The Challenge of Documenting Data on Mutations and The Diseases They Cause - The Human Variome Project

Richard G. H. Cotton and Timothy D. Smith
Human Variome Project

HISA Data Governance 2012
30 March 2012

Human Variome Project International Ltd is a Consultative Partner of the United Nations Educational, Scientific and Cultural Organisation
“It is a gruesome battle to take care of our sick sons and daughters, especially when their diseases are rare, and any available support to us is minimal. Very often we go into the fight alone. In some cases, a parent could not face the reality of raising two sick children, and has sadly chosen to abandon the whole family.”
What is the Human Variome Project

- An international consortium of scientists and health-care professionals
- All information on genetic variation should be collected, curated, interpreted and shared freely and openly.
- Focus on variations causing disease
- Aim to facilitate data collection in all genes worldwide
- Some members produce software/strategies
- 30-50 countries represented
- Initiated at a WHO hosted meeting in Melbourne in 2006
- Consultative Partner of UNESCO
How it is being done

• Coordination from Melbourne
• Biennial meetings; upcoming: 4th HVP Meeting, 11-15 June 2012, Paris (www.humanvariomeproject.org/paris/)
• Expert advisory committee and Board
• Setting of standards
• Data flow from Labs and Clinics to Country Nodes to Expert Databases to Central Databases.
• ‘Herding cats’!
Why is it being done?

• Health care professionals need instant access to complete collections of variations and their effects to:
  – Avoid time-wasting web searches, correspondences, phone calls
  – Need the experience of others to inform diagnosis, prognosis and therapy
  – To gather cohorts for trials
  – To contact others with similar experiences
A proposed pathway for collection

I Patient phenotype

II Patient genotype

III Pathology

IV Patient profile

V Clinic file

VI Country specific database

VII Curation

VIII Publication

IX Central database browsers

HOSPITAL

COUNTRY

INTERNATIONAL
Why is Gene/Disease Specific Collection Needed?

• Experts focused on one gene for quality curation
• To spread the load
• Ensure contact with other experts
• Consortium curation will be better able to spread the load and obtain funding
Why is Country Specific Collection Needed?

- Ensuring coherent genetic healthcare for family branches in different states
- Documentation of disease found
- Development of accurate and needed care strategies
- Development of relevant and economical diagnostic strategies for the dominant ethnic group and their diaspora around the world
- Spread the load of mutation collection
- Ensure world-wide complete collection
- Increase accuracy of pathogenicity prediction
- To generate a registry

Cotton et al *GIM* 11(12): 843-849, 2009
Governance

- Australian Public Company Limited by Guarantee
  - Human Variome Project International Limited
- Not-for-Profit, No shareholders, Registered as a “Health Promotion Charity”
- Members: Prof. Sir John Burn, Emeritus Professor Jean-Jacques Cassiman, Dr. David Rimoin, Prof. Richard Cotton, Mr. David Abraham, Prof. Eric Haan and Dr. Xitao Li
- Interim Board: Mr. David Abraham, Mr. Chris Arnold, Prof. Richard Cotton and Dr. Xitao Li
- Interim Scientific Advisory Committee - elected
- HVP Country Node and Gene/Disease Specific Database Advisory Councils
Interim Scientific Advisory Committee

Arleen Auerbach  USA
Marc Greenblatt  USA
Garry Cutting  USA
David Rimoin  USA
Mireille Claustres  USA
Mona El Ruby  USA
Finlay Macrae  USA
Yoichi Matsubara  USA
Gert-Jan B. van Ommen  USA
Johan T. den Dunnen (alternate)  USA
Mauno Vihinen  USA
Christine Van Broeckhoven  USA
Aida Falcon Vargas  USA
John Burn  USA
Richard Gibbs  USA

The Human Variome Project
Sharing data reducing disease
www.humanvariomeproject.org
Current Status

• Specialist groups in action:
  – Colon cancer/ InSiGH
  – Neurogenetics/ Mitochondrial-genetics
  – Nutrigenomics
  – Others

• Country Nodes:
  – Much activity already underway (Greece, Australia, Others)
  – 12 countries had already signed up and are committed (Australia, Austria, Belgium, China, Cyprus, Egypt, Greece, Kuwait, Malaysia, Vietnam, Spain, Nepal)
Current Status

• **Funding**
  - $2 million over 10 years (total of $300 million) from China
  - NeAT Scheme → NeCTAR (Australian Node)
  - Director of Australian Node at the Department of Pathology, the University of Melbourne
  - $1 million for developing countries (China)

• **Capacity and Facilities**
  - Software for gene databasing (LOVD)
  - Facility at NCBI
  - Australian pilot
  - Other collaborative projects (EC, USA, etc)
  - Expert working groups
  - Rare disease initiative
Problems/ Needs - Solutions

- Unprecedented – highest level of global support
- Global – WHO, UNESCO, EC, NIH involvement
- ‘Rare Diseases’ – common as a group and devastating
- Legacy Data – country coordinator
- New Data – effortless software
- Clinical Data – e-files, BioGRID
- No incentives - mandatory/ part of quality assurance
4th Human Variome Project Meeting

UNESCO Headquarters, Paris

11-15 June 2012

Details and program available from website:

www.humanvariomeproject.org/paris/
Acknowledgements

- EC
- UNESCO
- WHO
- March of Dimes
- CASS Foundation
- Federal Government
- Victorian State Government
- Ming Qi (China)
- Hundreds around the world
- Those in the Melbourne Office
The Human Variome Project

sharing data · reducing disease

www.humanvariomeproject.org

YOU CAN HELP

Join the Human Variome Project Consortium online
The Challenge of Documenting Data on Mutations and The Diseases They Cause - The Human Variome Project

Richard G. H. Cotton and Timothy D. Smith
Human Variome Project

Human Variome Project International Ltd is a Consultative Partner of the United Nations Educational, Scientific and Cultural Organisation
Human Variome Project

- An international consortium of scientists and health-care professionals
- All information on genetic variation should be collected, curated, interpreted and shared freely and openly
- Focus on variations causing disease
- sharing data · reducing disease

www.humanvariomeproject.org
Data – what does it look like

data change in this gene has this effect on patient X
Central Databases

- dbSNP
- ClinVar
- OMIM
- HGMD
- GWAS Central
Gene/Disease Specific Databases

- InSiGHT
- hbVar
- CFTR
- Leiden Muscular Dystrophy Pages

- ~1,800 databases
### Locus Specific Mutation Databases

**Last Update 23rd May 2011**

**IMPORTANT NOTE:**

Genes are in order of HUGO APPROVED GENE DESIGNATION not alias. e.g. "p53" will be found under "TP53" while "CD40L" or "TNFSF5" will be found under "CD40LG" and so on.

If you wish to find an Approved gene symbol please select HGNC Search.

If your gene is not in these lists, you may like to check the "Disease Centred", "Mitochondrial Mutations" or "Other mutation Databases" database links as it may be in one of those.

If you wish to add an LSDB please go to the LSDB Submission Page. Please note that some LSDBs have a Curator vacancy - if you would like to serve please contact the acting curator via that database.

Please select the first letter of the Gene: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

OR

Specify the HGNC Gene Symbol: [ ] Go to Gene!

<table>
<thead>
<tr>
<th>HGNC GENE SYMBOL</th>
<th>GBML NO</th>
<th>DATABASE NAME</th>
<th>INTERNET ADDRESS</th>
<th>CURATORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA12</td>
<td>608780</td>
<td>Mutations of the ATP-binding Cassette Transporter A 12 (ABCA12) associated with harlequin ichthyosis, congenital ichthyosiform erythroderma and lamellar ichthyosis</td>
<td><a href="http://www.derm.kobusei.gr/ABCA12">Link</a></td>
<td>Masashi Akiyama and Kaon Sakai</td>
</tr>
<tr>
<td>ABCA13</td>
<td>607857</td>
<td>LOVD - Leiden Open Variation Database</td>
<td><a href="http://www.LOVD.nl/ABCA13">Link</a></td>
<td>Johan T. den Dunnen and Ben Pickard</td>
</tr>
<tr>
<td>ABCA4</td>
<td>601995</td>
<td>Mutations of the ATP-binding Cassette Transporter Patina</td>
<td><a href="http://www.reina-international.org/es-new/patina.htm">Link</a></td>
<td>Reina International</td>
</tr>
<tr>
<td>ABCB4</td>
<td>171090</td>
<td>COHMC - Human Genetics Mutation Database</td>
<td><a href="https://research.chimc.org/gbml/home.php?set=linkABC11">Link</a></td>
<td>Ammar Husami, Brian Richardson, Edita Freeman, Kerry Spooner, Theoda Jacobs and Thenu A. Sivakumar</td>
</tr>
<tr>
<td>ABCB7</td>
<td>603155</td>
<td>ABCB7 database at LOVD</td>
<td><a href="http://www.LOVD.nl/ABCB7">Link</a></td>
<td>Johan T. den Dunnen Leiden Univ. Med Centre (acting), Curator vacancy</td>
</tr>
<tr>
<td>ABCD1</td>
<td>602324</td>
<td>Mutations of the Multidrug Resistance-associated Protein 6 (ABCCD1 MPRES1 1C14-6)</td>
<td><a href="http://www.reina-international.org/es-new/abc6.htm">Link</a></td>
<td>Reina International</td>
</tr>
<tr>
<td>Family ID</td>
<td>Gene</td>
<td>Variant</td>
<td>Age of onset (years)</td>
<td>Gender</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>14</td>
<td>MSH2</td>
<td>Ile54Met</td>
<td>36</td>
<td>F</td>
</tr>
<tr>
<td>1</td>
<td>MLH1</td>
<td>Ala29Ser</td>
<td>53</td>
<td>M</td>
</tr>
<tr>
<td>11</td>
<td>MLH1</td>
<td>Val506Ala</td>
<td>28</td>
<td>F</td>
</tr>
<tr>
<td>12</td>
<td>MLH1</td>
<td>Val506Ala</td>
<td>26</td>
<td>F</td>
</tr>
<tr>
<td>13</td>
<td>MSH2</td>
<td>Gly149Asp</td>
<td>37</td>
<td>M</td>
</tr>
<tr>
<td>9</td>
<td>MLH1</td>
<td>Arg226Leu</td>
<td>50</td>
<td>F</td>
</tr>
<tr>
<td>16</td>
<td>MSH2</td>
<td>Thr552Pro</td>
<td>40</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>MLH1</td>
<td>Gly1017ys</td>
<td>49</td>
<td>F</td>
</tr>
<tr>
<td>8</td>
<td>MSH2</td>
<td>Val185Gly</td>
<td>40</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>MSH2</td>
<td>Gly244Asp</td>
<td>52</td>
<td>M</td>
</tr>
<tr>
<td>19</td>
<td>MSH2</td>
<td>Ser577Leu</td>
<td>46</td>
<td>M</td>
</tr>
<tr>
<td>17</td>
<td>MSH2</td>
<td>Ala636Gly</td>
<td>52</td>
<td>M</td>
</tr>
<tr>
<td>5</td>
<td>MLH1</td>
<td>Thr117Met</td>
<td>41</td>
<td>F</td>
</tr>
<tr>
<td>6</td>
<td>MLH1</td>
<td>Thr117Met</td>
<td>45</td>
<td>F</td>
</tr>
<tr>
<td>7</td>
<td>MLH1</td>
<td>Thr117Met</td>
<td>47</td>
<td>F</td>
</tr>
<tr>
<td>18</td>
<td>MSH2</td>
<td>Thr671ys</td>
<td>50</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>MLH1</td>
<td>Gly67Arg</td>
<td>33</td>
<td>F</td>
</tr>
<tr>
<td>3</td>
<td>MSH2</td>
<td>Gly67Arg</td>
<td>30</td>
<td>M</td>
</tr>
<tr>
<td>15</td>
<td>MSH2</td>
<td>Arg338Gly</td>
<td>45</td>
<td>M</td>
</tr>
</tbody>
</table>

*The gene and missense variant identified in the family, the proband age of first CRC and gender are given. Deleterious and neutral variant classification using criteria specified in Table 1. Families below the horizontal line have both germline deletions and missense mutations (see Results and Discussion sections).

Table 5. Familial CRC Probands With MAPP-MMR Predicted Deleterious Variants*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Age</th>
<th>MSI status</th>
<th>MAPP-MMR</th>
<th>Location</th>
<th>Histology</th>
<th>Stage</th>
<th>Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLH1</td>
<td>Tyr293Asp</td>
<td>49</td>
<td>MSI-H</td>
<td>25.3</td>
<td>Colon transverse</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Glu268Gly</td>
<td>41</td>
<td>MSS</td>
<td>6.02</td>
<td>Colon transverse</td>
<td>Adenocarcinoma</td>
<td>Stage I</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Tyr289Asp</td>
<td>49</td>
<td>NA</td>
<td>27.48</td>
<td>Cecum</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>Thr117Met</td>
<td>43</td>
<td>NA</td>
<td>21.05</td>
<td>Colon transverse</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Moderate</td>
</tr>
<tr>
<td>MSH2</td>
<td>Thr441Pro</td>
<td>56</td>
<td>MSI-L</td>
<td>5.9</td>
<td>Cecum</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>Thr552Pro</td>
<td>40</td>
<td>MSI-L</td>
<td>8.71</td>
<td>Hepatic flexure</td>
<td>Adenocarcinoma</td>
<td>Stage III</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Thr552Pro</td>
<td>40</td>
<td>NA</td>
<td>8.71</td>
<td>Colon transverse</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Thr552Pro</td>
<td>41</td>
<td>MSI-L</td>
<td>8.71</td>
<td>Colon descending</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>Thr552Pro</td>
<td>42</td>
<td>MA</td>
<td>8.71</td>
<td>NA</td>
<td>Adenocarcinoma</td>
<td>Stage II</td>
<td>Poor</td>
</tr>
</tbody>
</table>

*Proband age at CRC onset, or Southern California Colorectal Cancer Cohort ascertainment. MSI status, and MAPP-MMR scores are indicated. CRC location in colorectum, histology, stage, and tumor type are also indicated.

DISCUSSION

More than 300 unique MLH1/MSH2 missense variants have been reported. There is consensus among cancer geneticists that many of these mutations are not pathogenic, and several studies have demonstrated that not all pathogenic variants in the MLH1/MSH2 genes are validated.
this change

- Protein, DNA, RNA level
- Coding DNA or genomic DNA
- rs#
- HGVS Nomenclature
- Old school methods
- Bespoke methods (BIC)
this gene

- HGNC name: MC1R, melanocortin 1 receptor, HGNC:6929
- Synonyms: MSH-R
- OrphaNet ID: ORPHA139778
- OMIM: 155555
- Entrez Gene: 4157
- Sequence Accession: NM_002386
- Chromosomal Location: 16q24.3
- Coordinates: 16:89,984,286 - 89,987,384
means this

- Simple
- Free text
- Coding
  - ICD-10 (9)
  - SNOMED
- Ontologies
  - HPO
  - GO
  - PATO
- 20,000 genes = 20,000 disorders
- All different
patient X

- Ethical, legal and social issues
- Global context
- Patient ID
- Assigned by LSDB
- Useful?
Gene/Disease Specific Databases

• All disorders are different
• Molecular data - the same
• Phenotype - different

• Needs to be defined and managed by stakeholders
Human Variome Project

- Working with existing G/DSDBs
- Encouraging new ones
- Facilitating communication
- Skills sharing and education
- Standards and guideline development
- Advisory Council
Curation

• Curation in our world
• Data governance in yours
• Equivalent concepts
• Lack of interaction
InSiGHT

- Incorporated society
- Organisational governance
- Liability, insurance
- Full time "curator"
- Interpretation committee
- Policies and procedures
G/DSDBs

- Most not like this
- Research focused
- Side project
- Literature based
- No technical sophistication
- Limited coverage
- Not "bus proof"
Solutions

- HGVS nomenclature
- Open Source DBMS products
- LOVD, UMD, MutBase
- Minimum Content
- Ethics
- Submission Forms
Need More

• Standards Development Process
Quality

- Once standards exist you can measure
- Transition from research grade to clinical grade
- These resources ultimately need to be useful as clinical decision support tools
Global Collection Architecture

- Part of the transition to clinical grade resources is the ability to collect clinical grade data
HVP Country Nodes

- Repositories of variation within a country
- Service with in-country benefits
- Diagnostic labs
- Clinics
- Policy making and healthcare delivery planning
- Registries
HVP Country Nodes (2)

- Facilitate collection
- ELSI
- Ownership/buy-in
HVP Country Nodes (3)

- Australia
- Austria
- Belgium
- China
- Cyprus
- Egypt
- Greece
- Kuwait
- Malaysia
- Nepal
- Spain
- Vietnam
HVP Country Nodes (4)

- ICCAC
- Standards and guidelines
- Education and training
- Capacity building
- Can't collect data that doesn't exist
NeAT
- Pilot Phase
- 4 labs, 3 diseases
- Launched April 2011
- Molecular Data Only

Collection Tool
Portal
Data Model
Ethics Processes
Access & Usage Policy
Data Sharing Agreements
NeCTAR

- By end of 2013
- 18 more labs
- All genes they test for
- Configuration Tool
- Clinical Data/Phenotype
- Transfer data internationally
Global Collection Architecture
Areas of Action

- Setting Normative Function
- Behaving Ethically
- Sharing Knowledge
- Building Capacity

Human Variome Project
4th Human Variome Project Meeting
UNESCO Headquarters, Paris
11-15 June 2012
Details and program available from website:

www.humanvariomeproject.org/paris/