A model for the identification of polymorphisms responsible for common genetic conditions and its relationship to multiple sclerosis

Ernest Lad Heisten IV

Ohio State University Alumni Association, Columbus, OH, USA

Article info

Introduction: Analyzing whether genetic polymorphisms proven responsible for known genetic conditions are equal in incidence to the prevalence of those conditions may demonstrate a means to identify polymorphisms responsible for genetic conditions without a known genetic basis. Performing the same analysis on the temperature sensing region of the TRPM8 receptor in relation to multiple sclerosis (MS) could provide evidence of a genetic basis for MS in the TRPM8 receptor.

Aim: To test if correlation between the incidence of a genetic polymorphism and the prevalence of a genetic-linked condition in a large population can identify a polymorphism that could be associated with MS.

Materials and methods: Prevalences of genetic conditions with known genetic polymorphism responsible for that condition were matched to the genetic incidence of the polymorphism proven to cause that condition. The model was then used to identify MS candidate polymorphisms in the S4-S5 region of TRPM8. The University of California Southern California Genome Browser was used for this comparison to identify polymorphism incidence rates.

Results and discussion: Two polymorphisms, Rs28902201 on exon 19 and Rs149328116 on exon 20, were identified as candidate polymorphisms for MS. Other polymorphisms at these loci can contribute to the overall prevalence of MS. Rare variants that have the same genetic effect as the predominant polymorphism responsible for a condition can affect local populations significantly.

Conclusions: Rs28902201 and Rs149328116 are candidate variants for MS based on their incidence in the population and the similarity of these incidences to MS prevalence rates.

Corresponding author: Ernest Lad Heisten IV, 3669 Dragonfly Dr., Columbus Ohio, 43204 United States. Tel.: +1614 260-0125.
E-mail address: heisten.1@osu.edu.
1. INTRODUCTION

Transient receptor potential cation channel subfamily V member 4 (TRPV4) is a heat sensing thermoreceptor that corresponds to mild heat in the same fashion that the transient receptor potential cation channel subfamily M member 8 (TRPM8) thermoreceptor corresponds to mild cold. It stretches from a range extending from where the TRPM8 receptor ceased to distinctively function (25°C–28°C) to 44°C, with a wide peak range of activity that extends from 37°C to 43°C. TRPV4 has the opposite effect on IL-6 that TRPM8 has: it causes the suppression of IL-6 in correspondence to its activity – otherwise referred to as an inverse relationship. This produces a range of IL-6 change that not only changes continuously from the activation of the TRPM8 receptor at 8°C but also continues in a downward trend until the peak activity of TRPV4. Additionally, this corresponds roughly to the range of Clostridium perfringens germination activity, which peaks roughly around 44°C in a growth curve. C. perfringens is a bacteria known for cold related conditions such as gangrene. The reason for why the body would need less energy invested in immune response to C. perfringens at the top of the curve is likely due to the fact that the body experiences a minimal amount of physiological stress during this temperature range. Since the body maintains an average temperature around 37°C regardless of environment it should be recognized that the energy required to maintain this temperature would limit the energy that could be invested in fighting an infection during cold weather, making a rapid response to any bacteria growth imminently necessary, and thus necessitating elevated IL-6 levels. It should be noted that although C. perfringens germination peaks at 43°C–44°C germination effectively stops rising after 36°C. It is possible that this demonstrates a range in which C. perfringens does not germinate as well at higher temperatures, and thus would make IL-6 decreases energetically beneficial to the body.

TRPV4 has demonstrated a range of channelopathies which include skeletal dysplasias (less commonly known as dwarfism), arteropathies, and peripheral neuropathies. Mutations associates with these effects have been documented all along the TRPV4 protein but there are particularly pronounced at the region between TMP4 and TMP5. In TRPM8 the S–S5 region – or TRP4–TRP5 – is recognized as the cold sensing region of the TRPM8 receptor and this corresponds by receptor structure to the heat sensing region of the TRPV4 receptor. This is not the only structure of the receptor that is conserved between TRPM8 and TRPV4: the ‘TRP box’ on the chain end after the TMP6 region is conserved across as thermoreceptive channels. This box is associated with the sensing of lipids in the blood which regulate it’s activation based on temperature. This prevents such receptors from being active during temperature ranges in which their function is unnecessary. It would stand to reason that if there was a negative effect associated with genetic differences associated with the TRPM8 receptor then it would be associated with a missense mutation that disables the temperature sensing region of receptor. Cold sensing is associated with a region of-specified in Sabnis et al. as ‘between-TRP4 and TRP5 subunits of the TRPM8 receptor. This region is associated with the inside of the cell and would potentially be activated by aminoclohexanecarboxylic acid derived from Leucine. This region likely is associated with bending the conformation of the channel in order to facilitate the opening of the channel. Defects in this region would be expected to cause inability of this cold sensing membrane. The region in question is located on exon 19 and 20, and the additional TRP region and the associated menthol/benzylic group binding region is on exon 22. This region would be distinctively associated with any polymorphism that could affect the cold sensing of the TRPM8 receptor.

Exon 7 is associated with trafficking of TRPM8 receptors, but not with cold sensing. Most of the associated single nucleotide polymorphisms (SNPs) that were found to be associated with the Potopova or Kozyreva studies were associated with this region. Exon 7 is only expressed in the brain, liver, and testis like the entirety of the gene is. It is noteworthy, however, that all of the genes that were associated with these studies were also expressed on intron 19, which is located between exon 19 and 20. Intron variants are not as noteworthy as exon variants because their changes are not expressed. However, it is possible that changes in the expression of an intron could affect splicing in its neighboring exons (particularly exon 20, in this instance). As a result intron variants in intron 19 could affect the functioning of the cold sensing region. Analysis of intron 19 reveals that virtually all intron variants in this region have very prolific rates of occurrence, potentially enough to account for the entirety of the variety of the expression of the TRPM8 receptors range of genetic variation.

Analysis of a previously determined polymorphism from TRPV4 reveals a potential methodology to find the gene or genes that may be responsible for MS: V617L is a polymorphism that has been directly associated (more than 95% of patients have the mutation) with the development of polycythemia vera (PV). Prevalence has been crudely estimated at 22–24/100 000 based on a single study of Massachusetts. If an SNP of this condition were to be identified it would at least match that prevalence rate: However, inheritance of this genetic disposition could be heterozygous in nature: thus it would need to be at least twice this rate to be able to match the incidence if that were the case. The associated SNP for this mutation is Rs35078611, which has an incidence for a mutation of 186/125 996, or 148/100 000. This amounts to a prevalence if heterozygous for this condition of 74/100 000. As it turns out this is an autosomal dominant condition, meaning that only one copy of the allele is necessary for the condition. Thus the gene in question for the condition has a prevalence rate that roughly corresponds to the demonstrated incidence rate for PV. Rs56177950 (v-I change) also may be a strong possible candidate gene for corresponding incidence comparisons: It’s rate of prevalence is 979/125 246 (or 782/100 000), which is markedly
more common than most disorders but markedly less common than typical genetic variation – which would usually correspond to between 20%–70% of the subjects tested for any given polymorphism. This polymorphism is associated with mutation V562I. V562I is associated with cystic fibrosis (CF)\(^1\) and that condition has an estimated prevalence of 797 / 100 000 in Europe.\(^2\) This matches the incidence rate of rs56177950 almost exactly.\(^3\) It is also worth noting that the United Kingdom and Ireland were shown to have the highest incidences of CF, which is noteworthy considered the UK was shown to have most of the highest incidences of MS. Additionally, Tay–Sachs disease has been shown to be associated with rs76173977 and has an incidence of 1/3500 newborns (which typically die by age five, thus making prevalence comparisons inappropriate).\(^4\) Rs76173977 has an incidence rate of 29/124924.\(^5\) Using this association between genetic incidence and prevalence of the condition in a large population could be used to identify the polymorphism responsible for MS.

2. AIM

To test if correlation between the incidence of a genetic polymorphism and the prevalence of a genetic-linked condition in a large population can identify a polymorphism that could be associated with MS.

3. MATERIAL AND METHODS

In addition to these proven incidences and corresponding gene associations there are two more possible associations that can be used to give two additional data points: hereditary spherocytosis (HS) and it’s similar condition hereditary elliptocytosis (HE) are conditions with known genetic contributions but no specific gene associated with their incidence. Numerous studies have shown numerous mutations associated with the conditions\(^6\)–\(^8\) but no specific gene that can be said to cause these conditions. Part of the theory regarding this model is the suggestion that there is one predominant mutation that can be used as a barometer for the prevalence of the condition even if other mutations can cause the same conditions by triggering the same genetic result – such as by a downstream frameshift mutation that ultimately produces the same result as a specific missense mutation at a location in the genome that is crucial to a functional gene.\(^9\) This has been demonstrated for HE, where a frameshift mutation at intron 36 causes the effective deletion of exon 37.\(^10\) Analysis of exon 37 reveals rs764095440 has a prevalence of 1/5748,\(^11\) which is very similar to the reported incidence of the condition of HE that is 1/5000.\(^12\) HS has a prevalence of 1/2000\(^13\) and also located on exon 37 is rs141683960, which has a prevalence of 1/2465.\(^14\) Neither allele has been associated with the condition the author has speculated they are indicative of but by adding these additional possible associations it can provide a second level of comparison for a possible trendline by also comparing two additional conditions which have been associated with a particular genetic condition using this model. It is important to note that all of these mutations are point mutations resulting in missense mutations\(^15\) – one codon is changed, and one amino acid is changed in response. None of these are complex frameshift or deletion mutations. This is what would be expected from a very common mutation instead of a much rarer – and potentially lethal – mutation such as deletion or frameshift. These conditions’ prevalence rates and the corresponding genetic incidence of their associated polymorphisms were used to construct a regression line analysis between these two variables – both with and without the unproven associations of HE and HS. The results were considered significant in \(P < 0.05\). Standard error will be multiplied by 100000 to reach the variation expected by incidence and a variation of two times standard error will be used to identify a gene with an expected genetic incidence.

4. RESULTS

The \(P\) value of this correlation was shown to be significant \((P < 0.043)\). When the additional speculated genetic mutants associated with HS and HE are added, then the significance is still the highest possible \((P < 0.0002)\) suggesting that this model is still effective when weighted with its own comparison figures. Standard deviation calculation of acceptable variance based on error means that we can expect a variance of 84.4/100 000 between the highest prevalences of MS documented and the incidence of a candidate polymorphism. A standard deviation means that we can expect a variance of 58.5 from the highest incidence. Since MS affects as many as 193/100 000 in the Shetland Islands – which when corrected for 125996 people in the sample of the PV gene is 243.17/100 000 – polymorphisms demonstrating a missense mutation affecting between 158.97 and 327.37 per 100 000 samples were examined in TRP4–TRP5 (Table 1). Two genes, Rs28902201 and Rs149328116, were identified which were located on exon 19 and 20 of TRPM8, respectively, which fell within that range when corrected for 100000 samples. Rs28902201 represents an asparagine to serine change and Rs149328116 represents a valine to methionine change.

Table 1. Identified alleles (SNPs) that show a strong candidate relationship for MS.

<table>
<thead>
<tr>
<th>Proposed SNP</th>
<th>Exon Location</th>
<th>Frequency</th>
<th>Previously researched SNPs</th>
<th>Exon location</th>
<th>Corresponding function and source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs28902201</td>
<td>19</td>
<td>0.239</td>
<td>rs11562975</td>
<td>7, intron 19</td>
<td>Cold sensitivity (Kozyreva et al. 2011)</td>
</tr>
<tr>
<td>rs149328116</td>
<td>20</td>
<td>0.172</td>
<td>rs11562975</td>
<td>7, intron 19</td>
<td>Cold sensitivity (Kozyreva et al. 2011)</td>
</tr>
</tbody>
</table>
5. DISCUSSION

Collective missense polymorphisms may explain MS incidence in other locations in Europe and elsewhere. These do not include the SNPs already recommended for testing. It is noteworthy that the greatest number of missense variants appear to be clustered on Exon 20, which is not only where one of the candidate genes is located but also where the greatest degree of variety could be expected from changes caused by intron variants in intron 19. It is notable that rs139759512 in exon 24 accounts for 90% of that exon’s variability and is also expressed in intron 17, suggesting that this SNP is also an important contributor to TRPM8 variation. Exon 24 is the final translational region before the 3’ UTR region of TRPM8, so this may constitute another much less significant contributor to variation in the TRPM8 gene. Despite this it should be recognized that all the missense mutations together are not collectively enough to account for MS prevalence rates in many locations. This reality has broader implications for the field of genetics because this could represent a new perspective in how genetic mutations that lead to medical conditions are viewed: while it is true that numerous very infrequent mutations could be associated with the development of numerous conditions the chances that they have a meaningful impact on the prevalence rate of any condition is unlikely. Such mutations may be better viewed as indirectly producing the same effects as a specific mutation that explains the prevalence rate of a condition overall.

Spinal muscular atrophy (SMA, also known as spinal muscular and bulbar atrophy – SMBA) is a condition of which 96% of all cases of this condition are due to a deletion in Exon 7 of the SMN1 gene. It is known to be the second most common pediatric neuromuscular disorder following Duchenne muscular dystrophy and the second most common lethal recessive disorder after CF. In addition to there being numerous Exon 7 and upstream deletions which can cause this condition other polymorphisms can also have the same effect on the functioning of SMN1. In addition to this there is the documented effect that tandem nucleotide repeats (TNRs) in the SMN2 gene have been shown to reduce the severity of the condition. Thus SMA is not only a condition which can be detected early based on deletions for the conditions, but also polymorphisms which reduce the severity of the condition. Not all deletions associated with SMA have been detected in all populations. This screening for each gene’s condition is necessary to better assess the phenotypic development of the condition in addition to recognizing the need to intervene in the gene’s development early. Like HS and HE, different genotypes have been associated with varying phenotypes of the condition in question, and the same is true to a lesser degree with deletions that cause SMA. MS is presumably due to mutations which prevent the normal functioning of the TRPM8 receptor. Like SMN2 repeats, polymorphisms that increase the ability of the receptor to open can be expected to decrease the severity or prevent the development of MS—primarily those that produce a functional TRPM8 channel. Like SMN1, those that delete the molecular means to open the TRPM8 channel could be expected to cause the development of MS.

Not all genetic polymorphisms are present in all populations. In the case of SMA several different polymorphisms can all cause the same deletion but not all of those will necessarily be present in a given population. In the cases of MS there have been polymorphisms documented that cause Voltage dependent failure of the TRPM8 receptor and there have been polymorphisms documented that have caused temperature dependent failure of the TRPM8 receptor. In addition to familial inheritance that have been well documented for the development of MS, variations based on populations have also been documented. The Sami in Norway have historically been known to be resistant to MS. The Nuoro in Sardinia, by comparison, have been consistently shown to have some of the highest rates of MS in Europe without the traditionally recognized latitudinal dependence. Both likely possess group preservation of polymorphisms which prevent and promote, respectively, the development of MS. In addition to these extreme example populations of people in any local area could be expected to preserve some polymorphisms and not possess others. This highlights the need to develop genetic testing that accounts for all polymorphisms known—even those which are apparently rare when tested in a large population. Such rare polymorphisms may well be the defining genetic polymorphism responsible for MS in a local population particularly if they have the very same effect as the more common polymorphism. Case in point is Rs200511441, which is a premature stop codon in front of Rs149328116 with a genetic incidence of 2/121 380. This would be expected to have the same effect as the more prolific polymorphism but not expected to be tested for in general. In a small local population, however, it could potentially be expected that the significantly more common polymorphism would be absent and the much rarer polymorphism would be the defining cause of MS. It should be recognized that while this methodology can identify a reasonably common polymorphism in a large population it is not able to identify a rare polymorphism in the same population. Adjusting this methodology to test a local population for a condition in which the prevalence of that condition is known in that population could be used to identify the polymorphism responsible in that population, even if that polymorphism is much less common. This can be done by identifying the prevalence of that condition in a given population and then repeating the correlation with a local genetic database.

6. CONCLUSIONS

The genetic incidence of Rs28902201 and Rs149328116 in a large population roughly correspond to the prevalence of MS in the highest prevalence locations. It would be realistic to test these polymorphisms for their association with MS. Other polymorphism which affect the specific exons associated with these polymorphism should also be consid-
er as contributing factors to overall MS prevalence and potentially more significant than these polymorphisms in local populations. Genetic testing for all known polymorphism within the TRP4–TRP5 region of the TRPM8 receptor could be used to screen for the risk of developing MS. The method used to locate these candidate genes could be used to identify genetic polymorphisms responsible for other conditions as well as other polymorphisms responsible for such conditions in local populations. Correction of this methodology for the prevalence of a genetic condition in a local population could be used to identify a rare polymorphism not generally responsible for the same condition in a larger population.

Conflict of interest
None.

Funding
None.

References