

How to Catch All Those Mutations—The Report of the Third Human Variome Project Meeting, UNESCO Paris, May 2010

Maija R.J. Kohonen-Corish,^{1*} Jumana Y. Al-Aama,² Arleen D. Auerbach,³ Myles Axton,⁴ Carol Isaacson Barash,⁵ Inge Bernstein,⁶ Christophe Bérout,⁷ John Burn,⁸ Fiona Cunningham,⁹ Garry R. Cutting,¹⁰ Johan T. den Dunnen,¹¹ Marc S. Greenblatt,¹² Jim Kaput,¹³ Michael Katz,¹⁴ Annika Lindblom,¹⁵ Finlay Macrae,¹⁶ Donna Maglott,¹⁷ Gabriela Möslein,¹⁸ Sue Povey,¹⁹ Raj Ramesar,²⁰ Sue Richards,²¹ Daniela Seminara,²² María-Jesús Sobrido,²³ Sean Tavtigian,²⁴ Graham Taylor,²⁵ Mauno Vihinen,²⁶ Ingrid Winship,²⁷ and Richard G.H. Cotton,²⁸ on behalf of contributors to the Human Variome Project Meeting

¹Cancer Research Program, Garvan Institute of Medical Research, St. Vincent's Clinical School, University of NSW, Sydney, NSW, Australia;

²Department of Genetic Medicine, Faculty of Medicine and Princess Al-Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia; ³Program in Human Genetics and Hematology, The Rockefeller University, New York, New York; ⁴Nature Publishing Group, New York, New York; ⁵Boston University, Department of Philosophy, Genetics, Ethics & Policy Consulting, Boston, Massachusetts;

⁶Department of Surgical Gastroenterology, Hvidovre Hospital, Copenhagen, Denmark; ⁷INSERM U827, Laboratoire de Génétique Moléculaire, Université Montpellier 1, Hôpital Arnaud de Villeneuve, Montpellier, France; ⁸Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, United Kingdom; ⁹European Bioinformatics Institute (EBI), Hinxton, Cambridge, CB10 1SD, United Kingdom; ¹⁰Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; ¹¹Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands;

¹²University of Vermont College of Medicine, Burlington, Vermont; ¹³Division of Personalized Nutrition and Medicine, U.S. FDA National Center for Toxicological Research, Jefferson, Arkansas; ¹⁴March of Dimes Foundation, White Plains, New York; ¹⁵Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; ¹⁶Department of Colorectal Medicine and Genetics, and University of Melbourne Department of Medicine, The Royal Melbourne Hospital, Parkville, Victoria, Australia; ¹⁷Honorary Secretary, International Society for Gastrointestinal Hereditary Tumours (InSiGHT);

¹⁸NCBI-National Center for Biotechnology Information, NLM/NIH, Bethesda, Maryland; ¹⁹HELIOS St. Josefs-Hospital, Bochum, Germany; ²⁰Division of Life Sciences, UCL, London, United Kingdom; ²¹MRC Human Genetics Research Unit, Division of Human Genetics, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, South Africa; ²²Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon;

²³National Cancer Institute, Bethesda, Maryland; ²⁴Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Center for Biomedical Network Research on Rare Disorders (CIBERER), Institute of Health Carlos III, Spain; ²⁵Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah; ²⁶St. James's University Hospital, Leeds, United Kingdom; ²⁷Institute of Medical Technology, University of Tampere, and Research Center, Tampere University Hospital, Tampere, Finland; ²⁸University of Melbourne, Royal Melbourne Hospital, Melbourne, Victoria 3050 Australia; ²⁹Genomic Disorders Research Centre, Carlton South, Melbourne, Victoria, Australia

Received 22 September 2010; accepted revised manuscript 1 October 2010.
 Published online 19 October 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/humu.21379

ABSTRACT: The third Human Variome Project (HVP) Meeting “Integration and Implementation” was held under UNESCO Patronage in Paris, France, at the UNESCO Headquarters May 10–14, 2010. The major aims of the HVP are the collection, curation, and distribution of all human genetic variation affecting health. The HVP has drawn together disparate groups, by country, gene of interest, and expertise, who are working for the common good with the shared goal of pushing the boundaries of the human variome and collaborating to avoid unnecessary duplication. The meeting addressed the 12 key areas that

form the current framework of HVP activities: Ethics; Nomenclature and Standards; Publication, Credit and Incentives; Data Collection from Clinics; Overall Data Integration and Access—Peripheral Systems/Software; Data Collection from Laboratories; Assessment of Pathogenicity; Country Specific Collection; Translation to Healthcare and Personalized Medicine; Data Transfer, Databasing, and Curation; Overall Data Integration and Access—Central Systems; and Funding Mechanisms and Sustainability. In addition, three societies that support the goals and the mission of HVP also held their own Workshops with the view to advance disease-specific variation data collection and utilization: the International Society for Gastrointestinal Hereditary Tumours, the Micronutrient Genomics Project, and the Neurogenetics Consortium.

Hum Mutat 31:1374–1381, 2010. © 2010 Wiley-Liss, Inc.

KEY WORDS: mutation; variation; genomics; genetic disease

Additional Supporting Information may be found in the in online version of this article.

Communicated by Mark Paalman

*Correspondence to: Maija R.J. Kohonen-Corish, Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, NSW 2010, Australia. E-mail: m.corish@garvan.org.au

Introduction

The history of human genetics features a strong tradition of collaboration, which has produced remarkable achievements. One such example is the unraveling of the complex genetics of the HLA system, which was built on sharing of reagents and unpublished data. A more recent example is the Human Variome Project (HVP) (www.humanvariomeproject.org), which stems from a similar need to understand genetic variation and its implications for the treatment of diseases that arise from mutations and variations in the human genome. However, in the era of high-throughput analysis as well as ultraspecialization, one challenge is to deal with the unprecedented data overload and harness it for the benefit of routine healthcare. HVP is not “collaborative and exclusive” like many other scientific endeavors but “exclusively collaborative.” Its global nature demands “local action for the global good” and aims to bring together key individuals who are currently working on the development of protocols and systems relating to the generation, interpretation, and storage of data on human genetic variation as well as those generating the data.

The major aims of the HVP are the collection, curation, and distribution of all human genetic variation affecting health. This global project involves community activity to collect this information for databasing. It is estimated that there will be at least 100 to 1,000 variations in each of the estimated 20,000 human genes. This ambitious project is the vision of Prof. Richard Cotton (Melbourne, Australia) and the attendees at the official launch in 2006 in Melbourne [Cotton et al., 2007; Ring et al., 2006]. The first meeting was sponsored by the World Health Organisation (WHO) and supported by the American College of Medical Genetics and the Genomic Disorders Research Center. It brought together representatives from the European Union (EU), United Nations Educational, Scientific, and Cultural Organization (UNESCO), Organization for Economic Co-operation and Development (OECD), National Center for Biotechnology Information (NCBI), and the European Bioinformatics Institute (EBI), as well as geneticists from 30 countries. In 2008, the HVP Planning Meeting was held in Spain, which formulated the actions needed to address the challenge and consolidated the 12 key areas that form the current framework of HVP activities [Kaput et al., 2009]. We have minimized reference to the publications of the substantial amount of work by the members of the HVP/ HGVS community to simplify this report. Reference to such work can be found in the following: The HGVS Website, Cotton et al. [2007], Kaput et al. [2009], and the HVP Website (www.humanvariomeproject.org/index.php/publications).

Third HVP Meeting Objectives

The third HVP meeting on “Implementation and Integration” was held under UNESCO Patronage in Paris, France, at the UNESCO Headquarters May 10–14, 2010. It was co-organized by the Genomic Disorders Research Center (the HVP Coordinating Office) and the Division of Basic and Engineering Sciences, the Natural Sciences Sector of UNESCO, in association with the American College of Medical Genetics. It was attended by over 150 registrants from 30 countries representing all continents. The aim was to develop procedures for the implementation of the recommendations and actions from the first two meetings in a global collaborative context and to prepare the systems necessary to routinely and systematically gather the growing number of disease mutations now being discovered.

The objectives were:

1. To ensure progression of the objectives of the Human Variome Project to “collect variations causing disease (mutations) in all genes world-wide.”
2. To gather experts and leaders in the strategies and systems required, together with those responsible for collecting data and its users.
3. To gather those who are already collecting mutations within their countries and others who wish to establish country or regional nodes, from both developed and developing countries.
4. To present projects currently underway or in the planning stage that contribute to the achievement of the objectives of the Human Variome Project.
5. To examine all areas of the implementation and integration of the overall and individual parts of the project.
6. To identify specific new projects and develop formal collaborative groups to undertake the work and to develop standards.
7. To gather those who are experienced in funding to consider optimal ways for the Human Variome Project to move forward and to identify potential funding sources for the projects.

“The Human Variome Project as a Natural Successor to the Human Genome Project”

The meeting was officially opened by Dr. Walter Erdelen, UNESCO Assistant Director-General for Natural Sciences, who welcomed the participants and expressed UNESCO’s support for the principles and mission of the Human Variome Project. He outlined UNESCO’s active involvement in setting up the first meeting of the Human Genome Project in Paris in 1989, in response to the “revolution in biology that occurred in the late 1970s and 1980s” and the promise of “genetic engineering to solve the problems of both old and new diseases.” He emphasized UNESCO’s role in facilitating international scientific cooperation in this effort, particularly among developing countries and between developing and developed nations, and concluded that he saw “the Human Variome Project as a natural successor to the Human Genome Project.”

“Failure to Apply Genetic Knowledge Kills People”

The Plenary Lecture was given by Sir John Burn, the Director of the Institute of Human Genetics at the Centre for Life in Newcastle upon Tyne, UK. Among his many contributions to clinical genetics, he was the Founding President of the International Society for Gastrointestinal Hereditary Tumours (In-SiGHT), which is closely associated with the Human Variome Project. He gave a spirited overview of the use of genetic knowledge in the current healthcare system, where gene mutation testing is not always applied when necessary. He outlined the extraordinary promise of Next Generation sequencing and other new technologies and their potential benefits for routine clinical practice. He predicted that diagnostic labs will continue to be important players in the discovery of less predictable intronic and promoter variants. However, the key driver for success is access to detailed phenotypic data with the mutation data, which will allow therapeutic intervention for mutation carriers before the onset of the disease. He concluded that in order to offer better treatment options for patients, genetic risk has to be quantified using mathematical models such as Bayes’ Theorem, which has already been applied for many unclassified variants in hereditary breast and colorectal cancers.

Summary of Discussions from the 12 Working Groups

The plans and projects are ready for specific grant applications to carry out the defined and agreed activities. These could be piloted by students. More detailed summaries from each working group can be found in the online Supporting Information.

Ethics

The Ethics Workgroup has produced detailed guidelines that are intended to help the curators of locus-specific or disease-specific variation databases (LSDBs) resolve ethical issues [Povey et al., 2010].

Box 1. Major Points to Consider Pertaining to Ethical Issues Arising in the Curation of Human Locus Specific Variation Databases (LSDBs)

- Clarify the main purpose of the particular database.
- Define database policy with respect to sources of data.
- Take specific communities/cultures into account.
- Take vulnerable persons into account.
- Create an ethics oversight committee.
- Remove identifying information before submission to database.
- Add further protection of confidentiality if needed.
- Allow no further disclosure without consent.
- Provide provision for removal of data from the database.
- Be cautious in response to requests to an LSDB curator for a private opinion.
- Limit links to other LSDBs.
- Consider carefully the transfer of publicly available data from LSDBs to genome browsers.

It was resolved to explore the diverse needs of different cultures with a view to modification of future guidelines, with particular emphasis on acceptable forms of consent (both written and unwritten), privacy, and rules for collection and display of clinical data. The group will work with the Workgroup for the Data Collection from Clinics to find a method for the ethical export of data, including the possibility that explicit consent from the donor may not always be required or feasible. The group also voted against restrictions being placed on the secondary use of data that have already been published and are electronically available, but will maintain awareness of these concerns. Another developing area is a trend toward “open consent,” where donors consent to acknowledging that although researchers will expend every effort to assure privacy, in reality such assurance cannot be guaranteed. In other words, participants agree that privacy violations are a reality of research, specifically in regard to future possible uses of data. The ethical rationale behind this notion is transparency, namely, that it is ethical to be utterly transparent about actual practice.

Nomenclature and standards

The meeting discussed the proposed changes to HGVS recommendations for the description of sequence variants, which have evolved considerably over time (www.HGVS.org/mutnomen). The meeting agreed with the most significant changes, such as the introduction of a versioning system, which allows users to indicate the version used, and the recommendation to use the Locus Reference Genomic (LRG) sequence format as the new standard for the description of sequence variants. Details on how to obtain an LRG can be found at the LRG Website (www.lrg-sequence.org). Other subjects were discussed in depth, for example, changing the symbol for a stop codon from X to a * in the description of

changes at the protein level, to follow existing guidelines (IUPAC/IUBMB). Additionally, it was agreed that the HGVS recommendations should be advertised more widely (journals, human genetics societies, granting organizations, etc.), and the HGVS should try to get an ISO certification. The meeting agreed that guidelines are needed regarding the description of copy number variant (CNV) genes and complex sequence variants in the human genome. Proposals will be available for comment on the HUGO Gene Nomenclature Committee (HGNC) Website (www.genenames.org/index.html) and the HGVS Website (www.HGVS.org/mutnomen/HGVS_extend_PT.doc). Finally, agreement was reached that HVP should implement a gene ontology and consider the VariO proposal, which is available for comment on the VariO Website (variationontology.org). Mauno Vihinen has undertaken the task of reviewing all the standards available and those needed and to develop them.

Box 2. Tasks from the Nomenclature and Standards Workgroup

- Seek official certification for the HGVS recommendations for the description of sequence variants (e.g. ISO).
- Add a versioning system to the recommendations for the description of sequence variants.
- Extend current recommendations to facilitate the description of complex sequence variants.
- The Locus Reference Genomic (LRG) sequence format is the new standard for the description of sequence variants.
- Support the development of ontologies to improve overall database quality (data storage & retrieval), incl. the description of copy number variant (CNV) genes and the consequences of variations (variationontology.org/).

Publication, credit, and incentives

One of the recommendations from the previous HVP Meetings was to encourage more submission of mutation data to databases, by developing incentives, such as microattribution to the individual submitting data [Axton, 2008]. This concept was further discussed and developed by the Workgroup and integrated with new initiatives, such as Open Researcher and Contributor ID: ORCID, www.orcid.org; DataCite, www.datacite.org; Bioresource Impact Factor, BRIF, and nanopublication. To test the idea of microattribution reviews [Axton, 2008], hemoglobin variants have been collectively annotated together with terms from the MeSH controlled vocabulary describing phenotypes, population frequency data, geographic origin in the HbVar database (globin.bx.psu.edu/hbvar), and Globin Gene Server LOVD (lovd.bx.psu.edu). Each entry in the review gives citation credit to the original observers and the curators, both for published and unpublished data. After publication, these data can be represented on a wiki browser to collect comments and future variant reports, and can form the substrate for Concept Web nanopublication constructs (RDF triples). Unique, persistent identifiers for bioresources will be piloted in several projects, including the ongoing GEN2PHEN project (www.gen2phen.org) and BioSHARE-EU, which will commence in early 2011. Tracking citations to these identifiers from various scholarly works will enable the application of bibliometrics to assess impact of individual publications, as well as larger data compilations (databases). Via the Datacite initiative, bioresource IDs can be equated with DOIs and cited in the same way as preprints and other nonreviewed public resources. Publication of exacting research and “mutation updates,” where deposition of variants is rewarded by authorship, has previously been restricted to Nature Genetics and Human Mutation.

Data collection from clinics

Clinics have a central role in the collection and utilization of genetic knowledge because of their direct interaction with the patient and with the asymptomatic family members who may request genetic testing. Therefore, accurate and standardized phenotype data must accompany genotype data to maximize their clinical benefit. The Workgroup voted unanimously that phenotype data should be provided by clinics despite the challenges posed, and agreed that advancing our understanding of genome variation will not proceed without correlation with human phenotypes. To accomplish this task, standardized phenotyping definitions need to be developed that have universal rather than disease-specific applications. Furthermore, bioinformatic systems are needed to assemble these data and to collate with published information. The issue of the role of the lab versus the clinic in acquiring phenotype information remains undecided. Regulation or legislation might be used to require either or both to participate in this process. However, there were serious misgivings that enforced collection would be difficult to supervise and might impede the provision of genetic tests, especially for rare disorders.

Box 3. Tasks and Recommendations from the Data Collection from Clinics Workgroup

- Voluntary submission of clinical data with every genetic test was deemed to be highly desirable but could not realistically be enforced as an absolute requirement.
- The development of bioinformatic tools for objective phenotype assessment was deemed important for differentiating phenotypes.
- Journals to consider requesting authors use standardized phenotype nomenclature.
- Support an ongoing dialogue between clinicians and laboratories to establish reasonable methods for pairing phenotype information with genetic test results.

Overall data integration and access—peripheral systems/software

This Workgroup discussed the challenges of data integration and access from heterogeneous systems outside the central databases, such as disease-specific mutation databases. The systems need to have a user-friendly interface and data need to be standardized. It was concluded that many of the available systems are unfamiliar to the individual researcher and professionals working in the healthcare system. It was resolved to establish a Webpage on the HVP Website, which gives a listing of all the available software/tools and further information, such as the IT language used for their development, availability, price, URL, contact persons or institutes, and the experience with their use and implementation. A discussion forum should be created for people to submit problems and queries. Because a related initiative has already been started in the GEN2PHEN project, HVP will interact with this subgroup to share the large amount of work. An upcoming special issue of *Human Mutation*, co-edited by Peter Robinson and Annika Lindblom, will discuss these systems in more detail.

Data collection from laboratories

There have been encouraging advances in this area in recent years, including (1) standard international nomenclature rules (www.hgvs.org), (2) professional Standards and Guidelines for Clinical Molecular Laboratories (see American College of Medical Genetics, www.acmg.net), (3) multiple gene-specific databases, and (4) algorithms using multiple predictive software tools, functional assays, and family studies that are available for prediction of pathogenicity of variants of unknown clinical significance. However, the collection of data remains a challenge. The barriers identified include lack of incentive for laboratories to provide genotypic data and for clinicians to provide phenotypic data, concerns of patient confidentiality and HIPPA regulations, and practical issues surrounding data transfer. Also, the trend toward clinical syndrome, exome or genome sequencing is going to increase the scale of data substantially, and gene-specific databases may not be appropriate for these data.

Box 4. Tasks from the Workgroup for the Data Collection from Laboratories

- WHO to endorse HGVS nomenclature rules with appropriate reference sequence.
- Development of international standards for databases, including a minimal set of requirements and use of standard ontology.
- Development of regulatory incentives for clinical laboratories to encourage collection of clinical data and reporting of variant and clinical information into public databases.
- A workgroup composed of ethicists and clinical laboratory representatives to develop an informed consent model addressing laboratories depositing both genotype and phenotype information into public databases.
- Development of a process for laboratories to submit data that has quality assurance of data content and is user friendly for laboratories, making the data entry process easy within the laboratory workflow.
- Development of a process for prospective data review and revision allowing laboratories to update their previous entries in the database with new findings.

Assessment of pathogenicity

Unclassified variants (UVs), such as missense mutations in disease-associated genes, are a major problem in the interpretation of clinical significance of mutations. For example, a very significant proportion of the described *BRCA1* and *BRCA2* sequence variants fall in this category. The overall aims of this group are (1) to classify the pathogenicity of genetic variants with sufficient confidence to use clinically, (2) to pursue classification using methods that incorporate multiple lines of evidence or data types, and (3) to establish standards, validation, quantification, and transparency in evaluating each data type. Since the 2008 HVP meeting, considerable progress has been made with assigning pathogenicity of UVs, particularly in hereditary breast and colorectal cancers by the International Agency for Research on Cancer (IARC) Unclassified Genetic Variants Working Group. This work and other important initiatives were discussed in the meeting.

Box 5. Tasks and Recommendations from the Assessment of Pathogenicity Workgroup

- HVP reiterates the recommendation from 2008 that variant classification must rely on ≥ 2 types of evidence for each variant, including at least one method that addresses the variant directly rather than the association of a disease with an allele. No single type of evidence is sufficient by itself to classify a missense variant. However, strong evidence of segregation can establish association (not causality) of a variant with disease.
- HVP endorses the use of an established classification system for the likelihood of pathogenicity of genetic variants, linked to probabilities of pathogenicity.
 - The 5-class IARC system is preferred because it links classes to probabilities.
 - Elements of other published systems, such as the 6-class ACMG 2007 system or the Manchester system, can be used if they are already established in a country. Users should recognize that the IARC system can be used to classify some new variants that remain unclassified within other systems. The relationships among the systems should be reconciled.
 - These systems should be used clinically and for research, and their performance evaluated, calibrated, and validated.
 - HVP endorses the use of the IARC system for a community prediction project, Critical Assessment of Genome Interpretation (CAGI) that is planned to identify the progress and challenges currently existing in predicting the effects of genetic variants (genomecommons.org/cagi).
- HVP endorses a classification system based on the probability that a variant is pathogenic.
 - *In silico* and *in vitro* methods may be used to help classify variants within a framework that also includes other evidence.
 - HVP urges validation of probabilistic classification frameworks (including e.g. Bayesian classification) for individual genes or syndromes by experts on those genes or syndromes before they are accepted as classifiers. Some individual analytic methods, such as i) segregation, ii) *in silico* algorithms that consider gene evolution, protein multiple sequence alignments, and/or protein structure, and iii) mRNA splicing assays, may be generic across most genes. Even so, these will have to be recalibrated for each gene or syndrome to which they are applied. Other methods, in particular *in vitro* protein functional assays, are likely to be specific to each system and are likely to require validation by experts in each relevant field.
 - For all methods, assessors will attempt to generate quantitative data expressed in probability of pathogenicity (or Likelihood Ratios or Odds Ratios for pathogenicity that convert easily to probabilities).
 - It is acceptable to use qualitative data to classify a variant, but only if explicit rules for classification are specified.
- The Assessment of Pathogenicity Group will write a paper discussing standards for data types, classification, integration and urging all fields of genetics to validate clinicopathologic, *in vitro*, and *in silico* methods of classification.

Country-specific collection

A unique set of challenges for the global aspirations of HVP are created by the diverse ethical and legal requirements of individual countries and cultural differences relating to human health. Therefore, a centrally mandated one-size-fits-all approach is not feasible, and HVP has developed its One Project-Two Channels-Multiple Locations strategy. Central to this strategy is the establishment of HVP Country-Specific Nodes, which can range from one large country having several nodes, or several countries grouping together to form a single node. The steps required to begin the process of creating a HVP Country-Specific Node are shown in Box 6.

Box 6. How to Create a HVP Country Specific Node

- Form a consortium in your country and initiate projects and systems to achieve the goals of the Human Variome Project.
- Ensure that the data collected can be integrated into current and future global systems.
- Collaborate with international activities and assist other projects where possible.
- Undertake that the data are freely available.
- Ensure that the country's ethical requirements are integrated into the collection systems at all levels.
- Provide a viable source of funding for each country node.

The Workgroup recommended that HVP commence a concerted worldwide educational campaign, which targets policy makers, healthcare providers, scientists, patients, families, patient support groups, and the public. This campaign should aim to explain the importance of data collection demonstrating the potential universal benefit, increase the number of efficient curators, and advise about setting up a country node and maintaining it. One of the objectives of HVP is to involve emerging countries in the generation of molecular data, by developing capacity and skill and funding training programs. The Workgroup recommended establishing further regional/specialist networks, such as the African Genome Initiative, Pan Asian Consortium, the Center for Arab Genomic Studies, and the emerging Ibero-American HVP network initiative. These networks and linkages ensure regional networking and sustainability. An article detailing the setting up of a country node is in preparation, and a pilot study has been funded in Australia to join other country initiatives, which need to be implemented.

Translation to healthcare and personalized medicine

This Workgroup discussed how to improve the clinical utility of genetic information stored in databases and assist translation of gene tests from research to clinical practice. The data collected should be complete enough to be clinically useful, including both genotype and phenotype information. This is particularly important as genetic testing is increasingly integrated into mainstream clinical services, because many medical professionals are likely to have less experience in the interpretation of genetic data than services provided by clinical geneticists. The Workgroup recommended that laboratories and clinics submit genotypes with succinct data sets of phenotypes and family data in coded form.

Box 7. Tasks and Recommendations from the Translation to Healthcare and Personalised Medicine Workgroup

- Submission of genotypes requires corresponding phenotypes and coded family data.
- Reports should stratify sequence variants to clearly pathogenic, clearly polymorphic or unclassified variants (VUS).
- HVP to endorse accreditation of laboratories and LSDBs.
- Achieve sustainable funding of LSDBs.
- Facilitate pharmacogenetics and personalized medicine.
- Carry out health economic evaluation of the aims of HVP.

Reports should try to include as much phenotypic data as possible to try to facilitate interpretation of whether a variant is clearly pathogenic, clearly neutral, or a Variant of Uncertain Significance (VUS, also called UVs). However, final classification of pathogenicity for sequence variants should be undertaken by expert panels, to avoid presenting conflicting conclusions to the biomedical and clinical communities. A pilot study is underway under the auspices of InSiGHT to classify UVs in the DNA mismatch repair genes responsible for Lynch syndrome. The need was identified for Web-based tools to be used for immediate classification of variants. The online resource should list a network of interested scientists offering *in silico* studies, functional studies, and association studies to classify UVs in the genes of interest. The Workgroup also recommended that HVP should endorse accreditation of labs with

standards used in each country and a system of a limited number of accredited LSDBs for high quality. HVP should have a way of certifying databases with a scoring system to meet minimum requirements.

Data transfer, databasing, and curation

There are 1,550 LSDBs currently listed by the HGVS (www.hgvs.org/dblist/glsdb.html). Major initiatives such as the EU Framework Project 7 funded GEN2PHEN project, have provided the infrastructure necessary to establish and maintain more LSDBs. Specialized software have been developed, which allow fully Web-based initiation and curation of LSDBs following HGVS recommendations and have led to a more uniform presentation and content. This Workgroup covered three important development areas for LSDBs: data transfer between databases, and from laboratories to databases, as well as database systems and curation.

Box 8. Tasks and Recommendations from the Data Transfer, Databasing and Curation Workgroup

- Databases should be accredited for quality control.
- Standards recommended for use by all variation resources
 - HGVS nomenclature
 - LRG (and/or RefSeqGene)
 - HGNC gene names
 - dbSNP reference numbers
 - ISO certified standards should be used
- LSDBs should be consistent with the LSDB object model developed at GEN2PHEN.
- Databases should allow easy integration with other resources.
- Software should be open source.
- The need for respecting privacy, especially for rare disorders, has to be recognized. Data that is displayed should have open access within the limits of the local privacy legislation. There should be a provision for different levels of privacy. In cases when the data cannot be released, aggregated data could be provided (i.e. SNP calls).
- Credit for the database curators should be given in published articles. Each database should obtain digital identifier. Instructions how to cite a database should be in the web site.
- Development of software tools to avoid human errors and for easy, reliable and high-quality submission and curation of LSDBs is essential.

As much more data are going to be available in the near future it is important to automate routine steps to allow curators to concentrate on development and quality control. All variation sources should use the same standards, such as HGVS nomenclature, LRG or RefSeqGene sequences, HGNC gene names, dbSNP reference numbers, and ISO certified standards. The Workgroup recommended that a rating system be adopted for database quality control, which should be reevaluated at frequent intervals, for example, every 3 years. A training module for LSDB curators should be established to educate about the quality requirements and standards. Database curators should be given credit in published articles and each database should obtain a digital identifier. LSDBs should be consistent with the LSDB object model developed at GEN2PHEN and allow easy integration with other resources.

Overall data integration and access—central systems

This session discussed how information related to human variants and associated phenotypes are being processed and

represented by central databases, and focused on recommendations for data elements that should be centralized. The EBI and NCBI are ready to accept submissions of consented summary data from LSDBs and diagnostic laboratories and have developed a clearly defined process to do so. The minimum dataset required for data submission is shown in Box 9.

Box 9. Information Needed to Accompany Variant Submission to Central Databases

- Mandatory minimum information:
 - HGVS name or equivalent database accession for variant including alleles
 - Reference sequence used (accession and version if not an LRG)
 - Submitter identification (ORCID ID or equivalent)
 - Number of observations (or defaults to 1)
 - Germline/somatic/unknown
- Optional recommended fields (additional fields can be discussed):
 - LSDB identifier, diagnostic laboratory name or submitter name (preferably from a future LSDB Registry)
 - URL for attribution
 - DOI of publication (or PubMed uid)
 - Individual or sample ID
 - Individual sample ethnicity
 - Individual gender or sample gender frequency
 - Individual genotype or sample genotype frequency
 - Individual or sample phenotype as “ontology name: ontology value” eg MIM number
 - Variant frequency
 - Variant pathogenicity
 - Variant technology/platform
 - Remarks

together with additional, optional recommended fields [Cotton et al., 2008]. For their part, the central databases are committed to allowing data to be imported and displayed on common genome browsers and to facilitate supporting bidirectional data exchange. They will validate HGVS names of the submissions, assign database identifiers where necessary, and provide the technical solutions to enable placement on current and future genome assemblies. They can provide an easy-to-download report with Database, PubMed and OMIM identifiers, allele frequencies from population studies, GWAS associations, representation on LRG/RefSeqGene sequences, and validation results. For data that are not consented for release, the EBI and NCBI provide archives for controlled access data (EGA/dbGaP). A number of specific actions were agreed upon, such as NCBI and EBI to develop standard exchange formats, in conjunction with GEN2PHEN. The central database work group recommends creating a single LSDB Registry in a central location and the use of a phenotype and variation ontology.

Funding mechanisms and sustainability

Expert data curation is necessary to ensure that the publicly available gene mutation data are accurate. No single body worldwide can be expected to bear all the cost. This session proposed various approaches to the sustainability of this program, examined the cost-effectiveness of using virtual genomic information as a tool to enhance disease screening (cancer), and finally proposed a global mode of cooperation to sustain the HVP effort. The existing sustainability models that have been successful for other scientific consortia include (1) mixed model of institutional/government funding and cost recovery; (2) partnership between academic and private sector; (3) distributed cost model, where cost of infrastructures and its administration are distributed across

different grants and contracts within an overall scientific and funding plan; (4) split cost by country, disease, gene, such as the “Adopt-a-GeneTM or Adopt-a-DiseaseTM approach”; (5) top-down (UNESCO/WHO imprimatur) and bottom-up (individual groups/consortia seeking and lobbying for funding) strategies to seek funding. It was recommended that HVP develops an overall scientific plan for specific research and translational projects that will incorporate infrastructure cost, following the approach that is being piloted by the InSiGHT group. HVP should maintain a running list of grants and research projects that rely on or benefit from the HVP initiative to provide leverage with funders and establish a consortium of global funders interested in the HVP goals and mission.

Satellite Meetings

Three societies that support the goals and the mission of HVP and that aim to spread activities to their genes of interest, also held their own Workshops with the view to advance disease-specific variation data collection and utilization. InSiGHT is the peak organization representing health professionals and researchers working on inheritable gastrointestinal cancers (www.insightgroup.org). InSiGHT has volunteered to be a pilot for collecting mutation data for all genes from all countries, and this system is intended to form the template for other disease-specific gene collections. The purpose of the InSiGHT Workshop was to build on the initiatives founded at the 2009 meeting in Duesseldorf between InSiGHT, the HVP, and the National Institutes of Health Colon Family Register [Kohonen-Corish et al., 2010], to understand the progress and direction of the LOVD Colon Cancer Gene Variant Databases, to establish a robust process for addressing unclassified variants, and to commission a phenotype template on the database.

Box 10. Tasks and Recommendations from the InSiGHT Satellite Meeting

- Establish a robust process for collection of phenotype information associated with MMR variants, with a view to quantitation of the information.
- Calibrate various indirect approaches to interpretation of missense variants with “gold standard” information that is derived from phenotype, e.g. segregation analysis of functional assay results with cancer incidence across the family.
- Develop a robust and reproducible algorithm for prediction of pathogenicity based on multiple domains of information applied through a Bayesian Likelihood Ratio approach.
- Apply for funding to support curation and interpretation of unclassified variants, and development of the website.

A summary of this meeting is included in the Supporting Information and a full report is being prepared for publication.

The Nutrigenomics Organization (www.nugo.org) organized a *Micronutrient Genomics Project (MGP)* Workshop, the fourth in a series of meetings to establish and organize the creation of an international micronutrient genomics knowledge base (www.nugo.org/micronutrients) and research effort. The MGP is planning a public bioinformatics resource available consisting of a knowledge base with integrated analytical tools and databases. A key distinction of this effort will be the ability for new research results to be stored, managed, and retrieved for analyses. The three components of the MGP knowledge base are (1) a genetic variation module for all micronutrient-relevant variations; (2) a micronutrient pathway module that links pathways and gene–nutrient interactions at the level of RNA, protein, and metabolites

(micronutrients.wikipathways.org); and (3) a database of omics data, phenotype, and study design. Working groups consisting of researchers and experts have been or are being established for each micronutrient. A summary of this meeting is included in the Supporting Information, and a full report is being prepared for publication.

The *Neurogenetics Consortium* held their second HVP Workshop for the implementation and improvement of mutation databases for genetic disorders of the nervous system. The specific challenges of this task were discussed, including clinical complexity and overlap, genetic heterogeneity, phenotype nomenclature, variant interpretation, informatics procedures, and ethical aspects. Specific case examples on inherited neuropathies, channelopathies, motor neuron diseases, Parkinson’s disease, and mitochondrial cytopathies were presented. This multidisciplinary meeting was attended by clinical neurologists, clinical geneticists, basic researchers, private companies, and informaticians. The meeting resolved to form international expert working groups with the aim to establish coordinated LSDBs on neurogenetic disorders such as spastic paraparesis, Charcot-Marie-Tooth, and mitochondrial diseases. Researchers and healthcare professionals with an interest in the field are encouraged to join these efforts. The main goals and specific tasks emanating from the meeting are summarized in Box 11.

Box 11. Tasks and Recommendations from the Neurogenetics Consortium Satellite Meeting

- Encourage the development of disease-centered (or syndrome-centered) working groups with multidisciplinary experts for the construction of thorough and reliable databases of variations causing neurological disorders.
 - Organize a starting meeting for a motor neuron disease mutation database.
 - Organize a database working group for mitochondrial mutations.
 - Organize a database working group for inherited neuropathies.
 - Identify other areas with more urgent need: Disorder/gene frequently tested for, absence of LSDB, difficult interpretation, etc.
- Encourage already existing neurogenetic databases to join efforts and integrate in case of overlapping information.
- Promote the continuation of a consortium of neuro-databases within the frame of the Human Variome Project.
- Continue to work on the problems that are common to all databases such as development of standards for mutation nomenclature, assessment of pathogenicity, phenotype ontologies, collection process, informatics structure and ethics, focusing on the issues more specific to neurological conditions.
- Disseminate the neurogenetics database initiative at national and international levels.
 - Invite relevant experts and groups to join both the specific syndrome database effort and the consortium.
 - Publish a meeting report.
 - Dissemination actions within administrations and funding bodies: regional, national and international level.
 - Involve scientific societies, patient advocates, relevant journals, and other relevant partners.

A more detailed report of the two Neurogenetics Consortium meetings will be published elsewhere.

HVP Mission

“The vision of the Human Variome Project is to be a catalyst for reduction in human disease in the 21st century by facilitating the establishment and maintenance of standards, systems, and infrastructure for the worldwide collection and sharing of all genetic variations effecting human disease. The HVP is an international consortium committed to reducing the burden of genetic disease on the world’s population. We believe that the collection of information on every instance of a genetic variation

and its affect on human health is the only way that our vision can be achieved. The sharing of information on genetic variation and its consequences allows existing treatments to be delivered more effectively to patients and new treatments and cures to be developed. To ensure the complete capture of all human genetic variation, the HVP is focused on collecting information through two separate, yet complementary, channels: country-specific collection and gene/disease specific collection.”

The full document “Project Roadmap 2010–2012,” which outlines all HVP activities and its governance structure, is publicly available (www.humanvariomeproject.org/index.php/publications/policy-documents).

How to Join the HVP Consortium

Any interested individuals can join the HVP Consortium by registering at www.humanvariomeproject.org. Countries, databases, organizations, and groups are encouraged to participate by becoming a Partner Initiative or an Affiliated Initiative and should apply in writing to the Scientific Advisory Committee (www.humanvariomeproject.org/index.php/about/scientific-advisory-committee) via the HVP Coordinating Office in Melbourne, Australia.

Conclusion

The collation of genetic variations began in the 1950s by researchers and clinicians who established LSDBs to assist research and clinical care. In the 1990s, the discovery of single nucleotide polymorphisms (SNPs) grew into a colossal effort, resulting in the dbSNP. Efforts within the field have achieved increasing alignment between the two approaches; however, their application and infrastructure still require greater consistency. As exemplified in this meeting, the HVP has drawn together disparate groups, by country, gene of interest, and expertise, who are working for the common good with the shared goal of pushing the boundaries of the human variome and collaborating to avoid the unnecessary duplication often prevalent in this area of healthcare. Consequently, more countries are expressing interest: those gathering and curating mutational data together with those with critical relevant technologies.

The 12 sessions dealt with the general aspects needed for phenotype and genotype documentation in inherited disease. In concert with these sessions, three “applied” satellites were convened, one of which addressed the initial HVP/InSiGHT pilot study into inherited colon cancer. Importantly, this gives the variome community a test bed for approaches chosen for their apparent ability to assist busy labs and clinical workers to submit and access data critical to their patients. The community represented at this meeting has focused on the projects that are critical to further their area and represent targets for collaborations, major projects, and student projects. This list is a substantial advance on the 96 recommendations outlined by the first meeting [Cotton et al., 2007]. Furthermore, many of these 96 recommendations have been acted upon during the past 4 years. It is hoped that meetings in specialized areas, such as pathogenicity of variants and disease-specific genes, will lead to similar advances in the next years.

Acknowledgments

Special thanks must go to those at UNESCO whose support allowed this special meeting to occur in such a high profile venue, especially to Maciej Nalecz and Walter Erdelen. The Patronage by UNESCO was particularly welcome. Dr. Erdelen is thanked for his wonderful introductory comments. Julia Hasler and Casimiro Vizzini are thanked for their careful attention to detail and day-to-day support, which allowed the meeting to run smoothly and productively. We are grateful to Heather Howard, Rania Horaitis, Tim Smith, and Lauren Martin for their diligence, and the excellent organizational work they put into the meeting. We thank Wiley–Blackwell Publishing for sponsoring the Editors Meeting and EBI for holding a training workshop as a satellite event. Finally, we thank the speakers and attendees who made the meeting such a successful event.

Disclaimer: The views discussed in this publication do not necessarily reflect those of the U.S. FDA.

References

- Axton M. 2008. Human variome microattribution reviews. *Nat Genet* 40:1.
- Cotton RG, Appelbe W, Auerbach AD, Becker K, Bodmer W, Boone DJ, Boulyjenkov V, Brahmachari S, Brody L, Brookes A, Brown AE, Byers P, Cantu JM, Cassiman JJ, Claustres M, Concannon P, Cotton RG, den Dunnen JT, Flicek P, Gibbs R, Hall J, Hasler J, Katz M, Kwok PY, Laradi S, Lindblom A, Maglott D, Marsh S, Masimirembwa CM, Minoshima S, de Ramirez HZ, Scriver CR, Sherry S, Shimizu N, Shimizu N, Stein L, Tadmouri GO, Taylor G, Watson M. 2007. Recommendations of the 2006 Human Variome Project Meeting. *Nat Genet* 39:433–436.
- Cotton RG, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, Carrera P, Cox DW, Gottlieb B, Greenblatt MS, Hilbert P, Lehtvasilho H, Liang P, Marsh S, Nebert DW, Povey S, Rossetti S, Scriver CR, Summar M, Tolan DR, Verma IC, Vihinen M, den Dunnen JT. 2008. Recommendations for locus-specific databases and their curation. *Hum Mutat* 29:2–5.
- Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso S, Aretz S, Auerbach AD, Bapat B, Bernstein IT, Bhak J, Bleoo SL, Blöcker H, Brenner SE, Burn J, Bustamante M, Calzone R, Cambon-Thomsen A, Cargill M, Carrera P, Cavedon L, Cho YS, Chung YJ, Claustres M, Cutting G, Dagleish R, den Dunnen JT, Díaz C, Dobrowolski S, dos Santos MR, Ekong R, Flanagan SB, Flicek P, Furukawa Y, Genuardi M, Ghang H, Golubenko MV, Greenblatt MS, Hamosh A, Hancock JM, Hardison R, Harrison TM, Hoffmann R, Horaitis R, Howard HJ, Barash CI, Izagirre N, Jung J, Kojima T, Laradi S, Lee YS, Lee JY, Gil-da-Silva-Lopes VL, Macrae FA, Maglott D, Marafie MJ, Marsh SG, Matsubara Y, Messiaen LM, Möslein G, Netea MG, Norton ML, Oefner PJ, Oetting WS, O’Leary JC, de Ramirez AM, Paalman MH, Parboosingh J, Patrinos GP, Perozzi G, Phillips IR, Povey S, Prasad S, Qi M, Quin DJ, Ramesar RS, Richards CS, Savage J, Scheible DG, Scott RJ, Seminara D, Shephard EA, Sijmons RH, Smith TD, Sobrido MJ, Tanaka T, Tavtigian SV, Taylor GR, Teague J, Töpel T, Ullman-Cullere M, Utsunomiya J, van Kranen HJ, Vihinen M, Webb E, Weber TK, Yeager M, Yeom YI, Yim SH, Yoo HS; Contributors to the Human Variome Project Planning Meeting. 2009. Planning the Human Variome Project: the Spain report. *Hum Mutat* 30:496–510.
- Kohonen-Corish M, Weber TK, Lindblom A, Macrae F; the Meeting Participants. 2010. Report of the Combined Meeting of the International Society for Gastrointestinal Hereditary Tumours, the Human Variome Project and the National Cancer Institute Colon Cancer Family Registry, Duesseldorf, Germany, 24 June 2009. *Fam Cancer* (in press).
- Povey S, Al Aqeel AI, Cambon-Thomsen A, Dagleish R, den Dunnen JT, Firth HV, Greenblatt MS, Isaacson Barash C, Parker M, Patrinos GP, Savage J, Sobrido M-J, Winship I, Cotton RG. 2010. Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs). *Hum Mutat* 31 (in press).
- Ring HZ, Kwok PY, Cotton RG. 2006. Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics* 7: 969–972.