

Prematür Yaşlanmayla İlişkili *RCE1* Geninde Meydana Gelen Yüksek-Riskli, Yanlış Anlamlı Tek Nükleotid Polimorfizmlerinin Tespiti İle Yapısal Tesirleri

Detection And The Structural Effects Of The High-Risk, Missense Single Nucleotide Polymorphisms That Occur In *RCE1* Gene Related To Premature Aging

Nazlı Irmak GİRİTLİOĞLU¹

ÖZET

Amaç: Bu çalışmanın amacı, *RCE1* geninde meydana gelen yüksek-riskli, yanlış anlamalı SNPlerin belirlenmesi ve aminoasit değişiminin *RCE1* protein ürünü üzerindeki yapısal tesirlerinin tahmin edilmesidir.

Yöntem: *RCE1* geninde meydana gelen yanlış anlamalı mutasyonlar, NCBI SNP veritabanında taranmıştır. Yanlış anlamalı olarak kategorilendirilmiş 329 SNP'nin patojeniteleri; SIFT, PROVEAN ve PANTHER cSNP çevrimiçi araçlarındaki algoritmalar ile hesaplanmıştır. "Zararlı" ve "muhtemelen zararlı" olarak belirlenmiş olan üç SNP, *RCE1* proteini üzerindeki yapısal tesirlerinin tahmini amacıyla I-Mutant ve HOPE ile değerlendirilmiştir.

Bulgular: SIFT, PROVEAN ve PANTHER cSNP ile belirlenmiş olan üç yüksek-riskli yanlış anlamalı SNP (rs201456801, UniProt protein id'leri: Q9Y256, E9PI08 ve E9PKA7), I-Mutant yazılımı sonuçlarına göre protein stabilitesini arttırmaktadır. HOPE veritabanına göre üç SNP'nin de yol açtığı aminoasit rezidü değişimleri, vahşi tip rezidülere göre daha hidrofobik ve daha büyüktür. Sadece Q9Y256'da vahşi tip rezidü korunmuşluk göstermektedir. T303M ve T180M değişimlerinin proteine zarar vermeme olasılığı mevcuttur.

Sonuç: SIFT, PROVEAN ve PANTHER cSNP çevrimiçi araçlarına göre *RCE1* geninde yüksek-riskli olması beklenen üç SNPden ikisinin; I-Mutant ve HOPE'a göre proteine zarar vermeme olasılığı bulunmaktadır. Araştırmanın ileride *in silico*dan *in vivo*ya genişletilmesi ve "tahmin" seviyesinden "kesin" sonuçlara erişilmesi hususunda bu çalışmanın iyi bir başlangıç olabileceği düşünülebilir.

Anahtar kelimeler: *RCE1*, tek nükleotid polimorfizmleri, yüksek-riskli SNP

ABSTRACT

Aim: The aim of this study is to identify high-risk, missense SNPs in the *RCE1* gene and to predict the structural effects of amino acid change on the *RCE1* protein product.

Method: Missense mutations in the *RCE1* gene were screened in the NCBI SNP database. Pathogenicity of 329 SNPs that were categorized as missense calculated by algorithms in SIFT, PROVEAN, and PANTHER cSNP online tools. Three SNPs that were determined as "deleterious" and "probably damaging" were evaluated by I-Mutant and HOPE to estimate their structural effects on the *RCE1* protein.

Results: Three high-risk missense SNPs (rs201456801, UniProt protein ids: Q9Y256, E9PI08, and E9PKA7) that identified with SIFT, PROVEAN, and PANTHER cSNP increase protein stability by the I-Mutant software results. According to the HOPE database, amino acid residue changes caused by three SNPs are more hydrophobic and bigger than wild-type residues. The wild-

¹ Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Moleküler Biyoloji ve Genetik Anabilim Dalı, İstanbul/TÜRKİYE. ngiritlioglu@outlook.com, ORCID: 0000-0001-5034-0984.

type residue is conserved in only Q9Y256. There is a possibility that T303M and T180M modifications will not damage the protein.

Conclusion: According to the SIFT, PROVEAN, and PANTHER cSNP online tools, two of the three SNPs that are expected to be high-risk in the *RCE1* gene, there is a possibility of not damaging the protein according to I-Mutant and HOPE. It can be considered, this study could be a good start in expanding the research from *in silico* to *in vitro* and *in vivo* and reaching "definitive" results from the "prediction" level.

Keywords: RCE1, single nucleotide polymorphisms, high-risk SNP

GİRİŞ

Tek nükleotid polimorfizmleri (SNPler), insan genomunda yüksek yoğunluk ve nispeten eşit dağılım gösteren genetik markörlerdir (Chen & Sullivan, 2003). %1'e eşit veya daha büyük bir frekansa sahip değişiklikler polimorfizm olarak tanımlanmaktadır. Mutasyonlar ise genel popülasyonun %1'i kadarını etkileyen, nükleotid sekansındaki değişikliklerdir. Mutasyonlar genellikle patolojik fenotiplerin gelişmesine yol açarken, polimorfizmlerin etkileri önemli ölçüde değişir. SNPler en yaygın varyasyonlardır ve bireyler arası çeşitliliğin ana kaynağıdır. Bununla birlikte çoğu SNP, kodlamayan genom bölgelerinde yer alır ve bu nedenle etkilerini tahmin etmeyi zorlaştırır (Leaché & Oaks, 2017). SNPler gen içi ve gen dışı bölgelerde bulunabilir ve genellikle bialleliktir. Büyük bir kısmı insan hastalıklarına doğrudan etki ettiğinden dolayı (Ramírez-Bello & Jiménez-Morales, 2017), moleküler tıpta büyük bir öneme sahiptir. Hastalıklara karşı duyarlılığımızdaki farklılıkların, hastalığın ciddiyetinin ve tedaviye verdiğimiz tepkilerin altında bu genetik varyasyonlar yatmaktadır (Shaw, 2013). SNPler, DNA'nın kodlama yapılan bölgesinde ve kodlama yapılmayan bölgesinde olmak üzere genel olarak iki kategoriye ayrılabilir. Kodlama yapılan bölgelerde gerçekleşen SNPler, eş anlamlı olan ve olmayan SNPlerdir. Yanlış anlamlı SNPler, eş anlamlı olmayan SNPler grubuna dahildir. Bu SNP türü, üçlü kodonda tek bir nükleotidin değişimi ile translasyonda farklı bir aminoasitin eklenmesine ve protein dizisinin değişmesine neden olmaktadır.

RCE1 geni, metaloproteinaz ailesinin bir üyesi olarak sınıflandırılan integral bir zar proteinini kodlar. Bu enzimin, CAAX (C: Sistein, A: Alifatik aminoasit, X: Herhangi bir aminoasit) tipi prenil proteinlerin korunmasında ve işlenmesinde fonksiyonu olduğu düşünülmektedir ("RCE1 Gene", GeneCards). CAAX motifi, protein matürasyonunda çok önemli bir adımdır. Prematür yaşlanmaya neden olan Hutchinson-Gilford Progeria Sendromu (HGPS) ve bazı progeroid sendromlar, lamin A öncüsü olan prelamin A'nın hatalı işlenmesinden kaynaklanır. CAAX motifinin modifikasyonu; i) CAAX sisteininin farnesil transferaz aracılığı ile farnesilasyonu, ii) son üç aminoasit (AAX) rezidüsünün endoproteolitik olarak uzaklaştırılması ve iii) farnesillenmiş C-terminal sisteininin izoprenilsistein karboksi-metil transferaz (ICMT) ile metillenmesini içerir.

Farnesil transferaz, çözünür bir enzim iken RCE1, ZMPSTE24 ve ICMT enzimleri, endoplazmik retikulumda ve ayrıca iç nükleer membranda bulunan integral membran proteinleridir. RCE1 ve ZMPSTE24, son üç amino asidin CAAX motifli proteinlerden uzaklaştırılması için önemlidir

(Kühnle, Dederer, & Lemberg, 2019). Prelamin A'dan modifiye edilmiş kuyruğun çıkarılmasının amacı net değildir, ancak gerçekleşmediği takdirde ciddiyeti nispeten hafif ile neonatal ölümcül arasında değişen progeroid bozukluklarla ilişkili olmaktadır. Şiddetin artışına göre bu hastalıklar metabolik bozukluk, mandibuloakral displazi tip B, HGPS ve neonatal ölümcül kısıtlayıcı dermopati olarak sıralanabilir (Barrowman, Hamblet, Kane, & Michaelis, 2012).

Biyoinformatik çerçevede yapılan çalışmalarda, herhangi bir spesifik gendeki veya bir hastalıkla ilişkili tüm yüksek-riskli SNPlerin belirlenmesinde; ücretsiz ve yüksek tahmin skorlarına sahip *in silico* çevrimiçi araçlar karşımıza efektif bir yol olarak çıkmaktadır. Ayrıca belirli bir SNP'nin protein üzerindeki yapısal değişimleri de tahmin edilebilmektedir. Güncel NCBI SNP veritabanında (dbSNP) (Sherry ve diğ., 2001), *LMNA* ve *ZMPSTE24* genlerine ait yanlış anlamlı mutasyonların patojenik gruplarına dair bilgiler yer alsa da *RCE1* geni için yüksek-riskli SNP'lere dair bir bilgi bulunmamaktadır. Ayrıca, PubMed veritabanında "RCE1" ile ilgili karşımıza sadece 133 sonuç çıkmaktadır. Çalışmada, öncelikle NCBI SNP veritabanından (dbSNP) SNP madenciliği ile RCE1 genine ait tüm yanlış anlamlı SNPler belirlenmiş; sonra Sorting Intolerant From Tolerant (SIFT) (Ng & Henikoff, 2003), Protein Variation Effect Analyzer (PROVEAN) (Choi & Chan, 2015) ve PANTHER coding SNPs (cSNPs) (Thomas et al., 2003) yazılımları aracılığıyla bu SNPler arasından zararlı olanlar, risk skorları ile birlikte tespit edilmiştir. Son olarak I-Mutant (Capriotti, Fariselli, & Casadio, 2005) ve HOPE (Venselaar, te Beek, Kuipers, Hekkelman, & Vriend, 2010) yazılımları ile zararın protein üzerindeki yapısal etkileri ortaya çıkarılmış ve tartışılmıştır.

MATERYAL VE YÖNTEM

Yanlış Anlamlı Mutasyonların NCBI dbSNPden Eldesi

NCBI dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>), genetik polimorfizmleri kamuya açık ve geniş bir arşiv olarak sunan bir veritabanıdır. Her bir dbSNP kaydında polimorfizmin sekans bağlamı, meydana gelme sıklığı ve varyasyonun test edilmesinde kullanılan deneysel yöntemler, protokoller ve koşullar vardır (Kitts & Sherry, 2011). dbSNP veritabanında *RCE1* için toplam 2085 SNP bulunmaktadır ve bunların 329 tanesi yanlış anlamlıdır. Bu 329 SNP, .txt uzantılı olarak ayrı bir dosya halinde kaydedilmiştir.



Yüksek-Riskli Yanlış Anlamli Mutasyonların Belirlenmesi

329 SNPnin kaç tanesinin yüksek-risk içerdiğinin belirlenmesi için web-tabanlı araçlardan faydalanılmıştır. Bu amaçla SIFT (<https://sift.bii.a-star.edu.sg/>), PROVEAN ve PANTHER çevrimiçi araçları kullanılmıştır. İlk olarak .txt uzantılı dosya SIFT dbSNP rsIDs (SIFT4G predictions)'ye yüklenmiş ve "DELETERIOUS" ve "DELETERIOUS (WARNING LOW CONFIDENCE)" olarak belirlenmiş polimorfizmler PROVEAN (<http://provean.jcvi.org/index.php>) ve PANTHER cSNP (<http://www.pantherdb.org/tools/csnpscoreform.jsp>) çevrimiçi araçlarında araştırılmak üzere ayrı bir liste olarak kaydedilmiştir. PROVEAN, SIFT ile benzer bir mantıkta işleyen bir yazılımdır ve SNPlere ilişkili olarak değişen amino asit rezidülerinin; ilgili proteinlerin biyolojik işlevleri üzerinde bir etkiye sahip olup olmadığını tahmin etmek amacıyla bu çalışmada kullanılmıştır. PANTHER cSNP de PROVEAN gibi, aminoasit rezidü değişiminin protein üzerinde fonksiyonel bir değişime sebep olup olmadığını tahmin eden web-tabanlı bir yazılımdır. Analiz ayrıca aminoasitin ilgili proteine giden "soy"sal yolda ne kadar süre korunduğunu (milyonlarca yıl) da gösterir. SIFT, PROVEAN ve PANTHER cSNP yazılımlarından elde edilen verilerde yüksek-riskli olan ortak SNPler kaydedilmiştir.

Yüksek-riskli SNPlere İlişkili Aminoasit Rezidülerinin Yapısal Tesirlerinin Tahmini

Yüksek-riskli olan ortak SNPlerin yapısal tesirlerinin tahmininde I-Mutant ve HOPE yazılımları kullanılmıştır. I-Mutant, tek nükleotid mutasyonlarının proteinin stabilitesi üzerindeki etkilerinin tahmininde kullanışlı bir yazılımdır. Protein stabilitesini artmış/azalmış şeklinde güvenilirlik indeksi (Reliability Index-RI) ile birlikte göstermektedir. HOPE, bir mutasyonun protein yapısı üzerindeki etkisini analiz etmek için bir dizi kaynaktan topladığı bilgileri birleştirerek sunmaktadır.

BULGULAR VE TARTIŞMA

NCBI dbSNPden ekstrakte edilen 329 SNPnin sadece üç tanesinin "zararlı" ve "muhtemelen zarar verici" oldukları SIFT, PROVEAN ve PANTHER cSNP çevrimiçi araçları ile saptanmıştır (Tablo 1). Bu üç yüksek-riskli SNP, I-Mutant yazılımına göre protein stabilitesini arttırmaktadır (Tablo 2). HOPE veritabanına göre üç SNPnin de yol açtığı aminoasit rezidü değişimleri, vahşi-tip rezidüleri göre daha hidrofobik ve daha büyüktür (Şekil 1). Vahşi tip rezidülerden yalnızca bir tanesi (rs201456801-Q9Y256) korunmuşluk göstermektedir (Tablo 3).

Tablo 1. RCE1 genindeki yüksek-riskli SNPlerin SIFT, PROVEAN ve PANTHER cSNP çevrimiçi araçları ile belirlenmesi, yüksek-riskli SNPlerin skorları ve patojenite tahminleri

SNP ID	NT DEĞİŞİMİ	AA DEĞİŞİMİ	BÖLGE	PANTHER KORUNMA SÜRESİ	SKOR		TAHMİN		
					SIFT	PROVEAN	PANTHER cSNP	SIFT	PROVEAN
rs201456801	C>T	T303M	CDS	1628	0	-5.361	Muhtemelen zarar verici	Zararlı	Zararlı
rs201456801	C>T	T282M	CDS	1628	0	-5.361	Muhtemelen zarar verici	Zararlı	Zararlı
rs201456801	C>T	T180M	CDS	1628	0	-5.561	Muhtemelen zarar verici	Zararlı	Zararlı

T303M değişiminin yol açtığı mutant rezidü, transmembran bölgede bulunmaktadır. Vahşi-tip ve mutant tip arasındaki boyut farkı, lipid membran ile olan teması etkileyebilir ve düzensizliklere neden olabilir. Mutant rezidünün daha hidrofobik olması ise membran lipidleri ile hidrofobik etkileşimler üzerinde değişikliklere yol açabilmektedir. Bu da hidrojen bağlarının kaybına ve/veya katlanmada bozulmaya yol açabilmektedir. Vahşi tip rezidü çok korunmuştur. Mutant rezidü, diğer homolog proteinlerde bu

pozisyonda gözlenen diğer rezidü türleri arasında olmamakta ve mutasyona uğramış rezidü ile ortak bazı özelliklere sahip rezidüleri gözlemlenmektedir. Bunun anlamı, proteine zarar vermeden mutasyonunun gerçekleşebileceğidir. T282M değişiminde mutant rezidü, bu pozisyonda diğer homolog dizilerde görülen rezidü türleri arasında değildir; bu da mutant rezidünün proteine zarar verebilme ihtimalinin bulunduğu göstermektedir.

Tablo 2. SNP ile tetiklenen aminoasit rezidü değişimlerinin protein stabilitesi üzerindeki etkisinin I-Mutant yazılımı saptanması

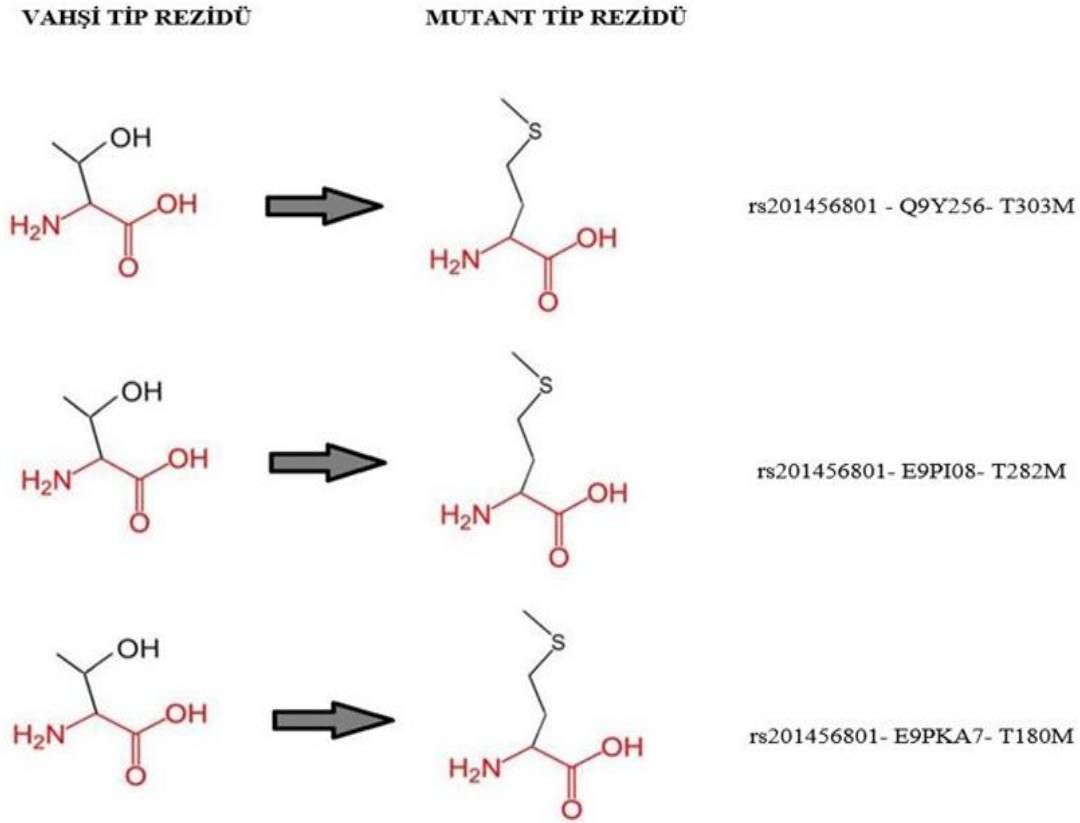
SNP ID	UniProt PROTEIN ID	I-MUTANT	
		RI	Protein Stabilitesi
rs201456801	Q9Y256	2	Artış
rs201456801	E9PI08	2	Artış
rs201456801	E9PKA7	2	Artış

T180M değişiminde, rezidünün pozisyonunda korunmuşluk yoktur. Diğer homolog dizilerde bu pozisyonda başka bir rezidü türü daha sık gözlemlenmektedir. Bu, diğer homolog proteinlerin, protein dizisindeki vahşi tip rezidüden ziyade diğer rezidü tipiyle var olduğu anlamına gelir. Ayrıca,

mutant rezidünün de bu pozisyonda bulunması, mutasyonun muhtemelen proteine zarar vermediğine işaret etmektedir. RI, I-Mutant yazılımının belirlediği, 0 ile 10 arasında bir değerdir ve I-Mutant yazılımı sonuçlarına göre her üç SNP için de RI değeri 2'dir.

Tablo 3. Yüksek riskli üç SNP'nin protein yapısı üzerindeki tesirleri

SNP ID	UNIPROT ID	VAHŞİ TİP REZİDÜNÜN ÖZELLİKLERİ		MUTANT TİP REZİDÜNÜN ÖZELLİKLERİ		VAHŞİ TİP REZİDÜ KORUNMUŞ/ KORUNMAMIŞ
		BOYUT	FİZİKSEL DURUM	BOYUT	FİZİKSEL DURUM	
rs201456801	Q9Y256	Daha küçük	Daha az hidrofobik	Daha büyük	Daha hidrofobik	Korunmuş
rs201456801	E9PI08	Daha küçük	Daha az hidrofobik	Daha büyük	Daha hidrofobik	Korunmamış
rs201456801	E9PKA7	Daha küçük	Daha az hidrofobik	Daha büyük	Daha hidrofobik	Korunmamış



Şekil 1. HOPE veritabanına göre vahşi tip ve mutant rezidülerin yapısal olarak gösterilmesi

SONUÇ

SIFT, PROVEAN ve PANTHER cSNP çevrimiçi araçlarına göre; bulgular ve tartışmada belirtilmiş olan ve RCE geninde yüksek-riskli olması beklenen üç yanlış anlamalı SNPden ikisinin; I-Mutant ve HOPE'a göre proteine zarar vermeme olasılığı bulunmaktadır. I-Mutant yazılımının algoritmasının, her üç SNPnin de protein stabilitesini arttırmış olarak göstermesi, RI'nın düşük olması dolayısıyla tam bir analiz sunmamakta; lakin HOPE veritabanının yapısal değerlendirme, I-Mutant'ın güvenilirliğini arttırmaktadır. Sonuç olarak, araştırmanın ileride *in silico*dan *in vivo*ya genişletilmesi ve "tahmin" seviyesinden "kesin" sonuçlara erişilmesi hususunda bu çalışmanın iyi bir başlangıç olabileceği düşünülebilir.

KAYNAKÇA

Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, 33(Web Server issue), W306-10. <https://doi.org/10.1093/nar/gki375>

Chen, X., & Sullivan, P. F. (2003, May 14). Single nucleotide polymorphism genotyping: Biochemistry, protocol, cost and throughput. *Pharmacogenomics Journal*, Vol. 3, pp. 77-96. Nature Publishing Group. <https://doi.org/10.1038/sj.tpj.6500167>

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics (Oxford, England)*, 31(16), 2745-2747. <https://doi.org/10.1093/bioinformatics/btv195>

Kitts, A., & Sherry, S. (2011). *The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation*. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK21088/>

Leaché, A. D., & Oaks, J. R. (2017). The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 48, 69-84. <https://doi.org/10.1146/annurev-ecolsys-110316-022645>

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), 3812-3814. <https://doi.org/10.1093/nar/gkg509>

Ramírez-Bello, J., & Jiménez-Morales, M. (2017). [Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases] - PubMed. *Gac Med Mex*., 153(2), 238-250. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/28474710/>

RCE1 Gene - GeneCards | FACE2 Protein | FACE2 Antibody. (n.d.). Retrieved October 28, 2020, from



<https://www.genecards.org/cgi-bin/carddisp.pl?gene=RCE1>

Shaw, G. (2013). Polymorphism and single nucleotide polymorphisms (SNPs). *BJU International*, 112(5), 664–665. <https://doi.org/10.1111/bju.12298>

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). DbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), 308–311. <https://doi.org/10.1093/nar/29.1.308>

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H.,

Karlak, B., Daverman, R., ... Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research*, 13(9), 2129–2141. <https://doi.org/10.1101/gr.772403>

Venselaar, H., te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*, 11(1), 548. <https://doi.org/10.1186/1471-2105-11-548>