DECIPHER Project Proposal

Aim

Rare genomic variants can occur anywhere in the genome, but changes occurring at a particular locus are very rare. The DECIPHER project was conceived as a clinical and research tool to:

1. Aid in the interpretation of data from genome-wide analyses eg. the differentiation between pathogenic and benign genomic variants
2. Utilise the nearly completed human reference sequence via Ensembl and other genome browsers to define which genes are involved in a specific copy number variant (microdeletion / microduplication) and for sequence variants, whether they are positioned within a gene or regulatory element.
3. Facilitate research into the study of genes that affect human health and development to improve diagnosis, management and therapy of rare diseases.

Background

There is a genetic basis in many cases of developmental delay, learning disability and / or congenital malformation. For example, Down syndrome is caused by having an extra chromosome 21. Conventional chromosome analysis, which involves looking at chromosomes under a light microscope after special staining, shows smaller abnormalities in other cases, eg. bits of a chromosome are deleted, or duplicated, or moved onto another chromosome (termed 'chromosome imbalance'). However, using these methods of analysis many cases of developmental delay, learning disability and / or congenital malformation remain unexplained, although there is a strong suspicion that genetic abnormality is responsible. Genomic array analysis is a very promising new technique for use in this situation, with much improved resolution and sensitivity when compared to conventional chromosome analysis, ie. it effectively takes a higher-powered 'photograph' of the chromosomes (Vissers, Shaw-Smith). The 1 Megabase array has more than 3,000 clones (meaning that the resolution of the chromosome picture can be imagined as having >3,000 'pixels' rather than 650-850 'pixels' of conventional techniques) and the "tiling array" has >32,000 'pixels'. This offers the potential for discovery of a huge number of different disorders of chromosome imbalance and polymorphic variants (changes observed in normal individuals). However, since individual microdeletions/microduplications identified by genomic array analysis are very rare, and knowledge of polymorphic variants at this level is in its infancy, an international collaborative effort that allows pooling of data will advance understanding more rapidly.

Exome or whole genome sequencing is now widely adopted in research practice and moving fast into clinical practice. Just as for the copy number variants described above there is a pressing need to share data on sequence variation in order to understand the significance of such variation for human health and development.

Collaboration

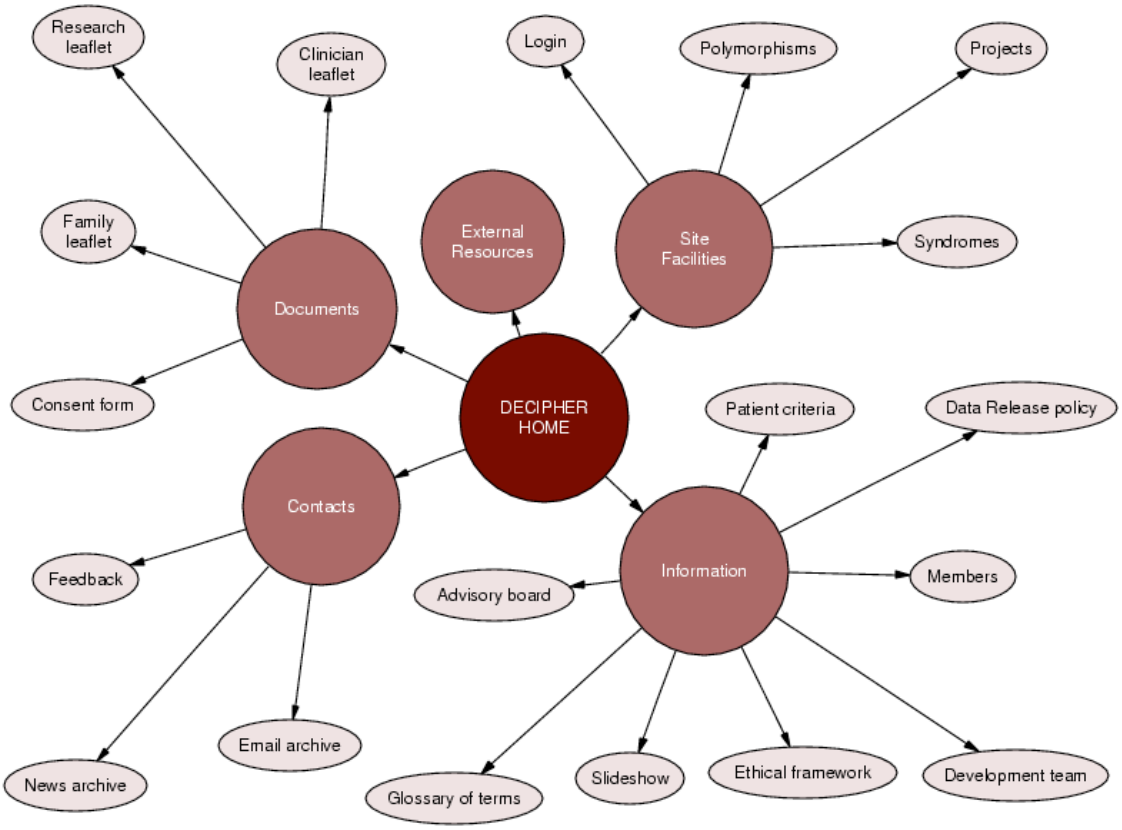
DECIPHER is a tool for collating information about very rare genomic variants. Results of genomic studies, which have been undertaken either during the clinical evaluation of a patient or as part of an approved research study or as part of the DECIPHER research array project, can be entered into DECIPHER. DECIPHER it is a vehicle to aid clinical geneticists in the interpretation of results. The introduction of exome and whole genome sequencing alongside genomic array technology to clinical practice over the next few years will frequently place clinicians in a position where they are trying to interpret novel results which they have not personally encountered before, and which have not been seen by other clinicians or scientists within their own department. It is imperative that clinicians and genetic scientists are able to draw on and share the experience of clinicians and scientists in other centres in order to:

- Aid in the differentiation of pathogenic (disease causing) and polymorphic (normal variant) genomic variants including copy number changes and sequence changes.
- Aid in the delineation of new genomic disorders to facilitate genetic counselling and patient management

DECIPHER has a global network of >200 academic centres of genetics contributing to the project

Design

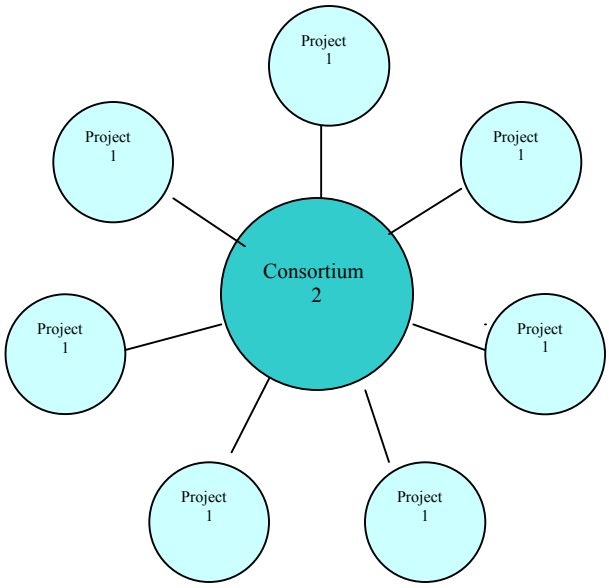
DECIPHER is an interactive web-site, which can be found at https://decipher.sanger.ac.uk/.



Ethical and legal framework

Each centre using DECIPHER must consider whether they require approval from a research ethics committee in their institution/country and must abide by any laws relating to data protection / consent / confidentiality and professional ethics in their institution / country when entering data into DECIPHER. Obtaining appropriate consent to enter phenotype data into DECIPHER is the responsibility of the submitting clinician. Information leaflets and sample forms drawn up for use in the UK may be downloaded from the homepage (http://decipher.sanger.ac.uk) but may need adaptation/translation to comply with regulations prevailing in individual institutions/countries outside the UK. We are happy to make any leaflets/consent forms which have been adapted/translated for specific projects available via the DECIPHER homepage.

Structure of DECIPHER



Project (1)

In individual project domains, DECIPHER is used exclusively as a clinical tool to aid in the interpretation of a high resolution genomic analysis eg. microarray analysis, exome sequencing or whole genome sequencing. The data is held in a linked anonymised form within the closed, password protected domain of an individual clinical genetics department that is affiliated to DECIPHER as a submitting centre. Data held within a project domain is only visible to nominated clinicians / cytogeneticists within that centre who have legitimate access to the data and are logged into DECIPHER or to nominated clinicians / scientists who are part of a datagroup within DECIPHER (eg. the NHS Datagroup)

Managed data access (2)

Linked-anonymised data is within DECIPHER where it becomes visible to some or all centres within the consortium to accelerate progress in the delineation of new syndromes and of gene function.

Public access (3)

Linked-anonymised data is released into DECIPHER-Ensembl view (and other genome browsers and genome variation databases) where it becomes visible to all projects and to clinicians and researchers worle-wide to accelerate progress in the delineation of new syndromes and of gene function.

The process

Guidelines for data entry and data sharing (Please read in conjunction with the data-flow chart)
Confidentiality and security

Great care has been taken in the design of DECIPHER to protect patient privacy. Broadly speaking, two levels of access are provided for. Parties issued a user name and password, which allows them to *log in* to the system, will have the highest level of access. They will be able to examine phenotype and genomic data, and identify which participating centre entered a particular phenotype. They will not be able to identify individual patients, except where it is a patient from their own clinic. Other members of the public will have low level access allowing them to *browse* phenotype and array data in DECIPHER-Ensembl view and similar genome browsers. At this level, patient information is linked-anonymized.

Only trusted individuals in trusted centres will be issued with usernames and passwords to log into DECIPHER. This information should not be shared with persons at other participating centres, nor disclosed to persons not authorised to log into DECIPHER. DECIPHER is served over an encrypted SSL

(secure socket layer) connection which is the industry standard for secure connections, similar to that used by banks and building societies to operate on-line banking systems.

Stage 1 – Genomic variant data

Genomic variant data generated from a sample by the local laboratory is entered into DECIPHER by the scientist or clinician in the participating centre for a given project, together with an external reference number (to enable decoding of the sample identity by the patient's own clinician/research team) and a DECIPHER ID is generated. The clinician contact in the participating centre enters the phenotype using restricted terms derived from the Human Phenotype Ontology (HPO) Access to the data table for an individual project is restricted; it is only possible for members of a given project who have a user name and password to log into DECIPHER. Data held in this form is highly anonymised and the potential for identification of an individual by a third party is negligible. Only those persons who have access to the encryption key would have the opportunity to identify the patient. Participating centres should ensure the encryption key is securely held by a person who is legally obliged to respect the confidentiality of the individual's identity and is otherwise authorised to view the information (for example, a member of the clinical team).

Stage 2 – Release into DECIPHER Ensembl view

Consent. When a genomic result has been communicated to the individual or their parent/guardian, the clinician or research nurse seeks consent for summary information regarding the genomic variant and the phenotype to be released into DECIPHER-Ensembl view. In addition to the Ensembl genome browser, such data may also be displayed in other genome browser eg. UCSC or shared with other genome variation databases eg. ClinVar. Consent should be sought from the person with legal authority to consent. In the case of a competent adult, this will be the person themselves. Where the person is a child, it may be the child or a person with responsibility for parenting depending on the child's age and level of competence. With older children it may be advisable to obtain the consent of both. The project may also be introduced by a research team or the patient support group 'Unique' www.rarechromo.org .The clinician, research nurse or support group officer should outline the aims of DECIPHER and give the individual and/or their parent/guardian a copy of the leaflet 'DECIPHER – information for families'. Written consent is then sought by the patient's clinician or an appropriately trained research nurse, which is stored in the patient record. If consent is not granted, and the consent box is not checked it is not possible to release data into DECIPHER-Ensembl view and other genome browsers. If consent is granted and the consent box is checked, the genotype and phenotype data will become visible in the DECIPHER track in Ensembl. It will then become possible to see that data (but no other information from that individual's patient record). It will also be possible to see whether other individuals with the same chromosomal change are known to DECIPHER, and to compare their phenotypes. The addition of brief phenotype information increases the possibility of subject identification which is why patient consent is sought at this stage. Individuals will have a right to withdraw their consent in accordance with the law and as indicated in the information leaflet 'DECIPHER – information for families'. Parents/guardians who have given consent for their child's record to be entered in DECIPHER are requested to ensure that when the child reaches the age of 16 yrs he/she is made aware of his/her DECIPHER entry so that he/she can make his/her own choice whether to continue or withdraw the entry.

Confidentiality. In order to minimise identification by a third party, only registered clinicians/clinical scientists and researchers who have log in rights to DECIPHER are able to see to which participating centre/project a patient belongs. Non-registered clinicians/scientists would thus have no access to information regarding the locality of the patient eg. if the patient data was submitted by a DECIPHER project in the UK, Europe or US. If a registered clinician wishes to follow-up a patient known to DECIPHER with a clinical member of the submitting project/participating centre he/she will use the (coded) DECIPHER ID. A registered clinician in the submitting project/submitting centre would be able to log-in to their own project, decode the DECIPHER ID and see which patient corresponds to the external reference number for their centre. They can then get back in touch with their own patient and seek their consent to provide more detailed information or consent for publication as appropriate. The facilitation of such contact between clinicians and scientists is essential to the delineation of new syndromes and proper sharing of clinical data.

Ensuring the quality of data entry in DECIPHER
Any data sharing project relies on the goodwill and professionalism of all contributors. The facility to enter data to DECIPHER is therefore restricted to an experienced clinical or research scientist and a clinician with expertise in genetics working within each participating centre. Usually, the clinical scientist will be responsible for entering the genomic data and the clinical geneticist for entering the phenotype data. DECIPHER has been specifically designed to enable data entry to be as quick and straightforward as possible. Once the system has been customised for your project/participating centre, data entry should take only a few minutes. For the benefit of all users of DECIPHER, data entry needs to be as complete and accurate as possible.

Patients' right to access data
Individual patients will be able to browse the DECIPHER database like any member of the public. In addition they will have the rights to access information as provided in law. For example, under the Data Protection Act 1998 (UK), which is based on European Directive on the protection of personal data (95/46/EC), a patient may request access to information that identifies them. To obtain access, the patient would need to contact their clinical geneticist, since only he/she has access to information that identifies them. The clinical geneticist is in the best position to explain the information to the patient, and to determine whether there are any reasons why access should not be granted. The clinical geneticist will be solely responsible for deciding whether to contact individual patients, if and when new genetic risk information is determined that might be relevant to the patient. The patient's interest in knowing and not knowing information about themselves should be balanced with reference to legal and ethical considerations.

Organisation of the DECIPHER CONSORTIUM
The DECIPHER database will be managed by person or persons nominated by the DECIPHER development team who will take advice from the 'DECIPHER Scientific Advisory Board'. The advisory body will be comprised of a representative of the development team and 4-6 people invited by the organisers. The organisers will endeavour to invite a person or persons trained in law, ethics or social science; a person with firsthand experience of genetic testing either as a patient, or a patient's carer; and a person trained in clinical genetics or employed by a participating centre.

Terms and Conditions of depositing and accessing data contained in DECIPHER
No monetary fee will be charged to access or deposit data in DECIPHER, in view of the public interest there is in better understanding gene function. However, the proper use of this resource is fundamental to its success, and users are required to note and observe the following terms and conditions:

- Persons depositing and accessing data must observe the guidelines above.
- Whilst the organisers of DECIPHER will take reasonable steps within the resources available to them to protect the privacy of patients' and participating centres and the quality of the data, they can accept no legal liability for the use or misuse of the information.
- Authors who propose to publish material which uses data obtained from the DECIPHER database should:
    - acknowledge the DECIPHER Consortium; and
    - contact the main contact for the project/participating centre that entered the data on any individual who they wish to include in their report (whether identified or not) and offer co-authorship to at least one representative from the project/participating centre (preferably the member who submitted the patient data).

The organisers reserve the right to unilaterally alter these terms and conditions, as necessary or desirable in the public interest or to protect the individuals whose information is included in the databank, [or for any other reason].

References
Vissers LEM, de Vries BBA et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet. 2003;73: PMID: 14628292

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, Bobrow M, Carter NP. Microarray-based comparative genomic hybridization (array-CGH) detects submicroscopic

chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 2004;41:241-8

Flint J, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. Nat Genet. 1995 Feb;9(2):132-40.

B B A De Vries, R Winter, A Schinzel and C van Ravenswaaij-Arts. Telomeres: a diagnosis at the end of the chromosomes (Review). J Med Genet 2003;40:385-98