EUDRAGENE: European collaboration to establish a case–control DNA collection for studying the genetic basis of adverse drug reactions

Mariam Molokhia[1†] & Paul McKeigue[2]

†Author for correspondence
1London School of Hygiene & Tropical Medicine, Non-Communicable Disease Epidemiology Unit, Keppel Street, London, UK
Tel.: +44 0207 927 2633;
Fax: +44 0207 580 6897;
E-mail: mariam.molokhia @lshtm.ac.uk
2University College Dublin, Conway Institute, Belfield, Dublin, Ireland

Type B adverse drug reactions (ADRs) are often serious, limit the usefulness of drugs that are otherwise effective, and increase the risks of drug development as they often lead to postmarketing withdrawal. There is evidence that susceptibility to at least some Type B ADRs is under strong genetic influence. Identifying genes in which variation influences susceptibility has obvious practical value for genetic testing and might also make it easier to screen molecules likely to cause ADRs at an early stage of the drug development process. Research in this area is hampered by the lack of a resource in which to study genetic determinants of susceptibility to Type B ADRs. As serious Type B ADRs are rare, case–control designs are the most frequently-used approach. The EUDRAGENE collaboration seeks to develop a resource using an international collaboration. This will provide a basis for adverse drug susceptibility genome association-wide studies using tag single nucleotide polymorphisms, or a direct approach using putative functional polymorphisms.

Most research on the genetic basis of susceptibility to adverse drug reactions (ADRs) has focused on studying dose-dependent reactions that are predictable on the basis of the drug's known pharmacological actions. These Type A reactions are relatively common, and there has been considerable progress in identifying genes that determine susceptibility through effects on drug metabolism [1,2]. There has been little research on the genetic basis of susceptibility to Type B reactions: idiosyncratic reactions that are not necessarily dose-dependent. Only a few genetic causes of such reactions have been identified [3]: mutations in the *RYR1* gene predict susceptibility to anesthetic-induced malignant hyperthermia, and a polymorphism in the *G6PD* gene predicts susceptibility to drug-induced hemolytic anemia. Serious idiosyncratic adverse reactions are necessarily rare: a drug that commonly causes such reactions will not be licensed. Of all ADRs occurring in hospitalized patients in the USA, approximately a quarter are Type B reactions [4].

Why studying the genetics of Type B adverse drug reactions is important

Although serious Type B adverse reactions are rare, they are of considerable importance to human health because they may severely limit the use of an otherwise effective drug (such as clozapine, which causes agranulocytosis in fewer than 1% of patients) or because they may lead to withdrawal of a drug (such as cerivastatin, associated with 52 reported deaths from rhabdomyolysis and for rofecoxib found to be associated with increased risk for myocardial infarction and stroke) after the costs of development and bringing it to market have been incurred. Pharmaceutical risk management is being implemented in the EU as a means to reduce the probability of unanticipated regulatory decisions through the continuous assessment of benefit/risk. Further possibilities include identifying susceptible individuals through genetic testing, and greater understanding of these disease syndromes when they are not drug induced.

Aims of the EUDRAGENE project

The objective of the EUDRAGENE project, funded by the European Commission 5th Framework program, is to establish a freely-shared case–control collection of DNA samples, as a multicenter European collaboration. This will act as a resource for studying genetic predictors of ADRs, using the existing system for spontaneous reporting of suspected ADRs to national or regional pharmacovigilance centers.

How many cases are needed?

As Type B reactions are rare (<1%), and families containing multiple affected individuals are even rarer, family linkage studies cannot be used to find susceptibility genes. Using large cohorts or registries to study ADRs as proposed [5]

would not yield sufficient cases. Case–control designs are a more efficient approach to studying genetic determinants of susceptibility to rare outcomes. A case–control study provides a valid estimate of risk or hazard ratio. As the exposure (genotype) is not affected by the onset of disease and can be measured without error, one limitation of case–control studies (biased measurement of exposure known as protopathic bias) does not apply. Selection bias is also less serious since the genotype does not influence referral or recruitment of cases and controls and can be overcome, as controls can be recruited from the same population at risk.

Simple power calculations are based on 90% power to detect association at p-values of less than 10^{-6}, using the indirect 'haplotype-tagging' approach, utilizing selected SNPs to capture most of the common variation in the gene [6]. To detect a disease-associated variant that accounts for at least 25% of all cases, around 200–250 cases of each ADR under study are required.

Why is international collaboration necessary?

As most such ADRs are rare, a case–control design is the most frequently used approach, and a multicenter European collaboration is necessary as no single country will generate a sufficient number of cases of any given ADR within a reasonable time. National and regional pharmacovigilance systems provide a systematic means of case ascertainment. Table 1 shows the numbers of events reported over 1997–2002 in Europe for six classes of ADR that are of current interest. Rates show considerable variation between countries, possibly explained by differences in exposure and/or in reporting.

Adverse reactions studied (initial set)

The adverse reactions studied (initial set) include:

- Torsades de pointes (associated with long Q-T syndrome [LQTS]) caused by several classes of drug, including anti-arrhythmic agents, antibiotics and antipsychotic agents
- Myopathy or rhabdomyolysis caused by statins and/or fibrates
- Agranulocytosis caused by all classes of drugs including thyroid inhibitors, sulphasalazine or clozapine
- Tendon rupture caused by fluoroquinolones
- Liver injury (using Council for International Organizations of Medical Sciences [CIOMS] and related criteria [7–9]) caused by all drug classes
- Mefloquine-induced psychosis

Case definitions

Case definitions for each class of ADR have been compiled by a working group of EUDRA-GENE investigators, in consultation with other experts, based on internationally agreed criteria. Events are classified as drug exposed only where the temporal sequence from drug intake to onset is clear and following the exclusion of other known medical or environmental causes, supplemented with Naranjo criteria for estimating ADR probability.

Case ascertainment

All European countries have a system for spontaneous reporting of adverse reactions, supplemented by reports from drug manufacturers to regulatory agencies. Although pharmacovigilance remains the responsibility of national drug regulatory agencies, national systems are being harmonized through the European Medicines Evaluation Agency, which is compiling a Europewide database of reports of ADRs [101]. Even for serious ADRs there is considerable under-reporting [10]; estimated reporting rates range from less than 10% [11,12] to less than 1% [13]. In some, but not all countries, algorithms are applied to each report to assess the causal role of the drug. French pharmacovigilance systems are especially well developed, with mandatory reporting by physicians and follow up of individual cases by regional pharmacovigilance centers. In the UK, the Medicines and Healthcare Products Regulatory Agency (MHRA) has recently completed a review of access to data from the yellow card system, and indicated that data will be made available for researchers in the near future.

For certain classes of ADR, case ascertainment can be supplemented through hospital records: for instance to search for cases of LQTS. Although the final sample of cases may not be representative of all cases occurring in the population (due to under-reporting, restrictive case definition and nonresponse), this selection bias will not affect genetic associations unless there is population stratification, which can be dealt with in the statistical analysis [14].

Collection of material from cases

For each report of a suspected ADR falling into one of the classes under study, a letter from the local investigator is forwarded to the physician who reported the ADR. This letter explains the objectives of the project, and asks the physician to help by inviting the patient to donate a blood sample for genetic studies. Patients are invited to complete a short questionnaire including demographic

Table 1. Numbers of reported adverse drug reactions in each class by country 1997–2002.

Center	Pop. (millions)	TdP	LQTS	MY	R	Liver injury		Agranulocytosis				Ten	Neuro
		All		Statins/fibrates		T	NSAIDS	All	TI	S	C	F	ME
Italy*	58	4	–	64	156	5	88	95	8	–	1	139‡	24
Veneto§	4. 5	–	3	13	10	6	29	49	5	–	7	47 (6 rupture)	11
Sweden	9	43	16	12	30	10	108	188	31	33	10	47	74
Germany	83	55	55	307			–		33			31	325
Belgium	10	10	14	7		0		51	3	2	4	361 (100 ruptures)	18
Spain¶	40	15	19	43	115	14	100	137	2	0	2#	66	2
France	60	127	118	939	282		300	1211	43	16	23	465 (87 rupture)	98
Netherlands**	16	27	3	5		14	–	42	15	4	8	114	833
Norway	4. 5	14		18			26	34	2	8	1	0	25
UK‡‡	60	32		1735			1107	989	67	285	3	723	106

*Italy data collected from 2001 to 2005.

‡The data about tendon rupture is not available.

§Veneto region data collected from 1997–2002.

¶Spain data collected from 1997–2001.

#Spain has a separate database for clozapine.

**Netherlands data collected from 1997–2001.

‡‡UK estimates combined from IMS MediPlus and THIN primary care databases; individual drug classes for agranulocytosis available from IMS only.

ADRS: Adverse drug reactions; C: Clozapine; F: Fluoroquinolones; LQTS: Long QT syndrome; ME: Mefloquine; MY: Myopathy; NEURO: Neuropsychiatric reactions; NSAID: Nonsteroidal anti-inflammatory drugs; R: Rhabdomyolysis; S: Sulphasalazine; T: Triazoles; TdP: Torsades de Pointes; TEN: Tendinitis/tendon rupture; THIN: The Health Improvement network; TI: Thyroid inhibitors.

information and medical history, to donate a blood sample, and to provide written consent for the use of medical records and the anonymized sample for genetic studies. Other data collected include the drug suspected to be the cause of ADR, the indication for which this drug was prescribed, a list of other medications taken concurrently, and relevant environmental exposures.

Selection of controls

Two possible sources of controls include general population controls and drug matched controls. One advantage of using population controls is the existence of several existing collections with consent to use of the anonymized DNA samples for genetic studies. A possible disadvantage of using population controls is that the case–control comparison is confounded by any genetic factors that are associated with factors for which the drug is prescribed. For example, individuals receiving statins are at higher risk of cardiovascular disease than population controls. In practice this is not a serious risk as most drugs are prescribed for broad indications, or there are several classes of drugs implicated that may cause the ADR. Databases of drug utilization and primary care have facilitated the identification of patients prescribed certain classes of drug [15].

The main confounder in case–control studies is population stratification; methods exist to control for this, allowing elimination of almost all other confounding [14,16,102]. Genes which metabolize external substrates may display large frequency differentials within Europe which may act as confounders, unless adjusted for [17].

Strategies for selecting candidate genes

Two main strategies are available for genetic association studies: direct (identifying polymorphisms likely to affect function) and indirect approaches (marker polymorphisms, and relying on detecting allelic associations with functional polymorphisms in the same gene).

The direct approach relies on the ability to identify libraries of functional polymorphisms using bioinformatics tools [103]. The indirect approach is efficient where allelic heterogeneity is low (only one or two disease-associated alleles in each of the genes that influence risk). One strategy for studies is based on the indirect approach, in which 4–10 SNPs in each gene are used to 'tag' the haplotypes and thus capture most of the common variation in the gene [6]. This approach exploits the genetic history of the European population, in which past constraint of population size has led to loss of haplotype diversity. The 'haplotype map'

project initiated by the US National Institutes of Health facilitates the identification of 'haplotype-tagging' SNPs [104]. Although tag SNPs can be used for multiple populations, different strategies will be required if haplotypes causing the ADR are extremely rare [18].

As genotyping costs fall, whole-genome association studies (WGA) are becoming more feasible, provided that there is adequate statistical power to detect associations at stringent levels of significance that adjust for the multiple tests performed [19,20]. Although technology has vastly improved in recent years, error rates remain constant. Ideally, cases and controls should be genotyped using the same platforms, to reduce differential error rates in genotype scoring that can bias results [21]. In practice it is often more practical to use controls that have already been genotyped, and new methods of statistical analyses may be required to overcome this problem.

For each class of ADR to be studied, there are several approaches to select candidate genes:

• Selection of functional polymorphisms in genes for drug transport and metabolism such as cytochrome P450 (*CYP*)*2D6* and adenosine triphosphate-binding cassette (*ABCB1 – MDR1*, encoding the drug transporter P-glycoprotein) genes [22,23]. One approach will explore SNPs in the *CYP2D6* gene for which assays have been set up and individuals can be classified into four main groups [Eichelbaum M. PERS. COMM.].

• Selection of genes causing Mendelian forms of the same syndrome; for example LQTS, for which causative genes have been identified. Common variants in these genes can be investigated as candidates for susceptibility to drug-induced forms of the same syndrome. At least six genes in which mutations cause Mendelian forms of LQTS have been identified by family linkage studies: five encoding potassium channels and one encoding a sodium channel. A recent study identifying cases of drug-induced QT_c prolongation and cardiac arrest in the Netherlands screened 4/45 eligible cases for genetic mutations and identified variants in potassium voltage-gated channel, subfamily H, member 2 (*KCNH2*), sodium channel, voltage-gated, type Vα (*SCN5A*) and potassium voltage-gated channel, ISK-related family, member 1 (*KCNE1*) [24]. Not all individuals who harbor disease-causing mutations are affected [25], and these mutations are rare in patients with drug-induced LQTS [26],

suggesting that the effects of these genes depend upon other modifier loci. Similarly, at least three different genes have been identified as causing Mendelian forms of agranulocytosis: elastase 2, neutrophil (ELA2) [27], growth factor independent 1 (GFI1 – protooncogene which targets ELA2) [28], granulocyte colony stimulating factor (GCSF) [29] and GCSF receptor (GSCFR) [30] and Janus kinase 2 (JAK2 – a tyrosine kinase involved in GCSF signaling) [31].

• Genetic determinants of immune function including polymorphisms of cytokines, interleukins and human leukocyte antigen (HLA) haplotypes associated with idiosyncratic reactions, such as drug-induced hepatotoxicity [32].

Compliance with guidelines for research using human DNA samples

This project is conducted in accordance with current ethical guidelines on human tissue and biological samples for use in research. Most European countries have now laid down guidelines similar to those formulated by the UK Medical Research Council [105]. Similar regulations are being followed by the Council of Europe's Steering Committee on Bioethics. A key requirement of these guidelines is that before undertaking open-ended genetic studies, DNA samples must be irreversibly anonymized, so genotype data cannot be linked to patients' identities.

Ethical approval was obtained from the London School of Hygiene & Tropical Medicine, the ethical review board for the European Commission, and all national and local ethical committee approvals were obtained from participating European centers.

Outlook: UK a future resource

This international collaboration has been established on the basis that the case–control resource will be freely available to researchers in the participating countries, in the both the academic and industry sectors. As a condition of access to the samples, investigators will be required to submit all their genotype data to the main database, so that subsequent research can make use of data already acquired. Over the next few years more countries are expected to join the EUDRAGENE network, allowing the rate of case accrual to increase. Even with a multicenter European approach, the main challenges will be identifying cases.

The computerization of primary care in the UK may facilitate the development of new and more rapid methods for detecting ADRs in clinical

Highlights

- European network of shared resources for case–control studies of genetic determinants of selected adverse drug reactions.
- The collection comprises clinical data and anonymized DNA.
- Ascertainment of cases is through pharmacovigilance and hospital networks.
- Cases can be stratified by phenotype or by drug class.
- A secure web-based interface has been developed for data management.
- The EUDRAGENE collaboration is a resource for academic and industry users.

pactice with the creation of primary care databases with longitudinal medical records and drug prescription data covering several million people [33]. In principle, it is possible to exploit these databases for a more direct approach to the detection of associations between drugs and adverse events, as ascertainment of adverse events in these databases can be almost complete in practices with good systems for collecting diagnostic data. For case–control comparisons, cases and controls from such databases can also be matched for clinical indication. In the longer term, the development of integrated health records by the National Health Service (NHS) could allow the methods to be applied to a much wider population and thus greatly improve the detection rates of ADRs.

The establishment of EUDRAGENE provides a unique case–control resource to study the genetic basis of ADRs, including meta-analyses. This collection will provide the basis for studies of association with candidate genes and future genome-wide association studies. We anticipate that the resource will continue for several years, based on the issues of current concern.

Acknowledgment

Bibliography

Papers of special note have been highlighted as of interest (•) to readers.

1. Wolf CR, Smith G: Pharmacogenetics. *Br. Med. Bull.* 55, 366–386 (1999).
2. Severino G, Del ZM: Adverse drug reactions: role of pharmacogenomics. *Pharmacol. Res.* 49, 363–373 (2004).
3. Nebert DW: Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin. Genet.* 56, 247–258 (1999).
4. Lazarou J, Pomeranz BH, Corey PN: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279, 1200–1205 (1998).
5. Roses AD: Genome-based pharmacogenetics and the pharmaceutical industry. *Nature Rev. Drug Discov.* 1, 541–549 (2002).
6. Johnson GC, Esposito L, Barratt BJ *et al.*: Haplotype tagging for the identification of common disease genes. *Nature Genet.* 29, 233–237 (2001).
   • **Explains how knowledge of common haplotypes and the single nucleotide polymorphisms (SNPs) that tag them can be used systematically for strategic genotyping and linkage disequilibrium mapping.**
7. Benichou C: Criteria of drug-induced liver disorders. Report of an international consensus meeting. *J. Hepatol.* 11, 272–276 (1990).
8. Benichou C, Danan G, Flahault A: Causality assessment of adverse reactions to drugs – II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. *J. Clin. Epidemiol.* 46, 1331–1336 (1993).
9. Garcia Rodriguez LA, Duque A, Castellsague J, Perez-Gutthann S, Stricker BH: A cohort study on the risk of acute liver injury among users of ketoconazole and other antifungal drugs. *Br. J. Clin. Pharmacol.* 48, 847–852 (1999).
10. Rawlins MD: Pharmacovigilance: paradise lost, regained or postponed? The William Withering Lecture. *J. R. Coll. Physicians Lond.* 29, 41–49 (1995).
11. Smith CC, Bennett PM, Pearce HM *et al.*: Adverse drug reactions in a hospital general medical unit meriting notification to the Committee on Safety of Medicines. *Br. J. Clin. Pharmacol.* 42, 423–429 (1996).
12. Stephens MDB: Introduction. In: *Stephens' Detection of New Adverse drug Reactions.* 5th edition. Talbot J, Waller P (Eds), John Wiley & Sons Ltd, Hoboken, New Jersey, USA, 10–17 (2005).
13. varez-Requejo A, Carvajal A, Begaud B, Moride Y, Vega T, Arias LH: Under-reporting of adverse drug reactions. Estimate based on a spontaneous reporting scheme and a sentinel system. *Eur. J. Clin. Pharmacol.* 54, 483–488 (1998).
14. Hoggart CJ, Parra EJ, Shriver MD *et al.*: Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* 72, 1492–1504 (2003).
15. Garcia Rodriguez LA, Perez GS: Use of the UK General Practice Research Database for pharmacoepidemiology. *Br. J. Clin. Pharmacol.* 45, 419–425 (1998).
16. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181 (2000).
17. Campbell CD, Ogburn EL, Lunetta KL *et al.*: Demonstrating stratification in a European–American population. *Nature Genet.* 37, 868–872 (2005).
18. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: Efficiency and power in genetic association studies. *Nature Genet.* 37, 1217–1223 (2005).
19. Davey SG, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR: Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 366, 1484–1498 (2005).
20. Hattersley AT, McCarthy MI: What makes a good genetic association study? *Lancet* 366, 1315–1323 (2005).
21. Clayton DG, Walker NM, Smyth DJ *et al.*: Population structure, differential bias and genomic control in a large-scale, case–control association study. *Nature Genet.* 37, 1243–1246 (2005).
    • **Differential bias in genotype scoring may lead to false positive associations; possible solutions are discussed.**
22. Gough AC, Miles JS, Spurr NK *et al.*: Identification of the primary gene defect at the cytochrome P450 *CYP2D* locus. *Nature* 347, 773–776 (1990).
23. Sakurai A, Tamura A, Onishi Y, Ishikawa T: Genetic polymorphisms of ATP-binding cassette transporters *ABCB1* and *ABCG2*: therapeutic implications. *Expert Opin. Pharmacother.* 6, 2455–2473 (2005).
24. De Bruin ML, van Puijenbroek EP, Bracke M, Hoes AW, Leufkens HG:

Pharmacogenetics of drug-induced arrhythmias: a feasibility study using spontaneous adverse drug reactions reporting data. *Pharmacoepidemiol. Drug Saf.* 15, 99–105 (2006).

25. Yang P, Kanki H, Drolet B *et al.*: Allelic variants in long QT disease genes in patients with drug-associated torsades de pointes. *Circulation* 105, 1943–1948 (2002).

26. Paulussen AD, Gilissen RA, Armstrong M *et al.*: Genetic variations of *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1*, and *KCNE2* in drug-induced long QT syndrome patients. *J. Mol. Med.* 82,182–188 (2004).

27. Horwitz M, Benson KF, Person RE, Aprikyan AG, Dale DC: Mutations in *ELA2*, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis. *Nature Genet.* 23, 433–436 (1999).

28. Person RE, Li FQ, Duan Z *et al.*: Mutations in proto-oncogene *GFI1* cause human neutropenia and target ELA2. *Nature Genet.* 34, 308–312 (2003).

29. Le Beau MM, Lemons RS, Carrino JJ *et al.*: Chromosomal localization of the human *G-CSF* gene to 17q11 proximal to the breakpoint of the t(15;17) in acute promyelocytic leukemia. *Leukemia* 1, 795–799 (1987).

30. Dong F, Hoefsloot LH, Schelen AM *et al.*: Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. *Proc. Natl Acad. Sci. USA* 91, 4480–4484 (1994).

31. Rauprich P, Kasper B, Tidow N, Welte K: The protein tyrosine kinase JAK2 is activated in neutrophils from patients with severe congenital neutropenia. *Blood* 86, 4500–4505 (1995).

32. Andrade RJ, Lucena MI, Alonso A *et al.*: HLA class II genotype influences the type of liver injury in drug-induced idiosyncratic liver disease. *Hepatology* 39, 1603–1612 (2004).

33. Majeed A: Sources, uses, strengths and limitations of data collected in primary care in England. *Health Stat. Q.* 5-14 (2004).

Websites

101. Eudravigilance website. www.eudravigilance.org/human/evGateway01.asp

102. ADMIXMAP. www.ucd.ie/genepi/software.html

103. Biodata website. www.biodata.ee/methods.htm

104. International HapMap Project. www.hapmap.org

105. MRC Ethics Series – Human tissue and biological samples for use in research. www.mrc.ac.uk/pdf-tissue_guide_fin.pdf

Appendix: contributors in alphabetical order

Dr Albert-Jan Aarnoudse
Dr Frank Andersohn
Professor Julio Benitez
Dr Martin Brunner
Dr Juan-Antonio Carillo
Professor Alfonso Carvajal Garcia-Pando
Dr Laetitia Caturla
Dr Sonia Colcera
Dr Anita Conforti
Professor Michel Eichelbaum
Dr Erik Eliasson
Professor Bruno Flamion
Professor Edeltraut Garbe
Dr Luisa Ibañez

Dr Kari Kivistoe
Dr Maryse Lapeyre-Mestre
Professor Håkan Melhus
Professor Jean-Louis Montastruc
Dr Yola Moride
Dr Markus Mueller
Dr Atul Pathak
Professor Ingemar Persson
Maria Sainz
Ines Salado
Professor Bruno Stricker
Dr Miriam Sturkenboom
Professor Giampaolo Velo
Dr Qun-Ying Yue