Deciphering Developmental Disorders (the DDD study) - a UK-wide project to transform the diagnosis of rare childhood disorders

Helen V Firth DM FRCP DCH
Cambridge, UK
DDD Study
Deciphering Developmental Disorders

Objectives:

• Understand genetic architecture of severe developmental diseases
• Optimise clinical implementation of genomic technologies for diagnosis

A UK-wide collaboration:

• 12,000 patients, their families, all NHS Regional Genetics Services (>150 clinicians) and the WTSI

Details:

• Record clinical phenotypes systematically
• Apply exon-array and exome sequencing
• Filter variants against panel of known genes and partition into pathogenic (clinical) and possibly pathogenic (ongoing research)
• Identify and feedback causal variants to NHS via DECIPHER (linked) >100 new diagnoses so far
• Display likely pathogenic variants in DECIPHER to facilitate ‘match-making’
• All data available for research, externally via EGA (anonymised)
Recruitment strategy

- Current diagnostic success in patients with developmental disorders is poor
- Many patients are undiagnosable with current methodology in routine use (~30%) – maximum benefit in this group
- DDD recruitment is based on severe/extreme phenotypes present from early childhood of unknown cause but high expectation of genetic basis
- Recruitment is primarily of trios (ie child and both parents)

SWAN children from www.undiagnosed.org.uk
Total patients: 23678 -
Accessible patients: 7980 -
Syndromes: 66 -
Participating centres: 252
DECIPHER-aided findings

**Syndromes**
- 12p13.33 microdeletion
- 14q11.2 microdeletion
- 5q14.3 microdeletion
- 17q21.31 microdeletion
- 3q13.31 microdeletion
- etc.

**Research Findings**
- Gene Haploinsufficiency predictions
- Mouse-human phenotype association studies
- Characterisation of copy-number stable regions

- 460+ DECIPHER-referenced publications since 2009

Proven track record in bringing clinical meaning to genomic variants of uncertain significance
Deciphering Developmental Disorders - Overview

Thousands of babies are born each year in the UK who fail to develop normally because of errors in their genetic makeup. Currently, diagnosis is restricted to a small minority of children and requires the clinician to recognise the appearance of the child and the pattern of symptoms, supplemented by the use of microscopes to identify large rearrangements of the genetic material in chromosomes. Research shows that the latest molecular testing methods identify previously undetectable changes in chromosomes allowing new diagnoses to be made. However, clinical use is hampered by the limited availability and inconsistent application of these technologies, and by lack of basic knowledge to link genetic changes directly to symptoms. The consequence is that clinical diagnoses remain impossible except for a small number of children.

We propose to apply state of the art molecular testing to 12,000 UK children with abnormal development. The results will provide a unique, on-line catalogue of genetic changes linked to symptoms that will enable clinicians to diagnose developmental disorders. Furthermore, we will design more efficient and cheaper diagnostic assays for relevant genetic testing to be offered to all such patients in the UK and so transform clinical practice for children with abnormal development.
Phenotype entry in DDD

• Accurate phenotyping essential for filtering of variants prior to reporting
• Uses restricted vocabulary based on Human Phenotype Ontology (HPO)
• Terms linked to OMIM, GO, MPO enabling advanced bioinformatic analysis
• Takes <2 mins per patient and done by patient’s consultant

Drag and drop entry for phenotype
Patient DDD-TEST255694

Question

General Information:

- Date of birth: dd.mm.yyyy
- Date at last clinical assessment: dd.mm.yyyy
- Chromosomal Sex: -select-
- Mother's age at birth of the child: ___ years
- Father's age at birth of the child: ___ years

Significant exposures in pregnancy:

- Anticonvulsant drugs: ○ Yes ○ No
- Maternal diabetes: ○ Yes ○ No

Details/Comments

- Birth and Newborn data
- Developmental Milestones
- Family History
- Growth Parameters
- Additional Phenotype
- Clinical summary
Deciphering Developmental Disorders Study Flowchart

1. Receive initial information brochure and invitation to take part in DDD
   - >1 month reminder
2. NHS: Return invitation
   - Receive and read patient information sheet. Talk with DDD contact before signing consent form
3. NHS: Return consent form
   - Receive sample collection kits, return envelopes (SAE) and instruction DVD
4. Watch videos online or on DVD
   - Within 2 weeks...
5. Collect saliva samples
6. Invitation returned
7. Consent returned
8. Samples returned
   - >1 month reminder
9. Post tubes in SAE
10. You are now in the DDD study
11. NHS: Final results sent to regional genetics service for validation and feedback – your geneticist will contact you if we find a diagnosis
   - And again at the end of the study (October 2015)...
12. NHS: Initial results sent to regional genetics service for validation and feedback – your geneticist will contact you if we find a diagnosis
   - Within 1 year...

Please contact your local DDD contact if you have any queries about the study. SWAN UK (www.swanuk.wordpress.com) or Unique (www.rarechromo.org) are also happy to provide further advice and support.

Your local DDD contact is: Tel:
Knowledge gap between genomic science and clinical medicine - OMIM statistics for April ‘13

• Estimated number of genes in human genome = ~21,000
• Phenotype description, molecular basis known = 3,751
• ~18% of genes have a known disease association in humans.
• Rare disorders and unusual phenotypes are often caused by mutation of highly penetrant genes.
• Highly informative for understanding biological processes eg. Li Fraumeni and p53
Identifying causal variants

• Need good coverage of target in order to identify heterozygous changes. Clinical exome usually ~30x
• The larger the target, the more stringent the filtering needed to parse the data to manageable numbers of variant.
• Common variants will be filtered, so need to focus on RARE diseases
• Targeting gene panels at analysis stage improves identification of causal variants

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Panel</th>
<th>Exome</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>~1kb</td>
<td>~50 kb</td>
<td>~60Mb</td>
<td>~3.2 Gb (3,200 Mb)</td>
</tr>
<tr>
<td>~2-3 variants</td>
<td>~100+ variants</td>
<td>~52,500 variants</td>
<td>~3,500,000 variants</td>
</tr>
</tbody>
</table>
Developmental Disorder Gene Database (DDG2P)

DD genes

- **Confidence:** confirmed / probable / possible
- **Allelic:** monoallelic, biallelic, etc
- **Mutation type:** LoF, Activating, etc
- **Diseases:** HPO phenotypes

Updated regularly

~10-15 new genes per month

Existing exome data parsed against new gene list on a 6-monthly basis to identify new causal variants ie new diagnoses.

Similar Gene Lists could be developed by other specialties focusing on genes relevant to their diseases.

Avoids interpretative and ethical dilemmas associated with incidental findings.
Data flow in DDD

1. Web-based recruitment and phenotyping (DECIPHER+HPO)
2. Sample tracking (LIMS)
3. Assays and analysis pipelines set-up (QC, variant calling)
4. Database of called variants and their annotations (exome & aCGH)
5. Variant filtering (novelty, pathogenicity, inheritance, phenotype matching)
6. Manual review and curation of ‘reportable’ variants
7. Automated generation, dissemination and notification of reports (DECIPHER)
Systematic Genomic Analyses

(1) High resolution exon-enriched aCGH of child
- **Structural changes** (gains and losses of genes)
- ~200 variants per person

(2) Exome sequencing of child-mother-father trios
- **Sequence changes** (alterations to genes)
- ~60,000 variants per person

Filter variants by:
- **Novelty** (remove common variants)
- **Overlap with known disease genes**
- **Effect** (coding change)
- **Inheritance** (de novo, mum/dad)
- **Phenotype similarity**
- **Validation**

Interpretation is still problematic…
Finding Likely Causal Variants

**Site** - *Likely importance:*
- Evolutionary conservation
- Position within protein

**Variant** - *Likely impact:*
- Truncating, missense, etc
- Pathogenicity Score
- Inheritance

**Gene** - *Likely relevance:*
- Known causal genes
- Functional models
- Expression patterns
- Literature mining

**Subjective weighing and weighting of evidence**

**Disease causing**
**Probably disease causing**
**Unclear significance**
**Probably benign variant**
**No plausible candidates**

**Patient** - *importance:*
- Family history
- Phenotype
Feedback of Results

- Feedback nothing
  - Standard for most research studies

- Feedback pertinent findings
  - Relevant to clinical enquiry

- Feedback clinically actionable findings
  - Clinical practice in some places – but more like screening?

- Feedback a mixture of findings

- Feedback everything
  - Standard for consumer genomics companies, including raw sequence

Public survey: www.genomethics.org
Avoiding losing patient benefit & research value in translation……

Getting maximum value from data requires near patient work to check:

• Does the genomic variant explain the phenotype
• Does the patient have other features suggested by the genetic diagnosis but not yet recognised clinically?
• Does the variant segregate appropriately in the family
• Time consuming, but necessary to get maximum value from the genomic data

**Decision Diagram (DD):**
- **DD causing**
- **Probably disease causing**
- **Unclear significance**
- **Probably benign variant**
DDD Reports

About DDD Reports

This summary Report relates to genetic variants in the proband that may play a causal role in analysis pipeline to date. The Report will be updated periodically in light of new discovery and intervals linked previously with developmental disorders. Please refer to our Feedback Policy for details of the decision process for reporting.

These research data were generated in a laboratory without CPA accreditation. The responsibility to the families lies with the referring clinician. It is strongly advised that a standard of interpretation of variants based on DDD research data is established locally and agreed by the multiple plausible disease-causing variants. For most DDD participants, partial research data reported before the complete data (including exome sequencing) will be available. Uncertain variants in known disease genes may subsequently be shown to play a minor causal role, if any, in interpretation of array or exome data in isolation.

Download Reports

- Report-261176_2012-05-16
- Report-261176_2012-04-13
DECIPHER with Sequence

Patient TGD269856

### Copy-number variant

<table>
<thead>
<tr>
<th>Location</th>
<th>Interval (Mb)</th>
<th>Mean Ratio</th>
<th>Genes</th>
<th>Inheritance</th>
<th>UCSC/e!</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:145386506-145948067</td>
<td>0.56</td>
<td>-1</td>
<td>17</td>
<td>Inherited from parent with similar phenotype to child</td>
<td></td>
</tr>
</tbody>
</table>

### Sequence variant

<table>
<thead>
<tr>
<th>Location</th>
<th>Gene</th>
<th>Allele</th>
<th>Transcript</th>
<th>Consequence</th>
<th>Inheritance Genotype</th>
<th>UCSC/e!</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:145507646-145507646</td>
<td>RBM8A</td>
<td>delG</td>
<td>ENST00000330165 c.-21delG</td>
<td>In annotated regulatory region (5 prime UTR variant, feature truncation)</td>
<td>Maternal Heterozygous</td>
<td></td>
</tr>
</tbody>
</table>
Additional Reference Resources

Ensembl Variant Effect Predictor (VEP)

1 to 10 of 12 features

<table>
<thead>
<tr>
<th>Ensembl Transcript</th>
<th>HGVS</th>
<th>Position</th>
<th>AA</th>
<th>SIFT ?</th>
<th>Polyphen ?</th>
<th>VEP Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENST00000330165</td>
<td>ENST00000330165:chr2:145507556-145513535</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5_prime_UTR_variant, feature_truncation</td>
</tr>
</tbody>
</table>
Sequence View

Sequence Variation

Context Menu

Incremental Rollout
Mid-April 2013

Population and Pathogenic Variation
Some de novo mutation diagnoses

<table>
<thead>
<tr>
<th>Gene</th>
<th>Freq</th>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANKRD11</td>
<td>1</td>
<td>KBG syndrome</td>
</tr>
<tr>
<td>ARID1B</td>
<td>3</td>
<td>ID/Coffin Siris syndrome</td>
</tr>
<tr>
<td>GJA8</td>
<td>1</td>
<td>cataract zonular pulverulent type 1</td>
</tr>
<tr>
<td>KCNQ2</td>
<td>2</td>
<td>benign neonatal epilepsy type 1</td>
</tr>
<tr>
<td>KIF1A</td>
<td>1</td>
<td>Neuropathy, hereditary sensory, type IIC</td>
</tr>
<tr>
<td>MLL2</td>
<td>1</td>
<td>Kabuki syndrome</td>
</tr>
<tr>
<td>NFIX</td>
<td>1</td>
<td>Marshall-Smith Syndrome</td>
</tr>
<tr>
<td>NRXN2</td>
<td>1</td>
<td>Autism</td>
</tr>
<tr>
<td>NSD1</td>
<td>1</td>
<td>Sotos syndrome</td>
</tr>
<tr>
<td>SCN2A</td>
<td>1</td>
<td>Benign Familial Neonatal Infantile Seizures</td>
</tr>
<tr>
<td>SCN8A</td>
<td>1</td>
<td>Epileptic encephalopathy, early infantile, 13</td>
</tr>
<tr>
<td>SMAD4</td>
<td>1</td>
<td>Myrhe syndrome</td>
</tr>
<tr>
<td>SMARCA2</td>
<td>1</td>
<td>Nicolaides-Baraitser Syndrome</td>
</tr>
<tr>
<td>STXBP1</td>
<td>1</td>
<td>epileptic encephalopathy early infantile type 4</td>
</tr>
<tr>
<td>SYNGAP1</td>
<td>3</td>
<td>mental retardation autosomal dominant type 5</td>
</tr>
</tbody>
</table>
**LMBRD1:**
- Founder mutation L352fsX18
- lysosomal cobalamin exporter
- Vitamin B12 metabolism disorder (Cobalamin F)
- Diagnosis confirmed by elevated homocysteine and methylmalonic acid concentrations in blood or urine samples.

**Action:** fast tracked feedback to clinician

**81 known treatable causes of ID**
DDD results: an iterative reporting cycle

- Exome analysis of known DD genes adds ~20%
- Analysis of remaining variants will increase yield further
- Both from continuing research on novel DD genes within the study and iterative analysis against newly published DD genes

![Diagram showing the iterative process of DDD results](image)
Maximising clinical and research benefit in the DDD Study

- Core DDD-research analysis at WTSI
- Complementary analysis proposals (CAPS) led by collaborating clinicians/scientists on DDD data
- CAPS studies are phenotype/genotype driven; >20 projects approved to date
- Improves chance of diagnosis for patients
- Sharing DDD reports through DECIPHER to enable ‘match-making’ of rare variants and ‘one-offs’
Acknowledgements

- Patients and their families
- Clinicians, laboratory scientists and research nurses in the 23 regional genetics services
- Sanger pipelines

The DDD study is funded by the Health Innovation Challenge Fund (HICF) with research nurse support from the National Institute of Health Research (NIHR)