Project Final Report

Grant Agreement number:	**279062**
Project acronym:	**EpiPGX**
Project title:	**Epilepsy Pharmacogenomics: delivering biomarkers for clinical use**
Funding Scheme:	**Collaborative Project**
Date of latest version of Annex I against which the assessment will be made:	**06-12-2011**
Period covered:	**From 01-11-2011 to 31-10-2015**

Name, title and organisation of the scientific representative of the project's coordinator:

Prof. Sanjay Sisodiya - University College London

Tel: + 44 20 3448 8612

Fax: + 44 20 3448 8615

E-mail: s.sisodiya@ucl.ac.uk

Project website address: www.epipgx.eu

Table of Content

Section 1 – Final publishable summary report

EpiPGX

Logo:

Project title: Epilepsy Pharmacogenomics: delivering biomarkers for clinical use

Website: www.epipgx.eu

Contractors involved (EpiPGX consortium):

The project is coordinated by

Prof. Sanjay Sisodiya, 01 UCL University College London, Great Britain

1, Gower Street

London WC1E 6BT

United Kingdom

Other partners and team leaders:

02 ULB (EpiPGX) Université Libre de Bruxelles, Belgium
03 IGG (EpiPGX) Istituto Giannina Gaslini, Italy
04 EKUT (EpiPGX) Eberhard Karls Universitaet Tuebingen, Germany
05 SEIN (EpiPGX) Stichting Epilepsie Instellingen Nederland, The Netherlands
06 UKB (EpiPGX) Universitätsklinikum Bonn, Germany
07 RCSI (EpiPGX) Royal College of Surgeons in Ireland, Ireland
08 BHSCT (EpiPGX) Belfast Health and Social Care Trust, Great Britain
09 deCODE (EpiPGX) Islensk Erfdagreining EHF, Iceland
10 UL (EpiPGX) Université du Luxembourg, Luxembourg
11 UMCU (EpiPGX) Universitair Medisch Centrum Utrecht, The Netherlands
12 ULIV (EpiPGX) The University of Liverpool, Great Britain
13 Imperial (EpiPGX) Imperial College of Science, Technology and Medicine, Great Britain
14 UGLA (EpiPGX) University of Glasgow, Great Britain
15 GABO:mi (EpiPGX) GABO:mi Gesellschaft für Ablauforganisation :milliarium mbH & Co KG GABO, Germany

1.1 Executive summary.

EpiPGX is working to find genomic variants that predict or explain clinically-relevant responses and adverse reactions to antiepileptic drugs. Establishing variants that are of clinical value will contribute to improved patient care, precision medicine, stratification for clinical trials and ultimately improved quality of life for people with epilepsy and the best use of increasingly stretched healthcare resources. Determining which drug might suit, and which drug should be avoided, for each individual person with epilepsy should enable the right drugs to be utilised early during disease course, which in turn should reduce some of the adverse outcomes of the epilepsies, and make best use of the global investment in drug development which has already taken place. It may even lead to the resurrected use of drugs that are currently out of favour because of their unpredictable but serious adverse effects and should also provide insights into both drug mechanisms of action and disease biology. EpiPGX therefore represents an important investment in the better understanding and management of the epilepsies.

Over the course of its funding lifetime, EpiPGX has moved significantly towards achieving these aims. We anticipated that the work would be challenging, and requiring painstaking attention to detail. With this in mind, we spent time generating carefully-considered definitions for the phenotypes to study, crafting an electronic case record form for use across the entire consortium through a common web platform, producing a system for the reporting, discussion and resolution of technical and conceptual issues, and checking the quality of our phenotyping. This robust infrastructure has proved critical in the working of EpiPGX and has contributed to the creation of probably the largest epilepsy pharmacogenomics database so far, to our knowledge, in the world, with data from over 12,000 people with epilepsy and over 32,000 individual drug response phenotypes. In a measure of the value of these efforts, our definitions and electronic case record form have been adopted by other projects and consortia across the world.

We have also gathered together a very large dataset of genetic information. Changing technologies and falling costs worked largely to our benefit, enabling the collection of more genotypic and sequencing data that we had envisaged. We did suffer some delays beyond our control during this process, and have now much more data than we had anticipated. This represents significant additional value for money from our initial funding. The additional data also mean that there is much more analysis possible, many more questions that can be answered, but also that all this work will take us beyond the official funding period, and is likely to generate results of value for some years to come.

EpiPGX has been a successful consortium, and has firmly established new collaborations across Europe. We are a close grouping, and look forward to working together for years to come, and to leverage value from EU funding by further grant applications at both national and international levels. We have recruited and trained the next generation of clinicians and researchers who will be invaluable to the community broadly as genomic and pharmacogenomic data become much more part of routine clinical and scientific work. We have established relationships beyond EpiPGX with several international consortia, and through shared analytic pipelines, we will further increase the value of EpiPGX: for example, we have contributed the largest additional cohort of genotyped patients to the International League Against Epilepsy Consortium on Complex Epilepsies ongoing meta- and mega-analyses, as well as leading efforts across consortia to identify variants that cause serious adverse reactions.

Data have and continue to be analysed. We have already made some important discoveries which we are in the process of confirming and validating. As these results are still formally confirmed and have not undergone the important process of peer review, we have not included formal results in this report, but anticipate full open access publication in due course. Work on the EpiPGX dataset continues, with all partners in EpiPGX keen to maximise what can be derived from our work over the last four years. To this end, we have set up systems to ensure the continuity of EpiPGX, to maintain the security and safety of its data, and to continue the spirit of goodwill and collaboration it has already fostered.

We anticipate that we will achieve the aims we proposed at the inception of EpiPGX, and in fact we hope to have achieved more over the years to come. EpiPGX has, and will continue to, demonstrate the importance and power of large-scale collaboration across Europe to work together for the benefit of its citizens.

1.2 Summary description of project context and objectives:

Background and Aims

The purpose of this project, EpiPGX, is to identify genome-based predictive biomarkers for use in routine clinical practice to personalise treatment of epilepsy with existing antiepileptic drugs (AEDs), and stratify people with epilepsy for clinical trials, aiming to maximise clinical effectiveness, avoid chronicity, prevent relapse and reduce adverse drug reactions (ADRs). Improving the use of existing treatments will benefit individuals and society alike. This is the first systematic attempt to identify such biomarkers in any complex, important neurological disease.

The need both for improved use of current treatments and for new treatments in epilepsy is undoubted and pressing. Epilepsy is a common serious condition, affecting 60,000,000 people of all ages worldwide. Epilepsy consists of many subtypes, which can be grouped together in various combinations for particular purposes. Epilepsy is associated with increased morbidity across all aspects of life, including a high risk of premature death: uncontrolled seizures lead directly to the death of over 12,000 Europeans/year. Over 20 antiepileptic drugs (AEDs) are licensed for epilepsy treatment. Seizures can be effectively controlled by AEDs in ~70% of people. Control of seizures leads to risk reduction for most of the consequences of epilepsy, improves quality of life, permits social re-integration and leads to direct economic benefits. However: a) in 30% of people with epilepsy, currently-available AEDs do not control seizures – recurrent seizures threaten life and impair its quality, and account for much of the €15.5 billion annual cost of epilepsy in the EU alone; b) there is currently no way to accurately predict which individuals will respond in any particular way to a specific or all AEDs; c) even in the 70% who do respond, only 50% respond to the first AED, and only ~15% respond to the second– whilst the correct drug is being sought, risks from seizures continue; we need to be able to predict the right (and wrong) drugs for an individual from the outset; d) there are individuals who do respond to the AED that happens to be chosen as the fourth or fifth (or so on) AED – these particular AEDs need to be identifiable from the outset; e) unrelated to responder status, AEDs can cause serious ADRs, and no predictors of any such ADRs exist; f) there is a clear need for novel means of discovery of new AEDs; biomarkers identified here will also provide insights to disease biology and thence novel treatment approaches. This is a proven approach for discovery of novel treatments.

There is clear evidence of heterogeneity of response to existing AEDs and a significant unmet need for effective intervention. Available data show genetics plays a role in variable AED response. We propose to employ pharmacogenomics in large cohorts to establish definitively the genetic contribution to variation in response to several established AEDs and improve the use of those AEDs. Apart from directly assisting people with epilepsy, the project will maximize benefits from genomic research in Europe, enhance the established competitiveness of Europe in pharmacogenomics and epilepsy treatments in particular, and provide important benefits to European SMEs through the development of clinical tests and expertise in interpretation. The project will directly lead to the development of an unrivalled computerised biobank of linked phenotypic and genotypic data from >10,000 people with epilepsy. This tightly-regulated resource will be of great benefit to biotechnology companies seeking to integrate pharmacogenomics into phase III and IV studies. During the early stages of development of a novel AED with known target(s), companies will also be able to see the landscape of real genetic variation on which the AED will act in a very large virtual cohort of people with epilepsy – for example, a company could explore (the functional consequences of) variation in a particular channel gene which encodes the target of their novel AED, thus enabling the company to develop ideas of potential responders/non-responders to the novel AED in computational models and/or experimental settings even before trials in man. This novel virtual test-bed will represent a scientific breakthrough, speed development of new AEDs and reduce clinical trial costs. It can also facilitate Phase IV studies. For people with epilepsy, it will provide pre-prescription genomic information. Only by collaboration of existing centers of expertise in epilepsy management with research-driven SMEs can the resources be gathered to undertake this project: the technology and expertise exist, coordination and funding are required.

Pharmacogenomics is well established in clinical practice: examples include (but are not limited to) i) for the prevention of life-threatening allergic reaction to the anti-HIV drug abacavir; ii) for predicting response to several anti-cancer drugs; and iii) for prevention of life-threatening allergic reaction to the AED carbamazepine in people of South Asian extraction; this was the only useful pharmacogenomic test in existence for epilepsy when EpiPGX started. As the much-stigmatised poor relation to almost every condition, epilepsy lags behind in pharmacogenomics, but breakthroughs are ready to be made and will be invaluable in clinic and will advance knowledge of disease mechanisms, as illustrated by Dravet syndrome. In this condition, genetics has identified the cause, and explained

response to certain AEDs. As a strong integrated network, EpiPGX has the clinical resources, technical ability, cooperative infrastructure, equipment, expertise and motivation to deliver a broad set of genome-based biomarkers for established AEDs. We anticipated that there would be significant challenges in bringing pharmacogenomics research findings to clinical practice. We note, equally, that changes in clinical practice with actual benefit from pharmacogenomics in fact stand out as real success stories in this regard. Frameworks already exists to evaluate how such challenges can be overcome for genetic tests in particular. The main challenges identified are: analytical validity, clinical validity, clinical utility and ELSI (ethical, legal, social issues). EpiPGX is structured to deal with the first two challenges; it will set up the fundamental discoveries and understanding of genetic findings to allow the third challenge, clinical utility, to be addressed as the next step after EpiPGX. We planned to be guided by an Ethics Advisory Board regarding the fourth challenge, which can only be fully addressed during clinical implementation, but throughout each project within EpiPGX we will pay attention to ELSI as findings emerge.

The variants we planned to identify would meet criteria to qualify as biomarkers (as defined by the U.S. Food and Drug Administration), and will supplement decision-making, in concert with clinical factors. It is important to note that NINDS in the USA consider epilepsy pharmacogenomics as an important and near term research target – our proposal is a timely response, and EpiPGX is currently probably the only European, or indeed global, effort capable of meeting this target (http://www.ninds.nih.gov/funding/research/epilepsyweb/2007_benchmarks.htm).

Work strategy and general description

The project started from what was a very limited set of baseline existing data. There were only three replicated pharmacogenomic predictors for AEDs: CYP2C9 variants for phenytoin pharmacokinetics; HLA-B*1502 for carbamazepine-induced rash in people of South Asian origin and HLA-A*3101 for carbamazepine-induced rash in people of Caucasian origin. We aimed to advance over this very limited state-of-the-art by discovering more genetic variants which could predict response to AEDs in general, to specific AEDs and for ADRs. We intended to use innovative schemes to achieve our aims, as set out in the WorkPackages, studying powerful cohorts from across Europe. The potential for exploitation is significant – for example, product labelling has already changed for carbamazepine regarding the HLA-B*15:02 variant, and is likely to do so also for the HLA-A*31:01 variant, with potential for designed, point-of-care tests, an ideal area for SME development.

Our proposal consisted of a series of logical and parallel steps, broken down into work packages (WPs). Components of WP ran in parallel, and certainly WPs were not themselves the logical sequential steps. Having formed the consortium, with an appropriate management structure and consortium agreement in place, the plan was first to confirm the phenotypes to be worked on, and to generate datasets for the training and periodic monitoring of performance of clinical researchers on the project. These datasets were to be derived from real-life cases from our own cohorts. Training and monitoring was to be organised as satellite sessions around Consortium meetings. This phenotype quality assurance would be the responsibility of WP01, also responsible for coordinating EpiPGX activities with other global pharmacogenomic efforts and iterative development of phenotypes as indicated, throughout the EpiPGX lifecycle. To facilitate data collation for analyses, WP10, working with WP01, was to disseminate existing in-house databases, to support work in all the other workpackages and facilitate for the work of WP08 in generating an in silico genetic landscape of people with epilepsy. We recognised that web-based databasing, and associated regulatory requirements, would probably have evolved further towards the conclusion of EpiPGX and planned to remain flexible about the exact strategies of making our data more widely accessible.

The definitions generated and refined by WP01 were to feed directly into the particular workpackages that would address the clinically-relevant phenotypes chosen. We intended to study several pharmacogenomic phenotypes of interest, including failure of the first AED (WP02), broad resistance to several AEDs irrespective of putative mechanism of action (WP03), subsequent responsiveness to one or more of the most-commonly used AEDs (WP04), serious, often use- or dose-limiting ADRs related to AEDs (WP05), and major congenital malformations caused by valproate (WP06). Phenotypes and quality assurance of phenotyping were to be provided and monitored by WP01. Each WP would have a nominated leader, and a nominated geneticist (in one case also the WP leader), who together would shape and direct work within a given WP. Responsibility for performance fell to the WP leader. There will regular meetings of the WP leaders and of the entire consortium planned, at which a Scientific Advisory Board and an Ethics Advisory Board would also be present. Clinical and genetic data were to be shared across WP as necessary.

There was to be cross-cutting WP activity throughout EpiPGX. Two SMEs helped form the overall structure and

direction of the consortium (directing WP08 and WP09). WP01 was tol bind together the entire consortium through discussions on phenotype and quality control of phenotyping. The remaining cross-cutting WP (WP07) was a bioinformatics core, whose responsibility it was be to consider multivariate analyses of data informed by the burgeoning progress in both data analysis and bioinformatic fields. The key SMEs were to be responsible for: a) the genotyping/ sequencing core, whose role in data acquisition would feed in to each WP, and was not specifically allocated to only one WP, and which would also lead translational development of tests; and b) the management SME, GABO:mi, whose role it was to provide an overall management infrastructure, coordination, dissemination and completion of necessary EU processes for the running of the project (WP09).

The combination of clinical and genetic strength in each clinically-directed WP, with cross-cutting WP was designed to ensure a resilient consortium structure capable of delivering our objectives through close collaboration, dissemination of databasing, research, findings and experience, and a unified motivation to find genome-based biomarkers that could inform clinical decision making, with benefit to people with epilepsy as our main aim.

We anticipated overall, therefore, the capacity to deliver the objectives we have set out. We built in flexibility and contingency planning, as well as collaborations with external agencies and consortia. We intended to pay particular attention to existing and developing ethical and regulatory issues, assisted by our Ethics Advisory Board. The diversity of national interpretations of European regulations had to be noted. We intended to take account of this as necessary at local and global levels.

Management structure and procedures

The Project Coordinator ensured the smooth operation of the project and guaranteed that all efforts were focused towards the objectives. The Coordinator submitted all required progress reports, deliverables, financial statements to the European Commission, and, with the assistance of GABO:mi, was responsible for the proper use of funds and their transfers to participants. The EpiPGX office was established by and based at the Coordinator base in London and at GABO:mi in Munich. The Project Office of the Coordinator was concerned with the scientific management and the co-ordination of all research activities. The Project Office at GABO:mi was responsible for administrative, financial and contractual management and the organisational co-ordination of the project activities.

The General Assembly was in charge of the political and strategic orientation of the project and acted as the arbitration body. It met once a year, and when the interest of the project required intermediate meetings. The Project Steering Committee consisted of all workpackage leaders and the Coordinator and was in charge of monitoring all activities towards the objective of the project in order to deliver as promised, in due time and in the budget. The Steering Committee met every six months during the funding period and had monthly telephone conferences. Furthermore, a scientific advisory board was implemented to ensure a high standard of research and monitor the progress of the project by taking part in the annual Governing Board Meetings.

Objectives of EpiPGX:

Our specific objectives are:

1. To bring together several existing European endeavours, datasets and collaborations in this area, harmonising phenotype definitions (WP01), and facilitating data access, so generating large interrogable datasets from which discoveries can proceed
2. To identify genome-based biomarkers of early treatment response in newly-diagnosed epilepsy (WP02) or to late response to specific AEDs (WP04), and for resistance to multiple AEDs (WP03). Both common and rare variants will be sought, moving beyond univariate analyses with novel bioinformatic methods, merging genomic information with clinical predictors (WP02, WP07)
3. To identify genome-based biomarkers of specific ADRs, for prevention of ADRs in clinical practice and improved targeting of AEDs (WP05).
4. To identify genome-based biomarkers of teratogenesis associated with AED, especially valproate, use in pregnancy, helping with the difficult process of AED choice in certain epilepsies in women of child-bearing age (WP06).
5. To establish an in silico virtual test-bed for new AEDs and facilitate Phase III and IV studies (WP08)
6. To initiate development of an 'EpiPGX chip' combining validated genome-based biomarkers from EpiPGX and other sources for eventual clinical use (WP08)
7. To promote effective management and resource usage (WP09) and dissemination & training (WP10), paying heed to societal factors and uptake of findings in health services and industry (all WP).

1.3 Description of the main S&T results/foregrounds of EpiPGX

EpiPGX has been a very successful project. We have brought together researchers and clinicians focussed on epilepsy from 15 different centres, and for each centre there have often been a number of other linked researchers and clinicians. We have also brought in new associate partners or collaborators and worked with several other international consortia. We have built an effective functioning partnership: that we intend to continue to work together beyond the official funding period is a mark of the success of our collaboration. The project has been challenging. We have had to manage delays beyond our control due to issues with supply of reagents for genotyping and sequencing. We have also learnt about the diversity of practice across Europe in evaluating epilepsy: we foresaw that this issue would need to be addressed, and planned phenotype quality control to a level not, to our knowledge, addressed in previous projects of this type. This was necessary as the phenotypes we were dealing with were more complex than simply disease descriptions: we had to formalise drug response phenotypes in ways that were both meaningful and achievable, permitting harmonisation across several sites and nations. But these challenges were expected and managed, such that comprehensive evaluation across a range of drug response phenotypes proved possible. Moreover, we were able to exploit technological developments and progress across the field of epilepsy and genetics research, benefitting from falling costs to type and sequence many more cases than we had anticipated, and taking advantage of new public resources across these types of genetic data. As a result, after four years of diligence, commitment and painstaking attention to detail, EpiPGX is poised to deliver pharmacogenomics results across a series of clinically relevant phenotypes. Much of the work is ongoing and will be submitted for publication. In order not to compromise the important processes of peer review and formal publication in the scientific literature, the data currently remain confidential, but will be made available in the public arena upon publication through Open Access support. The key achievements of EpiPGX are considered below, listed by our main aims.

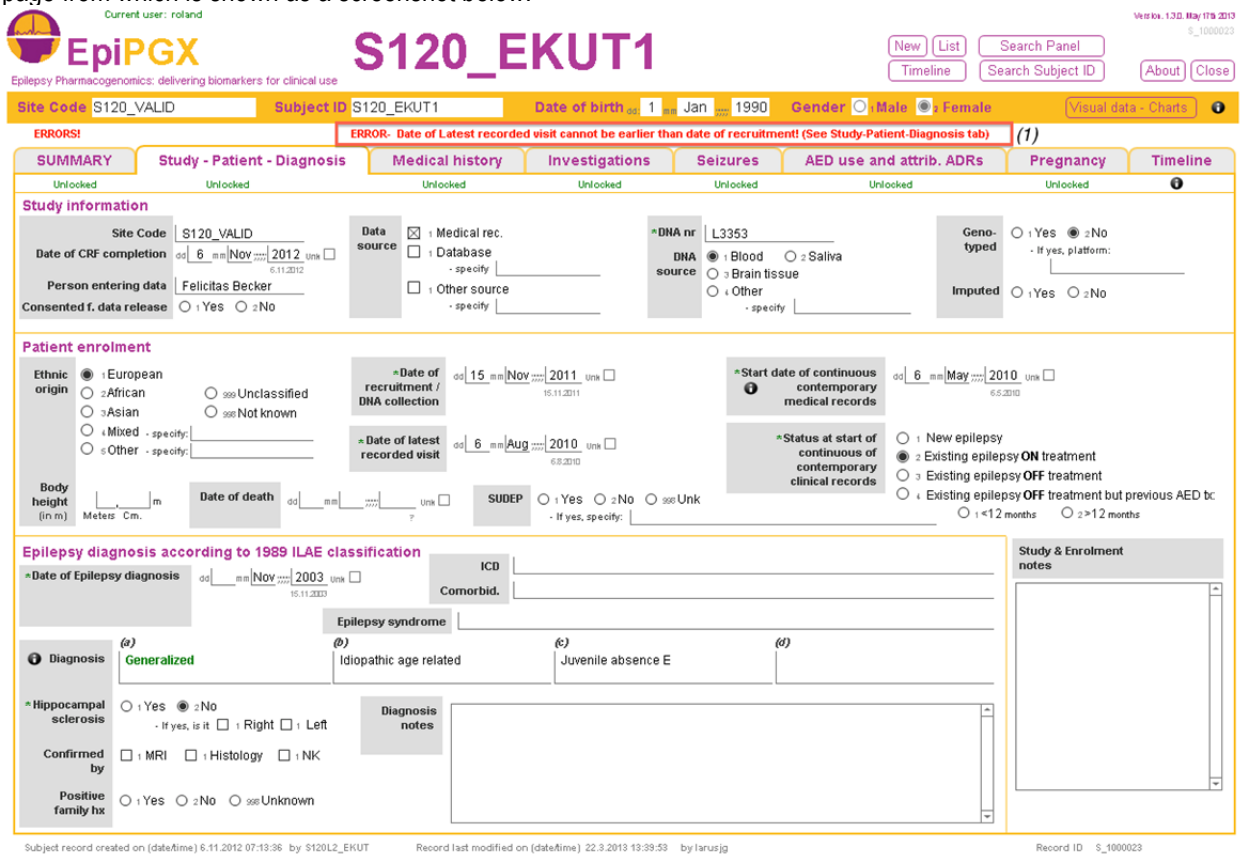
1. To bring together several existing European endeavours in this area, harmonising phenotype definitions, analyses and data access, so generating large interrogable datasets from which discoveries can proceed.

Detailed consensus phenotype definitions were generated by CR02, in close collaboration with the other CRs contributing to WP01. Definitions were based on evidence from the literature where available and further adapted to the specific needs of the project. We developed definitions for the following phenotypes: (a) remission with the first well-tolerated antiepileptic drug (AED); failure of first AED due to lack of efficacy and adverse drug reactions (ADRs), respectively (WP02); (b) broad AED resistance and responsiveness, and extreme AED resistance and responsiveness (WP03); (c) late response to and failure of specific AEDs in focal and generalized epilepsy; extreme late response to specific AEDs (WP04); (d) AED-specific ADRs (WP05); and (e) cases and controls for valproate-related teratogenesis (WP06). A first draft of the definitions was generated and circulated following discussion at the EpiPGX kick-off meeting in November 2011. Additionally, CR02 – in collaboration with CR01 and CR12 - developed an EpiPGX-specific case record form (CRF) following discussion at the kick-off meeting. Design of this document was based on several existing CRFs already in use by some of the CRs and further adapted specifically according to the phenotyping definitions in order to capture all the required clinical data. The CRF was circulated to all CRs in March 2012. Further improvements were subsequently made based on discussion with the different WP leaders and on initial phenotyping experience. The CRF also forms the basis for the electronic CRF (eCRF) developed by CR09 and CR10, and to which WP01 has provided major input and feedback.

CR02 organised a clinical training and phenotype workshop in Brussels on April 6th 2012. This workshop was attended by CR01, CR02, CR04, CR06, CR07, CR10 and CR12. At this interactive meeting we extensively discussed the phenotype definitions and provided suggestions for further improvement. Representatives from WP01-WP05 presented an overview of recruitment and phenotyping strategies at their centres and demonstrated their respective CRFs and databases, aiming to exchange experiences, to streamline patient recruitment and phenotyping and to highlight differences that might impact the project's outcome. CR01 and CR04 related their initial experience with the CRF and provided suggestions for further improvement as well as for eCRF development. CR10 further discussed eCRF development and potential integration of different existing database formats. CRs 01 and 02 then presented a number of their own, complex clinical cases in order to highlight potential differences in interpretation, provide further testing of the CRF and homogenize phenotyping. All CRs were urged to start phenotyping their cases using the CRF. Following the workshop, a summary section was added to the CRF (by WP01) to allow easy identification of cases that could be useful for the different work packages and thus reduce the amount of screening required to complete the cohorts.

We established two processes to evaluate use of the CRF and eCRF, determine cross-centre consistency of data entry and classification, understand inter-rater reliability of interpretation of the definitions, and to get an idea of cross-centre differences in epilepsy care & treatment. The first consisted of a phenotyping exercise, for which CR01 selected four sets of clinical records from patients with complex epilepsies, fully anonymized them and circulated them to the other CRs involved in phenotyping. Each of the CRs was then asked to phenotype these anonymized cases using the paper CRF and according to their own interpretation of the definitions, and send the results back to CR 01. CR 01 drafted a summary report based on phenotyping results received from CRs 02, 04, 07 and 08. These showed very good cross-centre agreement for recording most of the data, and highlighted some differences in interpretation that were addressed with centre-specific and WP-specific recommendations and guidelines. We decided that: (a) in order to reduce errors in interpretation, patients should be classified for suitability for different WP by the person who performed the phenotyping; (b) although phenotyping definitions should be applied strictly, the phenotyper would have discretion for particular aspects e.g. interpretation of AED doses in the context of patient age (e.g. paediatric cases) or co-medication etc.; judgement on end dates of ADRs; adequacy of AED trials in the absence of detailed data on seizure frequency, dates or dose; taking into account provoked seizures etc; all provided there is clear clinical evidence; (c) CR10 would work closely with WP leaders to set up semi-automated classification of patients in the eCRF; (d) phenotyping quality control was repeated at intervals.

WP07, working with partners CR01, CR04, CR07, CR09, CR10 and CR13 amongst others, generated the eCRF, a page from which is shown as a screenshot below.



Aided by several of the other participants, participant CR02 put together a manual for the eCRF, providing practical instructions (e.g. flowcharts) in order to facilitate and homogenize data entry by the different participants. This manual was completed and circulated to all participants in February 2014. For the sustainability of EpiPGX work beyond the funding period, both the eCRF and its user manual will be extremely helpful.
In addition, additional phenotyping, data collation and troubleshooting tasks were hosted and supported by a secure web platform with links to analysis tools, which was expanded to other reporting, e.g. on the state of genotyping.
Having established systems for phenotype collation, we have now finely honed these to the extent that they have become attractive to external groups, as have our case definitions, which seem set to establish standards within the community. We have undertaken additional unique exercises in cross-centre phenotyping, sharing fully anonymised raw data (case notes) to determine cross-centre reproducibility, and continued phenotype quality control through the project. These efforts will be valuable to a community moving increasingly to international collaborations. The results of previous phenotyping quality exercises have been further analysed and discussed extensively, yielding some further suggestions to improve phenotyping homogeneity. The phenotyping manuscript is currently being re-reviewed by a medical statistician for further analysis of the data.
Patient phenotyping has been impressive: numbers of patients in the eCRF increased from 7,150 in October 2014 to over 12,000 by the end of the funding period, with 4,800 fully phenotyped patients in October 2014 and 7,750 in October 2015. We generated monthly reports on phenotyping.
Our success has encouraged further large-scale multi-collaboration efforts, accelerating discovery. Our web-based database has been adopted by others, as have our sound, uniform phenotype definitions, facilitating further research beyond EpiPGX.

2. To identify genome-based biomarkers of response to the first, later and to multiple current AEDs (broad drug resistance).

Identifying biomarkers predicting the pattern of response to the first well-tolerated AED will be of high clinical importance for people with newly-diagnosed epilepsy. It is at this point that critical life choices have to be made by most people with epilepsy. Most importantly, were a set of variants to be identified at the outset predicting with high reliability that an individual was unlikely to respond to several AEDs, early referral for specialist evaluation would be possible. This was a large part of the project and was spread across three main workpackages that have undertaken mutually informative work and will continue to work together to generate novel pharmacogenomics findings. The main workpackages involved were WP02, WP03 and WP04.

WP02 is headed by CR12 (ULIV) and CR14 (UGLA) and CR13 (IMP), with work coordinated from CR12. The principal objectives for WP02 are to identify genomic biomarkers of clinically-relevant treatment outcomes following initial antiepileptic drug exposure in patients with newly-diagnosed epilepsy. WP02 investigators have previously accrued DNA samples and clinical information from large cohorts of epilepsy patients, many of whom have been followed up prospectively from initial diagnosis. A proportion of those cases had already undergone GWAS analysis at the Wellcome Trust Sanger Institute using funding from other sources. We also had access to resources, including genotypes and detailed phenotypes, from a unique, prospective epilepsy pharmacogenetics study at the University of Melbourne, Australia.
WP02 has been key in the coordination and completion of phenotyping required for analysis in all three tasks within WP02 and in other WPs within EpiPGX. This involved manual entry of data from hospital notes of more than 1,500 people with epilepsy in both Liverpool and Glasgow. In addition, data from a further 1,500 people with epilepsy which was held in existing electronic databases, constructed for the purposes of previous studies, has now been transferred to the EpiPGX eCRF. This onerous task required lengthy computer scripts to automate the transfer of data between databases and to ensure that clinical variables were accurately mapped and that coding of data was maintained. The EpiPGX eCRF contains more than 500 potential data-fields for every patient and each of those had to be translated individually from the existing databases, of which there were several and all in different formats. Data transfer was kindly supported by colleagues in WP07.
A key component of the intended analysis in WP02 was to first establish the relative influence of demographic (i.e. age, sex) and clinical (i.e. epilepsy type) factors in the variability in treatment outcomes in newly-diagnosed epilepsy. Understanding and adjusting for these factors (or covariates) should allow a more sensitive investigation of the relative contribution of genomic variants in the subsequent genome-wide association analyses. This essential work

was undertaken using a logistic regression method to identify significant non-genetic predictors of treatment outcome and to quantify the extent of their influence.

WP02 had three main tasks, all completed, with data analysis continuing and manuscripts in preparation.

Task 1: Identifying genome-based biomarkers of remission with first well-tolerated drug

All phenotype and genotype data for this analysis were assembled by end of 2014. Subsequent quality control checks were implemented in early 2015. These resulted in a loss of some cases due to either missing phenotype information essential to the determination of treatment outcome or missing genotype information above a pre-determined ceiling for inclusion. The final genome-wide association study (GWAS) was undertaken in a population of 1,514 individuals with newly-diagnosed epilepsy, who had been followed-up prospectively at a single epilepsy centre from initial diagnosis and treatment initiation and until such time that they reached an efficacy end-point associated with their first well-tolerated AED. Patients were stratified into those experiencing an immediate remission from seizures, those experiencing a later (or deferred) remission, and those who did not experience remission on their first AED. This population was then interrogated using a variety of statistical approaches in an effort to identify genomic variants associated with treatment outcome, using binary, multinomial and survival GWAS methods. The latter approach (survival GWAS) necessitated the development of novel statistical methodology – three-way mixed modelling GWAS – which was outlined in the previous periodic report and has been the subject of dissemination activities.

Task 2: Identifying genome-based biomarkers that distinguish general and selective drug responsiveness

Chronologically, this was intended as the final task for WP02. It was originally designed as a collaborative effort between WP02 and WP04 and reliant on the identification of genome-wide significant biomarkers of early and late remissions in WP02 Task 1 and WP04 Tasks 1 & 2, respectively. Developing our approach, we will also determine whether the underlying biology of drug response is such that it requires collaboration across WP02, WP03 and WP04 to explore response versus non-response to specific medications or classes of medication, with analyses adjusted according to whether response is early (first drug) or late (second or subsequent drugs).

Task 3: Identifying genomic biomarkers of first drug failure

The patient cohort for this analysis is essentially identical to that for Task 1. As described above, all genotypes and phenotypes had been assembled by end of 2014 and subsequent quality control checks were implemented in early 2015. Some cases were again excluded due to missing phenotype and/or insufficient genotype data. For this analysis, the final GWAS drew on a population of individuals with newly-diagnosed epilepsy who had been followed-up prospectively at a single epilepsy centre from initial diagnosis and treatment initiation and until such time as their first ever AED failed. Failure was defined as withdrawal of the first drug and/or addition of a second drug. Patients who did not experience drug failure were censored at the time of last recorded clinical visit. Patients who experienced failure of the first drug were stratified into those failing due to unacceptable adverse events and those failing due to inadequate seizure control. This population was then interrogated with a novel two-way competing risks GWAS approach, which was developed specifically for EpiPGX.

WP03 focussed on broad AED resistance. WP03 worked extensively with WP01, and the other work packages, to achieve consensus definitions for the phenotypes in EpiPGX, including the drug resistance phenotype. We based the definition of drug resistance on the International League against Epilepsy (ILAE) consensus proposal which defines drug resistance as a failure of adequate trials of two tolerated and appropriately chosen and used antiepileptic drug schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom. By this definition, individuals with very rare seizures (for example 1 seizure/year) can potentially be classified as drug resistant. To maximize the chances to identify biomarkers of clinically meaningful broad drug resistance, the members of the Consortium agreed to modify the definition as follows: seizures recurring at a frequency of ≥4/year over the year preceding the latest data entry, despite adequate trials of ≥2 tolerated and appropriately chosen and used AED schedules, whether as monotherapies or in combination. We defined drug responsiveness as freedom from seizures for ≥12 months up to the latest recorded visit.

The phenotyping workshop held in April 2012 provided an opportunity to foresee, discuss and solve some of the issues regarding the definitions and the case record form (CRF) used to collect phenotypic information. CR01's contribution at the workshop was the presentation of the first 50 fully-phenotyped cases. The risk of ambiguous interpretations of the same phenotype was considered, and further clarifications were provided with the aim of minimising differences in phenotyping across the centres. For WP03, it was agreed that individuals with non-epileptic seizures and individuals known to be non-adherent should be excluded. It was also agreed that cases which had undergone surgical treatment for drug-resistant epilepsy should be included in the drug-resistant group, provided they had fulfilled the criteria for drug resistance prior to surgery. Several improvements of the CRF were suggested to allow recording of essential information for WP03 and other work packages. As part of the detailed phenotyping, we

noted additional findings on the stability of phenotypes over time and on new methods for using longitudinal data for classification. CR01 contributed to the quality control phenotyping exercises in May 2013 and August 2014 by providing anonymised cases that were circulated to the other centres and phenotyped by several phenotypers. The outcome of the phenotyping exercises was presented at the General Assembly meetings in November 2012, October 2013 and October 2014. There was good agreement regarding classification for WP03, Task 1 across the consortium. Both phenotyping exercises provided the opportunity for the Consortium members to fine-tune the definitions and data collection.

WP03 had three main tasks, all largely completed, with data analysis continuing and manuscripts in preparation.

Task 1: Generation of cohorts of suitable cases and controls for analysis.

All cohorts for analysis have now been collated, with the benefit of the inclusion of additional cohorts, including from IGG (CR03), and from cases phenotyped and genotyped in a previous EU project, EpiCURE. In total, there are available for the different components of WP03: >2800 patients meeting criteria for drug resistance, >1800 patients meeting criteria for drug responsiveness, and >9000 unscreened genotyped controls from several different cohorts. More than 200 drug-resistant MTLEHS cases for Task 3 (many more than anticipated) have whole genome sequence data available for analysis (only whole exome sequence data was planned), through a separate project with an independent funding stream; 50 cases of MTLEHS that are drug-responsive will have whole exome sequence data available, also exceeding the number planned. This task is complete. As data will be meta-analysed, the option will remain open to add additional cohorts from outside EpiPGX if these become available.

Task 2. Identification of common genetic variants as biomarkers via genome-wide association study (GWAS) of phenotype (broad drug resistance) against genotype (SNP/CNV).

We have completed the first stages of the planned GWAS, comparing patients with drug-resistant epilepsy to healthy controls, and to patients with drug-responsive epilepsy. GWAS was undertaken according to established procedures, with enhancements that represent the current state-of-the-art and were not foreseen when EpiPGX was written and funded. We used a protocol published in the setting of common variant susceptibility studies in epilepsy, which will promote inter-operability of these datasets, so favouring further discovery (protocol published in International League Against Epilepsy Consortium on Complex Epilepsies. Lancet Neurol. 2014;13:893-903). For example, allowing greater coverage of common genetic variation, data from all cases and controls were imputed using IMPUTE2 with 1000 Genomes (July 2011 release) as the reference set, generating >5,000,000 SNPs/case overall before quality control. Healthy controls were matched for ancestry. For association analysis, we utilized a linear mixed model using FaSTLMM. The same protocol has been used to compare patients with drug-resistant epilepsy with those with drug-responsive epilepsy.

Task 3. Identification as genome-based biomarkers of shared rare variants and genes with increased burden of individual variants via high-throughput sequencing in a selected cohort of individuals with extreme phenotype.

We used the following definition for extreme drug resistance: seizures recurring at a frequency of ≥4/year over the year preceding the latest data entry, despite adequate trials of ≥5 tolerated and appropriately chosen and used AED schedules, whether as monotherapies or in combination. Drug responsiveness was defined as freedom from seizures for ≥12 months up to the latest recorded visit.

WP04 focussed on Pharmacogenomic biomarker discovery for late response to specific antiepileptic drugs. Its results will contribute to novel joint analyses with WP02 and WP03.

3. To identify genome-based biomarkers of variable response to selected individual AEDs.

The AEDs upon which we have chosen to focus are those most commonly prescribed in the EU, that feature most prominently as recommended for first-line therapies in partial, generalised or unclassified epilepsies, and that account for the largest AED prescription cost in the EU. The selected AEDs are often more cost-effective than newer drugs. Therefore, information allowing their improved use is likely to have a major impact on healthcare costs. There is currently no way to determine which AED might work for an individual; this is the aim of pharmacogenomics, and would be a major benefit of EpiPGX.

WP04 has mainly focussed on this task. The most important aspect of WP04 in the first phase of the project was the definition of phenotypes in collaboration with WP01 and the creation of an electronic database in collaboration with WP07. Both activities needed close collaboration with all clinical partners to consider all important aspects of phenotyping and to unify and harmonize the different views of all project partners. Every partner had been already

recruiting patients for many years and has had their own system of classifying pharmacological response and failure, their own databases and ethical considerations. Thus, harmonizing all participating groups was an important achievement for WP04 Tasks 1-3. Moreover, CR04 extended its recruitment campaign nationwide throughout Germany, including 18 centers (several new ones since start of EpiPGX) for systematic recruitment of patients with focal and idiopathic generalized epilepsies. WP04 established standardized recruitment procedures and obtained ethical approval in all those centers.
Subsequently, samples with relevant phenotypes were genotyped and with regard to the identification of genome-based biomarkers for response/non-response to specific AEDs, WP04 focussed on AEDs that are in broad use: i.e. lamotrigine (LTG), levetiracetam (LEV) and valproic acid (VPA) for idiopathic generalised epilepsies (IGEs; also known as genetic generalised epilepsies, GGE) and lamotrigine (LTG) for focal epilepsies (FEs). Furthermore it was decided to not only compare late responders and non-responders, but to increase the available analysis cohort by including all responders.
WP04 had three tasks.
Task 1: To identify genome-based biomarkers for late response/non-response to specific AEDs in focal epilepsies
The GWAS for the FE cohort has been completed for LTG response yielding several suggestive association hits A hypothesis-driven re-analysis for both early and late responders will be run with ADME genes and after all the exomes currently being sequenced have been analysed.
Task 2: To identify genome-based biomarkers for late response/non-response to AEDs in generalized epilepsies
The GWAS for the IGE cohorts have been run, yielding several suggestive association hits ($P<10^{-5}$). Samples from WP02 and WP03 were included in this analysis as planned. As for Task 1, WP04 will re-analyse data after with an approach focused on ADME genes and again when exome analyses have been completed.
Task 3. To identify common genome-based biomarkers for late drug response.
These analyses are also ongoing, including 1) IGE samples with response/non-response to LTG, VPA and/or LEV; 2) IGE and FE samples with late response to LTG, LEV and/or LCM.

4. To identify genome-based biomarkers of specific adverse reactions (ADRs).

Pharmacogenomic guidance leading to avoidance of serious, even life-threatening, ADRs is of obvious benefit to individuals. Pharmacoeconomic estimates are difficult to model precisely in the absence of data on predictive value, but we are aiming to identify high-risk variants such as those for carbamazepine-induced rash, which together with rapid inexpensive new methods for genotyping are likely to make discoveries from EpiPGX not only clinically important, but also economically viable. GWAS and exome sequencing studies are underway, and have started to identify variants of potential clinical utility.

WP05 led this work. The first step in this process was to agree case and control definitions for each of the individual adverse drug reactions (ADRs) that we had proposed to study. The phenotype definitions also provided for the design of the eCRF developed through WP07. The eCRF provides the electronic infrastructure for the secure, centralised storage of patient phenotypes.
Phenotyping has led to the identification of over 2,000 adverse drug reactions that satisfy each of our 'strict' criteria:

- Occur within 6 months of initiation of AED (N/A for visual field defects)
- Lead to dose reduction or withdrawal of AED where appropriate
- Reverse or improve after dose reduction or withdrawal, where appropriate (not for visual field defects)
- Are not attributed to another cause by treating/phenotyping clinician.

Relaxing the criteria to the ADR being considered attributable to a specific AED by the treating clinician or phenotyper (i.e. the 4th condition above only), led to identification of 4,845 ADR events.
In addition to phenotype definitions and sample collation, WP04 had three main tasks.
Task 2: Generation of genome-wide genotypic datasets.
This task has been completed. Patients with a report of an ADR matching our strict or loose criteria were selected for genotyping using a dense genechip platform. All new genotyping took place at deCODE using the Illumina OmniExpress 12v1.1 chip. Working closely with partner CR09/deCODE, all GWAS datasets have been imputed to the latest reference from the 1000 Genome Study, and the resulting data have been assembled on a server at partner CR10/LUX.
Task 3: Assessment of common genetic variants as biomarkers for each ADR via genetic mapping across

cases and controls.
This task is now complete. WP05 assembled the imputed EpiPGX dataset on a secure server at partner CR10/LUX, and matched these genotypes to various ADR-related phenotypes specific to WP05. Part of this work was conducted in collaborations arranged through the International League against Epilepsy Consortium on Complex Epilepsies. This was for work relating to a meta-analysis on aromatic AED-associated skin rash.
Task 4: Assessment of rare genetic variants as biomarkers for a specific ADR via high-throughput sequencing of selected cases.
Working with partner CR09/deCODE, WP05 generated exome sequence data on several hundred WP05 ADR cases, of whom many had experienced a severe cutaneous ADR to CBZ, LTG, PHT or OXC. WP05 also established a collaboration with the Canadian Pharmacogenetics Network for Drug Safety (CPNDS), to work on exome-sequence data related to severe cutaneous ADRs attributable to AEDs. In addition to cutaneous ADRs, we have generated exome sequence data on hundreds of patients experiencing vigabatrin-induced visual field defects, hepatoxicity (due to any AED), carbamazepine-induced hyponatraemia, topiramate-induced speech disorder, levetiracetam-induced behavioural disorders, neutropenia (due to any AED), topiramate-induced cognitive impairment and valproate-induced tremor. Screened, drug-tolerant controls will be selected from the overall pool of exome-sequenced cases. As with our GWAS datasets, we have developed a pipeline to process and store all historical and EpiPGX-generated exome datasets on the server at partner CR10/LUX.

5. To identify genome-based biomarkers of teratogenesis associated with valproate use in pregnancy, to inform the difficult decision-making process of drug choice in certain epilepsies in women of child-bearing age.

Recent guidance and regulations from the European Medicines Agency, in the light of new data and analyses of the risks associated with VPA use in pregnancy has added impetus to this component of EpiPGX research. Knowing which women are at higher risk for having an adverse pregnancy outcome (major congenital malformation, MCM) due to valproate would reduce fear of its use, so that more women could be seizure-free with all its benefits. In practice, this is a first step in the process of selection of women able to take valproate for epilepsy without running predictable high risk of an adverse outcome, as learning disability can occur without MCM. At a biological level, identified high-risk variants would reveal teratogenesis mechanisms, useful for the population as a whole.

This work was led by WP06. In order to establish a resource of DNA and clinical data of newborns suffering from AED-induced MCM and their parents for genetic and epigenetic analysis, WP06 lead (CR08 – Belfast Health and Social Care Trust (BHSCT)) first had to seek ethical approval that would permit us to recruit cases and controls, that had previously been recruited to UK Epilepsy and Pregnancy Register (www.epilepsyandpregnancy.co.uk), or which continue to be recruited. The majority of cases and controls for study in WP06 came from this already established resource. Having agreed a protocol for WP06, to include the phenotypic data to be collected, with our partners (CR01, CR02, CR07 and CR11), a letter of invitation for participants, patient information sheets, patient consent forms, a questionnaire for completion and covering letters for general practitioners and other healthcare workers were devised. Ethical approval was applied for, after sponsorship had been provided by the BHSCT, using the UK's Integrated Research Application System (IRAS).
In regards to phenotyping, working with partners CR01, CR02, CR05, and CR11, WP06 agreed those phenotypic details that would be included for study and agreed definitions that would be used for the project duration, to include AED-specific ADRs/malformations. This was been achieved during a combined phenotype/training workshop, an annual clinical workshop to ensure uniform phenotyping.
Within this workpackage, over 700 subjects were been identified for whom complete phenotypic data and DNA are available. Sequencing, typing and analyses are underway.

6. To establish a virtual test-bed for new AEDs/ Phase IV studies.

EpiPGX has delivered a huge curated dataset in epilepsy, with over 32,000 individual drug response phenotypes. As many people remain under active follow-up, we can return to the person for specific follow-up as indicated by the results, representing a virtual epilepsy cohort with real potential for testing of new and licensed AEDs in silico, without the risks/costs of exposure to people at an early stage. We have had further industrial interest from this perspective. As analyses are completed, we envisage further use of the collated data for these purposes.

7. To promote dissemination, in society, health services & industry.

With scientific progress, ethical, social & legal issues must also be considered. Epilepsy management must always be holistic. We are ideally placed to discuss important issues that arise, informing global debate on these issues given the rich dataset we will have and understanding we will develop. We hope to share impact with patients through our wide links with national & international patient organisations. The dissemination is detailed in Section 2.

1.4 The potential impact

Socio-economic impact and the wider societal implications of the project

Contribution to Community and social objectives

EpiPGX has been active in community objectives, using social media to spread information about its work and engage with the community.
For examples, see:
http://bit.ly/1k9rkKu
http://bit.ly/1jHESwo
and related comments.

Main dissemination activities and exploitation of results

EpiPGX has been active in dissemination. As noted above, results are only now beginning to come through, at the end of the funding period and we anticipate there will be clinical exploitation of the results for the benefit of people with epilepsy within the foreseeable future. In the meantime, our dissemination activities have been widespread and have helped us both develop the group and achieve its aims. Dissemination has been at the level of scientific organisations, as well as scientific meetings and open meetings to the general public.

We aimed to reach all the potential audiences, in particular patient organisations and public health authorities. As agreed with our Project Officer, we arranged an Open Day on EpiPGX for this purpose, reaching a broader range of interested parties than would have been possible at a conference satellite meeting (for example, the Open Day was free, whereas conferences require registration fees to be paid by attendees).

As detailed in each WP and the Dissemination details in Section 2, there has already been dissemination of information about EpiPGX to many audiences. Publicity has included:

- Conferences (scientific and open to lay people)
- Presentations
- Workshops
- Posters
- Tweets under the #EpiPGX hashtag
- An image video on youtube: https://www.youtube.com/watch?v=CwGtrz4XdOc
- Features on the channels of the Epilepsy Society on facebook, Twitter and youtube
- Website updates

Outlook and future research

The outlook for EpiPGX is very promising. We have established a well-functioning Consortium, with mutual understanding and cooperation in the search for genomic variants that could be used to improve treatment and care for people with epilepsy. By the official end of the project, the team had worked together through many difficulties and challenges, and had resolved them with input from the experience and expertise of the whole group. Data collection had taken longer than anticipated and delays beyond the Consortium's control in obtaining timely genotyping meant that data became available really only towards the end of the project, with momentum building right until the official end of the Project, and indeed beyond that time as the full scope and power of the data become apparent. Building on this momentum, the Consortium has chosen to continue as a collaborative entity, with a revised Consortium Agreement to supplement that already in existence, with the clear intention to pursue the original plans, but also to develop new research ideas based on the mass of data available. The Consortium also intends to apply for further funds from national and international funding agencies to capitalise on the value of the gathered data and experience. Part of this planning for the future includes ensuring the security, resilience and accessibility of the existing phenotypic, genotypic and sequencing data for the future use both of the Consortium and new collaborators. We hope that these efforts will promote continued discovery and further use of the valuable resources developed by EpiPGX.

Section 2 – Use and dissemination of foreground

Please see ECAS.

Section 3 – Report on societal implications

Please see ECAS.