translationally modifies α-dystroglycan, the cellular receptor for Lassa fever virus (as well as other arenaviruses), and the modification has been shown to be critical for virus binding[3]. The virus name is derived from Lassa, Nigeria, where the disease is endemic, with 21% of the population showing signs of exposure[18]. We also noted that the *DMD* locus is on our larger candidate list of regions, with the signal of selection again in the Yoruba sample. *DMD* encodes a cytosolic adaptor protein that binds to α-dystroglycan and is critical for its function. We hypothesize that Lassa fever created selective pressure at *LARGE* and *DMD*[12]. This hypothesis can be tested by correlating the geographical distribution of the selected haplotype with endemicity of the Lassa virus, studying infection of genotyped cells *in vitro*, and searching for an association between the selected haplotype and clinical outcomes in infected patients.

In the Asian samples, we found evidence of selection for non-synonymous polymorphisms in two genes in the ectodysplasin (EDA) pathway, which is involved in development of hair, teeth and exocrine glands[6]. The genes are *EDAR* and *EDA2R*, which encode the key receptors for the ligands EDA A1 and EDA A2, respectively. Notably, the EDA signalling pathway has been shown to be under positive selection for loss of scales in multiple distinct populations of freshwater stickleback fish[19]. A mutation encoding a V370A substitution in *EDAR* is near fixation in Asia and absent in Europe and Africa (Fig. 1e–h). An R57K substitution in *EDA2R* has derived-allele frequencies of 100% in Asia, 70% in Europe and 0% in Africa.

The *EDAR* polymorphism is notable because it is highly differentiated between the Asian and other continental populations (the 3rd most differentiated among 15,816 non-synonymous SNPs), and also within Asian populations (in the top 1% of SNPs differentiated between the Japanese and Chinese HapMap samples). Genotyping of the *EDAR* polymorphism in the CEPH (Centre d'Etudie du Polymorphisme Humain) global diversity panel[20] shows that it is at high but varying frequency throughout Asia and the Americas (for example, 100% in Pima Indians and in parts of China, and 73% in Japan) (Fig. 2, and Supplementary Fig. 7). Studying populations like the Japanese, in which the allele is still segregating, may provide clues to its biological significance.

EDAR has a central role in generation of the primary hair follicle pattern, and mutations in *EDAR* cause hypohidrotic ectodermal

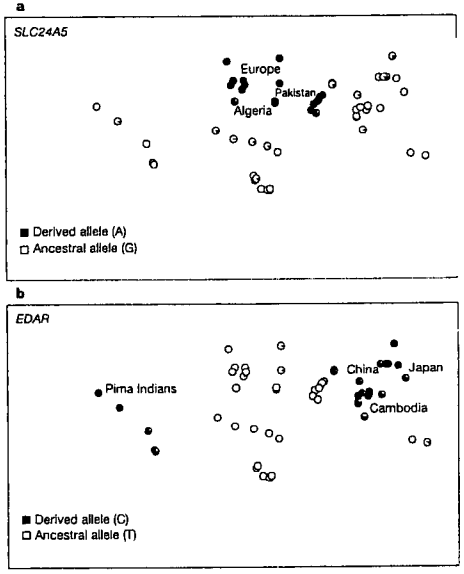


Figure 2 | Global distribution of *SLC24A5* A111T and *EDAR* V370A. Worldwide allele-frequency distributions for candidate polymorphisms with the strongest evidence for selection[20]. a, *SLC24A5* A111T is common in Europe, Northern Africa and Pakistan, but rare or absent elsewhere. b, *EDAR* V370A is common in Asia and the Americas, but absent in Europe and Africa.

dysplasia (HED) in humans and mice, characterized by defects in the development of hair, teeth and exocrine glands[6]. The V370A polymorphism, proposed to be the target of selection, lies within *EDAR*'s highly conserved death domain (Supplementary Fig. 8), the location of the majority of *EDAR* polymorphisms causing HED[21]. Our structural modelling predicts that the polymorphism lies within the binding site of the domain (Fig. 3).

Our analysis only scratches the surface of the recent selective history of the human genome. The results indicate that individual candidates may coalesce into pathways that reveal traits under selection, analogous to the alleles of multiple genes (for example, *HBB*, *G6PD* and *DARC*) that arose and spread in Africa and other tropical populations as a result of the partial protection they confer against malaria[2,12]. Such endeavours will be enhanced by continuing development of analytical methods to localize signals in candidate regions, generation of expanded data sets, advances in comparative genomics to define coding and regulatory regions, and biological follow-up of promising candidates. True understanding of the role of adaptive evolution will require collaboration across multiple disciplines, including molecular and structural biology, medical and population genetics, and history and anthropology.

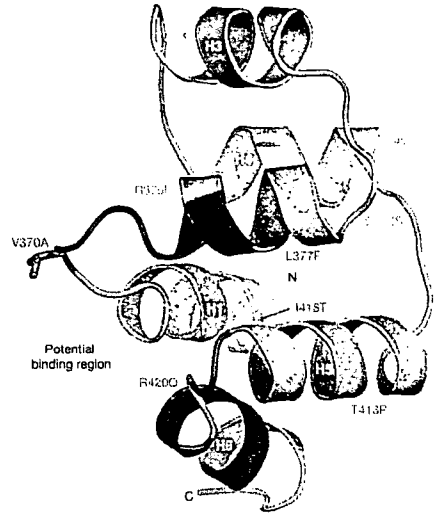


Figure 3 | Structural model of the EDAR death domain. Ribbon representation of a homology model of the EDAR death domain (DD), based on the alignment of the EDAR DD amino acid sequence (EDAR residues 356–431), with multiple known DD structures. The helices are labelled H1 to H6. Residues in blue (the H1–H2 and H5–H6 loops, residues 370–376 and 419–425, respectively) correspond to the homologous residues in Tube that interact with Pelle in the Tube-DD–Pelle-DD structure[24]. These EDAR-DD residues therefore form a potential region of interaction with a DD-containing EDAR-interacting protein, such as EDARADD. The V370A polymorphic residue (red) is located prominently within this potential binding region in the H1–H2 loop. Seven of the thirteen known mis-sense mutations in EDAR that lead to hypohidrotic ectodermal dysplasia (HED) in humans are located in the EDAR-DD: the only four mutations in EDAR that lead to the dominant transmission of HED (green) and three recessive mutations (yellow)[21]. Four of these mutations, R375H, L377F, R420Q and I418T are located in the vicinity of the predicted interaction interface.

METHODS SUMMARY

Genotyping data. Phase 2 of the International Haplotype Map (HapMap2) (www.hapmap.org) contains 3.1 million SNPs genotyped in 420 chromosomes in 3 continental populations (120 European (CEU), 120 African (YRI) and 180 Asian (JPT+CHB))[1]. We further genotyped our top HapMap2 functional candidates in the HGDR-CEPH Human Genome Diversity Cell Line Panel[20].

LRH, iHS and XP-EHH tests. The Long-Range Haplotype (LRH), integrated Haplotype Score (iHS) and Cross Population EHH (XP-EHH) tests detect alleles that have risen to high frequency rapidly enough that long-range association with nearby polymorphisms—the long-range haplotype—has not been eroded by recombination; haplotype length is measured by the EHH[8,9]. The first two tests detect partial selective sweeps, whereas XP-EHH detects selected alleles that have risen to near fixation in one but not all populations. To evaluate the tests, we simulated genomic data for each HapMap population in a range of demographic scenarios—under neutral evolution and twenty scenarios of positive selection—developing the program Sweep (www.broad.mit.edu/mpg/sweep) for analysis. For our top candidates by the three tests, we tested for haplotype-specific recombination rates and copy-number polymorphisms, possible confounders.

Localization. We calculated F_{ST} and derived-allele frequency for all SNPs within the top candidate regions. We developed a database for those regions to annotate all potentially functional DNA changes (B.F., unpublished), including non-synonymous variants, variants disrupting predicted functional motifs, variants within regions of conservation in mammals and variants previously associated with human phenotypic differences, as well as synonymous, intronic and untranslated region variants.

Structural model. We generated a homology model of the EDAR death domain (DD) from available DD structures using Modeller 9v1 (ref. 22). The distribution of conserved residues, built using ConSurf[23] with an EDAR sequence alignment from 22 species, shows a bias to the protein core in helices H1, H2 and H5, supporting our model.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 August; accepted 13 September 2007.

1. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* doi:10.1038/nature06258 (this issue).
2. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* 312, 1614–1620 (2006).
3. Kunz, S. *et al.* Posttranslational modification of α-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J. Virol.* 79, 14282–14296 (2005).
4. Graf, J., Hodgson, R. & van Daal, A. Single nucleotide polymorphisms in the *MATP* gene are associated with normal human pigmentation variation. *Hum. Mutat.* 25, 278–284 (2005).
5. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310, 1782–1786 (2005).
6. Botchkarev, V. A. & Fessing, M. Y. Edar signaling in the control of hair follicle development. *J. Investig. Dermatol. Symp. Proc.* 10, 247–251 (2005).
7. The International Haplotype Map Consortium. A haplotype map of the human genome. *Nature* 437, 1299–1320 (2005).
8. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837 (2002).
9. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72 (2006).
10. Kimura, R., Fujimoto, A., Tokunaga, K. & Ohashi, J. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE* 2, e286 (2007).
11. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, e171 (2007).
12. Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3, e90 (2007).
13. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74, 1111–1120 (2004).
14. Teshima, K. M., Coop, G. & Przeworski, M. How reliable are empirical genomic scans for selective sweeps? 16, 702–712 *Genome Res.* (2006).
15. Kuokkanen, M. *et al.* Transcriptional regulation of the lactase–phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut* 52, 647–652 (2003).
16. Miller, R. G. *Simultaneous statistical inference* XVI 299 (Springer, New York, 1981).
17. Soejima, M., Tachida, H., Ishida, T., Sano, A. & Koda, Y. Evidence for recent positive selection at the human *AIM1* locus in a European population. *Mol. Biol. Evol.* 23, 179–188 (2006).
18. Richmond, J. K. & Baglole, D. J. Lassa fever: epidemiology, clinical features, and social consequences. *Br. Med. J.* 327, 1271–1275 (2003).
19. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307, 1928–1933 (2005).

20. Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* 298, 2381–2385 (2002).
21. Chassaing, N., Bourthoumieu, S., Cossee, M., Calvas, P. & Vincent, M. C. Mutations in *EDAR* account for one-quarter of non-*EDI*-related hypohidrotic ectodermal dysplasia. *Hum. Mutat.* 27, 255–259 (2006).
22. Marti-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325 (2000).
23. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* 33, W299–W302 (2005).
24. Xiao, T., Towb, P., Wasserman, S. A. & Sprang, S. R. Three-dimensional structure of a complex between the death domains of Pelle and Tube. *Cell* 99, 545–555 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements P.C.S. is funded by a Burroughs Wellcome Career Award in the Biomedical Sciences and has been funded by the Damon Runyon Cancer Fellowship and the L'Oreal for Women in Science Award. We thank A. Schier, B. Voight, R. Roberts, M. Kreiger, A. Abzhanov, D. Degusta, M. Burnette, E. Lieberman, M. Daly, D. Altshuler, D. Reich, D. Lieberman and I. Woods for helpful discussions on our analysis and results. We also thank L. Ziaugra, D. Tabbaa and T. Rachupka for experimental assistance. This work was funded in part by grants from the National Human Genome Research Institute (to E.S.L.) and from the Broad Institute of MIT and Harvard.

Author Contributions P.C.S., P.V., B.F. and E.S.L. initiated the project. P.V., B.F. and P.C.S. developed key software. P.C.S., P.V., B.F., S.F.S., J.L., E.H., C.C., X.X., E.B., S.A.McC. and R.G. performed analysis. P.C.S., E.B. and E.H. performed experiments. P.C.S., E.S.L., P.V. and S.F.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.C.S. (pardis@broad.mit.edu).

The International HapMap Consortium (Participants are arranged by institution and then alphabetically within institutions except for Principal Investigators and Project Leaders, as indicated.)

Genotyping centres: Perlegen Sciences Kelly A. Frazer (Principal Investigator)[1], Dennis G. Ballinger[2], David R. Cox[2], David A. Hinds[2], Laura L. Stuve[2]; **Baylor College of Medicine and ParAllele BioScience** Richard A. Gibbs (Principal Investigator)[3], John W. Belmont[3], Andrew Boudreau[4], Paul Hardenbol[5], Suzanne M. Leal[3], Shiran Pasternak[6], David A. Wheeler[3], Thomas D. Willis[4], Fuli Yu[7]; **Beijing Genomics Institute** Huanming Yang (Principal Investigator)[8], Changqing Zeng (Principal Investigator)[8], Yang Gao[8], Haoran Hu[8], Weitao Hu[8], Chaohua Li[8], Wei Lin[8], Siqi Liu[8], Hao Pan[8], Xiaoli Tang[8], Jian Wang[8], Wei Wang[8], Jun Yu[8], Bo Zhang[8], Qingrun Zhang[8], Hongbin Zhao[8], Hui Zhao[8], Jun Zhou[8]; **Broad Institute of Harvard and Massachusetts Institute of Technology** Stacey B. Gabriel (Project Leader)[7], Rachel Barry[7], Brendan Blumenstiel[7], Amy Camargo[7], Matthew Defelice[7], Maura Faggart[7], Mary Goyette[7], Supriya Gupta[7], Jamie Moore[7], Huy Nguyen[7], Robert C. Onofrio[7], Melissa Parkin[7], Jessica Roy[7], Erich Stahl[7], Ellen Winchester[7], Liuda Ziaugra[7], David Altshuler (Principal Investigator)[7,9]; **Chinese National Human Genome Center at Beijing** Yan Shen (Principal Investigator)[10], Zhijian Yao[10]; **Chinese National Human Genome Center at Shanghai** Wei Huang (Principal Investigator)[11], Xun Chu[11], Yungang He[11], Li Jin[12], Yangfan Liu[11], Yayun Shen[11], Weiwei Sun[11], Haifeng Wang[11], Yi Wang[11], Ying Wang[11], Xiaoyan Xiong[11], Liang Xu[11]; **Chinese University of Hong Kong** Mary M. Y. Waye (Principal Investigator)[13], Stephen K. W. Tsui[13]; **Hong Kong University of Science and Technology** Hong Xue (Principal Investigator)[14], J. Tze-Fei Wong[14]; **Illumina** Luana M. Galver (Project Leader)[15], Jian-Bing Fan[15], Kevin Gunderson[15], Sarah S. Murray[1], Arnold R. Oliphant[16], Mark S. Chee (Principal Investigator)[17]; **McGill University and Génome Québec Innovation Centre** Alexandre Montpetit (Project Leader)[18], Fanny Chagnon[18], Vincent Ferretti[18], Martin Leboeuf[18], Jean-François Olivier[4], Michael S. Phillips[18], Stéphanie Roumy[19], Clémentine Sallée[19], Andrei Verner[18], Thomas J. Hudson (Principal Investigator)[20]; **University of California at San Francisco and Washington University** Pui-Yan Kwok (Principal Investigator)[21], Dongmei Cai[21], Daniel C. Koboldt[22], Raymond D. Miller[22], Ludmila Pawlikowska[21], Patricia Taillon-Miller[22], Ming Xiao[21]; **University of Hong Kong** Lap-Chee Tsui (Principal Investigator)[23], William Mak[23], You Qiang Song[23], Paul K. H. Tam[23]; **University of Tokyo and RIKEN Yusuke** Nakamura (Principal Investigator)[24,25], Takahisa Kawaguchi[25], Takuya Kitamoto[25], Takashi Morizono[25], Atsushi Nagashima[25], Yozo Ohnishi[25], Akihiro Sekine[25], Toshihiro Tanaka[25], Tatsuhiko Tsunoda[25]; **Wellcome Trust Sanger Institute** Panos Deloukas (Project Leader)[26], Christine P. Bird[26], Marcos Delgado[26], Emmanouil T. Dermitzakis[26], Rhian Gwilliam[26], Sarah Hunt[26], Jonathan Morrison[27], Don Powell[26], Barbara E. Stranger[26], Pamela Whittaker[26], David R. Bentley (Principal Investigator)[28]

Analysis groups: Broad Institute Mark J. Daly (Project Leader)[7,9], Paul I. W. de Bakker[7,9], Jeff Barrett[7,9], Yves R. Chretien[7], Julian Maller[7,9], Steve McCarroll[7,9], Nick Patterson[7], Itsik Pe'er[29], Alkes Price[7], Shaun Purcell[9], Daniel J. Richter[7], Pardis Sabeti[7], Richa Saxena[7,9], Stephen F. Schaffner[7], Pak C. Sham[23], Patrick Varilly[7], David Altshuler (Principal Investigator)[7,9]; **Cold Spring Harbor Laboratory** Lincoln D. Stein (Principal Investigator)[8], Lalitha Krishnan[6], Albert Vernon Smith[6], Marcela K. Tello-Ruiz[6], Gudmundur A. Thorisson[30]; **Johns Hopkins University School of Medicine** Aravinda Chakravarti (Principal Investigator)[31], Peter E. Chen[31], David J. Cutler[31], Carl S. Kashuk[31], Shin Lin[31]; **University of Michigan** Gonçalo R. Abecasis (Principal Investigator)[32], Weihua Guan[32], Yun Li[32], Heather M. Munro[33], Zhaohui Steve Qin[32], Daryl J. Thomas[34]; **University of Oxford** Gilean McVean (Project Leader)[35], Adam Auton[35], Leonardo Bottolo[35], Niall Cardin[35], Susana Eyheramendy[35], Colin Freeman[35], Jonathan Marchini[35], Simon Myers[35], Chris Spencer[7], Matthew Stephens[36], Peter Donnelly (Principal Investigator)[35]; **University of Oxford, Wellcome Trust Centre for Human Genetics** Lon R. Cardon (Principal Investigator)[37], Geraldine Clarke[38], David M. Evans[38], Andrew P. Morris[38], Bruce S. Weir[39]; **RIKEN Tatsuhiko Tsunoda** (Principal Investigator)[25], Todd A. Johnson[25]; **US National Institutes of Health** James C. Mullikin[40]; **US National Institutes of Health National Center for Biotechnology Information** Stephen T. Sherry[41], Michael Feolo[41], Andrew Skol[42]

Community engagement/public consultation and sample collection groups: Beijing Normal University and Beijing Genomics Institute Houcan Zhang[43], Changqing Zeng[8], Hui Zhao[8]; **Health Sciences University of Hokkaido, Eubios Ethics Institute, and Shinshu University** Ichiro Matsuda (Principal Investigator)[44], Yoshimitsu Fukushima[45], Darryl R. Macer[46], Eiko Suda[47]; **Howard University and University of Ibadan** Charles N. Rotimi (Principal Investigator)[48], Clement A. Adebamowo[49], Ike Ajayi[49], Toyin Aniagwu[49], Patricia A. Marshall[50], Chibuzor Nkwodimmah[49], Charmaine D. M. Royal[48]; **University of Utah** Mark F. Leppert (Principal Investigator)[51], Missy Dixon[51], Andy Peiffer[51]

Ethical, legal and social issues: Chinese Academy of Social Sciences Renzong Qiu[52]; **Genetic Interest Group** Alastair Kent[53]; **Kyoto University** Kazuto Kato[54]; **Nagasaki University** Norio Niikawa[55]; **University of Ibadan School of Medicine** Isaac F. Adewole[49]; **University of Montréal** Bartha M. Knoppers[14]; **University of Oklahoma** Morris W. Foster[56]; **Vanderbilt University** Ellen Wright Clayton[57]; **Wellcome Trust** Jessica Watkin[58]

SNP discovery: Baylor College of Medicine Richard A. Gibbs (Principal Investigator)[3], John W. Belmont[3], Donna Muzny[3], Lynne Nazareth[3], Erica Sodergren[3], George M. Weinstock[3], David A. Wheeler[3], Imtaz Yakub[3]; **Broad Institute of Harvard and Massachusetts Institute of Technology** Stacey B. Gabriel (Project Leader)[7], Robert C. Onofrio[7], Daniel J. Richter[7], Liuda Ziaugra[7], Bruce W. Birren[7], Mark J. Daly[7,9], David Altshuler (Principal Investigator)[7,9]; **Washington University** Richard K. Wilson (Principal Investigator)[59], Lucinda L. Fulton[59]; **Wellcome Trust Sanger Institute** Jane Rogers (Principal Investigator)[26], John Burton[26], Nigel P. Carter[26], Christopher M. Clee[26], Mark Griffiths[26], Matthew C. Jones[26], Kirsten McLay[26], Robert W. Plumb[26], Mark T. Ross[26], Sarah K. Sims[26], David L. Willey[26]

Scientific management: Chinese Academy of Sciences Zhu Chen[60], Hua Han[60], Le Kang[60]; **Genome Canada** Martin Godbout[61], John C. Wallenburg[62]; **Génome Québec** Paul L'Archevêque[63], Guy Bellemare[63]; **Japanese Ministry of Education, Culture, Sports, Science and Technology** Koji Saeki[64]; **Ministry of Science and Technology of the People's Republic of China** Hongguang Wang[65], Daochang An[65], Hongbo Fu[65], Qing Li[65], Zhen Wang[65]; **The Human Genetic Resource Administration of China** Renwu Wang[66]; **The SNP Consortium** Arthur L. Holden[15]; **US National Institutes of Health** Lisa D. Brooks[67], Jean E. McEwen[67], Mark S. Guyer[67], Vivian Ota Wang[67,68], Jane L. Peterson[67], Michael Shi[69], Jack Spiegel[70], Lawrence M. Sung[71], Lynn F. Zacharia[67], Francis S. Collins[72]; **Wellcome Trust** Karen Kennedy[61], Ruth Jamieson[58], John Stewart[58]

[1]The Scripps Research Institute, 10550 North Torrey Pines Road MEM275, La Jolla, California 92037, USA. [2]Perlegen Sciences, 2021 Stierlin Court, Mountain View, California 94043, USA. [3]Baylor College of Medicine, Human Genome Sequencing Center, Department of Molecular and Human Genetics, 1 Baylor Plaza, Houston, Texas 77030, USA. [4]Affymetrix, 3420 Central Expressway, Santa Clara, California 95051, USA. [5]Pacific Biosciences, 1505 Adams Drive, Menlo Park, California 94025, USA. [6]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [7]The Broad Institute of Harvard and Massachusetts Institute of Technology, 1 Kendall Square, Cambridge, Massachusetts 02139, USA. [8]Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 100300, China. [9]Massachusetts General Hospital and Harvard Medical School, Simches Research Center, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [10]Chinese National Human Genome Center at Beijing, 3-707 N. Yongchang Road, Beijing Economic-Technological Development Area, Beijing 100176, China. [11]Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China. [12]Fudan University and CAS-MPG Partner Institute for Computational Biology, School of Life Sciences, SIBS, CAS, Shanghai, 201203, China. [13]The Chinese University of Hong Kong, Department of Biochemistry, The Croucher Laboratory for Human Genetics, 6/F Mong Man Wai Building, Shatin, Hong Kong. [14]Hong Kong University of Science and Technology, Department of Biochemistry and Applied Genomics Center, Clear Water Bay, Knowloon, Hong Kong. [15]Illumina, 9885 Towne Centre Drive, San Diego, California 92121, USA. [16]Complete Genomics, 658 North Pastoria Avenue, Sunnyvale, California 94085, USA. [17]Prognosys Biosciences, 4215 Sorrento Valley Boulevard, Suite 105, San Diego, California 92121, USA. [18]McGill University and Génome Québec Innovation Centre, 740 Dr Penfield Avenue, Montréal, Québec H3A 1A4, Canada. [19]University of Montréal, The Public Law Research Centre (CRDP), PO Box 6128, Downtown Station, Montréal, Québec H3C 3J7, Canada. [20]Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 500, Toronto, Ontario, M5G 1L7, Canada. [21]University of California, San Francisco, Cardiovascular Research Institute, 513 Parnassus Avenue, Box 0793, San Francisco, California 94143, USA. [22]Washington University School of Medicine, Department of Genetics, 660 S. Euclid Avenue, Box 8232, St Louis, Missouri 63110, USA. [23]University of Hong Kong, Genome Research Centre, 6/F, Laboratory Block, 21 Sassoon Road, Pokfulam, Hong Kong. [24]University of Tokyo, Institute of Medical Science, 4-6-1 Sirokanedai, Minatoku, Tokyo 108-8639, Japan. [25]RIKEN SNP Research Center,1-7-22 Suehiro-cho, Tsurumi-ku Yokohama, Kanagawa 230-0045, Japan. [26]Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [27]University of Cambridge, Department of Oncology, Cambridge CB1 8RN, UK. [28]Solexa, Chesterford Research Park, Little Chesterford, nr Saffron Walden, Essex CB10 1XL, UK. [29]Columbia University, 500 West 120th Street, New York, New York 10027, USA. [30]University of Leicester, Department of Genetics, Leicester LE1 7RH, UK. [31]Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine, Broadway Research Building, Suite 579, 733 N. Broadway, Baltimore, Maryland 21205, USA. [32]University of Michigan, Center for Statistical Genetics, Department of Biostatistics, 1420 Washington Heights, Ann Arbor, Michigan 48109, USA. [33]International Epidemiology Institute, 1455 Research Boulevard, Suite 550, Rockville, Maryland 20850, USA. [34]Center for Biomolecular Science and Engineering, Engineering 2, Suite 501, Mail Stop CBSE/ITI, UC Santa Cruz, Santa Cruz, California 95064, USA. [35]University of Oxford, Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK. [36]University of Chicago, Department of Statistics, 5734 S. University Avenue, Eckhart Hall, Room 126, Chicago, Illinois 60637, USA. [37]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. [38]University of Oxford/Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. [39]University of Washington Department of Biostatistics, Box 357232, Seattle, Washington 98195, USA. [40]US National Institutes of Health, National Human Genome Research Institute, 50 South Drive, Bethesda, Maryland 20892, USA. [41]US National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, 8600 Rockville Pike, Bethesda, Maryland 20894, USA. [42]University of Chicago, Department of Medicine, Section of Genetic Medicine, 5801 South Ellis, Chicago, Illinois 60637, USA. [43]Beijing Normal University, 19 Xinjiekouwai Street, Beijing 100875, China. [44]Health Sciences University of Hokkaido, Ishikari Tobetsu Machi 1757, Hokkaido 061-0293, Japan. [45]Shinshu University School of Medicine, Department of Medical Genetics, Matsumoto 390-8621, Japan. [46]United Nations Educational, Scientific and Cultural Organization (UNESCO Bangkok), 920 Sukhumwit Road, Prakanong, Bangkok 10110, Thailand. [47]University of Tsukuba, Eubios Ethics Institute, PO Box 125, Tsukuba Science City 305-8691, Japan. [48]Howard University, National Human Genome Center, 2216 6th Street, NW, Washington, District of Columbia 20059, USA. [49]University of Ibadan College of Medicine, Ibadan, Oyo State, Nigeria. [50]Case Western Reserve University School of Medicine, Department of Bioethics, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. [51]University of Utah, Eccles Institute of Human Genetics, Department of Human Genetics, 15 North 2030 East, Salt Lake City, Utah 84112, USA. [52]Chinese Academy of Social Sciences, Institute of Philosophy/Center for Applied Ethics, 2121, Building 9, Caoqiao Xinyuan 3 Qu, Beijing 100067, China. [53]Genetic Interest Group, 4D Leroy House, 436 Essex Road, London N130P, UK. [54]Kyoto University, Institute for Research in Humanities and Graduate School of Biostudies, Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan. [55]Nagasaki University Graduate School of Biomedical Sciences, Department of Human Genetics, Sakamoto 1-12-4, Nagasaki 852-8523, Japan. [56]University of Oklahoma, Department of Anthropology, 455 W. Lindsey Street, Norman, Oklahoma 73019, USA. [57]Vanderbilt University, Center for Genetics and Health Policy, 507 Light Hall, Nashville, Tennessee 37232, USA. [58]Wellcome Trust, 215 Euston Road, London NW1 2BE, UK. [59]Washington University School of Medicine, Genome Sequencing Center, Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [60]Chinese Academy of Sciences, 52 Sanlihe Road, Beijing 100864, China. [61]Genome Canada, 150 Metcalfe Street, Suite 2100, Ottawa, Ontario K2P 1P1, Canada. [62]McGill University, Office of Technology Transfer, 3550 University Street, Montréal, Québec H3A 2A7, Canada. [63]Génome Québec, 630, boulevard René-Lévesque Ouest, Montréal, Québec H3B 1S6, Canada. [64]Ministry of Education, Culture, Sports, Science, and Technology, 3-2-2 Kasumigaseki, Chiyodaku, Tokyo 100-8959, Japan. [65]Ministry of Science and Technology of the People's Republic of China, 15 B. Fuxing Road, Beijing 100862, China. [66]The Human Genetic Resource Administration of China, b7, Zaojunmiao, Haidian District, Beijing 100081, China. [67]US National Institutes of Health, National Human Genome Research Institute, 5635 Fishers Lane, Bethesda, Maryland 20892, USA. [68]US National Institutes of Health, Office of Behavioral and Social Science Research, 31 Center Drive, Bethesda, Maryland 20892, USA. [69]Novartis Pharmaceuticals Corporation, Biomarker Development, One Health Plaza, East Hanover, New Jersey 07936, USA. [70]US National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Rockville, Maryland 20852, USA. [71]University of Maryland School of Law, 500 W. Baltimore Street, Baltimore, Maryland 21201, USA. [72]US National Institutes of Health, National Human Genome Research Institute, 31 Center Drive, Bethesda, Maryland 20892, USA.

J Hum Genet (2007) 52:675–679
DOI 10.1007/s10038-007-0170-1

ORIGINAL ARTICLE

Nationwide survey on predictive genetic testing for late-onset, incurable neurological diseases in Japan

Kunihiro Yoshida · Takahito Wada · Akihiro Sakurai · Keiko Wakui · Shu-ichi Ikeda · Yoshimitsu Fukushima

Received: 15 February 2007 / Accepted: 12 June 2007 / Published online: 5 July 2007

Abstract A nationwide survey was conducted for predictive genetic testing for late-onset, incurable neurological diseases. A questionnaire was sent to 125 university hospitals and national hospitals, and was returned by 69% of them. Of the 86 responding hospitals, 63 had genetic counseling clinics and answered the questions concerning predictive testing. Of these, 46 had experienced clients with an interest in or a request for predictive testing during the period from April 2004 to March 2006. A total of 322 clients were accumulated, the majority of which were interested in myotonic dystrophy ($n = 150$), followed by spinocerebellar ataxia ($n = 86$), spinal and bulbar muscular atrophy ($n = 40$) and Huntington's disease ($n = 31$). Most such clients were counseled by medical doctors, who had the "Japanese Board of Medical Genetics, Clinical Geneticist" certification, but others, including neurologists, nurses, clinical psychologists or genetic counselors also contributed, albeit to a lesser extent, to genetic counseling in Japan. Many respondents felt that a multidisciplinary approach by a counseling team consisting of a clinical geneticist, a neurologist, a genetic nurse, a clinical psychologist and a genetic counselor had not yet been established. There will be a great need for educated and trained non-medical doctor staff not only to improve the quality of genetic counseling and psychological support for such clients, but also to conduct the psychosocial research on Japanese clients requesting predictive genetic testing.

Keywords Genetic counseling · Predictive genetic testing · Neurological disease · Clinical geneticist

K. Yoshida (✉) · S. Ikeda
Department of Neurology and Rheumatology,
Shinshu University Hospital, 3-1-1 Asahi,
Matsumoto 390-8621, Japan
e-mail: kyoshida@hsp.md.shinshu-u.ac.jp

K. Yoshida · T. Wada · A. Sakurai · K. Wakui · Y. Fukushima
Division of Clinical and Molecular Genetics,
Shinshu University Hospital, Matsumoto, Japan

T. Wada · A. Sakurai · K. Wakui · Y. Fukushima
Department of Preventive Medicine,
Shinshu University School of Medicine,
Matsumoto, Japan

Introduction

In Japan, genetic counseling has not yet taken root as a common clinical practice; but genetic testing and genetic research for human diseases have become increasingly popular. Along with this trend, the establishment of a genetic counseling system is strongly needed to successfully conduct clinical practice and research for genetic diseases. In recent years, specific genetic counseling departments have been set up nationwide. Now more than 60 hospitals have opened genetic counseling clinics in Japan. The National Liaison Meeting for Clinical Sections of Medical Genetics has been held annually since 2003, where the persons responsible for genetic counseling departments have gathered from throughout the country and held discussions on several issues concerning the genetic counseling system in Japan.

The most controversial topic in clinical genetics is how we should handle the need for predictive and prenatal genetic testing, especially, predictive testing for late-onset, incurable neurological diseases such as Huntington's disease (HD) and spinocerebellar ataxia (SCA). It raises various ethical, legal and psychosocial implications (Brandt

1994; Evers-Kiebooms and Decruyenaere 1998; Chapman 2002; Taylor 2004); thus, pre-test genetic counseling and post-test follow-up should be carefully conducted for clients.

Predictive testing for late-onset, incurable neurological diseases has become popular in the United States, Canada, Australia and various European countries, and integrated guidelines and protocols for predictive testing have been established in these countries (Craufurd et al. 1992; Quaid 1992; International Huntington Association (IHA) and World Federation of Neurology (WFN) 1994; Benjamin et al. 1994; Decruyenaere et al. 1995; Mandich et al. 1998; Robins Wahlin et al. 2000). However, little is known about the actual circumstances surrounding this issue in Japan, and only a few studies on psychological aspects of clients seeking predictive testing have been published (Abe and Itoyama 1997; Muto 1998; Yoshida et al. 2002).

In this survey, we investigated the current situations and attitudes of genetic counseling departments in Japan concerning predictive testing for late-onset, incurable neurological diseases.

Procedure

We sent the questionnaire to 125 institutions (university hospitals and national hospitals) in Japan, most of which were highly advanced and specialized medical center hospitals under the supervision of the Ministry of Health, Labor and Welfare, to investigate the actual circumstances concerning predictive testing for late-onset, incurable neurological diseases. The diseases we indicated in the questionnaire included HD, SCA, amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), spinal and bulbar muscular atrophy (SBMA), myotonic dystrophy (DM1) and prion disease. Familial amyloid polyneuropathy (FAP) was also included, although it may no longer be incurable due to liver transplantation. Before submitting the questionnaire, we asked whether or not the hospital had a specific genetic counseling department. If not, the responding persons were instructed not to answer the questions about predictive testing. The questionnaire contained five questions as shown in Table 1. The chief person responsible for the genetic counseling department in each hospital was asked to fill out the questionnaire. In this paper, a clinical geneticist indicates a medical doctor who has the "Japanese Board of Medical Genetics, Clinical Geneticist" certification, and a genetic counselor indicates a non-medical doctor who has the "Japanese Board of Medical Genetics, Certified Genetic Counselor" certification. The certificates are given by the Japan Society of Human Genetics and the Japanese Society for Genetic Counseling.

Results

The questionnaire was returned by 69% ($n = 86$) of the hospitals surveyed. Of the 86 responding hospitals, 63 (73%) had clinics that specialized in genetic counseling and answered the questions concerning predictive testing as shown in Table 1. The distribution of these 63 hospitals is shown in Fig. 1. The remaining 23 hospitals did not have genetic counseling clinics. Of the 63 hospitals with clinics, 46 (73%) had clients with an interest in or a request for predictive testing for late-onset, incurable neurological diseases during the period from April 2004 to March 2006. Of these 46 hospitals, 30 (65%) were attached to national universities or national centers, 13 (28%) were private universities, and 3 (7%) were public universities. The remaining 17 hospitals had no requests or inquiries for predictive genetic testing during this period.

In total, 322 clients from 243 families (excluding FAP) were reported from the 46 responding hospitals (Table 1 Q2). National and public hospitals (total 33) had 194 clients, and private hospitals (total 13) had 128 clients. The maximum number of clients was 80 in a hospital located in the Kanto area. Eleven hospitals had more than 10 clients, and 27 had less than 5 clients during this period. The greatest interest was shown in DM1 (150 clients, 47%), followed by SCA (86 clients, 27%), SBMA (40 clients, 12%) and HD (31 clients, 10%). These clients were distributed nationwide, and more than ten of the hospitals had clients interested in each of the following four diseases (DM1: 31 hospitals; SCA: 30; SBMA: 10; HD: 14). Those interested in prion disease (13 clients, 4%) were much fewer and were found in only three hospitals; a single hospital reported 10 of the 13 clients for prion disease. The interest in ALS was very small (2 clients, <1%) and was found in only two hospitals. There were no clients reported to be interested in AD during this period. There was a significant number of clients (more than 13 clients from 11 families) interested in FAP, but the exact number was not obtained because the answer "a lot" was returned from a single hospital. In contrast to the clients for DM1 or SCA, those for FAP were exclusively concentrated in two hospitals that are close to the former endemic areas for this disease and that have actively conducted liver transplantations for the patients.

The professions of the participants in genetic counseling sessions are summarized in Table 1 (Q3). Clinical geneticists participated in 75–100% of the total counseling sessions in 35 of the hospitals (56%). The contributions of neurologists and nurses were much smaller; they participated in 75–100% of counseling sessions in just 14 and 13 hospitals, respectively. The participation of the other professions (psychiatrists, clinical psychologists, genetic counselors, etc.) was not common. Seventeen

Table 1 Questionnaire and results

Q1	**Have you had clients who visited your hospital with an interest in or a request for predictive genetic testing for late-onset, incurable neurological diseases during the period from April 2004 to March 2006? (total responding = 63)**	
	Yes; (go to Q2-5)	46
	No; (go to Q4-5)	17
Q2	**What kinds of diseases were your clients interested in? Please indicate the numbers of families (clients) for each disease. (total responding = 46)**	
	Myotonic dystrophy	110 (150)
	Spinocerebellar ataxia	68 (86)
	Spinal and bulbar muscular atrophy	32 (40)
	Huntington's disease	26 (31)
	Prion disease	5 (13)
	Amyotrophic lateral sclerosis	2 (2)

Q3 **How much did the following professions contribute to the counseling sessions of such clients? Please select the most appropriate percentage among the following: a: 0%, b: 0-25%, c: 25-50%, d: 50-75%, and e: 75-100% ('0%' indicates no participation and '100%' indicates full-time participation in the counseling sessions.)**

	0%	0-25%	25-50%	50-75%	75-100%
Clinical geneticist (46)#	2	1	4	4	35
Neurologist (41)	10	7	7	3	14
Psychiatrist (33)	26	3	3	0	1
Nurse (40)	13	5	6	3	13
Clinical psychologist (37)	17	6	6	0	8
Genetic counselor (28)	26	0	0	1	1

Q4 **If such clients visit your department in the future, how will you handle them? Please select the most appropriate answer from the following: (total responding = 63)**

You will provide genetic counseling in your own department	36
You will refer the client to another hospital without providing genetic counseling yourself	2
You will make your decision (1 or 2 shown above) depending on the situation and/or disease of interest of the client	23
You will reject the client	0
Not answered	2

Q5 **What do you think are the most important requirements for maintaining genetic counseling for such clients? Please select the three most important requirements from the following and rank them in the order of their importance, (1st) being most important. (total responding = 63)**

	1st	2nd	3rd
Genetic counseling staff	33	12	7
Time and space for clinical practice of genetic counseling	5	21	6
Administrative support for genetic counseling	10	8	8
Genetic testing system	5	9	17
Cooperation with the other hospitals	1	7	12
Understanding and recognition of clinical genetics by the general	8	3	7
Others	1	3	6

#Numbers in parentheses indicate those of responding hospitals

hospitals where nurses and/or clinical psychologists participated in 75–100% of counseling sessions had 201 clients (62%), while 20 hospitals where both of these professions participated in only 0–25% had 60 clients (19%).

Concerning the attitudes toward such clients in the future (Table 1 Q4), 97% of the responding hospitals ($n = 61$) showed positive attitudes toward these clients. Thirty-six hospitals (57%) indicated that they would conduct genetic counseling in their own hospitals. Two (3%) would refer the

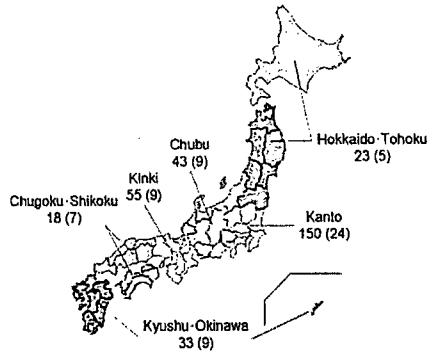


Fig. 1 The distribution and numbers of clients and responding hospitals. The numbers of clients shown here indicate the sum of those with an interest in or request for predictive testing for DM1, SCA, SBMA, HD, prion disease and ALS, but not for FAP. The numbers of responding hospitals in each district are indicated in the parentheses

clients to other hospitals without counseling them themselves, and 23 (37%) would make these decisions based on the individual clients.

We inquired about the needs of persons involved in the genetic counseling of such clients (Table 1 Q5). The most important requirement was an improvement in the counseling staff, followed by improvements in the environment of clinical practice for genetic counseling and the establishment of genetic testing systems. Administrative or financial support for genetic counseling and cooperation with other hospitals in genetic counseling were also strongly needed.

Discussion

The number of genetic counseling departments in hospitals has been increasing in Japan in recent years. This was the first nationwide survey of such departments conducted concerning predictive testing for late-onset, incurable neurological diseases. Inherent in such procedures are several ethical and social issues (Brandt 1994; Evers-Kiebooms and Decruyenaere 1998; Chapman 2002; Taylor 2004), and little information on the present conditions surrounding this issue has been accumulated in Japan (Abe and Itoyama 1997; Muto 1998; Yoshida et al. 2002).

It is obvious that a significant number of the clients with an interest in or a request for predictive testing for late-onset, incurable neurological diseases visited hospitals over last 2 years. The clients were not concentrated in specific districts of Japan, but were distributed nationwide. We should keep in mind that the numbers of clients reported in this paper may represent an underestimate, because not all hospitals and clinics were surveyed. Even in the hospitals we did survey, some respondents indicated that neurology departments in their hospitals have such clients independently from genetic counseling departments. Considering that several delicate and complicated issues are involved with predictive testing for the diseases under study, however, it is likely that the clients have concentrated in the hospitals that have specialized genetic counseling clinics. Thus, we presume that the numbers of clients in this paper roughly reflect the current situations on this matter.

There was a marked difference in the number of clients among hospitals we surveyed. The number of clients may be influenced by several factors. They include the locations of hospitals, the attitudes of counseling staff, the activities of genetic counseling departments and cooperation and specialties of neurologists involved. Academic interest of the neurology department in the university may also be closely associated with the numbers and kinds of diseases of the clients. A representative example for this is seen for FAP, as shown in Results. Of 11 hospitals that had more than 10 clients during the period of interest, nurses and/or clinical psychologists participated 75–100% of the time in genetic counseling sessions in 7 hospitals, whereas, neither of them participated at all in 2 hospitals. As far as predictive testing for late-onset, incurable, neurological diseases, therefore, the participation of non-medical doctor (non-MD) staff in the counseling session does not seem to directly parallel the number of clients. But we can say that requests of clients for predictive testing might be a driving force for having nurses or clinical psychologists involved in genetic counseling sessions. The type of hospitals (national, public or private) had no effect on the number of clients.

It is reasonable that the number of clients interested in DM1 was much greater than those interested in either HD or SCA. The major reason for this is that predictive testing for DM1 often may be linked with reproductive decision making, based on the fact that prenatal genetic testing for DM1 has often been performed in Japan. Many clinical geneticists and neurologists also feel that the matter of predictive testing for DM1 is far different than for HD or SCA because DM1 can be diagnosed by clinical and laboratory findings other than genetic testing, even when a person is not aware of the symptoms. Thus, a significant portion of DM1 clients who were likely to test negative may have already had genetic testing in Japan. This may also be the case for SBMA, because SBMA can be detected by electromyography at the preclinical stage. Thus, careful consideration should be given to whether different predictive testing procedures should be applied to DM1 or SBMA than to HD, SCA or AD.