

APROXIMACIÓN *IN SILICO* PARA LA PREDICCIÓN DE EFECTOS FUNCIONALES PROTEICOS A PARTIR DE LA DETECCIÓN DE VARIANTES SECUENCIALES DE DNA

IN SILICO APPROACH FOR THE PREDICTION OF PROTEIN FUNCTIONAL EFFECTS
FROM THE DETECTION OF SEQUENCE VARIANTS OF DNA

Adalberto Sánchez Gómez¹

¹ Profesor Asistente, Departamento de Ciencias Fisiológicas, Facultad de Salud, Universidad del Valle.
email: adalberto.sanchez@correounivalle.edu.co

Recibido: 10 de Junio de 2014

Aceptado: 08 de Agosto de 2014

Correspondencia del autor: E-mail adalberto.sanchez@correounivalle.edu.co
Facultad de Salud, Universidad del Valle. Cali, Colombia

RESUMEN

El desarrollo de tecnologías de secuenciación masiva de regiones codificantes del DNA en los últimos años ha generado un impacto en el área de la patología molecular. Sin embargo, los resultados de este proceso han sobrepasado las capacidades para la predicción y asociación de variantes de secuencia de DNA con efectos de tipo patológico por alteraciones en proteínas vinculadas a procesos fisiológicos, anatómicos y metabólicos. Para aliviar este efecto, se han desarrollado recientemente una serie de herramientas computacionales que permiten traducir los resultados de variantes secuenciales en efectos funcionales que explicarían la etiología de muchas enfermedades clasificadas como mendelianas. En este trabajo se evalúan dos de esas herramientas, Sift y PolyPhen-2. Para esta evaluación se utilizaron los datos de secuenciación exómica de un grupo de 2 pacientes diagnosticados con alguna de las patologías clasificadas en el grupo de la mucopolisacaridosis, provenientes del suroccidente colombiano. En el estudio se registraron en promedio un número de 60000 variantes secuenciales por paciente, siendo al menos la mitad de ella clasificadas como no registradas en la literatura. A partir del análisis de estos datos se registraron al menos 3 variantes con efecto funcional que estarían explicando el fenómeno patológico de los pacientes. Se propone entonces la utilización de este tipo de herramientas computacionales para agilizar la interpretación de los resultados de secuenciación exómica con el propósito de realizar un proceso de consejería terapéutica en pacientes con enfermedades de tipo mendeliano y además permitir la aplicación de nuevas estrategias para estas enfermedades como es el caso de la terapia de reemplazo enzimático.

Palabras clave:

ABSTRACT

The development of technologies of mass sequencing of coding regions of DNA in recent years has had an impact in the field of molecular pathology. However, the results of this process have surpassed the capabilities for the prediction and association of DNA sequence variants with type pathological effects of alterations in proteins linked to physiological, anatomical and metabolic processes. To alleviate this effect, a series of computational tools have been developed to translate the results of sequence variants in functional effects that explain the etiology of many diseases classified as Mendelian. In this paper we evaluate two of these tools, Sift and PolyPhen-2. For this evaluation, exomic sequencing data of a group of 2 patients diagnosed with any of the diseases classified in the group of mucopolysaccharidosis, from southwestern Colombia were used. In the study a total of 60,000 sequence variants per patient were recorded on average, at least half of it classified as not registered in the literature. From the analysis of these data, it were recorded at least 3 variants whose effect would explain the pathological phenomenon of patients. The use of this type of computational tools is then proposed to improve the interpretation of the results of exome sequencing. Therefore, this improving would allow a better therapeutic process, including counseling on patients that suffer from Mendelian diseases and also enable implementation of new strategies for these diseases as is the case of enzyme replacement therapy.

Keywords:**Introducción**

La predicción *in silico* de cambios en las estructuras proteicas ha sido un área de particular interés en el campo de la patología molecular. Desde las primeras aproximaciones realizadas en la década de los setentas, el reconocimiento de plegamientos tridimensionales de estas biomoléculas mediante algoritmos, ha permitido una rápida asociación entre variantes secuenciales y diversas patologías en seres vivos. En la actualidad, se han desarrollado nuevas herramientas para la detección de este tipo de variantes secuenciales pero a partir de las secuencias de ADN que codifican por la proteína de interés, una de ellas es la secuenciación del genoma completo que ofrece un panorama general del contenido informativo de un individuo. Sin embargo sus costos han impedido una implementación para la masificación de su uso en estudios de asociación patológica Lacey *et al*(1). Recientemente, se ha propuesto la secuenciación de la porción codificante del genoma, exoma, ya que los costos asociados al proceso están alrededor de una tercera parte de la secuenciación completa del genoma y se estima que alrededor del 80% de las variantes secuenciales que se registran en esta proceso se asociación a enfermedades Lalonde *et al*(2). Estas variantes secuenciales de tipo no sinónimo se relacionan directamente con modificaciones de la estructura proteica que se tra-

duce en efectos funcionales que afectan a un tejido, órgano o inclusive al sistema completos, Ng *et al*(3); Prada *et al*(4) .

Varios estudios con este tipo de variantes, como los realizados por Koboldt(5) y Peloso(6), han demostrado que las denominadas raras, osea aquellas que a nivel poblacional presentan una Frecuencia Alélica Mínima ($MAF \leq 1$) presentan una mayor prevalencia sobre aquellas que tienen valores $MAF \geq 1$ y $MAF \geq 5$. Esta tendencia nos relaciona además, el origen de muchas de estas variantes, encontrándose que un gran número son de novo o que aún no han sido reportadas, y que poseen una altísima probabilidad de estar asociadas con alteraciones funcionales que expliquen fenómenos de tipo patológico Kong *et al*(7); Sanders *et al*(8). La literatura relacionada registra entonces la necesidad de verificar este tipo de comportamientos epidemiológicos y la búsqueda de herramientas con suficiente poder de predicción para verificar la asociación propuesta. Con este propósito se han diseñado una serie de protocolos computacionales que combinan los depósitos de las bases de datos de estructura tridimensional proteica con las plataformas de análisis de variante genómico que tienen el poder de asignar un puntaje que califica la asociación de las variantes secuenciales a nivel del DNA con alteraciones funcionales que se clasifican en dos tipos: deletéreas

y benignas según Wu *et al*(9). Dos de los protocolos más utilizados y que serán evaluados en este estudio son las plataformas de análisis en línea SIFT de Kumar *et al.*(10) y PolyPhen2 de Adzhubei *et al.*(11), existiendo otras alternativas disponibles que ofrecen resultados muy similares. En términos generales estos protocolos, permiten con un cierto grado de precisión caracterizar dominios proteicos de interés funcional y evaluar el efecto de cambios a partir de las variantes secuenciales exómicas en estos dominios, Freddolino *et al.*(12).

En este estudio se presentan los resultados del análisis de un grupo de variantes secuenciales exómicas encontradas en pacientes que padecen enfermedades de depósito lisosomal agrupadas en una familia conocida como mucopolisacaridosis, que se han registrado en la literatura como asociadas a un gen y de las cuales existe un inventario de polimorfismos que están asociados a alteraciones funcionales a nivel proteico que explican la etiología molecular de estas enfermedades.

Métodología

Población de estudio, extracción del DNA y secuenciación de exomas

La muestra de estudio para este trabajo comprometió a un grupo de 2 pacientes diagnosticados con alguna de las mucopolisacaridosis descritas en la literatura y provenientes del Suroccidente Colombiano y sus padres. A este grupo de pacientes se les tomó muestra de epitelio bucal y se envió a los laboratorios de la compañía FamilyTreeDNA donde se procedió con la extracción y secuenciación del DNA. En ambos casos se realizaron las secuenciaciones con métodos que garantizan una saturación promedio de al menos 20X el genoma de estudio. Para el alineamiento y comparación de las variantes secuenciales se utilizó el genoma humano de referencia hg19 (o GRCh37) y los resultados se compilaron en un archivo de salida por paciente del tipo Variant Call Format (VCF), en donde se registran el número total de SNPs, deleciones, inserciones, variantes estructurales, entre otros tipos de variantes y sus respectivas anotaciones. Con estos resultados se procedió a realizar el análisis exómico y de predicción de efecto funcional todos los análisis del presente estudio.

Análisis exómico

Para reconocer las variantes secuenciales los datos

consignados en los archivos vcf fueron analizados utilizando la plataforma web del Variant Effect Predictor (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP?db=core) ubicada dentro de la rutina Ensembl, desarrollado por McLaren *et al.*(13), para este propósito se usó el genoma de referencia GRCh37.

Análisis de efecto funcional

Con los resultados de variantes reconocidas se utilizaron los algoritmos SIFT, Kumar *et al.*(10) y PolyPhen2 Adzhubei *et al.*(11), para predecir el efecto que puedan tener las variantes de los pacientes del estudio. Para clasificar el efecto de las variantes de DNA se utilizan escalas numéricas de cero a uno para asignar un puntaje a la predicción de cada variante. Para la interpretación de los resultados con el algoritmo SIFT, la variante es considerada patogénica (deleterious) si el puntaje es menor a 0,05; en el caso de Polyphen, la variante puede ser clasificada como posiblemente dañina (possiblydamaging) si toma valores entre 0,446-0,908 o probablemente dañina (probablydamaging) si está entre 0,909-1. El Variant Effect Predictor también integra bases de datos como el NCBI (<http://www.ncbi.nlm.nih.gov/>) y 1000 Genomes (<http://www.1000genomes.org/>), las cuales permitieron rastrear si la variante encontrada está previamente reportada, la significancia clínica que pueda tener y la frecuencia con la que se ha encontrado en las poblaciones.

Resultados

Se reconocieron un grupo de aproximadamente 60000 variantes secuenciales a lo largo de los diferentes exomas, de las cuales un 17% no han sido registradas en la literatura (fig. 1A) y dentro de estas (fig. 1B) un 50% son de tipo sinónimo (S), hallazgo que es muy común en este tipo de recorridos secuenciales exómicos y que confirma la predominancia de los eventos de deriva genética en las poblaciones sometidas a análisis de secuencia codificante. El porcentaje restante se distribuye así: un 47% de mutaciones de sentido errado (SE), un 1% de alteraciones en el marco de lectura (CML); un 1% de variantes asociados a la aparición de nuevos codones de parada (NCP) y un 1% de inserciones (INS). Se destaca entonces la ausencia de deleciones (DEL), variantes desconocidas (VD) y de pérdidas de codones de parada (PCP); ya que en ninguno de los exomas analizados se registró su presencia.

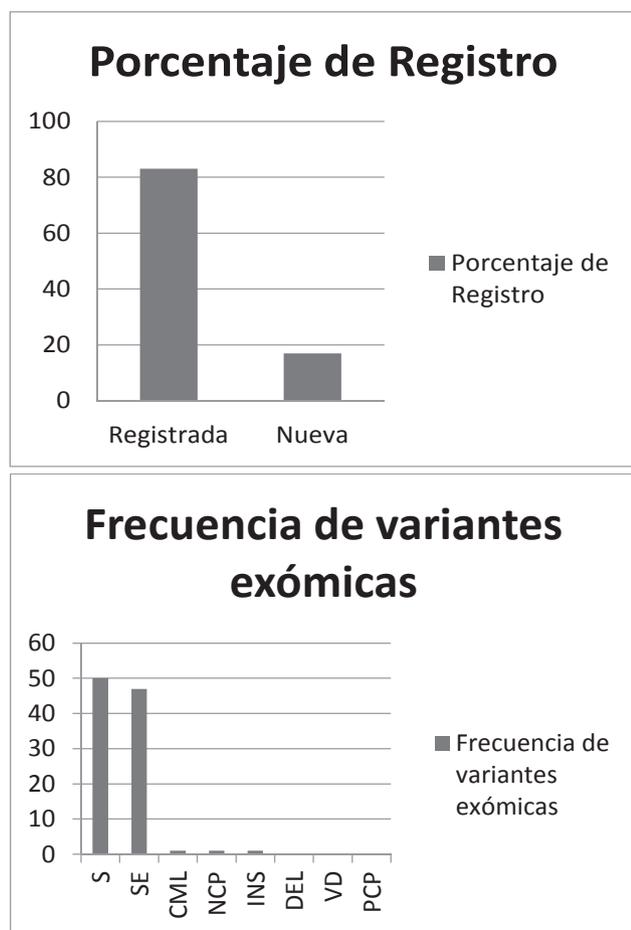


Figura 1. (a) Porcentaje de variantes exómicas nuevas y reportadas en los dos pacientes afectados por mucopolisacaridosis. (b) Frecuencia de variantes codificantes sinónimas (S), de sentido errado (SE), de cambio de marco de lectura (CML), de nuevos codones de parada (NCP), inserciones (INS), deleciones (DEL), variantes desconocidas (VD) y pérdida de codones de parada (PCP), en ambos pacientes.

Los ensayos diagnósticos realizados en los pacientes se registró deficiencia para uno de ellos en el gen ARSB asociado con el síndrome de Maroteaux-Lamy o MPS VI (#253200), mientras que el segundo paciente registró deficiencia del gen IDS que se asocia con el síndrome de Hunter o MPS II. En contraste, en los resultados del análisis de efecto funcional realizado sobre las variantes secuenciales asociadas a los 11 genes implicados en los diversos tipos de mucopolisacaridosis se identificaron cuatro variantes codificantes con efecto patológico en la paciente MPS IV, una en el gen SGSH y tres en el ARSB. Simultáneamente, el paciente MPS II presentó cinco variantes con efecto patológico, una en cada uno de los genes IDUA, IDS y ARSB respectivamente y dos en el gen SGSH.

Es de anotar que en el caso del paciente MPS II las rutinas SIFT y PolyPhen2 no determinaron el efecto deletéreo de la variante secuencial c.783-784G>GT (p.L189LX) del gen IDS, debe atribuirse el efecto patológico a esta mutación de cambio de marco de lectura (CML), ya que fue la única determinada para este gen en el paciente, además el análisis exómico mostró para esta variante una dosis de tipo homocigota lo que corroboraría el resultado en la actividad enzimática de este paciente. El análisis exómico realizado en los padres de este paciente arrojó una dosis heterocigota para la madre y para el padre no se registró variante, lo que permitiría indicar que un alelo mutante en el paciente es de origen materno y el otro se explicaría por un fenómeno *de novo*. En este paciente se registró también una segunda mutación *de novo* en c.1208A>C (p.T374P) para el gen IDUA, con una dosis heterocigota. Por otro lado, ambos pacientes presentaron las variantes c.226A>C (p.T47P) en el gen SGSH y c.2358G>A (p.V358M) en el gen ARSB.

Para la paciente MPS VI la mutación c.2358G>A (p.V358M) del gen ARSB es la única variante secuencial en dosis de homocigota para este gen, lo que la convierte en candidata para explicar el evento patológico siguiendo el modelo de herencia autosómica recesiva. Como hallazgo particular la madre de la paciente es homocigota para esta variante y no registra ninguno de los síntomas asociados a la enfermedad, adicionalmente en las bases de datos, esta variante identificada como dbSNPrs1065757 ha sido relacionada como benigna. La paciente además, registra las mutaciones c.2626G>T (p.C447F) con registro en la literatura y la c.2622G>A (p.G446S) que es una mutación no registrada. Las dos variantes se presentan en dosis heterocigota en la paciente y solo una de ellas, c.2622G>A, registró indicadores con valores de deletéreas con el análisis de las rutinas SIFT y PolyPhen-2. Se debe mencionar que para los padres de la paciente, estas mutaciones se presentan de dosis de heterocigosis de manera alterna en cada uno de ellos. Estos resultados se consolidan en la tabla 1.

En los dos casos del estudio es muy interesante el hallazgo de una acumulación de alelos deletéreos para el conglomerado de genes asociados a las mucopolisacaridosis, a pesar de un diagnóstico específico en cada uno de ellos. Esto nos alerta de la necesidad de utilizar métodos diagnósticos de naturaleza exómica a instancias de un simple proceso de amplificación de gen específico según la patología de interés.

Tabla 1. Caracterización de variantes secuenciales deletéreas según los algoritmos SIFT y PolyPhen2 para el conglomerado de genes implicados en mucopolisacaridosis. Para el hijo (H), madre (M) y padre (P) de cada familia, se ha determinado la cigosidad según el alelo mutante (-) y el alelo referencia (+).

	Gen	Chr:pos.	Mutación (cDNA/aa)	Reporte/ Re-fSeq	SIFT/ PolyPhen2	Cigosidad		
						H	M	P
Fam 1	SGSH	17:78190941	c.226A>C (p.T47P)	-/ NM_000199.3	0/0,997	+/-	+/-	+/-
	ARSB	5:78076482	c.2626G>T (p.C447F)	Garrido et al. (2007)/ NM_000046.3	0.69/0.827	+/-	+/-	+/+
	ARSB	5:78077675	c.2622G>A (p.G446S)	-/ NM_000046.3	0,03/0,982	+/-	+/+	+/-
	ARSB	5:78181477	c.2358G>A (p.V358M)	rs1065757/ NM_000046.3	0,09/0,931	-/-	-/-	+/-
Fam 2	IDUA	4:996204	c.1208A>C (p.T374P)	-/ NM_000203.3	0/0,986	+/-	+/+	+/+
	IDS	X:148579779	c.783-784G>GT (p.L189LX)	-/ NM_000202.5	-	-/-	+/-	+/+
	SGSH	17:78184644	c.1203G>A (p.M372I)	rs58786455/ NM_000199.3	0,01/1	+/-	+/-	+/+
	ARSB	5:78181477	c.2358G>A (p.V358M)	rs1065757/ NM_000046.3	0,09/0,931	+/-	+/+	+/-

Discusión

Los hallazgos producto de la secuenciación del exoma de los pacientes del estudio fueron comparados con resultados previos en la literatura, así por ejemplo, la variante c.2626G>T (p.C447F) encontrada en la paciente MPS VI, ha sido previamente descrita por Garrido *et al.*(14) en dos pacientes de origen español, en presentación de dosis heterocigota y que mostraban fenotipos de grado severo e intermedio para la patología, y que correspondía a rangos de la actividad enzimática de la proteína alterada de naturaleza nula comparada con la silvestre. Hasta el momento no es muy claro el efecto sobre la función de la proteína ARSB, el análisis de plegamiento tridimensional arroja una predicción de proximidad con un sitio potencial de glicosilación, que pudiera revertir en un efecto sobre la afinidad del receptor de manosa-6-fosfato y la recaptación de la enzima por distintos tipos celulares Fuller *et al.*(15); Garrido *et al.*(14); Garrido *et al.*(16)

También, Karageorgos *et al.*(17), registró la variante secuencial c.2622G>C, en una región terminal 3' del exón 7, que corresponde a una zona transicional de unión exón/intrón y con efecto codificante. Estos au-

tores, utilizando la metodología de RT-PCR hallaron un efecto de delección del exón 7 lo que generaría una proteína truncada y que explicaría un efecto funcional asociado a la patología MPS VI. Sin embargo, si se proyecta la lectura codificante de esta variante se obtendría una alteración p.G446R, que aunque registrado en la literatura, no se ha asociado hasta el momento un efecto. En la paciente del estudio se detectó una variante nueva en la misma posición c.2622G>A que generaría un cambio proteico p.G446S, que al ser analizado mediante los algoritmos SIFT (0,03) y PolyPhen2 (0,982), permitiría predecir un efecto deletéreo que explicaría un resultado patológico. No debe descartarse sin embargo una delección del exón 7 como lo mostró Karageorgos *et al.*(17), pero el mismo debe corroborarse mediante un ensayo de RT-PCR en conjunto con ensayos de concentración y actividad enzimática.

Con referencia a la variante c.2358G>A(p.V358M) del gen ARSB que se detectó en los dos pacientes del estudio, Karageorgos *et al.*(17), la registrarón como un polimorfismo asociado a un nivel de actividad enzimática neto de 42% aproximadamente comparado con la variante normal y por lo tanto considerándose

como benigno. Se presume entonces que esta variante, c.2358G>A(p.V358M), no está asociado como agente causal de la MPS VI en la paciente estudiada y que más bien es la conjugación de las dosis heterocigotas para las variantes c.2626G>T (p.C447F) y c.2622G>A (p.G446S), las que explicarían el efecto funcional causante del fenotipo patológico. Este supuesto estaría soportado por los análisis de SIFT y PolyPhen2 y por el registro individual en los padres de la paciente de una de las dos variantes en dosis heterocigota sin registro de un efecto fenotipo notorio. Este hallazgo tiene además un soporte en literatura con el hallazgo de Jin *et al.*(18) quienes registraron el caso de un paciente heterocigoto para dos mutaciones codificantes de sentido truncado, que exhibió un fenotipo MPS VI de clasificación leve y con un 7% de la actividad enzimática normal. Brooks *et al.*(19) también describieron una mujer sin signos sintomáticos del síndrome de Maroteaux-Lamy, clasificándose como sana, pero que presentaba dos variantes secuenciales en el gen ARSB, una de ellas asociada a un fenotipo severo en otros pacientes y otra nueva que estaría atenuando el efecto de la primera, permitiendo detectar niveles de actividad enzimática del 5% comparado con las variantes normales. Recientemente, Jurecka *et al.*(20) registraron un paciente de origen polaco con un fenotipo moderado para la MPS VI, con dos mutaciones codificantes en el gen ARSB en un modelo de dosis heterocigoto. Finalmente, a partir de los estudios de Jin *et al.*(18) y Brooks *et al.*(19) se encuentran inconsistencias entre los niveles de la actividad enzimática de las proteínas alteradas y los fenotipos MPS VI registrado en los pacientes, encontrándose en algunos casos fenotipos más leves que otros, pese a registrar actividades enzimáticas menores. Este tipo de discrepancias, podrán ser explicadas por efectos adicionales en otros genes del conglomerado de las mucopolisacaridosis, como se encontró en los dos pacientes del estudio, que intensificarían el efecto producido por la acumulación progresiva del dermatán sulfato.

Para el caso del paciente MPS II, se puede determinar que la mutación c.566-567T>TT (p.D190G, CML*9)

del gen IDS causa una alteración en el codón de parada, nueve codones corriente abajo, debido al cambio de marco de lectura (CML) generado. La mutación está ubicada en la posición nucleotídica 783-784 del cDNA del exón 5, el cual va de la posición 725 a 925nt según la secuencia referencia NM_000202.5, esto indica que la proteína que se está sintetizando está truncada casi desde el comienzo del exón 5, por lo que hay una delección completa de los exones 6, 7, 8 y 9. Esta tipo de cambio secuencial se puede asociar con la reducida actividad enzimática de 1,9 μ mol/1/h encontrada por fluorometría (ref. \geq 4 μ mol/1/h) y con los datos clínicos del paciente como: discapacidad cognitiva severa, enfermedad intersticial pulmonar difusa, hemorragias alveolares difusas, soporte ventilatorio de oxígeno permanente, grave impedimento de movilidad articular con dependencia completa de silla de ruedas, etc. Dados estos argumentos es posible concluir que esta mutación aun no reportada c.566-567T>TT (p.D190G, CML*9) configura un fenotipo severo del síndrome de Hunter o MPS II.

Conclusión

Los resultados de este estudio sugieren que el uso de algoritmos de predicción de efecto funcional es de particular importancia en el campo de la patología molecular, ya que permite una rápida interpretación del hallazgo de múltiples variantes secuenciales a partir de procesos de secuenciación exómica. Sin embargo, estos resultados deben ser comparados y corroborados con otro tipo de análisis, incluyendo la búsqueda de registros previos en literatura, para lograr un panorama más claro del efecto de los mismos en la patogénesis de la enfermedad, especialmente cuando se abordan problemas de cierta complejidad, como es el caso de las mucopolisacaridosis.

Agradecimientos

El autor reconoce y agradece la participación de los pacientes del estudio, así como también el de los integrantes del grupo de investigación Laboratorio de Biología Molecular y Patogénesis de la Facultad de Salud de la universidad del Valle. Este trabajo se encuentra financiado a través de fondos de la Vicerrectoría de Investigaciones de la Universidad del Valle.

BIBLIOGRAFÍA

1. Lacey, S., Chung, J.Y. and Lin, H. (2014), "A comparison of whole genome sequencing with exome sequencing for family-based association studies", *BMC Proceedings*, Vol. 8 Suppl. 1, S:38.
2. Lalonde, E., Albrecht, S., Ha, K.C., Jacob, K., Bolduc, N., Polychronakos, C., Dechelotte, P., Majewski, J. and Jabado, N. (2010), "Unexpected allelic heterogeneity and spectrum of mutations in Fowler síndrome revealed by next-generation exome sequencing", *Human Mutation*, Vol. 31, pp. 918-23.
3. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J. and Bamshad, M.J. (2010), "Exome sequencing identifies the cause of a Mendelian disorder", *Nature Genetics*, Vol. 42 No. 1, pp. 30-35.
4. Prada, C.E., Gonzaga-Jauregui, C., Tannenbaum, R., Penney, S., Lupski, J.R., Hopkin, R.J. and Sutton, V.R. (2014), "Clinical utility of whole-exome sequencing in rare diseases: galactosialidosis", *European Journal of Medical Genetics*, Vol. 57, pp. 339-44
5. Koboldt, D.C., Larson, D.E., Sullivan, L.S., Bowne, S.J., Steinberg, K.M., Churchill, J.D., Buhr, A.C., Nutter, N., Pierce, E.A., Blanton, S.H., Weinstock, G.M., Wilson, R.K. and Daiger, S.P. (2014), "Exome-based mapping and variant prioritization for inherited mendelian disorders", *The American Journal of Human Genetics*, Vol. 94, pp. 373-84.
6. Peloso, G.M., Auer, P.L., Bis, J.C., Voorman, A., Morrison, A.C., Stitzel, N.O., Brody, J.A., Khetarpal, S.A., Crosby, J.R., Fornage, M., Isaacs, A., Jakobsdottir, J., Feitosa, M.F., Davies, G., Huffman, J.E. et al. (2014), "Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks" *The American Journal of Human Genetics*, Vol. 94, pp. 223-32.
7. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Aslaug, J., Jonasdottir, A., Wong, W., Sigurdsson, G., Walters, G.B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D.F., Helgason, A., Magnusson, O.T., Thorsteinsdottir, U. and Stefansson, K. (2012), "Rate of *de novo* mutations, father's age, and disease risk", *Nature*, Vol. 488 No. 7412, pp. 471-75
8. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Wilsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Sestan, N., Lifton, R.P., Gunel, M., Roeder, K., Geschwind, D.H., Devlin, B. and State, M.W. (2012), "*De novo* mutations revealed by whole exome sequencing are strongly associated with autism", *Nature*, Vol. 485 No. 7397, pp. 237-41.
9. Wu, J., Li, Y. and Jiang, R. (2014), "Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies", *PloS Genetics*, Vol. 10 No. 3, e:1004237.
10. Kumar, P., Henikoff, S. and Ng, P.C. (2009), "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm", *Nature Protocol*, Vol. 7, pp. 1073-81.
11. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S. R. (2010), "A method and server for predicting damaging missense mutations", *Nature Methods*, Vol. 7 No. 4, pp. 248-49.
12. Freddolino, P.L., Harrison, C.B., Liu, Y., and Schulten, K. (2010). "Challenges in protein folding simulations: Timescale, representation and analysis". *Nat. Phys*, Vol 6, pp. 751-758.
13. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010), "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor", *Bioinformatics*, Vol. 26 No. 16, pp. 2069-70.

14. Garrido, E., Chabás, A., Coll, M.J., Blanco, M., Domínguez, C., Grinberg, D., Vilageliu, L. and Cormand, B. (2007), "Identification of the molecular defects in Spanish and Argentinian mucopolysaccharidosis VI (Maroteaux-Lamy syndrome) patients, including 9 novel mutations", *Molecular Genetics and Metabolism*, Vol. 92, pp. 127-30.
15. Fuller, M., Hopwood, J.J., Anson, D.S. (1998), "Receptor mediated binding of two glycosylation forms of N-acetylgalactosamine-4-sulphatase", *Biochim. Biophys. Acta*, Vol. 1406, pp. 283-90.
16. Garrido, E., Cormand, B., Hopwood, J.J., Chabás, A., Grinberg, D. and Vilageliu, L. (2008), "Maroteaux-Lamy syndrome: functional characterization of pathogenic mutations and polymorphisms in the arylsulfatase B gene", *Molecular Genetics and Metabolism*, Vol. 94, pp. 305-12.
17. Karageorgos, L., Brooks, D.A., Harmatz, P., Ketteridge, D., Pollard, A., Melville, E.L., Parkinson-Lawrence, E., Clements, P.R. and Hopwood, J.J. (2007), "Mutational analysis of mucopolysaccharidosis type VI patients undergoing a phase II trial of enzyme replacement therapy", *Molecular Genetics and Metabolism*, Vol. 90, pp. 164-170.
18. Jin, W.D, Jackson, C.E., Desnick, R.J. and Schuchman, E.H. (1992), "Mucopolysaccharidosis type VI: identification of three mutations in the arylsulfatase B gene of patients with the severe and mild phenotypes provides molecular evidence for genetic heterogeneity", *American Journal of Human Genetics*, Vol. 50, pp. 795-800.
19. Brooks, D.A., Gibson, G.J., Karageorgos, L., Hein, L.K., Robertson, E.F. and Hopwood, J.J. (2005), "An index case for the attenuated end of the mucopolysaccharidosis type VI clinical spectrum", *Molecular Genetics and Metabolism*, Vol. 85, pp. 236-38.
20. Jurecka, A., Rozdzynska, A., Marucha, J., Czartoryska, B., Wegrzyn, G. and Tylki-Szymanska, A. (2011), "Natural history of Polish patients with mucopolysaccharidosis type VI", *Central European Journal of Medicine*, Vol. 6 No. 2, pp. 163-71.