女性が罹る疾病と病状に関する事実（米国）

神経系疾患

- 米国民のうち**アルツハイマー病**罹患者数は450万人。治療法や予防法が発見できなければ2050年までにはこの数字は1,600万人にまで達するものと予測される。[21] 女性は男性よりもアルツハイマー病を発症しやすい（男性の2～3倍）。80歳を超える女性のうち50%～70%がアルツハイマー病を罹患している。[22]
- 女性の**アルツハイマー病**罹患率は上昇している上、アルツハイマー病患者の介護者の大半は女性である（80%）。介護者は重度のストレスを感じることが多い。[23]
- 2001年、**アルツハイマー病**により男性の2倍強の女性が死亡した。女性3万8,090人、男性1万5,762人である。[1]
- **片頭痛**の発症者は2,800万人に達するものと見られる。そのうち70%は女性。片頭痛を抱えた人の半数（51%以上）以上が、50%以上の生産性低下を報告している。また、66%が家庭における生産性の低下が50%以上に及ぶと報告している。

産科／婦人科系症候

- 18歳～50歳の女性450万人強が少なくとも年1回慢性的**婦人科疾患**を報告している。
- 女性の27%～60%で膣**カンジダ症**が見られる。出産可能年齢女性の75%で膣カンジダ症を罹患し、40%で再発を経験する。[25]
- **クラミジア**は米国で最も多く報告される**細菌性性感染症（STD）**である。1999年、66万件が報告された。だが、現実のクラミジア感染者は年間300万人と推定される。女性の75%および男性の50%が感染し、無症状である。[26]
- 現在、クラミジアは最も脅威的な**STD**sの一つに数えられる。女性の場合、クラミジアの治療を受けないと、骨盤内炎症性疾患（PID）を患う確率は40%にまで達する。そして、PIDを抱えている女性5人につき1人が**不妊**となる。クラミジアは30歳未満の若年女性では多く見られ、性に対して積極的な女性の50%は生涯のうちに一度はクラミジアに罹ることが判っている。[26]
- **クラミジア**に感染した女性が**ヒト免疫不全ウイルス（HIV）**と接触した場合、HIV感染の確率は3～5倍高くなる。
- 15歳～44歳までの女性の64%は、何らかの**避妊法**を試みている。1982年には56%であった。最も普及している避妊法は女性の不妊手術（18%）で、次いで経口避妊薬（17%）である。[3]
- 出産可能年齢女性が**子宮内膜症**に罹る割合は20%にまで達し、慢性的な痛みや不妊を引き起こす恐れがある。重度な生理痛を抱えている女性では、子宮内膜症の発症率は25%～35%である。[3]
- 全国的に12歳以上の4,500万人、または全青年および成人の5人に1人が、**陰部ヘルペス（単純ヘルペスウイルス2型）**に感染している。陰部ヘルペスは男性よりも女性に多く見られ、罹患率は男性では5人に1人、女性では4人に1人。[26]
- 性に対して積極的な米国成人のおよそ1%に、**陰部疣贅**が見られる。感染率は、大学の健康センターで治療を受けた女子学生の1.5%から性病専門家クリニックの13%までと推測される。[26]
- 現在、およそ2,000万人が**ヒト乳頭腫ウイルス（HPV）**に感染している。米国疾病対策予防センター（CDC）の調査によると、HPV16型の有疾率は、女性は男性の2倍以上である。全人種の女性のHPV16型の有疾病率は17.9%、男性8%。20歳～29歳のアフリカ系米国人では36%とHPV16型の有疾病率が最も高い。[26]
- **HPV16型**は**子宮頸がん**の50%強および高度異形成の原因である。子宮頸がんの80%は、18型、31型、45型と共にHPV16型による感染である。[26]
- 約610万人が**不妊症**を患っている。これは出産年齢にある女性の約10%に相当する。[27] 不妊の全症例のうち、約20%は男性に起因し、38%は女性に起因する。15%は男女いずれに原因があるか特定できない。不妊症の要因として排卵障害（18%）および**子宮内膜症**（9%）が挙げられる。[28] 不妊症カップルの4組に1組は**性感染症**の結果、不妊となっている。[3]
- 現在、毎年200万人弱の女性が50歳に達している。2000年には50歳以上の米国人女性（約4,200万人）のうち大半がそれまでに更年期障害を経験している。米国女性3人に1人が罹っている計算である。[29] **更年期障害**を経験している女性の約75%が、一過性熱感を経験する。[30]
- 女性の約43%が**性機能障害**を患っている。障害の大半は**月経閉期**後に発生することが判っている。ホルモン産生量が低下し、尊管に関する症状がより多く見られる。[31]

女性が罹る疾病と病状に関する事実（米国）

産科／婦人科系症候（続き）

- 現在、6,500万人超が不治の**性感染症**を抱えている。25種の疾病（**HIV/AIDS、陰部疱疹、クラミジア、HPV**など）は主として性行為により広がる。毎年、新たに1,500万人のSTDが発生する。[26]
- 極度の早産および死産の30%～40%は**性感染症**または細菌性膣炎が原因と見られる。[26]
- 1994年において、主要な**性感染症（単純疱疹（ヘルペス）、ヒト乳頭腫ウイルス、子宮頸がん、性感染によるHIV**など）およびその合併症による直接的および間接的な年間想定総費用は170億ドルである。[28]
- 女性の10%～20%が子宮筋腫に罹っている。主として出産年齢の女性が発症する。アフリカ系米国人女性では、白人女性の2～3倍の高さで子宮筋腫と診断される。[3]
- 年間行われる子宮摘出60万件の半数以上は、**子宮内膜腫**または**子宮筋腫**によるものである。子宮筋腫は子宮摘出の主要因であり、年間15万件～17万5,000件の手術が行われている。[3]

精神障害

- 毎年、**抑鬱性障害**に罹る女性の数（1,240万人）は男性（640万人）のほぼ2倍。女性の罹患率（12.0%）は男性（6.6%）の2倍。毎年、**大うつ病性障害**に罹る女性の数（670万人）は男性320万人のほぼ2倍。女性の罹患率（6.5%）は男性（3.3%）2倍。[3]
- **うつ病性障害**は自殺の危険性を高める。自殺による死亡者は男性では女性の4倍となっているが、自殺の試みは女性では男性の3～4倍となっている。世界の5歳以上の女性において、自傷行為による怪我は病気の苦しみの主要因（自殺など）のうち上位9位までを占める。[32]
- 成人400万人が**全般性不安障害（GAD）**を患っており、女性は男性の約2倍。**不安障害（パニック障害**、強迫性障害、心的外傷後ストレス障害、恐怖症、全般性不安障害など）は、年間でみると18歳～54歳の米国人の13.3%が経験している。または、これらの年齢層において成人1,910万人が不安障害を抱えている。強迫性障害と社会恐怖症を除きすべてのカテゴリーで女性は男性を上回っている。この2つでは、男女ともほぼ同数がこれらの障害を抱えている。[32]
- 日常の活動に支障をきたすほど深刻な月経前症候群（PMS）を抱える女性の数は全体の40%に達する。女性の約7%は支障をきたすようなPMSを抱えており、その度合いの深刻さにより**「月経前不機嫌性障害」**と名称される。[3]

敗血症

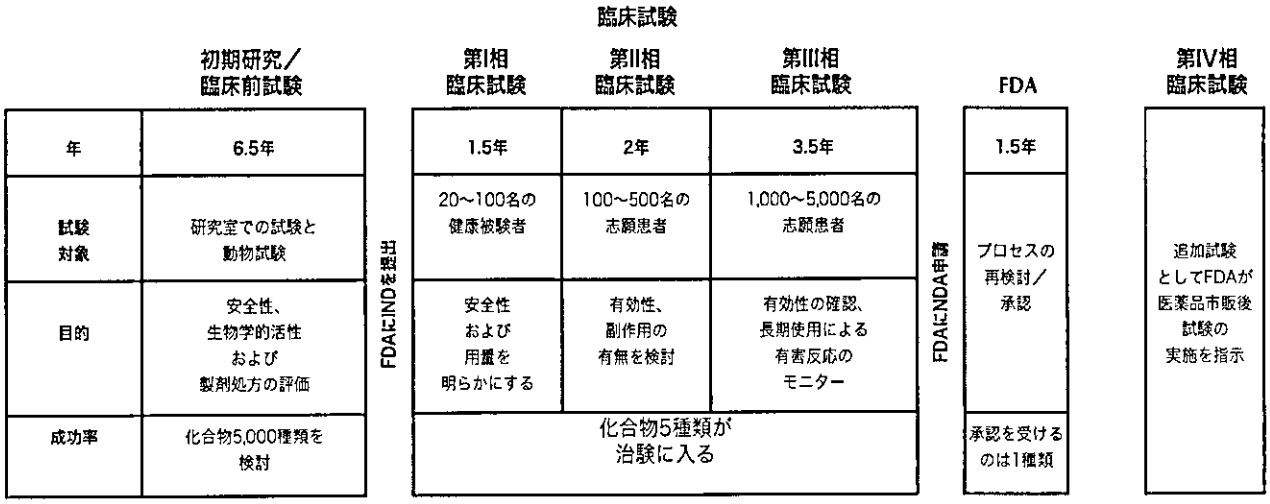
- 毎年、750万人が深刻な**敗血症**を発症し、そのうち20万人強が命を落としている。33入院患者100人につき2人が敗血症に罹る。[34]
- 2001年では、敗血症による死亡者は男性1万4,307人、女性1万7,931人。[1]

出所：

1. 全米保健統計センター（www.cdc.gov/nchs）
2. 女性健康研究協会（www.women-health.org）
3. 米国保健社会福祉省の女性保健局のープロジェクトである全米女性の健康情報センター（www.4women.gov）
4. 米国骨粗鬆症財団（www.nof.org）
5. 米国NIH関節炎・筋骨格皮膚疾患研究所（www.niams.nih.gov）
6. 米国自己免疫疾患協会（www.aarda.org）
7. 全米繊維筋痛パートナーシップ（fmpartnership.org）
8. 米国狼瘡財団（www.lupus.org）
9. 米国乾癬協会（www.psoriasis.org）
10. 米国癌協会（www.cancer.org）
11. 米国糖尿病協会（www.diabetesjournals.org）
12. 米国糖尿病協会（www.diabetes.org）
13. アメリカン・ジャーナル・オブ・オフサルモロジー（www.ajo.com）
14. アメリカ盲目予防協会（www.preventblindness.org）
15. 米心臓病協会（www.americanheart.org）
16. 米国泌尿器科学会（www.urologyhealth.org）
17. 米国排泄障害協会（www.nafc.org）
18. 全米泌尿器科協会（www.afud.org）
19. 国立糖尿病・消化器系・腎臓病疾患研究所（www.niddk.nih.gov）
20. 米国肺協会（www.lungusa.org）
21. アメリカ・アルツハイマー病協会（www.alz.org）
22. ワシントン大学（www.depts.washington.edu）
23. チャリティ・ワイヤ（www.charitywire.com）
24. 米国頭痛協会（www.headaches.org）
25. カリフォルニア大学サンフランシスコ校HIVインフォメーションセンター（www.hivinsite.ucsf.edu）
26. ガバメント・ガイド（www.governmentguide.com）
27. 米国生殖医学会（www.asrm.org）
28. UpToDate_患者情報（www.patients.uptodate.com）
29. 米国国立老化研究所（www.nia.nih.gov）
30. ママズヘルス（www.mamashealth.com）
31. ユーロロジー・チャネル（www.urologychannel.com）
32. 国立精神衛生研究所（www.nimh.nih.gov）
33. コンセプトコミュニケーションズ（消費者用健康報告）（www.stayinginshape.com）
34. ドクタークープ（www.drkoop.com）

医薬品の開発と承認プロセス

新薬が米国の患者の治療に利用されるまでには平均して10年～15年の期間が必要である。また、臨床前試験に入る5,000種類の化合物のわずか5種類の化合物だけがヒト試験に進む。これらヒト試験を行う化合物5種類のうち、承認されるのは1種類だけである。



新薬の開発と承認過程

米国における新薬承認制度は、おそらく世界で最も厳格なものである。タフツ大学新薬研究センター（Tufts Center for the Study of Drug Development）によれば、1994年から1998年に承認された新薬が、研究室から米国人患者の手に届くまでには平均して10～15年を要しているとのことである。前臨床試験からヒトを対象にした治験に入るのは、5,000種類もの化合物のうち5種類だけである。さらに、この5種類のうち1種のみに市販の許可が与えられる。

タフツ大学新薬研究センターの2001年11月の報告によれば、研究室から米国人患者に新薬が提供されるまでには、新薬1種類あたり8億200万ドルを要するとのことだ。

研究室で新化合物が同定されると、新薬は次のような経過で開発される。

前臨床試験：製薬会社では実験室での試験および動物実験を実施し、対象となる疾病に対する化合物の生物活性を立証し、次にその化合物について安全性を検討する。

新薬臨床試験開始届（IND）：前臨床試験の終了後、製薬会社では米国食品医薬品局（FDA）に対して新薬臨床試験開始届（IND）を申請し、ヒトを対象に目的とする薬剤の試験を開始する。30日以内にFDAから不承認がなければ、INDは有効となる。このINDの段階では、これまでの実験結果、新薬の試験方法、研究機関、実施者、また、化合物の化学構造、体内での考えられうる作用機序、動物実験での毒性の有無、およびその化合物の製造方法について明らかにする。全ての臨床試験は、治験が実施される施設内治験審査委員会（IRB）により、審査を受け承認されなければならない。臨床試験に関する進捗状況報告書をFDAおよびIRBに、少なくとも年に一度は提出するように求められる。

臨床試験、第I相：この段階の試験には、約20人から100人の健常者が選ばれる。この試験では安全投与量範囲をはじめ、薬剤の安全性の概要が試験される。また、薬剤の吸収、分布、代謝、排泄に関する機序並びに薬効の持続期間を試験する。

臨床試験、第II相：この段階では、約100人から500人の被験者（対象とする疾病に罹患している患者）を使った対象試験により、薬剤の有効性を評価する。

臨床試験、第III相：この段階の試験では、病院および医療機関において、通常1,000人から5,000人の患者を対象とする。治験担当医師により厳密に患者のモニターが行われ、薬効の確認と有害事象の同定が実施される。

新薬承認申請（NDA）：臨床試験の第III相までの全試験の終了後、全ての治験データを分析し、データから安全性と有効性のいずれもが立証されれば、FDAに対してNDAを申請する。このNDAには、その製薬会社が収集した全ての科学情報が含まれる。通常、NDA申請書類は10万ページにも及ぶ。2003年度に承認された新規分子物21種の平均NDA審査期間は、16.9カ月であった。

承認：NDAがFDAにより承認されれば、医師により新薬が提供され、処方が可能となる。製薬会社は、副作用に関する全ての症例および適切な品質管理記録を含めた報告書を引き続き定期的にFDAに提出しなければならない。薬剤によっては、長期効果を検討するために追加試験（第IV相）がFDAによって求められることもある。

安全で効果的な新薬を発見・開発することは、長く険しく費用のかかる過程である。PhRMA会員企業では、2003年度、研究開発費に332億ドル以上を投資した。

Medicines in Development for Women is presented by PhRMA in cooperation with the following organizations:

Allergy & Asthma Network Mothers of Asthmatics
American Academy of Allergy, Asthma and Immunology
American Academy of Physician Assistants
American Association for Cancer Research
American Autoimmune Related Diseases Association
American Cancer Society
American College of Allergy, Asthma and Immunology
American College of Obstetricians and Gynecologists
American Foundation for Urologic Disease
American Infertility Association
American Menopause Foundation
American Nurses Association
American Society for Bone and Mineral Research
American Society for Reproductive Medicine
American Society of Clinical Psychopharmacology
Anxiety Disorders Association of America
Association of Community Cancer Centers
Association of Women's Health, Obstetric, and Neonatal Nurses
Asthma and Allergy Foundation of America
Burger King Cancer Caring Center
Cancer Action
Cancer Care
Cancer Research and Prevention Foundation
Cancer Research Institute
Center for Women Policy Studies
Chemotherapy Foundation
Depressive and Bipolar Support Alliance
Endometriosis Association
Fertility Research Foundation
Global Alliance for Women's Health
Infectious Diseases Society of America
Interamerican College of Physicians & Surgeons
MAGNUM, The National Migraine Association
Multiple Sclerosis Foundation
National Alliance for Hispanic Health
National Alliance for the Mentally Ill
National Alliance of Breast Cancer Organizations
National Association for Continence
National Association of Anorexia Nervosa and Associated Disorders (ANAD)
National Black Nurses Association
National Cancer Institute
National Chronic Fatigue Syndrome and Fibromyalgia Association
National Council of Negro Women
National Eating Disorders Association
National Foundation for Depressive Illness
National Foundation for Infectious Diseases
National Headache Foundation
National Institute of Arthritis and Musculoskeletal and Skin Diseases
National Medical Association
National Mental Health Association
National Multiple Sclerosis Society
National Sarcoidosis Resource Center
National Women's Health Resource Center
North American Menopause Society
Office of Research on Women's Health, National Institutes of Health
Oncology Nursing Society
Planned Parenthood Federation of America
Prevent Blindness America
Scleroderma Research Foundation
SHARE: Self-help for Women with Breast Cancer or Ovarian Cancer
Simon Foundation for Continence
Sjogren's Syndrome Foundation
Society for Women's Health Research
The V Foundation for Cancer Research
Y-ME National Breast Cancer Organization

Being listed in this report in no way implies that the above-mentioned organizations endorse or recommend the use of any of the products in development contained in this publication. For further information, patients should consult their physicians or health care providers.

PhRMA
New Medicines. New Hope.

Pharmaceutical Research and Manufacturers of America
1100 Fifteenth Street, NW
Washington, DC 20005

www.phrma.org

1/04

＜参考資料３＞

ヒトゲノム研究に関する国際的動向を踏まえた今後の方向性、評価のあり方に関する研究

参考資料

1．Allen D.Roses:Nature　Reviews,1,541-549,2002.

2．Allen D.Roses，et.al: Drug discovery today,10,3,2005

3．Allen D.Roses:Nature　Reviews,5,645-656,2004

4．Lawrencs J．Lesko et.al:Nature　Reviews,1,541-549,2002.

5．Position Paper on Terminology in Pharmacogenetics／EMEA/CPMP/307/01

6．DC Anderson、et.al；2,284-292,2004.

7．Discover Genetics(bookmark)，グラクソスミスクライン社

8．Genetics eLearning @ your computer(card)，グラクソスミスクライン社

9．Studying the Genetics of Medicine Response(leaflet)，グラクソスミスクライン社

10．Finding Clues to the Right Medicine for the Right Patient(brochure)，グラクソスミスクライン社

11．ファーマコジェネティクスについて（小冊子），グラクソスミスクライン社

12．S.Gretarsdottir，et.al.: Nature Genetics, 35, 2, 131-138, 2003.

13．H. Stefansson，et.al.: Am. J. Hum. Genet. 71, 877-892, 2002.

14．A. Helgadotir, et.al.: Nature Genetics, 36, 3, 233-239, 2004.

15．S. Jonsson, et.al.: JAMA, 292,24,2997-2983,2004.

16．A. Kong, et.al.: Nature Genetics, 36, 11, 1203-1206, 2004

17．L. T. Amundadottir, et.al.: PLoS Medicine, 1, 3, e65, 0001-0008, 2004.

18．A. Kong, et.al.: Nature Genetics, 31, 241-247,2002.

19．H. Stefansson，et.al.: Nature Genetics Advance Online Publication,16 January 2005; doi:10.1038/ng1508

GENOME-BASED PHARMACOGENETICS AND THE PHARMACEUTICAL INDUSTRY

Allen D. Roses

Pharmacogenetic capabilities have changed markedly since The SNP Consortium made a dense single-nucleotide polymorphism (SNP) map freely available in 2001. For more than 40 years, pharmacokinetics and pharmacodynamics of drug-metabolizing molecules were the focus of practical applications. Today, it is possible to use SNP-mapping technologies to create a genetic profile of each individual that can be used to identify patterns of susceptibility genes for common diseases as well as genetic risk/efficacy factors that are related to the effects of drugs.

SINGLE-NUCLEOTIDE POLYMORPHISM (SNP). A specific location in a DNA sequence at which different people have different DNA bases. This can change the protein sequence, leading to disease (for example, sickle-cell disease), or have no known consequences.

Genetics Research, GlaxoSmithKline, Five Moore Drive, Research Triangle Park, North Carolina 27709, USA.
e-mail: adr69412@gsk.com
doi:10.1038/nrd840

Pharmacogenetics is no longer limited to providing molecular descriptions of variants of particular genes and how they might affect pharmacokinetics and pharmacodynamics[1–5]. Rapid advances in SINGLE-NUCLEOTIDE POLYMORPHISM (SNP) mapping, high-throughput genotyping, statistical analyses and bioinformatic processing can now be applied to characterize the individual patients who suffer an adverse event (AE) to a drug, or those for whom a drug shows efficacy[6–8]. It is now possible not only to determine groups that share particular phenotypes (for example, efficacy or AE), but also to identify and classify the individuals at risk[9,10]. As a result, sensitive tests to predict who would have a high probability to develop a phenotype as a consequence of exposure to a drug can be studied rapidly in the context of a clinical trial, or as the follow-up surveillance of approved medicines[11].

The unpredictable occurrence of adverse drug reactions is a risk in the development and subsequent clinical use of virtually all medicines — recent and past[12]. Preclinical safety procedures, in particular toxicity testing in animals, are required in drug development. Such data are reviewed carefully by regulators of the industry. Toxicity accounts for the largest proportion of attrition of candidate molecules in drug discovery. However, in addition to the failure of drug candidates, unpredictable AEs can occur during drug development in clinical trials and post-marketing, which also lead to product attrition[13]. Attrition can occur after product launch and marketing of medicines, when rarer AEs that could not be observed or appreciated during clinical development can occur, at a particularly high cost[14,15]. The unpredictable nature of drug reactions must always be weighed against the need to treat the disease and to achieve acceptable risk/benefit ratios, which is the main activity of regulating bodies worldwide. This review includes an example of SNP MAPPING of common disease genes, an examination of high-throughput CANDIDATE-GENE ASSOCIATIONS with AEs and efficacy, and the current status of whole-genome SNP mapping with completed studies and work in progress. Moreover, a standard SNP genetic-profiling system can be foreseen for targeting effective medicines to responsive patients, as well as predicting which individuals carry genetic risk factors for AEs. A systematic framework for efficacy and AEs would provide for more efficient drug regulation[16].

SNP mapping of common disease genes

In April 1999, the concept of identifying 300,000 SNPs and mapping 150,000 of them was a challenging task. Ten pharmaceutical companies and The Wellcome Trust came together to form The SNP Consortium (TSC), a not-for-profit company with a goal to contract academic sequencing laboratories to identify and map the SNPs[17]. The aggressive private sequencing initiative stimulated further incentive, which led to the rapid

progress of the public sequencing effort. As a result, it became possible to adjust strategy and technologies to introduce more than 2 million mapped SNPs into the public sector, providing accelerated opportunities for common, standardized SNP maps to be developed and applied for pharmacogenetic purposes. TSC was an extraordinary success as a public–private consortium, and far exceeded its objectives, both on schedule and under budget. Other projects, such as the creation of a SNP-LINKAGE MAPPING PANEL and quality-control experiments, have continued past the March 2001 completion of the initial mapping objectives.

With the expectation of new information from The Human Genome Project, it became possible to imagine a broader application of genetics to medicine[18]. It was predicted that sequentially ordered, high-density SNP maps could be used in new ways to identify inherited profiles that were statistically associated with drug efficacy or AEs. The high-density SNP-mapping hypothesis was tested using the identification of susceptibility genes for common diseases. The first experiment to test the hypothesis was designed to confirm the previous discovery of the apolipoprotein E (APOE) ε4 allele (*APOE4*), the susceptibility gene variant that is responsible for common, late-onset Alzheimer's disease (AD)[19]. The *APOE4* variant of the *APOE* gene had been widely confirmed in several ethnic groups to increase the risk and lower the mean age of onset of AD[20–24]. A high-density SNP map with an average SNP distance of 15 kilobases (kb) was constructed across a 4-megabase (Mb) region that encompassed the *APOE* gene on chromosome 19 (REF. 19). A simple experimental question was asked. Could the *APOE4* polymorphism be identified as the risk factor for AD within a much larger fragment of chromosome 19 if it was not known to be there? A small region of highly significant LINKAGE DISEQUILIBRIUM (LD) was clearly identified, which confirmed the already established chromosomal region and provided proof of concept for the identification of susceptibility variants using high-density SNP mapping[25,26]. The LD region contained only two genes, *APOE* and *APOC1*. This technique has since been used to identify susceptibility genes within small regions of linkage for several other diseases, including migraine with aura, psoriasis and Crohn's disease[27–29].

The ability to identify and study susceptibility genes for common diseases has now expanded considerably[8]. Using high-density SNP mapping as a tool, it is technically possible and feasible to study several genetic factors for common diseases simultaneously, a marked improvement on the previous historic approach of searching for highly PENETRANT, single point mutations that 'cause' uncommon or rare diseases. Several independent gene variants across the genome can each contribute, by either increasing or decreasing the susceptibility to develop a particular disease. A similar genetic basis might then be expected for susceptibility to variable clinical responses to external chemical toxins, including to those chemicals with uncommon AEs that we call 'drugs'. A clearly defined environmental factor and a subsequent unwanted phenotype is the ultimate environment–genetic interaction. This, too, is a common experience. Some people react negatively to poison ivy, whereas others do not. In fact, similar to observations in large, common-disease susceptibility studies, members of the same family are more likely to manifest a similar pattern of differential response to a toxic agent[30,31].

After the work on sickle-cell disease haemoglobinopathy more than 60 years ago that established gene mutations that are associated with disease, it has been easy to track inherited (usually rare) disease loci through families with several affected individuals, particularly those inherited as AUTOSOMAL-RECESSIVE, or highly penetrant AUTOSOMAL-DOMINANT or X-linked traits. With highly penetrant genetic diseases, the polymorphisms can be identified from tissue samples, such as blood obtained for genetic mapping. A genetically based drug reaction is more difficult to map as a genetic phenotype because only selected members of families might be exposed to the particular drug during their lifetime. Ascertainment of family members who might be put at risk by challenge with a drug that provides no therapeutic benefit is not ethically acceptable.

Having a SNP map is a starting point. Selecting informative SNPs that are spaced periodically along the genome, validating primer sets, and developing technologies to measure rapidly and reliably many thousands of SNPs, have followed on from the efforts of TSC[32,33]. GlaxoSmithKline has initiated a collaboration with several biotechnology companies, including Orchid BioSciences, Sequenom and Illumina, to produce genome-wide SNP-scanning marker sets. The most serious obstacle to rapid development is establishing an economic basis for the industrialization and increased expense of each project. Academic laboratories that are searching for disease-susceptibility genes do not provide the volume necessary to sustain commercially viable SNP-map development. Applied pharmacogenetics provided the rationale for the creation of TSC, and now provides the commercial promise of SNP mapping for the development of safer and more efficient medicines, and strategies to respond to undesirable drug effects by identifying patients who are most at risk of experiencing an AE.

Pharmacogenetic profiles versus diagnostic profiles. Diagnostic tests of highly penetrant diseases can be extremely accurate and represent as little as one base change (mutation) in the genome. Individual patient diagnosis is routinely possible. Many common diseases are now thought to contain several susceptibility variants (polymorphisms) that are dispersed across the genome, which together can affect the onset, severity or susceptibility to disease. For AEs, in a similar fashion, a panel of independently assorting variant loci, each of which contributes in some manner to the phenotypic reaction, could represent susceptibility or genetic risk factors that underlie a reaction to a particular drug. If this hypothesis is accepted as correct, then the most difficult experimental challenge would be how to recognize a diagnostic pattern of variants in as few patients as possible who experience the AE.

SNP MAPPING
A linear map of SNPs across the genome allows genetic traits to be localized by statistical association to the specific region of the genome that is marked by the SNP or several SNPs nearby. In the cases of complex diseases or responses to medicine, multiple SNP markers in several specific regions of the genome act as markers that are associated with the phenotype (disease or drug response).

CANDIDATE-GENE ASSOCIATIONS
A candidate gene is frequently indicated by previous hypotheses or genetic-linkage studies. Gene variants can cause or enhance susceptibility to a disease or drug-response phenotype.

SNP-LINKAGE MAPPING PANEL
The collection of SNPs that defines the linear map of the genome that is being studied. To localize the *APOE* gene, these SNPs would be located on chromosome 19. For HLA-B57, the SNPs are located on chromosome 6. To find other unknown regions, a SNP mapping panel that covers the whole genome can be used.

LINKAGE DISEQUILIBRIUM
(LD). Defines a region of allele (or inherited-variation) sharing along a relatively small length of the genome. If there is an association with one marker, then the likelihood of association with another marker within that region is increased.

PENETRANT GENETIC DISEASES
Describes the expression of the phenotype that results from the inheritance of disease-associated or phenotype-associated alleles. Some diseases have relatively highly penetrant mutant alleles, such as Huntington's disease. Penetrance of other traits, such as eye colour or late-onset Alzheimer's disease, might be due to combinations of less-penetrant polymorphic alleles found at several locations along the genome.

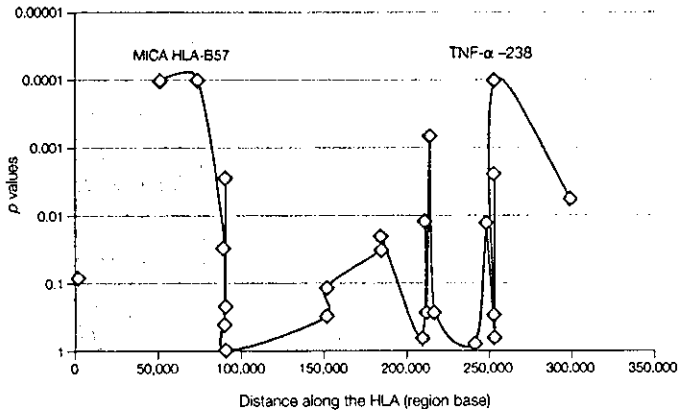


Figure 1 | **Genetic markers in the HLA-B region that are associated with hypersensitivity to abacavir.** The TNF-α -238 and the HLA-B57 polymorphisms were both associated with susceptibility to hypersensitivity. These markers are found to be in close proximity to each other on the same chromosome. Several other intervening markers had varying degrees of positive association. The TNF-α -238 and the HLA-B57 polymorphisms are not within the same linkage disequilibrium (LD) blocks, but show by overlapping patient sets that detection for pharmacogenetics can exceed the LD blocks, which indicates that fewer than 200,000 single nucleotide polymorphisms (SNPs) will be necessary to define SNP profiles. HLA, major histocompatibility complex (MHC) locus; MICA, MHC class I chain-related gene A; TNF-α, tumour-necrosis factor-α.

If ascertainment of hundreds of AE patients and controls were necessary, practical usefulness in clinical situations would be minimized.

Once again, it is useful to examine routinely the blood-typing model of individual identification in a relatively accurate manner. Indeed, the polymorphic variation of each blood-type antigen is far less informative than that available within LD blocks across the genome. So, in transfusion medicine with comparatively little information, it is routinely possible to recognize individual patient variants that put a patient at risk. It is also anticipated that relatively few patients will be needed to recognize and confirm a SNP PRINT™ for an AE. For example, the forensic identification of suspects in criminal investigations represents an accurate identification of individuals on the basis of a handful of highly polymorphic markers. The use of LD markers for identifying individuals who are at risk for AEs would be similar, except that the loci used for diagnosis would have to be identified for each drug, rather than a standard few loci that can distinguish one individual from another. The existence of forensic examples provides a familiar rationale for the hypothesis that AE-susceptible individuals can be characterized easily with great accuracy once a standard LD diagnostic panel is agreed.

Successful post-marketing example: abacavir. 'Idiosyncratic', 'idiopathic' or 'random' are words that are descriptive but not explanatory. Idiosyncratic AEs to specific drugs, or combinations of several drugs, are the ultimate highly specific environmental–genetic interactions. A patient receives a prescription for a new drug for an appropriate clinical indication, but could react with a stereotyped AE that might be observed in a small percentage of other patients who receive the same drug. Without experiencing the drug (environment), there is no adverse phenotype. Depending on the frequency and severity of the AEs, drug development molecules frequently fail at the clinical-trial phase. Safer molecules, which include those with an AE rate so low that AEs do not occur during clinical trials (frequently limited to less than 10,000 subjects) are approved for the market, and then might be used for the first time by hundreds of thousands of patients. Relatively rare, although sometimes severe, AEs might be observed for the first time when the drug is widely used. The balance between risk and benefit is an unpredictable factor that concerns individuals who are taking medicines, public-interest groups, regulators and the pharmaceutical industry. Evaluating the benefit of the drug, the severity of the illness and the nature of the AE determines whether an effective medicine should remain available for the vast majority of patients in whom no AE occurs.

Hypersensitivity reaction (HSR) to abacavir, a reverse-transcriptase inhibitor that is used to treat human immunodeficiency virus (HIV), occurs as idiosyncratic AEs and is characterized by fever, rash, gastrointestinal symptoms, malaise and respiratory symptoms. HSR is a well-described clinical syndrome that affects ~4% of patients who are treated[34]. It starts rather consistently during the initial weeks of exposure to abacavir, with more than 90% of patients presenting with symptoms within six weeks, and a mean time to onset of 11 days. The incidence of hypersensitivity among 30,595 patients in clinical trials and expanded-access programmes was 4.3% (REF. 35). The symptoms usually disappear promptly when the drug is stopped. Re-initiation of abacavir in those patients who have previously shown hypersensitivity can result in a rapid return of severe hypersensitivity symptoms, often within hours. The re-challenge reaction is usually more severe than the initial reaction, and can be associated with hypotension, renal impairment and death. It is therefore contraindicated to re-initiate abacavir therapy for patients who have shown hypersensitivity. Current clinical follow-up surveillance programmes have proved to be quite effective in recognizing HSR early as well as avoiding re-challenge, but a sensitive predictive test for HSR would add to the safety profile of the medicine.

Abacavir was selected for the initial pharmacogenetic proof-of-concept experiment for whole-genome SNP mapping for purely practical reasons. HSR has a stereotypical clinical phenotype, and occurs frequently enough to allow identification of sufficient numbers of HSR patients and matched controls for research purposes. The goal of the experiment would be to identify a pattern of chromosomal regions defined by SNPs that are associated with the AE in a reproducible, small number of patients (see below). To test empirically and confirm those specific aims, a larger number of AE patients needed to be identified and collected. The experimental strategy was to collect clinical data and DNA samples from sufficient numbers of well-characterized patients and matched abacavir-treated controls for both test and

AUTOSOMAL RECESSIVE
A mode of inheritance that requires that the mutation of a single gene is present on both paternally and maternally derived alleles for the clinical phenotype to be expressed.

AUTOSOMAL DOMINANT
A mode of inheritance that requires that the mutation of a single gene is present on either of the paternally and maternally derived alleles for the clinical phenotype to be expressed.

SNP PRINT™
Describes the pattern of inheriting SNP variations from a panel of SNPs. In the case of SNP mapping along the whole genome, it is analogous to the detection of highly private sets of traits, such as those that determine fingerprints, which make individuals unique. In the case of pharmacogenetics, it is the specific set of measured SNPs that define a drug response, and which can therefore be abstracted from the whole map to develop a more simple medicine-response profile.

conformation series in a reasonable timeline, in this case 2 years. HSR seems to be an immunological reaction[36]. Susceptibility factors were unknown. Unlike common side effects from medications, HSR to abacavir seems to be relatively dose independent. In a clinical research study that used twice the current recommended dose of abacavir, there was no increase in the rate of subjects who developed this reaction[37]. As the risk of HSR occurs in the first six weeks of exposure to abacavir, it does not seem to be related to cumulative exposure to the drug, as is the case with side effects that are associated with other antiretroviral agents, such as mitochondrial toxicity or lipodystrophy[38].

Abacavir: HSR candidate-gene results. Before a whole-genome SNP map was constructed from the data that were generated by TSC and a series of validated primer sets had been chosen and tested, a preliminary candidate-gene experiment was done. An arbitrary list of potential candidate genes for an immunological reaction was determined. The initial candidate-gene analysis tested for more than 100 polymorphisms from candidate genes that were identified from the literature or proposed within GlaxoSmithKline. In addition, the SNP analyses allowed direct comparisons of genotyping technologies that were being developed in our laboratories, and the evaluation of several biotechnology companies and academic laboratories for potential platform technologies. All polymorphic markers were genotyped initially using the 5′ nuclease assay (Taqman, Applied Biosystems) as a standard of accuracy for alternative, faster and less expensive methods of genotyping. In addition, major histocompatibility complex (HLA) typing for loci A, B and DR was performed[39].

The initial association data that were derived from the candidate genes, which were selected from the list of genes that were thought to be associated with immunological reactions, showed an unexpected and startling result[39]. Polymorphisms from two of the candidate genes that were chosen for study before the sequenced genome was available were found to be located very close to each other on chromosome 6, and were both highly associated with HSR to abacavir (FIG. 1). Results were completed in a series of 85 HSR patients and 115 matched abacavir-treated controls. The striking part of the results indicated that the two genes that were located very close to each other on the same chromosome were not in complete LD, as was commonly surmised when the sequenced genome became available. The −238 polymorphism of the tumour-necrosis factor-α (TNF-α) gene and the HLA-B57 polymorphism were both highly associated in overlapping groups of HSR patients. FIGURE 1 presents the association data. In the case of HLA-B57, the population frequency in Caucasians is known to be ~5%, but was greater than 40% in HSR patients. Similar data were obtained for TNF-α −238, without adding extra patients, which indicates that approximate LD can be detected in heterogeneous populations over much greater distances than precise LD blocks. In fact, these association data were independently confirmed in a smaller, more localized series of patients[40]. This point has important practical consequences for LD SNP mapping. It has been assumed in theoretical models that associated polymorphic variants needed to be within limited regions of LD[32]. However, these data show that significant associations can extend well beyond LD blocks. This means that informative SNPs can be located much further apart than the number estimated by calculations based on exact LD blocks, which indicates that far less than 200,000 SNPs will be needed for whole-genome scanning to detect chromosomal regions with multiple phenotype-associated variations.

The clinical utility of the candidate-gene AE data during the interim year before the whole-genome mapping is scheduled for completion is still being debated by clinicians who care for HIV patients. Between the two published data sets, a diagnostic accuracy between 30–70% of white males might be tested successfully[39,40]. All cases would still require the successful clinical risk-management programme to be continued. Therefore, with the expectation that SNP-defined MEDICINE-RESPONSE PROFILES will lead to much greater accuracy across ethnic groups, use of the current data as a diagnostic might be premature. More clinical research by experienced HIV physicians will be necessary to determine whether these data are considered to be clinically useful. At present, negative data would not change the current risk-management procedures in common use by clinicians. The HSR candidate-gene experiment provided the first independently confirmed example that there is an LD region for an adverse reaction that can be identified with multiple polymorphic markers within the practical spacing limitations (every 15 kb) of the SNP genome-mapping screen. Although the arbitrary candidate-gene-variation panel that was chosen was not expected to show any positive associations, the data from these experiments provide strong support for the hypothesis that other regions of LD that contain susceptibility-gene variants that are associated with HSR will be identified in a whole-genome screen. These independently assorting regional variants could well provide a highly sensitive and selective prediction of patients who are susceptible to HSR from abacavir therapy.

Pharmacogenetic success during a clinical trial: example. Tranilast is a drug that was in Phase III clinical trials for the treatment of restenosis after percutaneous transluminal coronary revascularization. The trial involved 11,500 patients, a comparatively large number for a Phase III clinical trial. During the course of the double-blind trial, ~10% of patients in the trial developed HYPERBILIRUBINAEMIA (FIG. 2). In these patients, bilirubin levels were increased significantly, with only variably mild elevations of other liver-function enzymes. No patients progressed to hepatic failure.

The experiment was clear: could the genetic susceptibility to Tranilast that causes hyperbilirubinaemia be identified, and a test be developed rapidly that would predict which patients were at risk? The experiment had to be carried out during the limited, defined time frame of the continuing clinical trial, so that regulators could

MEDICINE-RESPONSE PROFILE (MRP). A test or set of tests that indicates the likely response of a patient to a medicine. In the context of developing a test from a SNP Print™, the test might be the selection of SNPs from the genome scan that are associated with the phenotypic response; for example, adverse event.

HYPERBILIRUBINAEMIA Bilirubin is a metabolite of cholesterol, and its elevation above normal serum levels is called hyperbilirubinaemia. This state can be caused by several liver- and gall-bladder-related problems.

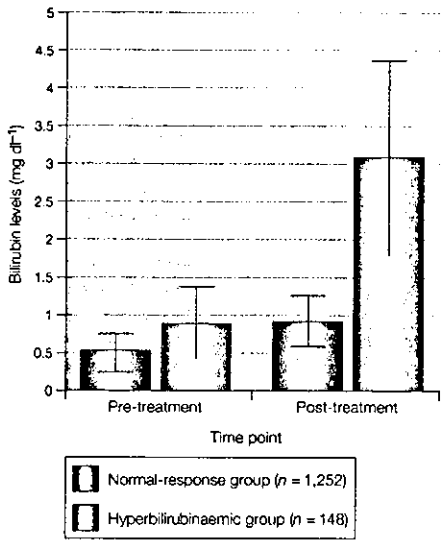


Figure 2 | **Bilirubin levels pre- and post-treatment in 1,400 individuals from a Phase III clinical trial of Tranilast.** Out of the individuals used in the trial, 1,252 show no evidence of elevated bilirubin levels post-treatment (89.4%), whereas 148 individuals show levels greater than 2 mg dl^{-1} post-treatment (10.6%). Patients with a slightly higher pre-treatment bilirubin level were more likely to have higher post-treatment bilirubin levels.

examine the data as part of the drug-registration package. A severe hepatic warning could prevent the success of the medicine.

As previously stated, in 2001, a whole-genome SNP map was not yet available. Therefore, we set out to try to identify candidate genes for screening in those patients who developed hyperbilirubinaemia in the clinical trial compared with matched control patients in the trial who did not develop hyperbilirubinaemia. As the clinical trial was double blind, the choice of subjects was also double blind with respect to whether an individual received a drug or a placebo.

The UDG-glucuronosyltransferase 1 (*UGT1*) gene had been included as one of the candidate genes that was associated with hyperbilirubinaemia. Mutations in the *UGT1A1* gene had been shown to cause Crigler–Najjar syndrome type I, an autosomal-recessive disorder that is associated with hyperbilirubinaemia, as well as other signs and symptoms[41,42]. A polymorphic $(TA)_n$ repeat element in the TATAA box of *UGT1A1* was a known polymorphism that was previously reported by Bosma and co-workers[43]. The specific repeat variant was simple: whether there were six or seven repeats (6,6; 6,7 or 7,7). Homozygosity for the $(TA)_7$ repeat had also been identified as a major susceptibility factor in Gilbert's syndrome, a spontaneously occurring benign form of hyperbilirubinaemia[44]. The bilirubin level of patients with both Crigler–Najjar syndrome and Gilbert's syndrome increased with each dose of the $(TA)_7$ element. Was this risk factor for hyperbilirubinaemia that was known to be associated with defined syndromes also a risk factor for the hyperbilirubinaemia AE of Tranilast?

Some patients in the Tranilast trial with the 7,7 genotype did not become hyperbilirubinaemic. However, the predictive experimental analyses could be carried out only after the clinical trial was completed and the patients who were receiving a drug or a placebo could be identified. Approximately 50% of those patients with the 7,7 genotype who received Tranilast, but not those who received placebo, developed hyperbilirubinaemia. As the key to the double-blind data was not available during the candidate-gene polymorphism study, and the code was broken only after the clinical trial, the choice of Tranilast-treated 7,7 patients and placebo-treated 7,7 patients was also random. FIGURE 3 represents the proportion of Tranilast-treated patients and their relative bilirubin levels as a function of their genotype. As can be seen from FIG. 3, because some 6,7 heterozygous patients also show elevated bilirubin levels, it indicates that the *UGT1A1* homozygous 7,7 genotype is not the only genetic factor that contributes to the elevated bilirubin levels as is the case in spontaneously occurring diseases[41,42].

These data provided the opportunity to predict which patients were susceptible to Tranilast-induced hyperbilirubinaemia. Having such a test could possibly add great commercial and patient value. By further determining that Tranilast could induce Gilbert's syndrome in identifiable susceptible patients, it was possible to define a benign condition that was associated with hyperbilirubinaemia, rather than patients who were susceptible to more severe hepatic complications. As a result, a drug might avoid an unwanted 'black box' warning on its label in the drug-registration process, and a commercially successful drug would be 'safer', with few, if any, warnings of possible unknown hepatic AEs. Attrition of drugs during development due to safety concerns is one of the main problems in the pharmaceutical pipeline. The cause of concern for Tranilast was defined during the time course of a Phase III trial to a genetic risk factor that was associated with a form of clinically benign Gilbert's syndrome.

From a pharmacogenetic perspective and under the constrained time limits demanded, this experiment was a great success. Unfortunately for the company, Tranilast was not found to have efficacy in preventing restenosis of coronary vessels, and the development programme was terminated. However, this experiment clearly showed the power of genetic-variance screening to rapidly identify susceptibility risk factors that lead to a defined AE that arises during drug development.

In 2002, with the implementation of whole-genome SNP mapping and new high-throughput SNP analytical technologies, the role of genetics during the development process has become more efficient. Standardized whole-genome LD mapping should generate hypothesis-independent association data during the course of virtually all drug development programmes selected for analyses in much less time. In fact, such data would be very valuable in assessing clinical heterogeneity if there were several AE responses to a drug candidate.

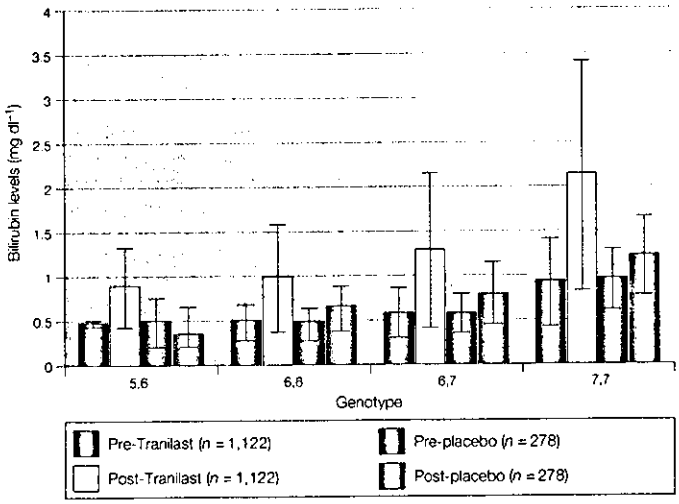


Figure 3 | **Relationship between genotype and response to Tranilast or placebo.** Analysis of the relationship between the *UGT1A1* genotype — $(TA)_n$ — and response to Tranilast treatment in 1,400 individuals from a Phase III clinical trial of Tranilast. Individuals that were homozygous for the 7,7 genotype show evidence of hyperbilirubinaemia post-Tranilast treatment but not post-placebo treatment. *UGT1A1*, UDP glucuronosyltransferase 1, polypeptide A1.

Whole-genome SNP mapping

The drive to create safe and effective medicines, as well as reduce asset attrition, has catalysed the industry to lead the way to timely development of expensive and experimental pharmacogenetic methodologies. These methods and concepts are in largely uncharted waters. Many scientists have been sceptical of anticipated outcomes. Within the pharmaceutical industry, creating a safer product is becoming even more necessary for products to survive and compete in a highly regulated market. Abacavir, which was selected for the first whole-genome AE pharmacogenetics study and was referred to earlier in this review, is an example.

Standardized LD mapping. The first SNP panel that has been constructed for use in whole-genome mapping consists of more than 200,000 SNPs that can be arranged and analysed in the order in which they occur across the human genome[8]. This density is thought to be redundant of the number of markers that might eventually constitute a specific LD map. It is estimated that considerably fewer markers would be needed if a set of common SNPs were identified that marked the various specific regions of LD across the genome for several ethnic groups. So, in the future, the possibility of making a standardized panel of SNPs for profiling susceptibility loci due to disease or reactions to any toxin (that is, a specific drug) would create the opportunity for a single lifelong pattern (SNP Print™) to be determined that could be accessed appropriately for personal medical care[8,18]. The ordered pattern of variance of LD regions across the genome of a person becomes the standard profile template against which any medicine with known AE or efficacy profiles can be tested in patients *in silico*.

CLUSTER ANALYSIS
A formal method of statistical analysis that can be used to determine the significance of clusters of data points. For example, the clustering of disease-associated SNP variants in a region of LD, or the identification of multiple regions of nearby SNP variants (~LD) at particular locations along the genome that defines a high probability that a specific drug response will occur.

Genome profiling. The key point that forms the basis for whole-genome SNP profiling is that SNPs are arrayed in the same order throughout the genome for everyone. To maximize how informative profiling is, SNPs can be chosen for inclusion in screening panels by their allele frequencies in various ethnic groups. There is then an almost infinite panel of combinations. Forensic genetic profiling is usually carried out with a limited number of highly polymorphic markers (usually 10–15) that are unlinked to each other, and are chosen from different chromosomal locations. The odds of two samples being identical are usually incredibly high. The statistical model for forensic identity involves a relatively small number of highly polymorphic variants from single individuals. The essential problem that will need to be determined empirically is how few patients of a defined drug-related phenotype will be necessary to define a confirmed pattern of regions of phenotype-associated SNPs. By analogy to forensic testing, the complexity is provided by the large number of independently variable SNPs, and not simply the number of control subjects and patients. Once the pattern is determined (analogous to the standard polymorphisms that are used for forensic testing), then the identity of SNP pattern that predicts the phenotype — for example an AE — would be expected to result in significantly higher odds as a diagnostic than more typical clinical quantitative traits. The experimental plan behind SNP profiling is quite straightforward: detect several sets of SNPs that are sufficiently close to each other and define several specific DNA regions that include genetic risk factors that are related to a particular phenotype. Rather than single-marker associations, linear arrays of nearby ordered SNPs that roughly define LD regions allow CLUSTER-ANALYSIS algorithms to be standardized. So, SNP profiles (SNP Prints™) that define genetic-susceptibility responses to medicines can be used to identify patients at risk before the medicine is prescribed to those individuals[18].

Can these techniques actually work? There are several examples in which high-density SNP mapping of disease linkage regions have defined susceptibility loci for common complex diseases — despite theoretical predictions that the experiments would not succeed. The average distances between SNP markers in successful disease-susceptibility-gene experiments — ~15 kb — led to the density estimates that were used for whole-genome screening (3 billion bp/200,000 SNPs = 15 Kb). With regard to AE data, there is already an example of LD defining a region that has significant predictive value for AE susceptibility to abacavir[39,40].

Success of proteomic profiling. Despite published opinions to the contrary, there are several recent examples that use cluster analyses and genetic algorithms for diagnosis or prediction. A particularly clear example of cluster analysis of biological data is described in a recently published paper that studied proteomic patterns in serum to identify early ovarian cancer[46]. Whether peptide-mass data from serum samples are used, as was done by Petricoin *et al.*[46], or ordered, digitized SNP data across the genome, the resulting spectra can be based on

similar genetic algorithms (FIG. 4). AE pharmacogenetic studies have the advantage of a large number of standardized, ordered data points to determine genetic risk factors in diverse situations using common tools, in contrast to the empirical experimental peptide profiles for specific diseases. From a practical perspective, AE pharmacogenetic studies have the disadvantage of requiring that patterns be developed from a relatively small number of patients to define the SNP Print™. (If too many AEs affect the risk/benefit ratio, then the medicine is removed from the market, whereas cancer diagnostic studies, unfortunately, have a larger number of subjects available.) By using a large, redundant set of AE patients in the abacavir experiment, studies of the minimal number of patients to define and confirm a risk pattern can be carried out with actual SNP Print™ data.

Disease diagnostics versus pharmacogenetic uses. A frequent concern about the proposal for widespread use of pharmacogenetic profiling often comes from those who commonly deal with highly penetrant genetic-disease ethics[7]. Abuse of any genetic data is certainly possible, particularly if those data contain diagnostic information that is related to disease. Collateral information for relatives of tested individuals might also be anticipated to be an ethical problem[47,48]. Those circumstances can be greatly minimized by separating disease-diagnostic panels from pharmacogenetic-variant panels. In fact, it would be quite easy and practical to select polymorphisms with no known relationship to any disease diagnosis for the panel of pharmacogenetic variants used for LD mapping of drug reactions. With the millions of SNPs that are now well documented across the genome, replacing a SNP in the testing panel that was known to be related to disease diagnosis, or one that will eventually be useful for disease diagnosis or association, isolates pharmacogenetic profiling from disease-related information. There will always be the chance of a particular variant being identified unintentionally, but, if the magnitude and severity of drug-induced AEs is considered, the benefits far exceed the hypothetical and controllable risks.

Efficacy and market segmentation

Patients, physicians, health-care providers, regulators and pharmaceutical executives are all favourably inclined towards making safer medicines available more rapidly[15,49,50]. The concept of market segregation for product efficacy seriously concerns the commercial rationale that is necessary to sustain the development of a new medicine. Therefore, determining which segments of the population might experience drug efficacy is sometimes viewed sceptically or dismissed outright by many marketing executives in the pharmaceutical industry. Conversely, the possibility of resurrecting failed commercial compounds can be viewed quite positively, especially in a strict regulatory environment.

In fact, efficacy pharmacogenetics could considerably lower the cost of Phase III clinical trials if randomization could be applied to a population of patients who could be selected for drug efficacy in Phase II. How to determine the efficacy genomic profile for patients early in the development process is frequently overlooked in most scenarios. In Phase II trials, it would need to be determined if a drug is effective for an individual who carries a specific genotype in order to achieve anticipated major economies of smaller, faster, more focused Phase III trials. Usually, the reported caveat to a double-blind Phase III protocol that includes only patients who are enriched for efficacy is that regulatory agencies require safety information for the whole population. However, in an era of AE pharmacogenetics, any AE profiles that were identified during the efficacy-enriched trials could be confirmed during a post-marketing, well-designed surveillance protocol[51]. Such a surveillance population would be less costly than much larger Phase III trials — and would provide more accurate results from larger populations[52].

In general, patient efficacy enrichment would not be as definitive as the case for AE pharmacogenetics. AEs can generally be defined clearly, and are recognized easily during clinical trials and post-marketing surveillance research for which the purpose is the identification of AEs. The size of a placebo effect on AEs would be negligible, as patients usually do not want to experience AEs. By contrast, during Phase II clinical-efficacy trials, there is usually a finite, sometimes large and unpredictable placebo effect. This placebo effect would add experimental noise to any determination of genetic-efficacy profiles. Technically, patients who experience placebo efficacy would affect the data in a similar manner to patients with a heterogeneous form of genetic disease. A 'false-positive' test for efficacy would simply prolong current practices of medicine: a patient could be prescribed a medicine and not be guaranteed efficacy.

Phase II trials would need to be larger and have more objectively defined clinical efficacy or surrogate end points. Expensive Phase II trials for a large number of potential drug assets would be necessary, with no guarantee of making it to the market. Therefore, efficacy pharmacogenetics would be applied only to selected assets. During Phase II drug development, attrition of drug candidates is far greater than would be expected of registered medicines. Maintaining a medicine in the market with improved safety has a more immediate appeal to all segments of medical care.

Unless there is a clear-cut benefit for some small segment of patients who are suffering from a major unmet medical need, it is difficult to imagine routinely using efficacy pharmacogenetic analyses. Special cases that are characterized by complicated, expensive Phase III clinical trials with potentially large commercial evaluations would be preferred applications. An example of such a special case might be illustrated by a hypothetical drug candidate that clearly improves ~10% of patients with AD, and has the potential to slow the progression of the disease in that segment of patients. The targeted market would still be large, because the disease is prevalent, severe and a major health problem. At the

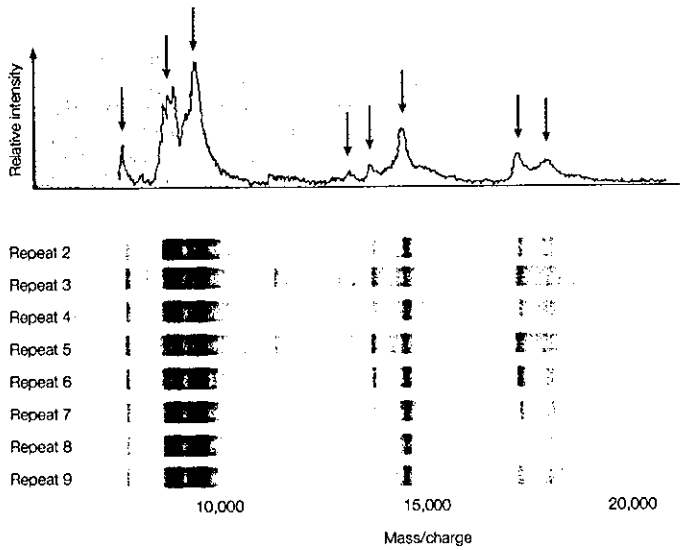


Figure 4 | **Example of between-chip reproducibility of proteomic mass spectra.** Serum from an unaffected female control was individually applied to a single bait surface region on 100 separate C18 chips and analysed by SELDI-TOF. Nine randomly obtained spectra from the 100 that were used in the analysis are shown. The eight proteins with the highest consistent amplitudes (arrows) were used as a surrogate for reproducibility by calculation of the coefficient of variance of the normalized peak amplitudes for each of the eight. This methodology was used to determine proteomic patterns for patients with ovarian cancer and was then tested in a confirmation series of patients with suspected ovarian cancer. Again, the specificity and pattern of the serum peptide peaks that were associated with a defined phenotype (ovarian cancer) were diagnostic, even without determination of the identity of each peak. Reprinted from REF. 46 © (2002), with permission from Elsevier Science. SELDI, surface-enhanced laser desorption and ionization; TOF, time of flight.

present time, trying to register a drug candidate with only 10% efficacy would need large, lengthy and expensive clinical trials to detect the effect. For example, cognitive end points would need to be followed for at least 6–12 months.

Should the responding 10% of AD patients be identified and enriched by SNP Prints™, then the size of Phase III trials could be decreased enormously. Should individuals at risk be identifiable by SNP Prints™ before they become symptomatic, then parallel preventive outcome trials would be faster and less expensive as well. The outcome trials for disease prevention could expand the number of years that an effective medicine could be used. Unless the patients who could benefit from the medicine that is being tested could be identified, the rationale (and market estimates) for developing such a medicine depends on the whole population of AD patients being treated to experience efficacy for 10% of the population. Although that scenario might still look promising as a commercial proposition at present, it can also be expected that the first time that a competitor can segregate and identify patients for efficacy treatment or prevention, the commercial viability of a less-specific medicine will plummet. Payers would prefer to support highly effective medicines for a premium in an identifiable proportion of the affected population in preference to paying for a medicine for 100% of patients that would be expected to work in only 10%.

So, there are several pitfalls to the rapid development of efficacy pharmacogenetics that differentiate it with regards to the clinical and commercial value of AE pharmacogenetics. For the present, owing to the enormous costs that are associated with applying whole-genome SNP scanning, immediate commercial benefits could be derived by keeping registered medicines in the market and not focusing efforts on stages in the development pipeline at which considerable attrition is expected.

As time goes by, genome profiling for pharmacogenetics will no doubt show its scientific and economical effectiveness. Regulators have recently embraced more specific uses of pharmacogenetics in drug-trial consultations[16]. Data for predicting adverse events and increased safety profiles will become more common and aid the sale of new medicines. The real question becomes one of public health rather than single-product acceptance. Can the practice of medicine be changed so that safety profiles can be generated quickly, with relatively few patients experiencing AEs from newly marketed medicines, so that drugs with high efficacy can remain available to the majority of patients who can be identified as not at risk of the AE?

How much pain and suffering can be avoided by developing a standardized genome map that can be economically justified for all people and legally safeguarded for access and use in pharmacogenetics? In the future, when standardized SNP-based LD maps are successfully tested, and the cost of repetitive genotyping of a large number of variants for each individual are significantly reduced, the opportunity for public-health safety will become more evidence based and highly effective. We are at the initial stages of this science, with considerable debate about the scientific and ethical variances, but we should remember that hundreds of thousands of lives are lost at the moment as a consequence of drug complications, at a cost of tens of billions of dollars per year. Successful disease management with safer drugs can solve uncounted quality-of-life issues. The problem is huge, and will not be solved by making effective medicines difficult to obtain but rather to predict safety and efficacy using safeguarded, private *in silico* information banks, to anticipate and avoid potential complications. Although it is frequently stated by watchdog groups that "everyone is at risk for AEs", everyone is not at equal risk. There is now the opportunity to define those who are at unacceptable risk before submitting each patient to the experiment of medicine prescription.

There are always reasons for not pursuing new science. Debates about new ideas or processes are healthy for creating information from data, and knowledge from information. As you take your new prescription into the pharmacy, and you worry about a possible unwanted side effect, the relevance of applied pharmacogenetics will take on a whole new meaning.

1. Motulsky, A. J. Pharmacogenetics and ecogenetics in 1991. *Pharmacogenetics* **1**, 2–3 (1991).
2. Pfost, D. R., Boyce-Jacino, M. T., & Grant, D. M. A SNPshot: pharmacogenetics and the future of drug therapy. *Trends Biotechnol.* **18**, 334–338 (2000).
3. Edwards, R. & Aronson, J. K. Adverse drug reactions: definitions, diagnosis, and management. *Lancet* **356**, 1255–1259 (2000).
4. McLeod, H. L. & Evans, W. E. Pharmacogenomics: unlocking the human genome for better drug therapy. *Annu. Rev. Pharmacol. Toxicol.* **41**, 101–121 (2000).
5. Sadee, W. Pharmacogenomics: using genetic information to optimize drug therapy. *Medscape Pharmacotherapy* [online] (cited 10 Jun. 02) <http://www.medscape.com/Medscape/pharmacology/journal/2000/v02.n06/mp7377.sade/mp7377.sade-01> (2000).
6. Kurth, J. H. Pharmacogenomics: future promise of a tool for identifying patients at risk. *Drug Info. J.* **34**, 223–227 (2000).
7. Roses, A. D. Pharmacogenetics and the practice of medicine. *Nature* **405**, 857–865 (2000).
An early review of high-density SNP-mapping technologies, with discussion of ethical considerations of pharmacogenetic SNP mapping compared with disease-gene identification.
8. Roses, A. D. Pharmacogenetics. *Hum. Mol. Genet.* **10**, 2261–2267 (2001).
9. Evans, W. E. & Johnson, J. A. Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu. Rev. Genomics Hum. Genet.* **2**, 9–39 (2001).
10. Pirmohamed, M. & Park, B. K. Genetic susceptibility to adverse drug reactions. *Trends Pharmacol. Sci.* **22**, 298–305 (2001).
11. Fijal, B. A., Hall, J. M. & Witte, J. S. Clinical trials in the genomic era: effects of protective genotypes on sample size and duration of trial. *Control. Clin. Trials* **21**, 7–20 (2000).
12. Lazarou, J., Pomeranz, B. H., & Corey, P. N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* **279**, 1200–1205 (1998).
13. Jeffries, D. B., Leakey, D., Lewis, J. A., Payne, S. & Rawlins, M. D. New active substances authorized in the United Kingdom between 1972 and 1994. *Br. J. Clin. Pharmacol.* **45**, 151–156 (1998).
14. Pirmohamed, M., Breckenridge, A. M., Kitteringham, N. R., & Park, B. K. Adverse drug reactions. *BMJ* **316**, 1295–1298 (1998).
15. Wieczorek, S. J. & Tsongalis, G. J. Pharmacogenomics: will it change the field of medicine? *Clin. Chim. Acta* **308**, 1–8 (2001).
16. Lesko, L. J. & Woodcock, J. Pharmacogenomic-guided drug development: regulatory perspective. *Pharmacogenomics J.* **2**, 101–103 (2002).
An influential article, which was written by FDA Directors, which invited proposals using SNP-mapping pharmacogenetic technologies, and changed the perception of a widely held view outside the FDA that regulators would not consider such applications.
17. The SNP Consortium. <http://snp.cshl.org/index.html> (2002).
18. Roses, A. D. Pharmacogenetics place in modern medical science and practice. *Life Sci.* **70**, 1471–1480 (2002).
19. Lai, E., Riley, J., Purvis, I. & Roses, A. D. A 4-Mb high-density single nucleotide polymorphism-based map around human *APOE*. *Genomics* **54**, 31–38 (1998)
The first high-density SNP map that was constructed as a proof of principle for LD mapping around a known susceptibility gene — to test whether the *APOE4* polymorphism, which is associated with an increased risk of AD, would have been identified more rapidly by using SNP LD-association mapping.
20. Roses, A. D. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* **47**, 387–400 (1996).
21. Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
22. Tang, M. X. *et al.* The *APOE* ε4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *J. Am. Med. Assoc.* **279**, 751–755 (1998).
23. Graff-Radford, N. R. *et al.* Association between apolipoprotein E genotype and Alzheimer disease in African American subjects. *Arch. Neurol.* **59**, 594–600 (2000).
24. Kardaun, J. W. *et al.* Genotypes and phenotypes for apolipoprotein E and Alzheimer disease in the Honolulu–Asia aging study. *Clin. Chem.* **46**, 1548–1554 (2000).
25. Martin, E. R. *et al.* SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around *APOE* in Alzheimer disease. *Am. J. Hum. Genet.* **67**, 383–394 (2000).
A detailed analysis of the first high-density SNP map around the *APOE* gene, which was carried out as a proof of principle in reference 19.
26. Martin, E. R. *et al.* Analysis of association at SNPs in the *APOE* region. *Genomics* **63**, 7–12 (2000).
27. McCarthy, L. C. *et al.* Single-nucleotide polymorphism alleles in the insulin receptor gene are associated with typical migraine. *Genomics* **78**, 135–149 (2001).
The insulin receptor was the first susceptibility gene for a complex disease that was identified using high-density SNP mapping of a genetic-linkage region.
28. Hewett, D. *et al.* Identification of a psoriasis susceptibility candidate gene by linkage disequilibrium mapping with a localized single nucleotide polymorphism map. *Genomics* **79**, 305–314 (2002).
This research identifies a gene that contains specific SNP associations with psoriasis, and defines a gene that was not a previous disease susceptibility candidate.
29. Rioux, J. D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature Genet.* **29**, 223–228 (2001).
The identification of a region using SNP mapping that contains a gene cluster that is associated with susceptibility to Crohn's disease.
30. Bergstresser, P. R. Sensitization and elicitation of inflammation in contact dermatitis. *Immunol. Ser.* **46**, 219–245 (1989).
31. Fleming, C. J., Burden, A. D. & Forsyth, A. The genetics of allergic contact hypersensitivity to nickel. *Contact Dermatitis* **41**, 251–253 (1999).
32. Ardlie, K. G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nature Rev. Genet.* **3**, 299–309 (2002).
33. Cantor, C. Whole genome SNP scans: the reality. *The Human Genome Organisation* [online] (cited 31 May 2002) <http://hgm2002.hgu.mrc.ac.uk/Abstracts/Publish/Symposia/Symposium04/hgms0013.htm> (2002).
34. Foster, R. H. & Foulds, D. Abacavir. *Drugs* **55**, 729–736 (1998).
35. Symonds, W. *et al.* Risk factor analysis for hypersensitivity reactions to abacavir. *Clin. Ther.* (in the press).
36. Hetherington, S. W. *et al.* Hypersensitivity reactions during therapy with abacavir: analysis of 636 cases for clinical presentation and risk factors. [online] (cited 31 May 2002) *7th Conference on Retroviruses and Opportunistic Infections* <http://www.retroconference.org/2000/abstracts/60.htm> (2000).
37. Brew, B. J. *et al.* Phase III, randomized, double-blind, placebo controlled, multicentre study to evaluate the safety and efficacy of abacavir (ABC 1592) in HIV-1 infected subjects with AIDS dementia complex (CNA3001). *12th World AIDS Conference* Abstract 32,192 (1998).
38. Brinkman, K., Smeitink, J. A., Romijn, A., & Reiss, P. Mitochondrial toxicity induced by nucleoside-analogue reverse transcriptase inhibitors is a key factor in the pathogenesis of antiretroviral-therapy-related lipodystrophy. *Lancet* **354**, 1112–1115 (1999).
39. Hetherington, S. *et al.* Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* **359**, 1121–1122 (2002).
This paper shows that several polymorphisms in a region of extended LD confer susceptibility to the HSR to abacavir, using a list of candidate genes that are thought to be involved in immunological disorders. By chance, SNPs from two distinct genes were in extended LD, and so defined a larger-than-expected region, which shows that estimates of the number of SNPs that are needed for whole-genome screening profiles might be too large.
40. Mallal, S. *et al.* Association between the presence of HLA-B*5701, HLA-DR7, and HLA-GQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
These independent experiments look at specific HLA regions as candidate sets of polymorphisms for association to the HSR to abacavir and, along with reference 39, provide immediate published confirmation of this LD region.
41. Monaghan, G., Ryan, M., Seddon, H. R. & Burchell, B. Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet* **347**, 578–581 (1996).
42. Aono, S. *et al.* Analysis of genes for bilirubin UPD-glucuronosyltransferase in Gilbert's syndrome. *Lancet* **345**, 958–959 (1995).
43. Bosma, P. J. *et al.* The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *N. Engl. J. Med.* **333**, 1171–1175 (1995).
44. Bosma, P. J. *et al.* Sequence of exons and the flanking region of human bilirubin–UDP-glucuronosyltransferase gene complex and identification of a genetic mutation in a patient with Crigler–Najjar syndrome, type I. *Hepatology* **15**, 941–947 (1992).
45. Roses, A. D. Pharmacogenetics and future drug development and delivery. *Lancet* **355**, 1358–1361 (2000).
46. Petricoin, E. F. *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577 (2002).
47. Buchanan, A. *et al.* Pharmacogenetics: ethical issues and policy options. *Kennedy Inst. Ethics J.* (in the press).
48. Robertson, J. A. Pharmacogenetic challenges for the health care system. *Health Affairs* (in the press).
49. Wolf, C. R., Smith, G. & Smith, R. L. Science, medicine and the future: pharmacogenetics. *BMJ* **320**, 987–990 (2000).
50. Ingelman-Sundberg, M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J. Int. Med.* **250**, 186–200 (2001).
51. Park, K. B. & Pirmohamed, M. Toxicogenetics in drug development. *Toxicol. Lett.* **120**, 281–291 (2001).
52. Cardon, L. R. *et al.* Testing drug response in the presence of genetic information: sampling issues for clinical trials. *Pharmacogenetics* **10**, 503–510 (2000).

Acknowledgements

A.D.R. acknowledges the GlaxoSmithKline (GSK) department heads and project leaders whose teams contributed to the experiments and examples described in this review: D. Burns, D. Campbell, R. Dement, A. Hughes, E. Lai, P. Manasco, A. McCarthy, I. Purvis, J. Riley, A. Saunders, S. Sehgal, Y. Smithies and N. Spurr. Also, the many scientists involved in creating the SNP mapping strategy and capability within GSK; T. Isenhour and E. McPherson, who assisted in preparing and editing the manuscript; A. Holden and D. Wang, who accelerated The SNP Consortium to be on time, under budget and to exceed all expectations; and M. McPeak, E. Haner, M. Beadle and S. Groh, who contributed support for team coordination. J. Niedel, T. Yamada, A. Hannah, J. Palmer, J. Robinson, B. Koch and K. Spitz provided essential early and continued support for pharmacogenetics and pharmacogenomics in the Genetics Research Directorate in GSK.

Online links

DATABASES

The following terms in this article are linked online to:

Cancer.gov:
http://www.cancer.gov/cancer_information
ovarian cancer

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
APOC1 | APOE | HLA-A | HLA-B | HLA-DR | TNF-α | *UGT1*

Medscape DrugInfo:
http://promini.medscape.com/drugdb/search.asp
abacavir

OMIM: http://www.ncbi.nlm.nih.gov/Omm
Alzheimer's disease | Crigler–Najjar syndrome type I | Crohn's disease | Gilbert's syndrome | psoriasis | sickle-cell disease

FURTHER INFORMATION

The Human Genome Project:
http://www.nhgri.nih.gov/HGP

The SNP Consortium: http://snp.cshl.org

The Wellcome Trust: http://www.wellcome.ac.uk

Access to this interactive links box is free online.

Making the right choice of targets at the beginning of the pipeline will be the first step down the long road of creating innovative medicines

Keynote review:

Disease-specific target selection: a critical first step down the right road

Allen D. Roses, Daniel K. Burns, Stephanie Chissoe, Lefkos Middleton and Pamela St. Jean

Relevance of a drug target for a disease is often inferred with strong belief but fragile evidence. Here, a program for early identification of human disease-specific drug targets using high-throughput genetic associations is described. Large numbers of well-characterized patients (>1000) and matched controls are screened for genetic associations using several thousand (>7000) single nucleotide polymorphisms from more than 1500 genes. The genes were selected because they are members of target classes for which there are precedents for high-throughput chemical screening technology. This review summarizes the methods and intensive data analyses leading to target gene identification for type 2 diabetes mellitus, including the statistical permutation methodology used to correct for many variables.

Discussions of target identification for drug discovery have become technically oriented and complicated over the past decade [1,2], a period of time that coincides with decreases in productivity across the pharmaceutical industry [3–6]. The latest technologies for selecting targets can be fascinating and imaginative, but are these targets relevant to treating human diseases [7,8]? Although methods for high-throughput chemical screening and for optimizing lead molecules have undoubtedly made major advances, target selection remains a crucial step [9]. Currently, there is no strategy of equally high-throughput for the selection of targets that are directly associated with human diseases. This need becomes even clearer when animal models are considered the ultimate target validation [7,10].

The drug discovery process should be clear-cut – identify the best molecule, for the most effective treatment, as fast and efficiently as possible. Although high-throughput molecular methods have progressed, matching the molecule to the appropriate clinical indication

Allen D. Roses*
Daniel K. Burns
Stephanie Chissoe
Pamela St. Jean
GlaxoSmithKline R&D, Research Triangle Park, NC 27709, USA
*e-mail: allen.d.roses@gsk.com
Lefkos Middleton
GlaxoSmithKline R&D, London, UK

ALLEN D. ROSES

Allen D. Roses was appointed Senior VP for Genetics Research at GlaxoSmithKline (GSK) in 2000. Before joining GlaxoWellcome in 1997, Roses was at Duke University, where he was the Jefferson Pilot Professor of Neurobiology and Neurology and Director of the Center for Human Genetics. Roses led the team that identified apolipoprotein E as a major, widely confirmed susceptibility gene in common late-onset Alzheimer's disease. While at GSK, he was charged with organizing genetic strategies for susceptibility gene discovery, developing and implementing pharmacogenetics approaches and integrating genetics into medicine discovery and development; translation of genetic and genomic research into pathway analyses, drug discovery and pharmacogenetics in development is ongoing at GSK.

DANIEL K. BURNS

STEPHANIE CHISSOE

LEFKOS MIDDLETON

PAMELA ST. JEAN

1359-6446/04/$ – see front matter PII: S1359-6446(04)03321-5

for development is still an arduous task: disease-relevant target identification can now be an important criterion for determining the correlation between molecule and indication. The foundation of the successful reputation of the 1980s and early 1990s, when newly discovered biological pathways and receptors were selected as targets, was a rich biological literature identifying the 'low-hanging fruit' [11]. Biology has continued to direct targets to a small extent, which is probably best illustrated by biopharmaceuticals [12]. For small-molecule screening over the past decade, the selection of targets using animal models, simple organism genomics and genetically manipulated animals has a less than successful track record [13,14] – why not return to patients to select targets for their diseases?

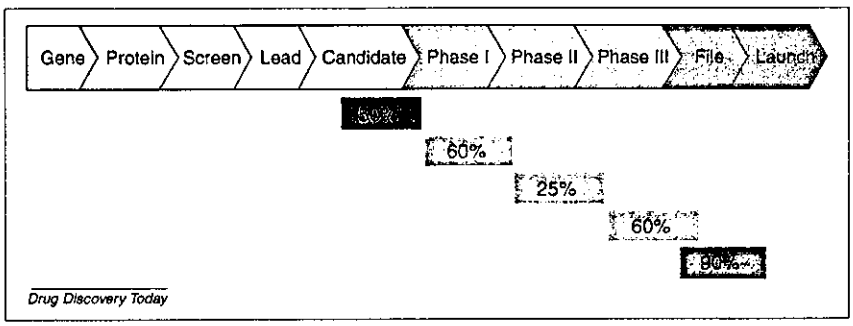
If the speed and efficiency of identifying targets that are relevant to human diseases were improved, then the technical advances for lead validation, high-throughput chemical screening and lead optimization could be applied to molecules with a greater probability of success. For example, polymorphic variants of target genes might have different interactions with lead molecules. Studies of mechanism, chemical lead validation and on- or off-target effects can be conducted in directed mouse models [15]. Although leads from mouse model targets are frequently ineffective in human clinical trials (this is particularly true in cancer and CNS diseases [2]), for most discovery research, animal models set a gold standard for target selection [7,16,17]. A complimentary approach would be to identify first those targets that are genetically associated with human diseases and then create appropriate knockout and conditional knockin models with target gene variants. This suggestion differs significantly in the scope and depth of human phenotyping from proposals to phenotype a broad catalogue of knockout mice to suggest targets [10].

Here, early (but not preliminary) data are presented from the application of a high-throughput human disease-specific target program (called HiTDIP within GlaxoSmithKline) that focuses on the association of tractable targets with specific patient groups. The principles of complex gene association studies and disease susceptibility will be addressed, particularly with respect to matching targets with clinical indications. Several reviews have recently been published that cover genetic association studies [18–26], and these broad, complex, and sometimes esoteric, disciplines are not discussed here. Rather, the focus of this article is the strategy and process of genetic association studies to match pharmaceutical targets with clinical indications related to several human diseases.

The pharmaceutical industry and its business analysts track attrition at various stages in the drug discovery and development pipeline. For example, 95% of candidate quality leads fail to produce a medicine. Of the molecules that enter Phase I clinical trials after surviving preclinical testing, only 21.5% reach the market [27]. Furthermore, the number of new molecular entities (NMEs) – drugs with a novel chemical structure – submitted to the FDA over the past decade has decreased (Figure 1) [13].

Candidate leads evaluated at Phase IIA for efficacy represent products of target selection generated during the past ten years, generally before the possible contributive effects of the completion of the human genome sequence could be realized. The major sources of attrition after entering Phase I trials are toxicity and lack of efficacy. One method of decreasing attrition at an early clinical stage is to apply prospective efficacy pharmacogenetics (PGx) at Phase IIA [3]. Another strategy is to select the right target initially.

The pharmaceutical applications of HiTDIP methodologies were adopted as a response to the apparent increased failure rate of clinical development that has plagued the industry over the past decade. A basic hypothesis for this attrition might be stated simply as: 'A reason that many molecules are ineffective in clinical trials is that the selection of the originally screened target was based on data and rationale that are actually irrelevant to the etiology or pathogenesis of the human disease'. For a drug to be successful in treating a disease, two variables must be matched – the target and the right therapeutic indications. The success rate of discovery of molecules from screens is therefore directly related to target choice, and targets identified using animal models and sophisticated genomic analyses have so far provided fewer molecules for NME submission than anticipated [28,29].



NCE success ratios: probability of progressing through each phase. Only one in 25 NCE candidate compounds is approved by the regulators (Table 1). Note that only 25% of those molecules that survive Phase II successfully pass into Phase III. Molecules derived from genetically-associated targets could increase the success rate. A small increase would have a significant effect on approvals. The addition of safety pharmacogenetics can also contribute to decreased attrition during development [60; http://csdd.tufts.edu/NewsEvents/RecentNews.asp?newsid=4].

A critical view of genomic applications for target identification

Analysis of the human genome sequence promised to bring a flood of new targets,

TABLE 1

The pharmaceutical pipeline: definition by milestone

R&D stage (milestone passed)	Description
From target identified to screening hit and/or lead compound	Success* in identifying a compound with the desired pharmacological activity at the desired molecular, cellular or mechanistic target: this compound might not have all the characteristics required to be a viable drug
From lead compound to drug candidate	Success in identifying a viable drug candidate by optimizing the characteristics of the initial screening hit and/or lead; typically, this requires appropriate potency, selectivity and efficacy but can also involve other criteria such as bioavailability, metabolic stability and preliminary safety screening
From drug candidate to FTIH and/or Phase I	Success in progressing a drug candidate into initial studies in humans, usually (Phase I) in healthy volunteers
From Phase I to entry into Phase II	Success in progressing a clinical development candidate into small-scale exploratory studies in patients that have the targeted disease
... to proof of concept	Usually considered to be the point at which a drug candidate has demonstrated efficacy in its intended patient population, typically within Phase II: therefore, project attrition can be measured for progression to or from this milestone, as well as to or from the more traditional clinical development milestones (Phases I–III)
From Phase II to entry into Phase III	Success in progressing a medicinal candidate into large-scale (pivotal) clinical trials suitable for registration
From Phase III to regulatory filing	Success in progressing a data package into a regulatory submission
From regulatory review to approval	Success in gaining regulatory approval
From regulatory approval to launch	Success in launching an approved product, added indication and/or label-change

*Project progression can be quoted as attrition (failure) or success. Abbreviation: FTIH, first time in human.

and therefore increase the potential throughput of the pharmaceutical pipeline [28–31]. This scenario has also proved disappointing in not reaching the projected goals. What factors have limited target selection and drug discovery productivity? Although HTS technologies were successfully implemented and spectacular advances in mining chemical space have been made, the universe for selecting targets expanded, and in turn almost exploded with an inundation of information. Perhaps the best explanation for the initial modest success observed was the dramatic increase in the 'noise-to-signal' ratio, which led to a rise in the rate of attrition at considerable expense. The difficulty in making the translation from the identification of all genes to selecting specific disease-relevant targets for drug discovery was not realistically appreciated. There was certainly a large pharmaceutical industry investment in the intellectual property speculation surrounding the sequencing of all possible genes, with the hope that a stream of new targets for drug discovery would result. Senior R&D scientists recently recognized the need for a 'quantal step-up in discovery' [32]. To feed a high-throughput pipeline, a high volume flow of specific, disease-relevant targets is necessary. Whether or not individual researchers believe in a particular disease hypothesis, in the specific relevance of a target class to some aspect of human disease pathogenesis or in particular animal models of human disease, the evidence that 'validates' (substitute believe in, consensus view or champion, among others) the choice of a target molecule for a potential therapeutic strategy in humans is crucial to starting down the right road.

Target validation is one of those terms that scientists use in multiple ways. With respect to target identification and selection in the pharmaceutical industry, validation is interpreted as providing increased confidence to initiate expensive chemical screens and subsequent discovery programs. On occasion, support of possible relevance comes from the sheer weight of 'potentially' (substitute believed in, validated, rational or accepted, among others) relevant information. For example, a target gene could be determined to be expressed in the tissue that is affected pathologically by a particular disease, differentially expressed in disease-relevant tissues or have a visible effect in animal models when manipulated [33]. These data might provide modest human disease-specific support as the starting point for a drug development program to treat a particular disease. A putative target located on neuronal surfaces could be relevant to a human neurological or psychiatric disease – but which one? Proof of concept in humans can only occur on completion of preclinical testing and Phase I safety studies of an optimized lead candidate. A Phase II clinical trial is an extremely costly hurdle with which to justify the target choice after many years of confidence-building research. The situation could be compounded when a molecule with exceptional drug qualities, but an unclear clinical indication, is tested in several clinical trials involving multiple clinical endpoints.

The success of a target is judged after many years – usually in hindsight by counting marketed products. The success rate of efficacy (proof of concept) studies is probably a much earlier indicator of pipeline health. If the right target was selected more often and this led to the selection of effective lead candidates more frequently, then attrition would be reduced. Shots on goal are good, but center forwards who miss 99% of the time (and take all the shots) are not hired by professional teams. The pharmaceutical

industry must move away from the numbers game and strive for specificity and speed at the earliest stages of the pipeline. Targets for specific diseases that are chosen based on strongly held beliefs have a significant probability of being the totally wrong target. Current 'knowledge' (substitute accepted beliefs, strongly held views or reasonably good ideas, among others) might not define disease-relevant hypotheses accurately. It is the result – an effective and safe medicine – that is of importance to the patients, physicians and industry. Indeed, the mechanisms of action of many successful medications are still unknown.

Can target molecules be directly associated with human diseases that have highly statistically significant data? Yes. Can chemical leads be produced from screening these targets that can enter the pharmaceutical pipeline? Yes. Will lead candidates produce a higher rate of future success in demonstrating efficacy compared with current metrics? A decision on this aspect has yet to be reached – more time is needed to study the flow through the pipeline into human testing. However, when a genetically associated target gene has already been screened chemically, there can be a rapid progression from identification of the target to the entry of lead molecules into clinical development.

Gene-specific target association study design

With the human genome sequenced, it is possible to define virtually all genes belonging to the known target classes using analogous sequence regions that define specific structures or functions. However, there are few real indications as to which gene might be specifically associated with a particular disease. It is possible to test each gene individually for disease relevance using genetic association studies, but only if sufficiently large and well-characterized patient and control groups are available. What of high-throughput association studies of many sequence variations within all genes of each target class? As an example hypothesis, assume there is a G-protein-coupled receptor (GPCR, sometimes referred to as a 7-transmembrane repeat) target class gene variant that is associated pathologically with Alzheimer's disease. Which GPCR is it? If there are almost 500 known GPCR genes, then is the Alzheimer's disease-specific GPCR the third on the list? Is it number 222, or maybe number 407? High-throughput, gene-based single nucleotide polymorphism (SNP) genotyping technologies provide the opportunity for the rapid testing of each of the GPCR variants for disease association. Genetic association studies provide an evidence-based opportunity to inform target choice rapidly and more specifically.

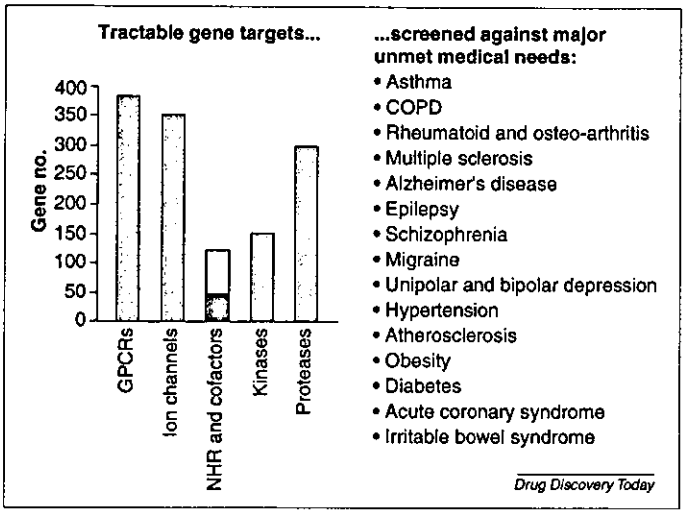
There is an implied crucial assumption when using a gene-disease association strategy: disease-specific associations might be identified for genes that are selected simply because it is known how to screen them against large chemical libraries. Seven years ago, the scientific community was reluctant to take such a risk. However, more recently, retrospective data were published that support this assumption. Goldstein *et al.* [34] examined 42 sequence variants of genes that had been associated with a drug response at least twice. These investigators found that 21 of the 42 variants were in the target or in a known pathway of the target. It is therefore reasonable to propose that the probability of identifying a disease-relevant target would be increased by screening all potential targets for genetic association with well-defined diseases. That is to say, there are now data available to support the hypothesis that candidate leads derived from genetically associated targets can increase the probability of success (decrease attrition) at Phase II or Phase III of clinical trials.

Another important scientific contribution to gene-disease association studies has come from advancements in the fields of genetic epidemiology and statistical genetics. These are specialized disciplines to most of the pharmaceutical industry, but the ability to analyze rapidly many clinical traits simultaneously in a high-throughput fashion, and with appropriate methods to define statistical significance, is probably the most useful contribution to such studies other than the sequenced genome template. By 2002, this advancement made it possible to test therapeutic class genome-wide variants for statistical associations with particular human diseases. Additional support for a disease gene-association strategy can be found in the plethora of recent studies linking specific gene variants to particular diseases [35–40]. When the effect of the variant results in expressed clinical disease, it is generally viewed as a disease mutation. Where multiple variants of several genes contribute to the expression of a disease, they are now commonly referred to as susceptibility genes. The practical problem of solely studying disease genetics to generate targets is that most susceptibility genes are not drug targets, and therefore the high-throughput methodologies currently available cannot be used to screen for the formation of chemical interactions [41].

Would the initial identification of high-throughput targets with human disease-specific associations result in a more efficient pipeline with less attrition? Because there was great confidence that the human genome projects (both public and private) would eventually provide gene sequencing, and some variant, information, it was anticipated that a large, low-throughput resource would be needed; that is, prospective, well-phenotyped patient collections and appropriate controls, each of which would have consent for commercial applications.

The patients define the relevance

To study the association of gene variants with the clinical expression of human diseases, a large number of consenting patients and controls must be carefully examined and the data placed into accessible databases; in addition, DNA must be collected and stored. Because the clinical examination of patients is performed one patient at a time, the generation of large patient collections that are



HiTDIP: genetic associations between 'tractable' targets and major diseases. High-throughput analysis of approximately 7000 polymorphisms in 1800 candidate genes (numbers of validated SNPs and candidate target genes have increased over time) are screened for association in common diseases, such as asthma, schizophrenia, depression, osteoarthritis, Alzheimer's disease, metabolic syndrome, hypertension, acute coronary syndrome and others. The approximate numbers of genes in various target-classes used in these experiments are also indicated. Genes of marketed products and other target genes such as enzymes and protein ligands are also included in the target classes. As more target groups become tractable, they are added to the screen. Between 2002 and 2004, the gene list has expanded from approximately 1450 genes to >1800. For the NHR and cofactor column, the yellow color represents the number of NHR cofactor gene targets. Abbreviations: COPD, chronic obstructive pulmonary disorder; NHR, nuclear hormone receptor.

suitable for disease association studies using multiple markers is time- and resource-intensive, and consequently is not high throughput. Anticipatory clinical research of this type requires sustained access to clinical expertise, extensive supporting resources and commitment over a period of years. Few such prospective collections exist and, in those academic environments that possess such databases, informed consent for commercial uses is usually absent.

In 2005, the genome is sequenced, target class gene variants are known, rapid genotyping technologies are available and interactive databases with analytical capabilities have been built. Since 1997, GSK has organized external clinical specialists, who are expert in more than a dozen important diseases, and has accumulated more than 80,000 patients and controls, each examined with a prospectively standardized protocol, informed consent and stored DNA samples. Several association experiments were completed that provided exciting new putative targets for pipeline consideration. The leading edge of chemical leads has begun to enter the pipeline.

As molecules resulting from insights of HiTDIP reach the published portfolio of GSK over the next few years, there will be a relatively straightforward method for the comparison of attrition with historical metrics. With respect to the gene variants identified for early drug discovery, specific disclosure of early targets and leads in the pipeline could be limited by regulatory and commercial concerns. However, the best and most rapid biological and genetic validation of gene variants associated with disease can occur where the data are available for confirmation by academia and industry. It is therefore planned that these large association experiments will be published in scientifically reviewed journals to enable the full weight of academic and industry disease-specific target validation to be focused in this area. Furthermore, pharmaceutical companies principally share many of the same targets – why not compete on the screening and lead chemistry of disease-relevant targets and well-designed drug development?

High-throughput disease-specific target discovery – the experiment

To perform this experiment for the identification of the targets associated with human disease, three major components are required: (i) selection of the gene targets to be screened; (ii) well-characterized clinical data; and (iii) genetic data generation and statistical analyses.

The targets

Before the sequencing of the human genome, the pharmaceutical industry knew of perhaps 500 targets [42]. Widely appreciated target classes, for example, nuclear receptors, kinases and GPCRs (Figure 2) now constitute ~1200 genes, and there are many additional enzymatic screens that bring one estimate of the total druggable genome to over 3000 genes [43]. One method of dealing with screenable targets is to create knockout or other modified mice and search for phenotypes. This undoubtedly provides some additional support, but the phenotypic correlation between mouse and man can be difficult to interpret [17]. It would seem that direct associations with human disease phenotypes would be a promising and efficient point to search for knockout and conditional knockin models. Kola and Landis [2] listed five places along the pharmaceutical pipeline where attrition might be managed. Their first point was that 'building the need to get very strong evidence for proof of mechanism into the discovery paradigm is critical'. In the past, this was possible because an extensive literature had developed as scientists concentrated on biochemistry, physiology, pharmacology and other disciplines before suggesting targets. If the 100 best-selling drugs are examined retrospectively, the targets were initially selected because of strong and confirmed (in the literature) biology, including studies in man [17]. The prediction that knockout models will have the same efficiency prospectively as that provided by retrospective analyses will only be possible with extensive and specific phenotyping of each knockout. Screening the phenotypes of many knockouts might be much more superficial [16]. Perhaps, the most efficient approach would