MEETING REPORT

Human Mutation

OFFICIAL JOURNAL

HGVS
HUMAN GENOME
VARIATION SOCIETY
www.hgvs.org

Deciphering the Colon Cancer Genes—Report of the InSiGHT-Human Variome Project Workshop, UNESCO, Paris 2010

Maija R.J. Kohonen-Corish,[1]* Finlay Macrae,[2] Maurizio Genuardi,[3] Stefan Aretz,[4] Bharati Bapat,[5] Inge T. Bernstein,[6] John Burn,[7] Richard G.H. Cotton,[8] Johan T. den Dunnen,[9] Thierry Frebourg,[10] Marc S. Greenblatt,[11] Robert Hofstra,[12] Elke Holinski-Feder,[13] Ilkka Lappalainen,[14] Annika Lindblom,[15] Donna Maglott,[16] Pål Møller,[17] Hans Morreau,[18] Gabriela Möslein,[19] Rolf Sijmons,[12] Amanda B. Spurdle,[20] Sean Tavtigian,[21] Carli M.J. Tops,[22] Thomas K. Weber,[23] Niels de Wind,[24] and Michael O. Woods,[25] on behalf of Contributors to the InSiGHT-HVP Workshop

[1]Cancer Research Program, Garvan Institute of Medical Research; St. Vincent's Clinical School, University of NSW, Sydney, NSW, Australia; [2]Depatment of Colorectal Medicine and Genetics, and University of Melbourne Department of Medicine, and The Royal Melbourne Hospital, Parkville, Victoria, Australia; Honorary Secretary, International Society for Gastrointestinal Hereditary Tumours (InSiGHT); [3]Medical Genetics, Department of Clinical Pathophysiology, University of Florence; Fiorgen Foundation for Pharmacogenomics, Sesto Fiorentino, Italy; Tuscany Tumor Institute; [4]Institute of Human Genetics, University of Bonn, Germany; [5]Samuel Lunenfeld Research Institute, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; [6]HNPCC-registret, Hvidovre University Hospital, Copenhagen, Denmark; [7]Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, United Kingdom; [8]Genomic Disorders Research Centre, Carlton South, Melbourne, Victoria 3053, Australia; [9]Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; [10]Inserm U614, Faculty of Medicine, University of Rouen, Rouen, France; [11]University of Vermont College of Medicine, Burlington, Vermont; [12]University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [13]University Hospital of the Ludwig-Maximilians-University, Munich, Germany; [14]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; [15]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; [16]NCBI-National Center for Biotechnology Information, NLM/NIH, Bethesda, Maryland; [17]Section for Inherited Cancer, Department of Medical Genetics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway; [18]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands; [19]HELIOS St. Josefs-Hospital, Bochum, Germany; [20]Genetics and Population Health Division, Queensland Institute of Medical Research, Herston, Brisbane, Australia; [21]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah; [22]Human and Clinical Genetics, Laboratory for Diagnostic Genome Analysis, Leiden University Medical Center, Leiden, The Netherlands; [23]Department of Surgery, State University of New York at Downstate, New York, New York; [24]Department of Toxicogenetics, Leiden University Medical Center, Leiden, The Netherlands; [25]Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada

ABSTRACT: The Human Variome Project (HVP) has established a pilot program with the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) to compile all inherited variation affecting colon cancer susceptibility genes. An HVP-InSiGHT Workshop was held on May 10, 2010, prior to the HVP Integration and Implementation Meeting at UNESCO in Paris, to review the progress of this pilot program. A wide range of topics were covered, including issues relating to genotype–phenotype data submission to the InSiGHT Colon Cancer Gene Variant Databases (chromium.liacs.nl/LOVD2/colon_cancer/home.php). The meeting also canvassed the recent exciting developments in models to evaluate the pathogenicity of unclassified variants using in silico data, tumor pathology information, and functional assays, and made further plans for the future progress and sustainability of the pilot program. Hum Mutat 32:491–494, 2011. © 2011 Wiley-Liss, Inc.

KEY WORDS: colon; cancer; HNPCC; Pathogenicity; variant; InSiGHT; HVP

*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

The Human Variome Project—InSiGHT Pilot Program

The Human Variome Project (HVP) aims to collect all human genetic variation affecting health to a central resource that is freely available, and mobilize the scientific and medical community worldwide to participate in this effort. The International Society for Gastrointestinal Hereditary Tumours (InSiGHT, www.insight-group.org) established a pilot program with the HVP in 2007 to collect all inherited variation affecting colon cancer susceptibility genes. Since then, multiple gene mutation/variant repositories have been merged or linked forming the InSiGHT Colon Cancer Gene Variant Databases. Locus-specific databases (LSDBs) for nine colon cancer susceptibility genes—APC, EPCAM, MLH1, MLH3, MSH2, MSH6,

MUTYH, PMS1, and PMS2 (chromium.liacs.nl/LOVD2/colon_cancer/home.php)—are now housed on the Leiden Open Variation Database (LOVD) platform. Although this has resulted in a dramatic increase of variants listed, there are still major resources and initiatives that need to be incorporated and implemented.

This report provides a summary of the InSiGHT Workshop, which was held prior to the HVP Integration and Implementation Meeting in Paris [Kohonen-Corish et al., 2010b]. The purpose of this Workshop was to (1) understand the progress and direction of the InSiGHT Colon Cancer Gene Variant Databases, (2) establish a robust process for evaluating the clinical significance of unclassified variants in Mismatch Repair (MMR) genes, (3) commission a phenotype template for data submission, and (4) facilitate cost-effectiveness studies and interconnectivity.

The Workshop Convenor and Secretary of InSiGHT, Finlay Macrae, updated the meeting on the developments supporting the HVP-InSiGHT pilot program. The InSiGHT Council has endorsed the inclusion of pathogenicity assignments associated with particular variants on the Database. This carries with it a degree of medicolegal risk, as assignment of pathogenicity, even after interpretation by an expert panel, still carries some risk of being inaccurate due to new information becoming available, from which adverse health outcomes may follow. This has led to the decision by Council to incorporate the organization, providing some degree of protection to the Society and its officebearers. At the same time, Council is positioning the organization as a charity, allowing it to attract donations in a tax-deductible fashion worldwide.

Work in progress includes continuing encouragement for submission of gene variants to the Database, activation of the Interpretation Committee, support for the work of Sean Tavtigian and collaborators in defining the Bayesian Likelihood Ratio approach to assigning pathogenicity, defining a set of missense variants as clearly pathogenic and clearly nonpathogenic for the purposes of calibrating ancillary approaches such as functional assays and in silico analyses, finalizing a phenotype dataset that can be analyzed qualitatively and quantitatively to assist the Bayesian approach, and appointment of a full time curator through the generosity of the Melbourne-based Hicks Foundation. Other work includes fostering the worldwide MMR consortium, which is an initiative between InSiGHT and the NCI Colon Cancer Family Registry, born at the Washington meeting of the two groups on April 27, 2010.

The InSiGHT Databases now have over 3,800 unique variants listed. Initially, gene-specific data from the original InSiGHT Mutation database [Peltomäki and Vasen, 2004] was merged with two other MMR gene databases from Newfoundland, Canada [Woods et al., 2007] and from The Netherlands [Ou et al., 2008]. In 2009, APC and MUTYH were added to the LOVD platform for mutations in familial adenomatous polyposis (FAP) and MUTYH-associated polyposis (MAP). Also, a LOVD was recently established for EPCAM, the new susceptibility gene for Lynch Syndrome. Progress with collecting data for the FAP and MAP genes was discussed (Stefan Aretz and Carli Tops). For MUTYH, it was noted that a large proportion of the "unpublished" variants in the database are of unknown clinical significance. Therefore, expert panels and algorithms for assessment of variant pathogenicity are also needed for MUTYH and submitters are encouraged to provide as much phenotype data as possible.

Phenotype Issues With the Submission of Variants to the InSiGHT Colon Cancer Gene Databases

Unclassified variants continue to be a major challenge for providing reliable information to members of hereditary colon cancer families about their genetic risk. Accurate phenotype data are important in assisting with interpreting the probability of pathogenicity. The meeting discussed, led by Gabriela Möslein, the urgent need to settle the dataset relating to the phenotype of MMR gene mutation carriers. Some divided opinion was expressed about the feasibility and value of annotating phenotype to the database, but after in-depth discussion the meeting concluded that accompanying phenotype information was absolutely essential for the genotypic data to be useful. It did, however, need to be quality data and the logistics of gathering the relevant data are not trivial. The MMR gene Lynch Syndrome phenotype descriptors endorsed by the InSiGHT-HVP-NCI-CCFR meeting 2009 in Duesseldorf were presented again—a five-tier system of phenotype submission from minimal to comprehensive data on pedigrees, the latter allowing linkage analysis—all depending on the comfort of the submitter [Kohonen-Corish et al., 2010a]. A preferred format is that developed by Wijnen et al. [1998] or Barnetson et al. [2006], to predict whether a MMR mutation would be present in the family. It was, however, recognized that the currently available family–history-based algorithms have been developed via the ascertainment of colon cancer probands, and may not accurately reflect all the other cancers found in Lynch syndrome. Finally, the need to prepare an application for funding from the NCI through the NCI Colon Cancer Family Registry (NCI-CCFR) was recognized.

Models to Evaluate the Clinical Significance of MMR Gene Variants

Assessment of variant pathogenicity is one of the 12 key areas of the current HVP activities, to which InSiGHT is making an important contribution. The meeting was given an update of this work by Sean Tavtigian and Amanda Spurdle, who presented progress on developing a suitable algorithm for interpretation of MMR gene variants. The proposed algorithm is based on the previously established integrated evaluation or multifactorial likelihood model used for BRCA1/2 unclassified variants [Goldgar et al., 2004]. This approach can incorporate several independent variables relevant to pathogenicity assessment, such as segregation data, personal and family history of cancer, pathology data (e.g., histology, MSI, IHC, BRAF), in silico assessment of substitution effects and evolutionary conservation, and predicted or demonstrated RNA effects [Spurdle, 2010].

Sean Tavtigian presented pilot data to show the value of in silico analysis of MMR gene missense substitutions for estimation of prior probabilities of pathogenicity. As a starting point to this process, 111 missense variants considered to be clearly pathogenic ($n = 69$) or neutral/low clinical significance ($n = 42$) were identified from the literature [Arnold et al., 2009; Barnetson et al., 2008; Chao et al., 2008], or via questionnaire sent by Robert Hofstra, Rolf Sijmons, Niels de Wind, and Lene Rasmussen to all InSiGHT members in 2009. Sean Tavtigian then built a graded classifier of MMR gene variant pathogenicity using bioinformatic algorithms (Align-GVGD, SIFT, PolyPhen), and assigned prior probabilities in favor of pathogenicity to eight classes ascending from most likely neutral to most likely pathogenic.

Amanda Spurdle leads a project to classify MMR gene variants in the NCI Colon Cancer Family Registry, and presented results from analysis of this dataset to show that MSI status is highly predictive of MMR gene mutation status, and can now be incorporated into the MMR gene multifactorial likelihood model. These two new components will add to the existing method for

assessing variant causality by tracking segregation in families [Arnold et al., 2008], to build the baseline MMR gene multi-factorial model [Spurdle, 2010].

One of the main aims of the session was to define and endorse a set of variants for the calibration of functional assays. These are variants that can be confidently classified as clearly pathogenic or clearly neutral/low clinical significance (i.e., Class 5 or Class 1 in the IARC classification system) [Plon et al., 2008] based on other sources, such as segregation analysis, MSI, or IHC status, and RNA analyses. Following discussion led by Maurizio Genuardi, it was agreed that the set of 111 missense variants already used for in silico evaluation could also be used for functional assay calibration. Due to financial constraints, it is not currently possible to test the whole set. In addition, as shown by different groups, the pathogenicity of some variants may be mainly exerted through effects on RNA processing. Therefore, the number of variants to be tested will be scaled down by selecting a representative group of missense substitutions. Pathogenic variants will then be subjected to RNA analysis (in silico and minigene splicing assays), in order to select those that are most likely to exert their functional effect at the protein level.

It was also agreed that, although the amount of evidence for/against pathogenicity is sufficiently large, the classification process needs to be clearly established, taking into account a number of different factors (i.e., ascertainment mode, specificity, and sensitivity of RNA assay, etc.). To this purpose, a working group for the definition of classification criteria was established upon Finlay Macrae's suggestion. The group includes Bharati Bapat, Thierry Frebourg, Maurizio Genuardi, Marc Greenblatt, Robert Hofstra, Steve Lipkin, Pål Møller, Rolf Sijmons, Amanda Spurdle, Sean Tavtigian, and Michael Woods. It was suggested by Amanda Spurdle that the five-class qualitative scheme she developed for consistent classification of all MMR gene variants in the Colon CFR dataset could be used for standard calibration of the MMR missense dataset.

Functional Assays of Unclassified Variants of the MMR Genes

This discussion was led by Robert Hofstra and focused on the progress made with assigning pathogenicity to unclassified variants using splicing and other functional assays. Niels de Wind presented his laboratory's work on the development, for each MMR protein, of so-called "reverse diagnosis maps" that indicate the positions of amino acid residues that are critical for the function of the protein in vivo [Drost et al., in preparation]. This will enable easier interpretation of the pathogenicity of unclassified variants affecting these amino acids. The second approach to test MMR protein unclassified variants comprises a biochemical assay that directly measures the MMR activity of the protein in vitro [Drost et al., 2010] [Drost et al., in preparation]. This assay is completely cell-free, including the production of the mutant MMR gene and the encoded protein, and will be well suited for use in a clinical diagnostics setting. Evidence was also presented on the importance of RNA splicing analysis (Elke Holinski-Feder, Thierry Frebourg) and quantitation of allelic imbalance at the RNA level (Bharati Bapat). A large proportion of unclassified variants in MLH1 and MSH2 have been found to produce a strong splicing alteration and/or nonsense mediated decay even though in silico splice-site analysis does not predict changes. Many of these unclassified variants can thus be classified as pathogenic and these assays should be considered for integration into clinical practice.

Central Database Resources for the Colon Cancer Variome Community

The Central Databases, National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBI), are active supporters of the Human Variome Project and their contribution was also discussed at the InSiGHT Workshop (Donna Maglott, Ilkka Lappalainen). Clear guidelines have been published for the submission of any variant data, by individual researchers or LSDBs, to the Central Databases and the assistance that will be provided with this process [Kohonen-Corish et al., 2010b]. This will complement the work done by the LSDBs and can support the InSiGHT community by providing a distribution channel and data integration with other resources. Both EBI and NCBI can archive data that have been consented for research but not for fully public dissemination, such as identifiable data from patients. The European Genome–Phenome Archive (EGA; www.ebi.ac.uk/ega) and dbGaP (www.ncbi.nlm.nih.gov/dbgap) provide a service for secure archiving, processing, and dissemination of these data in a manner that respects the original informed consent. NCBI can provide deidentified summary data to gene-specific viewers, such as the LOVD platform. EBI can link variants and their associated phenotype information to the Ensembl Genome Browser (www.ensembl.org) by using the Locus Reference Genomic (LRG, www.lrg-sequence.org) sequences.

The meeting was also given an update on the progress made with virtual pathology databasing of hereditary diseases (Hans Morreau on behalf of Frederik Hes, Marcus Breemer, and others). This comprises high definition histological scanning and E-book documentation. The aims of this effort are to promote the recognition of hereditary causes of cancer, guide gene-finding strategies through pathology, provide guidelines and resources for education and experts.

Conclusion

This was a productive meeting for participants as it increased the awareness of the work completed, or in progress, regarding the collection of colorectal cancer gene variants and the wide range of strengths available in InSiGHT. It also allowed discussion on pertinent issues such as unclassified variants, and enabled participants to lay some plans for further progress and for grant applications. Most importantly, the stakeholders were able to advance the common purpose of understanding the variation in the genes predisposing to colorectal cancer, for the benefit of our patients. Also, the wider HVP community was interested to hear about the outcomes of strategic planning, which is progressively being implemented by InSiGHT.

References

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Goldgar DE, Young JP, Spurdle AB. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757–770.

Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME, Dunlop MG. 2008. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat 29:367–374.

Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. 2006. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 354:2751–2763.

Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H, Lynch H, Lipkin SM. 2008. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). Hum Mutat 29:852–860.

Drost M, Zonneveld JB, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat 31:247–253.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. 2004. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 75:535–544.

Kohonen-Corish M, Weber TK, Lindblom A, Macrae F; the Meeting Participants. 2010a. 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 9:705–711.

Kohonen-Corish MR, Al-Aama JY, Auerbach AD, Axton M, Barash CI, Bernstein I, Beroud C, Burn J, Cunningham F, Cutting GR, den Dunnen JT, Greenblatt MS, Kaput J, Katz M, Lindblom A, Macrae F, Maglott D, Möslein G, Povey S, Ramesar R, Richards S, Seminara D, Sobrido MJ, Tavtigian S, Taylor G, Vihinen M, Winship I, Cotton RG; Human Variome Project Meeting. 2010b. How to catch all those mutations—the Report of the Third Human Variome Project Meeting, UNESCO Paris, May 2010. Hum Mutat 31:1374–1381.

Ou J, Niessen RC, Vonk J, Westers H, Hofstra RM, Sijmons RH. 2008. A database to support the interpretation of human mismatch repair gene variants. Hum Mutat 29:1337–1341.

Peltomäki P, Vasen H. 2004. Mutations associated with HNPCC predisposition—Update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 20: 269–276.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282–1291.

Spurdle AB. 2010. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. Curr Opin Genet Dev 20:315–323.

Wijnen JT, Vasen HF, Khan PM, Zwinderman AH, van der Klift H, Mulder A, Tops C, Møller P, Fodde R. 1998. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 339:511–518.

Woods MO, Williams P, Careen A, Edwards L, Bartlett S, McLaughlin JR, Younghusband HB. 2007. A new variant database for mismatch repair genes associated with Lynch syndrome. Hum Mutat 28:669–673.