

Missense Mutations in Cancer Predisposing Genes: Can We Make Sense of Them?

Rodney J. Scott^{1,2}, Cliff J. Meldrum²

¹Discipline of Medical Genetics, School of Biomedical Sciences, Faculty of Health, University of Newcastle and The Hunter Medical Research Institute;

²Division of Genetics, Hunter Area Pathology Service, Lookout Road, New Lambton, NSW 2305, Australia

Key words: missense mutations, unclassified variants, segregation, clinical utility

Corresponding author: Rodney J. Scott, Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Lookout Road, New Lambton NSW 2305, Australia. E-mail: rodney.scott@newcastle.edu.au

Submitted: 18 August 2005

Accepted: 23 August 2005

Abstract

In the analysis of genes associated with predispositions to malignancy the causative status of mutations can be made relatively easily where it is obvious that there is a clear disruption in the coding sequence of the gene. Difficulties arise, however, if missense mutations are identified, as these are not easily categorised into genetic variants that are not associated with disease risk or into clearly causative changes that impart a significant risk of disease.

As more individuals are subject to DNA sequence analysis for the identification of causative changes in genes associated with cancer predisposition, an increasing number of missense mutations are being identified. Causative status assignment to missense mutations continues to be problematic especially where no functional assessment of the alteration can be made. As more information is gathered on missense mutations our predictive ability to assign significance will improve.

In this report we review, in broad terms, what measures can be undertaken to categorise missense mutations into those that are clearly causative, probably causative and most likely not causative.

Introduction

There are many facets that need to be considered when determining whether a missense mutation is indeed causative or just a common polymorphism that has no impact on the functional activity of the gene in question. The definition of a missense mutation is 'a single base pair substitution that results in the translation of a different amino acid at that position' [1]. This differentiates missense changes from silent polymorphisms where there is no apparent change in the protein product of the gene.

When considering what effect a missense change may have, several aspects need to be considered. 1) Is the change associated with a change in polarity of the

corresponding amino acid, and if it is what effects is this likely to have on the function of the protein? If the change does not alter the polarity of the amino acid, is there any observable effect (i.e. change in function)? 2) How common is the change in the population? Do different populations harbour different frequencies of the change, and if so are these associated with disease? 3) Does the missense change occur in association with another change, and if so is it clearly deleterious (e.g. a nonsense change), other missense changes or more cryptic changes that relate to gene modification?

In determining the effect of missense changes a number of avenues of investigation can be followed. These include the association of missense changes with disease (segregation analysis), the evolutionary

Table 1. Percentage of missense mutation variants identified in the Human Genome Database (to May 2005) [7]

Disease/mutation	Frequency of missense Mutation Variations
Breast Cancer	15.1% BRCA1 (99) 13.9% BRCA2 (62)
FAP	3.1% APC (19)
HNPCC	27.4% hMLH1 (102) 24.6% hMSH2 (75) 32.9% hMSH6 (23)
PJS	19.7% LKB1/STK11 (23)
TOTAL	403 missense variations

conservation of the region within the gene and the presence of loss of heterozygosity in tumour samples from patients harbouring the change.

Missense mutation frequency in different inherited predispositions

In relation to inherited predispositions to cancer there have been a considerable number of missense mutations submitted to the Human Genome Database as shown in Table 1. The number to date reflects a minor proportion but nevertheless some trends are emerging with respect to the likelihood of identifying a missense change in genes associated with predispositions to malignancy. On examination of Table 1 there appears to be a significant difference between the number of missense changes identified in the APC gene compared to any of the other genes. This is most likely due to there being a selection bias for causative mutations identified by the protein truncation test (PTT), which has been used by many groups to identify mutations in exon 15 of the APC gene due to its large size (~6,800 bp). A second reason for the bias comes from understanding the biology of the APC gene, which suggests that missense mutations are unlikely to be associated with disease. Currently, the primary function of the APC gene appears to be the regulation of β -catenin. The APC protein harbours a series of β -catenin binding sites and β -catenin binding and down-regulation domains. A single missense mutation is unlikely to knock out this function since there are multiple domains which harbour these functional moieties [2]. Support of this notion comes from studies that indicate that the site of the primary mutation in the APC gene appears to result in a selective advantage for particular types of mutation in the remaining wild type allele [2]. However, there is little information as to whether missense mutations in the APC gene also result in selective advantages. The difference in frequencies of missense

mutations in the other genes listed in Table 1 can also be attributed to the methods employed to analyse the respective genes. For example, both BRCA1 and BRCA2 contain large exons that comprise several thousand base pairs, which similar to APC lend themselves for analysis by the PTT. This test can only recognise changes that result in premature termination codons and cannot be used to identify any missense mutation. Therefore, the relative percentage of missense changes will be lower than that observed in genes screened by other methodologies, which interrogate DNA more specifically (such as direct sequencing or denaturing high performance liquid chromatography). This is observed in hereditary nonpolyposis colorectal cancer (HNPCC) and Peutz Jeghers Syndrome (PJS), where there is a greater proportion of missense mutations identified in the three genes listed in Table 1. In conclusion, the likelihood of identifying missense mutations in genes associated with inherited predispositions to cancer is influenced by how the majority of mutation analyses are performed and to a lesser extent by the function of the resultant protein product.

Can missense mutations be classified in the absence of functional studies?

In the absence of functional studies a prediction, with a relatively high degree of accuracy, can be made if a series of circumstantial evidence can be taken into account.

Neutral variants: Evidence is accumulating that suggests that there are approximately 3 million single nucleotide polymorphisms (SNPs) spread across the entire human genome, some of which would be described as missense mutations when they occur within coding sequences of DNA. Such changes can be deemed to be neutral variants if they occur frequently in the general population and cannot be linked to disease. One such example is the D1822V change in the APC gene, which was initially believed to be causative but in population studies occurred frequently and was not linked to an increased likelihood of colorectal cancer. Indeed in our own studies the D1822V change was identified at a frequency of approximately 40% in an unselected population [3]. Furthermore, to support this evidence, extensive family studies failed to reveal any segregation with disease. Together, this information can be used to assign causal status of the missense mutation.

In contrast, some missense variants will, using this approach, be considered as causal. For example,

changes that are only observed in rare susceptibilities to disease and are not observed in population-based studies are highly suggestive of causality. One such change, for example, is the L285P missense mutation identified in a PJS family that is rare and segregates exclusively with disease.

This approach, however, is not absolutely reliable. It is possible that mutations exist which fulfil the above criteria but cannot be specifically assigned as a causative change. In one Birtt-Hogg-Dube family the affected proband harboured a S128R change, which appeared to be responsible for the patient's renal cell cancer but when the father (also suffering from renal cell cancer) of the proband was tested for this change the result was negative. Instead, the healthy mother of the individual harboured the change, suggesting either that the change was neutral or that it was associated with highly variable disease penetrance.

Co-occurrence in trans with another deleterious mutation

A missense mutation may be classified as neutral if it is observed in trans with another causative mutation in the germline. This assumes that homozygosity for a causative mutation or compound heterozygosity for two causative changes in any of the common genes associated with cancer predisposition are likely to be lethal conditions. To determine that a missense mutation is in trans with another deleterious mutation requires access to suitable databases that contain full sequencing data on the gene in question. At present, these are limited and probably one of the best sources for such information in relation to cancer genetics is that of Myriad Genetics™, which has sufficient information to assess such changes in BRCA1 and possibly BRCA2.

A single observation of co-occurrence of a missense mutation with a causative mutation is sufficient to classify a missense variation as non-causative. However, classification of missense mutations as causative by this method is more problematic requiring the study of large numbers of individuals to obtain a significant measure of non-occurrence. In summary, this approach to assessing the relevance of a missense mutation is likely to result in the classification of a number of missense mutations as non-causative but overall there is a low power to show causation.

Evolutionary conservation

Evolutionary conservation represents possibly a relatively easy and effective measure by which to assess the causality of missense changes. The advantage of this approach is that it can be applied to every missense change but, and this cannot be overstated, only indirectly relates to disease risk and as such always requires additional information to provide any reliability in prediction.

In using this approach for missense change evaluation it is not sufficient to only compare the human sequence to one other species. It is better to undertake multiple homologue comparison as the further in evolutionary time apart the sequence similarities are the more likely the sequence is absolutely required for the function of the encoded protein. A lack of observed evolutionary change may indicate insufficient time for change rather than sequence conservation through evolution. One significant drawback to this approach is the availability of sequence information for a large enough variety of organisms, as without it the reliability of the result is likely to be lower than that required.

When using this approach, it must be remembered that this method of prediction will only be effective for conserved sequences and if the sequence change occurs in a segment of the gene that is not crucial for its function then it will be unlikely that this approach will be useful in determining causality.

Function

The evaluation of a variant's effect on protein and cellular function probably represents the ultimate assessment for the determination of causality. Any effect of a missense mutation that results in the abrogation or loss of function of the encoded protein is most likely to be causative. However, some caution needs to be taken in extrapolating functional changes to disease states. The relationship of a particular protein function to disease needs to be taken into account.

Missense changes that are different to those reported to alter protein function but occur at sites that have been reported for other changes cannot automatically be assumed to be causative unless functional studies have been undertaken. As mentioned above, not all functional assays, however, may be related to cancer causation and as such care needs to be taken in interpreting some alterations in protein activity. Examples of this occur in the transcriptional activation domains within the genes BRCA1, BRCA2 and the mismatch repair capacity of hMLH1 or hMSH2.

One of the major drawbacks in the assessment of protein function is the lack of functional assays for a significant proportion of genes associated with predispositions to cancer. Alternative methods are being developed to assess the role of missense changes in disease and several new algorithms are being developed for predictive purposes [4]. At the time of writing this review a new SNPWEB is being developed at the UCSF (the server is currently under construction). The aim of this web base is to determine how SNPs affect protein domain structures by comparing the wild type 3D protein model to that containing the SNP [4]. The result of this analysis provides a prediction of the effect of the SNP on the 3D model and how this may relate to a change in protein function.

In conclusion, functional studies that incorporate variations are probably the best evidence for causality, but since functional analyses are often difficult to develop it is more likely that appropriate algorithms will be developed that will have a high degree of predictability and that these will be used with an increasing frequency in the future.

Current practical approaches to assigning causal status to missense mutations

From the above description of what can be done to determine the causality of missense mutations it is obvious that a variety of approaches can be taken in determining whether or not they are linked to disease. The best approach would be to use all of the approaches together to obtain a result that has high predictive power. Unfortunately, not all of the approaches are feasible in all situations and some of them await continued refinement. The following example from Thompson et al [5] indicates how a combined approach can lead to an accurate prediction of causality.

The D2723H missense mutation in BRCA2 has been reported 24 times in the breast cancer information core (BIC) database but has never been shown to be a proven deleterious mutation.

However, the original data in the BIC database suggest at best the odds in favour of causality being at 2:1. Complete cosegregation analysis with disease using 10 pedigrees results in odds in favour of causality being 13,731:1. Combining the odds of both of these factors results in a change in favour of causality to 57,000:1 [5]. When this information is taken together with the knowledge that the change occurs in a conserved region of the genome and that this change has been associated with mis-localisation of the protein

[6] which results in a reduced DNA repair capacity, it is almost certain that this change is causative.

This approach clearly demonstrates the power of applying an integrated approach to determining the causality of a given missense mutation. There are, however, significant improvements that need to be made so that the causality of a missense change can be readily estimated without searching in several different databases and the requirement of functional studies.

In summary, Goldgar et al [6] have proposed an integrated approach to the classification of missense mutations incorporating the following points:

- 1) first step in developing an integrated approach
 - more data required to refine model,
- 2) classification based on:
 - complete neutrality OR causality comparable to known mutations (i.e nothing in between),
- 3) classification ideally based on clinical observation since directly related to cancer risk and more quantifiable,
- 4) of genomic data evolutionary conservation was a useful predictive factor severity of amino acid substitution data was not,
- 5) provides a standard system for evaluation of unclassified variants,
- 6) avoids over-reliance on any single method.

Further refinement is required to optimise this approach and an outcome of the approach taken above has been that the methods already used, which include cosegregation, family history, pathology and other clinical observations, are probably the most useful. Quantitation of the above should allow better standardisation and comparison and, finally, genomic data can strengthen the argument for or against but is not usually sufficient stand-alone evidence.

Whatever approach is used it is clear that at the present time a more systematic approach to the interpretation of missense mutations in genes predisposing to cancer is required. This should include a system for estimating odds ratios for segregation studies so that comparisons can be made between studies and reports of missense changes. This could extend to the inclusion of these estimates in databases that report missense changes, thus allowing the combination of various studies potentially leading to classification of a missense change when significance is reached.

References

1. Genetics Home Reference.
<http://ghr.nlm.nih.gov/ghr/glossary/missensemutation>

2. Albuquerque C, Breukel C, van der Luijt R, Fidalgo P, Lage P, Slors FJ, Leitao CN, Fodde R and Smits R. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet* 2002; 11 (13): 1549-1560.
3. Scott, RJ, Crooks R, Rose L, Attia J, Thakkinstian A, Thomas L, Spigelman AD and Meldrum CJ. Germline missense changes in the APC gene and their relationship to disease. *Hered Cancer Clin Prac* 2004; 2: 81-91.
4. Mirkovic N, Marti-Renom MA, Weber BL, Sali A and Monteiro NA. Structure-based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer predisposition. *Cancer Res* 2004; 64: 3790-3797.
5. Thompson D, Easton DF and Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* 2003; 73: 652-655.
6. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ, and the Breast Cancer Information Core (BIC) Steering Committee. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* 2004; 75: 535-544.
7. Human Genome Database. <http://www.hgmd.cf.ac.uk/hgmd0.html>