Published in final edited form as:

Bone. 2016 November ; 92: 196–200. doi:10.1016/j.bone.2016.09.005.

Assessing the general population frequency of rare coding variants in the *EXT1* and *EXT2* genes previously implicated in Hereditary Multiple Exostoses

Diana L. Cousminer^{1,2}, Alexandre Arkader³, Benjamin F. Voight^{2,4,5}, Maurizio Pacifici³, and Struan F. A. Grant^{1,6,7}

¹Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia

²Department of Genetics, University of Pennsylvania, Philadelphia

³Division of Orthopedic Surgery, The Children's Hospital of Philadelphia, Perelman School of Medicine at University of Pennsylvania

⁴Department of Systems Pharmacology and Translation Therapeutics, University of Pennsylvania, Philadelphia

⁵Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia

⁶Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia

⁷Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia

Abstract

Hereditary multiple exostoses (HME) is a rare childhood-onset skeletal disease linked to mutations in exostosin glycosyltransferase 1 (*EXT1*) or 2 (*EXT2*). Patients are heterozygous for either an *EXT1* or *EXT2* mutation, and it is widely assumed that exostosis formation and associated defects, such as growth retardation and skeletal deformities, require loss-of-heterozygosity or a second hit in affected cells. However, the relevance and phenotypic impact of many presumed pathogenic *EXT* variants remain uncertain. We extracted all amino acid-altering (missense) and loss of function (LoF; nonsense, frameshift, or splice-site) variants from the Exome Aggregation Consortium (ExAC), a large population-based repository of exome sequence data from diverse ancestries that has screened out severe pediatric disease, to assess the overall mutation spectrum of

Corresponding authors: Diana L. Cousminer, Division of Human Genetics, Children's Hospital of Philadelphia, 2615 Civic Center Boulevard, Room 1103F, Philadelphia, PA 19104, Phone: 267-426-2262, Fax: 267-426-0363, cousminerd@email.chop.edu, Struan F. A. Grant, Division of Human Genetics and Endocrinology, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Room 1102D, Philadelphia, PA 19104, Phone: 267-426-2795, Fax: 267-426-0363, grants@email.chop.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Web references

Exome Aggregation Consortium browser: <http://exac.broadinstitute.org/>

Multiple Osteochondromas database: <http://medgen.uantwerpen.be/LOVDv.2.0/>

Ensembl genome browser: <http://useast.ensembl.org/index.html>

SIFT: <http://sift.jcvi.org/>

PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2/>

predicted protein-damaging variants across these two genes in the general population. We then determined whether clinically-identified, presumably pathogenic variants implicated in HME exist among healthy individuals. We found six *EXT1* and four *EXT2* missense mutations in ExAC, suggesting that these mutations have either been misclassified as pathogenic or are not fully penetrant. Furthermore, *EXT1* is heavily selectively constrained, while *EXT2* is more tolerant to protein-damaging variants, especially at its C-terminus, possibly explaining the genotype-phenotype correlation that *EXT1* variants usually result in more severe disease. In conclusion, population-based exome data is a useful filter for determining whether clinically detected variants are likely pathogenic, as well as reveal biological insight into rare disease genes such as *EXT1* and *EXT2*.

Keywords

hereditary multiple exostoses; exome; exostosin-1; exostosin-2; osteochondroma

Introduction

Hereditary multiple exostoses (HME), also called multiple osteochondromas (MO), is a rare childhood-onset disease characterized by the development of benign bone tumors characterized by cartilage-capped bony exostoses adjacent to the growth plates of the long bones and flat bones (1). Because of their location and potentially large number, the lesions can lead to pain, decreased range of motion, skeletal deformities, short stature and other issues later in life, such as malignant degeneration (2–4). Most HME patients bear a heterozygous mutation in the genes encoding exostosin glycosyltransferase 1 (*EXT1*) or 2 (*EXT2*), which are Golgi-associated glycosyl polymerases responsible for heparan sulfate (HS) synthesis, leading to a systemic HS deficiency of about 50% (5). While approximately 1 in 10 patients have *de novo* mutations, the hereditary form is thought to be autosomal dominantly inherited with a penetrance very close to 100% (6–8), with the consequence being that a disease-associated mutation almost invariably leads to the clinical manifestation of the disease. The vast majority of presumed pathogenic variants are nonsense, frameshift, or splice-site mutations, which are predicted to result in a protein with diminished activity or complete loss of function (LoF) (8). Human and mouse studies have indicated that a heterozygous mutation may not be sufficient to trigger the disease manifestation, but loss-of-heterozygosity or a second hit are necessary in affected cells, further reducing HS levels (9–16).

The MO mutation database (MOdb) for HME (8) was initiated in 2009 and currently catalogs 432 and 223 unique variants detected in clinical sequencing studies of *EXT1* and *EXT2*, respectively. The pathogenicity and phenotypic impact of many of these variants remain uncertain. The clinical severity of HME varies greatly amongst individuals with the same *EXT* mutation, even among family members, implying a degree of variable penetrance, and to date, there is no reliable and widely accepted genotype-phenotype correlation to explain disease variability or help diagnosis and prognosis (3,17,18). Because a relatively small number of *EXT* variants have been directly tested for functional impact, it is possible that certain mutations may not be detrimental to enzymatic activity than

otherwise predicted by protein modeling algorithms. Indeed, the overall prevalence of presumed protein-altering mutations in *EXT1* and *EXT2* in the general population has yet to be systematically assessed. With the increasing availability of genomic sequence data from large numbers of individuals derived from the general population, reassessment of previously presumed pathogenic variants is now possible and can help guide clinical sequencing studies, providing an additional filter on the penetrance and pathogenic impact of variants in unaffected individuals.

One such resource, the Exome Aggregation Consortium (ExAC) (19), has amassed genomic sequences from the protein-coding portions of the genome, known as the exome, for over 60,000 individuals without severe pediatric disease (specifically screened out) and representing diverse populations, including European, African, Latino, and Asian ancestries. Using this resource along with MODO, we extracted all amino acid-altering (missense) and LoF (nonsense, frameshift, or splice-site) variants, i.e. those that are likely to directly impact the structure and function of the protein products of *EXT1* and *EXT2*, in order to assess the overall mutation spectrum of putative protein-damaging variants across these two genes in the general population. This allowed us to determine whether clinically-identified, presumably pathogenic variants from the MODO exist among healthy individuals (i.e., without overt clinical manifestations) from this large collection of individuals from diverse ancestral backgrounds.

Materials and Methods

We downloaded all missense and LoF (nonsense, splice acceptor, and splice donor) variants residing in the coding regions of *EXT1* and *EXT2* from the ExAC browser (Cambridge, MA (URL: <http://exac.broadinstitute.org>) [accessed May, 2016]), as well as their distributions across each respective gene. We similarly downloaded all missense and LoF variants listed in the MODO (8) (URL: <http://medgen.uantwerpen.be/LOVDv.2.0/> [accessed May, 2016]). Using these data, we cross-referenced the putatively pathogenic mutations from the MODO with the ExAC population to investigate if any were observed in apparently unaffected individuals.

Leveraging SIFT (<http://sift.jcvi.org/>) (20) and PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) (21) scores, which predict whether an amino acid-changing mutation is likely to be damaging to protein structure and function, we assessed overlapping mutations found in both MODO and ExAC from the Ensembl genome browser (<http://ensembl.org/> [accessed May, 2016]) using the Variant Table feature. Selective constraint metrics, which indicate the likelihood that each gene is intolerant to protein-altering mutations and is thus evolutionarily constrained, were accessed from the ExAC browser for *EXT1* and *EXT2* specifically.

Results

Presumed pathogenic *EXT1* and *EXT2* mutations in the general population

We screened *EXT1* and *EXT2* mutations from two databases, one cataloging variants from a large non-HME population (ExAC) and one curating mutations from HME cases (MODO), and identified those variants that were found in both. In total, eight *EXT1* and four *EXT2*

missense mutations from MODO were observed among the apparently non-HME ExAC subjects. Of these, six of the *EXT1* mutations and all of the *EXT2* mutations were annotated as “probably pathogenic” or “pathogenic” in MODO (Table 1). Several of the *EXT1* variants also displayed marked population specificity in the ExAC database (Table 2). For example, the rare minor allele of rs142122090 (p.Asn373Asp) was observed 81 of 82 times in Latinos within ExAC. Additionally, 24 out of 28 instances of the rare minor allele of rs188859975 (p.Ala486Val) was in East Asians.

Predicted deleteriousness of overlapping variants

SIFT and PolyPhen-2 scores both predict the effect of an amino acid substitution on the function of a protein. We observed a marked difference between these *EXT1* and *EXT2* variants in terms of predicted deleteriousness by both of these scores (Table 1). All six *EXT1* variants were predicted as likely tolerated or benign, while the four *EXT2* variants were predicted to be deleterious to the protein structure. One exception was rs76901081 (p.Arg299Lys) that was predicted deleterious by SIFT and benign by PolyPhen-2.

Overall distribution of missense and LoF variants across *EXT1* and *EXT2* within ExAC

In the ExAC browser, the distribution of sets of variants can be visualized across the exons of each gene of interest. Amino acid-altering variants were relatively common across both genes, with a total of 205 *EXT1* variants and 286 *EXT2* missense or stop variants observed within the ExAC dataset among all known transcripts of each gene.

A previous study based on MODO presumed that pathogenic mutations in *EXT1* were relatively uniformly scattered throughout the gene (8), which is consistent with the distribution of missense variants reported in ExAC. Additionally, although only two LoF mutations were reported in the ExAC database for the canonical *EXT1* transcript, they were found at either end of the gene (Fig. 1A).

For *EXT2*, the same report claimed that pathogenic mutations are concentrated at the N-terminus of the protein (8). In the general population, missense and LoF variants were found throughout the gene, and indeed the majority of LoF variants were concentrated at the C-terminus (Fig. 1B). Conversely, the missense variants found to overlap between ExAC and MODO were located towards the N-terminus.

Selective constraint of *EXT1* and *EXT2*

The number of expected variants in a gene can be calculated based on the probability of mutation at each base pair, the regional divergence between humans and macaques, and a correction for variations in sequencing depth, as described in detail previously (22). By comparing the number of observed variants to the number of expected variants per gene, ExAC provides a gauge of selective constraint, or how tolerant each gene is to certain types of variants. We observed that *EXT1* and *EXT2* were notably different in their tolerance to missense or LoF variants, as evidenced by these selective constraint metrics calculated by ExAC (Table 3) (22).

The first metric is a z-score of the difference between the number of observed versus expected variants located in the exonic region of the given gene. For *EXT1*, the predicted number of missense variants was 278.6, while the number observed in the canonical transcript was almost half at 198, yielding a constraint z-score of 2.36, indicating positive constraint or intolerance to variation. The second metric, the probability of being LoF intolerant (pLI), is indicative of extreme constraint when its value is > 0.9. For *EXT1*, 25.2 LoF variants were expected, while only one was observed (a frameshift mutation), yielding a probability of 1 being pLI, further indicating that this gene is extremely intolerant to these types of variants.

In contrast, these metrics revealed that *EXT2* is tolerant to protein-altering variants; 267.5 missense variants were predicted, while a very comparable 251 were observed, resulting in a constraint z-score of only 0.49, while 26.8 LoF variants were expected and 15 were observed, yielding a constraint pLI of 0.00.

Discussion

In this study, we utilized large-scale, publicly available data of missense and LoF variants in *EXT1* and *EXT2*, derived from the exome or coding sequences of in excess of 60,000 population controls of diverse ancestries (and free of severe pediatric disease), to assess presumed pathogenic variants for HME from a clinical mutation database. We found that six *EXT1* and four *EXT2* presumed pathogenic variants were in fact present among these individuals. Although these variants were all present in the MODOB and were annotated as “probably pathogenic” or “pathogenic,” the level of evidence for each varied; several were presumed to be causal for HME, several were suspected polymorphisms or variants found in the general population, and several remain unpublished. Additionally, we leveraged the ExAC data to characterize the spectrum of amino acid-altering variants in the general population in these two HME genes, and found that *EXT1* is extremely intolerant to protein-damaging mutations, while this was not the case for *EXT2*.

It is a widely held view that a ‘second hit,’ or somatic disruption of the functional allele, is necessary for HME manifestation (2,18), based on observations of local loss-of-heterozygosity within osteochondromas (9) and mouse studies in which stochastic, chondrocyte-specific inactivation of *Ext1* resulted in osteochondroma development and bone deformities, closely mimicking the human disease manifestation (13,16). Likewise, mice with a heterozygous *Ext1*^{+/-} or *Ext2*^{+/-} genotype are essentially normal, whereas compound heterozygous *Ext1*^{+/-}; *Ext2*^{+/-} mice and conditional *Ext1*-null mice have multiple exostoses as in human HME (14,15). For the *EXT1* and *EXT2* mutations thought to be pathogenic that we detected in the presumably healthy ExAC population, it is possible that (1) that these variants are not pathogenic and have been misclassified; (2) that they are not completely penetrant; or (3) that they are pathogenic, but the timing of the ‘second hit’ may alter disease severity. Thus, carriers of these mutations may have mild HME, but not severe enough to warrant removal from the ExAC population, which was screened for pediatric diseases. If this is the case, these findings challenge the idea that HME is fully penetrant, as previously suggested (23). On the other hand, the timing of the ‘second hit’ may be especially important, in that if a second hit occurs early in embryogenesis or development,

osteochondromas may be present at large numbers throughout the body, whereas a ‘second hit’ later during postnatal life would have more isolated, less severe consequences. Finally, it is possible that protective mechanisms exist elsewhere in the genome that mitigate the severity of the disease, especially given the large range of variation in disease severity often seen in a single family, but these have not been studied or detected.

In this study, the variants detected among the normal population may fall into any of these categories. For example, two of the *EXT1* variants, rs188859975 (p.Ala486Val) and rs119103287 (p.Arg340His), highlight the usefulness of the ExAC data for filtering out presumed pathogenic variants. For rs188859975 (p.Ala486Val), a variant originally detected in a study of Chinese patients, the ExAC data strongly suggests that this variant is a rare polymorphism predominantly found in the East Asian population (allele count = 24; frequency in East Asians = 0.003) and is therefore not likely to be pathogenic given this frequency (Table 2). Alternatively, for rs119103287 (p.Arg340His), this amino acid appears to be essential to the protein’s function, with the mutation at this residue resulting in defective heparan sulfate expression, although the protein retains the ability to form an oligomeric complex (24,25). Additionally, this mutation is among the most frequently seen in HME patients, reportedly found in 12% of patients with an *EXT1* mutation (8). Considering that it was detected only once in the ExAC data, this variant remains a strong candidate as pathogenic, but its detection in ExAC suggests that it is not completely penetrant, that the individual carrying the mutation never experienced the ‘second hit’ necessary to lead to an osteochondroma, or that the ‘second hit’ occurred later in life with only mild symptoms.

Many of the other variants also displayed marked population specificity (Table 2), most notably rs142122090 (p.Asn373Asp), which was seen 81 times in Latinos (population frequency of 0.007), and once in the “other” ancestral category, which could in theory also represent a person of Latino ancestry. Such population-specificity may have contributed to variant misclassification as pathogenic if appropriate matched-population controls were not assessed in the clinical studies that identified them.

Considering the spectrum of mutations across these two genes, *EXT1* is markedly intolerant to protein-damaging mutations, evidenced by the constraint metrics calculated by ExAC. Additionally, we found that clinically identified *EXT1* variants that occur in the general population were predicted to be tolerated or benign by two different algorithms. This lends further support to the possibility that not all *EXT1* amino-acid altering mutations lead to disease, with some resulting in an EXT1 protein remaining functional, and that these mutations must result in very slight, if any, changes to the protein function, given the strong selective constraint that has removed deleterious alleles in the general population. Taken together, these findings are consistent with data that has shown that *EXT1* mutations are more likely to result in severe HME than *EXT2* mutations (17,26).

Conversely, the protein encoded by *EXT2* is less biologically constrained, potentially helping to explain why mutations result in less severe phenotypes (17,26). Further evidence that an EXT2 protein that has sub-optimal function can be tolerated in healthy individuals is our finding of potentially damaging amino-acid altering mutations, as predicted by SIFT and

PolyPhen-2, among the apparently non-HME population. There may, however, be different constraints based on where in the gene the mutation occurs, as most presumed pathogenic mutations have been found in the N-terminal part of the protein (8), while in the general population, LoF variants were concentrated in the C-terminus. As such, one could conclude that the amino acids at the C-terminus can tolerate damaging variation without resulting in HME, while N-terminal LoF mutations are pathogenic.

One limitation of this study is that we were unable to examine detailed phenotype data or family histories of the individuals in ExAC, nor re-sequence them to confirm the mutations detected. However, given that the overall prevalence of HME is estimated at 1/50,000, we would expect ~1 HME case among the 60,000 ExAC samples. Still, even this low frequency should be an over-estimate given that the ExAC samples were all screened for pediatric diseases, which should have excluded any obvious cases. Finally, extensive quality control procedures were in place during the ExAC sequencing pipeline, and all variants presented here passed these quality filters.

Conclusions

Utilizing a large-scale population exome sequence collection, we identified ten missense mutations collectively across *EXT1* and *EXT2* that may be inaccurately classified as pathogenic for HME or, at least, may be only partially penetrant for the disease. Furthermore, through computational assessment, *EXT1* is heavily selectively constrained, while *EXT2* is more tolerant to protein-damaging variants, helping explain the genotype-phenotype correlation that *EXT1* variants result in more severe disease. Population-based exome data can thus be a useful filter for determining whether missense variants are likely to be pathogenic, as well as reveal biological insight into rare disease genes such as *EXT1* and *EXT2*.

Acknowledgments

Funding

The project described was supported by the National Center for Research Resources, Grant UL1RR024134, and is now at the National Center for Advancing Translational Sciences, Grant UL1TR000003. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Supported in part by the Institute for Translational Medicine and Therapeutics' (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics (D.L.C, B.F.V, and S.F.A.G). M.P. was supported by the NIH grant R01 AR061758.

Abbreviations

HME	hereditary multiple exostoses
MO	multiple osteochondromas
EXT1	exostosin glycosyltransferase 1
EXT2	exostosin glycosyltransferase 2
ExAC	Exome Aggregation Consortium
MOdb	Multiple Osteochondromas database

LoF	loss of function
pLI	probability of being LoF intolerant

References

1. Bovée JV. Multiple osteochondromas. *Orphanet J Rare Dis.* 2008; 3(1):3. [PubMed: 18271966]
2. Huegel J, Sgariglia F, Enomoto-Iwamoto M, Koyama E, Dormans JP, Pacifici M. Heparan sulfate in skeletal development, growth, and pathology: The case of hereditary multiple exostoses: Heparan Sulfate in Skeletal Dysplasia. *Dev Dyn.* 2013 Sep; 242(9):1021–1032. [PubMed: 23821404]
3. Arkader A. Multiple hereditary exostoses: its burden on childhood and beyond: commentary on an article by A.L. Goud, MD, et al.: “Pain, physical and social functioning, and quality of life in individuals with multiple hereditary exostoses in The Netherlands. A national cohort study.”. *J Bone Joint Surg Am.* 2012 Jun 6; 94(11):e81. [PubMed: 22637221]
4. Goud AL, de Lange J, Scholtes V a. B, Bulstra SK, Ham SJ. Pain, physical and social functioning, and quality of life in individuals with multiple hereditary exostoses in The Netherlands: a national cohort study. *J Bone Joint Surg Am.* 2012 Jun 6; 94(11):1013–1020. [PubMed: 22637207]
5. Anower-E-Khuda MF, Matsumoto K, Habuchi H, Morita H, Yokochi T, Shimizu K, et al. Glycosaminoglycans in the blood of hereditary multiple exostoses patients: Half reduction of heparan sulfate to chondroitin sulfate ratio and the possible diagnostic application. *Glycobiology.* 2013 Jul 1; 23(7):865–876. [PubMed: 23514715]
6. Schmale GA, Conrad EU, Raskind WH. The natural history of hereditary multiple exostoses. *J Bone Jt Surg Am.* 1994 Jul 1; 76(7):986–992.
7. Hecht JT, Hogue D, Strong LC, Hansen MF, Blanton SH, Wagner M. Hereditary multiple exostosis and chondrosarcoma: linkage to chromosome II and loss of heterozygosity for EXT-linked markers on chromosomes II and 8. *Am J Hum Genet.* 1995 May; 56(5):1125–1131. [PubMed: 7726168]
8. Jennes I, Pedrini E, Zuntini M, Mordenti M, Balkassmi S, Asteggiano CG, et al. Multiple osteochondromas: mutation update and description of the multiple osteochondromas mutation database (MOdb). *Hum Mutat.* 2009; 30(12):1620–1627. [PubMed: 19810120]
9. Reijnders CMA, Waaijer CJF, Hamilton A, Buddingh EP, Dijkstra SPD, Ham J, et al. No Haploinsufficiency but Loss of Heterozygosity for EXT in Multiple Osteochondromas. *Am J Pathol.* 2010 Oct; 177(4):1946–1957. [PubMed: 20813973]
10. Hecht JT, Hayes E, Haynes R, Cole WG, Long RJ, Farach-Carson MC, et al. Differentiation-induced loss of heparan sulfate in human exostosis derived chondrocytes. *Differ Res Biol Divers.* 2005 Jun; 73(5):212–221.
11. Hall CR, Cole WG, Haynes R, Hecht JT. Reevaluation of a genetic model for the development of exostosis in hereditary multiple exostosis. *Am J Med Genet.* 2002 Sep 15; 112(1):1–5. [PubMed: 12239711]
12. Zuntini M, Pedrini E, Parra A, Sgariglia F, Gentile FV, Pandolfi M, et al. Genetic models of osteochondroma onset and neoplastic progression: evidence for mechanisms alternative to EXT genes inactivation. *Oncogene.* 2010 Jul 1; 29(26):3827–3834. [PubMed: 20418910]
13. Jones KB, Piombo V, Searby C, Kurriger G, Yang B, Grabellus F, et al. A mouse model of osteochondromagenesis from clonal inactivation of Ext1 in chondrocytes. *Proc Natl Acad Sci U S A.* 2010 Feb 2; 107(5):2054–2059. [PubMed: 20080592]
14. Sgariglia F, Candela ME, Huegel J, Jacenko O, Koyama E, Yamaguchi Y, et al. Epiphyseal abnormalities, trabecular bone loss and articular chondrocyte hypertrophy develop in the long bones of postnatal Ext1-deficient mice. *Bone.* 2013 Nov; 57(1):220–231. [PubMed: 23958822]
15. Zak BM, Schuksz M, Koyama E, Mundy C, Wells DE, Yamaguchi Y, et al. Compound heterozygous loss of Ext1 and Ext2 is sufficient for formation of multiple exostoses in mouse ribs and long bones. *Bone.* 2011 May 1; 48(5):979–987. [PubMed: 21310272]
16. Matsumoto K, Irie F, Mackem S, Yamaguchi Y. A mouse model of chondrocyte-specific somatic mutation reveals a role for Ext1 loss of heterozygosity in multiple hereditary exostoses. *Proc Natl Acad Sci.* 2010 Jun 15; 107(24):10932–10937. [PubMed: 20534475]

17. Pedrini E, Jennes I, Tremosini M, Milanese A, Mordenti M, Parra A, et al. Genotype-phenotype correlation study in 529 patients with multiple hereditary exostoses: identification of “protective” and “risk” factors. *J Bone Joint Surg Am*. 2011 Dec 21; 93(24):2294–2302. [PubMed: 22258776]
18. Jones KB. Glycobiology and the Growth Plate: Current Concepts in Multiple Hereditary Exostoses. *J Pediatr Orthop*. 2011; 31(5):577–586. [PubMed: 21654469]
19. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016 Aug 18; 536(7616):285–291. [PubMed: 27535533]
20. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4(7):1073–1081. [PubMed: 19561590]
21. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Editor Board Jonathan Haines AI. 2013; Chapter 7(Unit7.20)
22. Consortium EA, Lek M, Karczewski K, Minikel E, Samocha K, Banks E, et al. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*. 2015 Oct 30.30338
23. Legeai-Mallet L, Munnich A, Maroteaux P, Merrer ML. Incomplete penetrance and expressivity skewing in hereditary multiple exostoses. *Clin Genet*. 2008 Jun 28; 52(1):12–16.
24. Cheung PK, McCormick C, Crawford BE, Esko JD, Tufaro F, Duncan G. Etiological point mutations in the hereditary multiple exostoses gene EXT1: a functional analysis of heparan sulfate polymerase activity. *Am J Hum Genet*. 2001 Jul; 69(1):55–66. [PubMed: 11391482]
25. McCormick C, Leduc Y, Martindale D, Mattison K, Esford LE, Dyer AP, et al. The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nat Genet*. 1998 Jun; 19(2):158–161. [PubMed: 9620772]
26. Francannet C, Cohen-Tanugi A, Merrer ML, Munnich A, Bonaventure J, Legeai-Mallet L. Genotype-phenotype correlation in hereditary multiple exostoses. *J Med Genet*. 2001 Jul 1; 38(7): 430–434. [PubMed: 11432960]
27. Raskind WH, Conrad EU, Matsushita M, Wijsman EM, Wells DE, Chapman N, et al. Evaluation of locus heterogeneity and EXT1 mutations in 34 families with hereditary multiple exostoses. *Hum Mutat*. 1998; 11(3):231–239. [PubMed: 9521425]
28. Signori E, Massi E, Matera MG, Poscente M, Gravina C, Falcone G, et al. A combined analytical approach reveals novel EXT1/2 gene mutations in a large cohort of Italian multiple osteochondromas patients. *Genes Chromosomes Cancer*. 2007 May; 46(5):470–477. [PubMed: 17301954]
29. Xu L, Xia J, Jiang H, Zhou J, Li H, Wang D, et al. Mutation analysis of hereditary multiple exostoses in the Chinese. *Hum Genet*. 1999 Aug; 105(1–2):45–50. [PubMed: 10480354]
30. Lonie L, Porter DE, Fraser M, Cole T, Wise C, Yates L, et al. Determination of the mutation spectrum of the EXT1/EXT2 genes in British Caucasian patients with multiple osteochondromas, and exclusion of six candidate genes in EXT negative cases. *Hum Mutat*. 2006 Nov.27(11):1160. [PubMed: 17041877]
31. Dobson-Stone C. Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of E... - PubMed - NCBI [Internet]. [cited 2016 Jun 8] Available from: <http://proxy.library.upenn.edu:2084/pubmed/?term=10713884>.
32. Tsuchiya T, Osanai T, Ogoe A, Tamura G, Chano T, Kaneko Y, et al. Methylation status of EXT1 and EXT2 promoters and two mutations of EXT2 in chondrosarcoma. *Cancer Genet Cytogenet*. 2005 Apr 15; 158(2):148–155. [PubMed: 15796962]

Highlights: 3-5 bullet points

- Several missense variants in *EXT1* and *EXT2* presumed to be pathogenic for hereditary multiple exostoses were found, sometimes with multiple recurrences, among a large collection of exome sequences from the general population.
- *EXT1* is under high selective constraint as evidenced by a depletion of protein-damaging variants among the general population, while variants predicted as protein-damaging can be found in *EXT2* among apparently healthy individuals.
- Large-scale public sequencing data can be informative for clinical sequencing studies, helping to prioritize potentially damaging mutations and characterize the variant spectrum of disease-causing genes.

A. *EXT1* B. *EXT2*

Figure 1. Distribution of variants in the general population for the canonical transcripts of *EXT1* and *EXT2*

Loss of function (LoF) variants (red circles) found in the ExAC database in the exons (rectangles) of A) *EXT1* and B) *EXT2*. The approximate location of variants found in both the ExAC and MObd are marked with arrows. Protein domains are derived from the Protein Families (Pfam) database (<http://pfam.xfam.org/>) and were accessed through the Ensembl genome browser.

Table 1
Variants found in both the MODb and the ExAC databases
Only variants predicted to be pathogenic (+) or probably pathogenic (+?) in the mutation database are presented here.

<i>EXT1</i>	ExAC Consortium population data					Predicted consequence		MODb mutation database		
Variant	Exon	Transcript variant	Protein variant	Allele count	Total allele count	Allele Frequency	SIFT	PolyPhen	Pathogenic	Reference
rs150818931	01	c.214 G>T	p.Glu72Lys ^b	9	121332	7.42E-05	tolerated	benign	+?	Jennes, N.P.
rs119103287 ^a	02	c.1019G>A	p.Arg340His	1	121188	8.25E-06	tolerated	benign	+?	Raskind (27)
rs142122090	03	c.1117A>G	p.Asn373Asp	82	121304	0.000676	tolerated	benign	+?	Bakker, N.P.
rs201504622	05	c.1360G>A	p.Val454Ile	4	121244	3.3E-05	tolerated	benign	+?	Signori (28)
rs188859975	06	c.1457C>T	p.Ala486Val	28	121208	0.000231	tolerated	benign	+	Xu (29)
rs144397063	07	c.1609 G>A	p.Val537Ile	16	121412	0.000132	tolerated	benign	+?	Pedrini, N.P.
<i>EXT2</i>										
Variant	Exon	Transcript variant	Consequence	Allele Count	Allele Number	Allele Frequency	SIFT	PolyPhen	Pathogenic	Reference
rs200613371	02	c.382C>T	p.Arg128Cys	8	121002	6.61E-05	deleterious	possibly damaging	+?	Lonie (30)
rs757297287 ^a	03	c.544C>T	p.Arg182Trp ^b	5	121290	4.12E-05	deleterious	possibly damaging	+?	Dobson-Stone (31)
rs76901081	05	c.896G>A	p.Arg299Lys	1	121164	8.25E-06	deleterious	benign	+?	Tsuchiya (32)
rs762292898	06	c.1073G>A	p.Arg358Gln ^b	7	121410	5.77E-05	deleterious	possibly damaging	+?	Pedrini, N.P.

^aThese mutations are among the most frequently reported in HME patients (Jennes, et al. 2009).

^bThese mutations were reported as resulting in a stop (X) codon in the MODb but reflect the same base pair change. No homozygotes were detected for any of these variants in the ExAc database. N.P., not published.

Table 2
Distribution of the *EXT1* and *EXT2* variants across the populations included in the ExAC database

Finnish were also included in ExAC, but none of these alleles were detected in the Finnish population (data not shown).

SNP	Protein variant	Total		European (Non-Finnish; n~66700)		South Asian (n~16500)		Latino (n~11550)		African (n ~10400)		East Asian (n~8650)		Other (n~900)	
		Allele Count	Frequency	Allele count	Frequency	Allele count	Frequency	Allele count	Frequency	Allele count	Frequency	Allele count	Frequency	Allele count	Frequency
EXT1															
rs150818931	p.Glu72X	9	7.42E-05	8	1.20E-04	0	0	0	0	1	9.62E-05	0	0	0	0
rs119103287	p.Arg340His	1	8.25E-06	1	1.50E-05	0	0	0	0	0	0	0	0	0	0
rs142122090	p.Asn373Asp	82	6.76E-04	0	0	0	0	81	7.01E-03	0	0	0	0	1	1.10E-03
rs201504622	p.Val454Ile	4	3.30E-05	2	3.00E-05	2	1.21E-04	0	0	0	0	0	0	0	0
rs188859975	p.Ala486Val	28	2.31E-04	2	3.00E-05	1	6.06E-05	0	0	1	9.67E-05	24	2.78E-03	0	0
rs144397063	p.Val537Ile	16	1.32E-04	14	2.10E-04	2	1.21E-04	0	0	0	0	0	0	0	0
EXT2															
rs200613371	p.Arg128Trp	8	6.61E-05	3	4.51E-05	1	6.06E-05	1	8.68E-05	3	2.89E-04	0	0	0	0
rs757297287	p.Arg182X	5	4.12E-05	4	6.00E-05	1	6.06E-05	0	0	0	0	0	0	0	0
rs76901081	p.Arg299His	1	8.25E-06	0	0	0	0	1	8.67E-05	0	0	0	0	0	0
rs762292898	p.Trp358X	7	5.77E-05	6	8.99E-05	1	6.06E-05	0	0	0	0	0	0	0	0

Table 3
Conservation metrics for the canonical *EXT1* and *EXT2* transcripts based on the predicted and observed variant counts in the ExAC database
Z, z-score for the deviation between the observed and expected counts; pLI, probability of being LoF intolerant.

EXT1			EXT2			
Variant type	Expected variant count	Observed variant count	Constraint	Expected variant count	Observed variant count	Constraint
Synonymous	127.7	119	Z = 0.48	111.4	122	Z = -0.62
Missense	279.6	198	Z =2.36	267.5	251	Z = 0.49
LoF	25.2	1	pLI = 1.00	26.8	15	pLI = 0.00