Journal of Genetics and Genomes

Mottes and Malerba, J Genet Genom 2017, 1:1

Editorial Open Access

Hunting Novel Human Disease Genes in the Next Generation Sequencing Era: Lessons from Osteogenesis Imperfecta

Monica Mottes* and Giovanni Malerba

Department of Neurosciences, Biomedicine and Movement Sciences Biology and Genetics Section, University of Verona, Italy

Osteogenesis Imperfecta: A Brief Discussion

The identification of causative genes in Mendelian disorders has been achieved in the past thanks to traditional approaches, with fairly good results (~ 3000 disease genes identified) [1].

Different combined strategies have been employed: the candidate gene approach was applied whenever knowledge of the physiological/biochemical bases of the disease was available. Linkage studies with polymorphic markers within families allowed positional mapping, i.e., the identification of candidate regions, which often contained many genes. Thus chances for a successful hunt depended mostly on spotting a most likely candidate gene within the identified region; characterized animal models (e.g., knockout mice) have often provided excellent hints for this. Studies on large pedigrees with high rates of consanguinity have been crucial in the case of recessive diseases, as well as studies of multiple generations' pedigrees with dominantly transmitted phenotypes [2]. Nevertheless the above mentioned approaches could not be applied across the board to all Mendelian genes; at the beginning of the twenty-first century it was clear that additional high-throughput strategies were badly needed in order to fill the gap. The Next Generation Sequencing (NGS) technology was introduced in 2005 and has, since then, revolutionized and suddenly accelerated the discovery of novel Mendelian disease loci. NGS allows sequencing of millions of fragments in a massively parallel fashion at affordable costs; an entire human genome can be sequenced within twenty-four hours. The agnostic approach of Whole Genome Sequencing (WGS), unlike the candidate gene approach, can be applied to any phenotype. A major problem consists in interpreting the overwhelming number of variants revealed by WGS. A widely used approach exploits Genome-Wide Association Studies (GWAS). Genotypes can be generated using SNP (Single Nucleotide Polymorphisms) arrays in order to localize the disease locus within one (or more than one) region of the genome, which will then be sounded out by targeted sequencing of candidate genes. The GWAS approach, compared to traditional linkage studies, allows localization of the sought-after causal mutation in a much smaller region (few kilobases, instead of megabases).

Thanks to commercially available whole exome-enrichment kits, NGS can also be employed for Whole Exome Sequencing (WES). Exome represents <2% of the genome, i.e., the protein-coding portion, where ~85% of mutations for Mendelian diseases occur. WES may be very useful also in molecular diagnostics, since it allows the discovery of new, rare pathological variants in single patients; these variants would otherwise get missed by ready-made screening arrays. Disease-gene hunters must anyway be aware that exome sequencing alone cannot reveal deep intronic mutations or causative variants in 5'/3'regulatory regions. Positional mapping data can be in any way very useful, whenever a causal mutation is not found: we must be aware of limitations in currently available sequencing techniques (none covers 100% of the human genome). Powerful positional mapping derives from the analysis of many phenotypically similar individuals taken singularly and/or within families; SNP-autozygosity mapping (homozygosity due to identical ancestral alleles) combined with exome sequencing allows successful identification of rare recessive disease loci even when small numbers of highly inbred families are available [3].

From here on the editorial will try to illustrate how all the different gene identification strategies described above have been applied in a thirtyfive years' time frame, for the discovery of seventeen different loci involved in Osteogenesis Imperfecta (OI). This Mendelian disorder, mainly characterized by bone fragility and skeletal deformities ranging in a broad phenotypic spectrum, has been known for a long time (it was first described clinically in 1883 by Lobstein). Hundred years later, thanks to strong biochemical evidence, a candidate gene approach allowed researchers to associate a case of lethal OI with a molecular defect in COL1A1 gene, which encodes alpha 1 chains of the hetrotrimeric alpha1$(I)_2$ alpha2$(I)_1$ Type I collagen [4]. Collagen I is the most abundant protein in bone Extracellular Matrix (ECM); qualitative and quantitative integrity of collagen fibrils is required in order to ensure normal ECM mineralization. As expected, in the following years, hundreds of different OI-causing mutations have been found in both type I collagen genes (COL1A1 and COL1A2) [5,6]. Four clinical phenotypes were defined in 1979 by Sillence. For decades OI has been considered an Autosomal Dominant (AD) collagen disorder, linked to two loci. However, increasing clues suggested that other unknown loci were to be discovered: i) OI patients found in highly inbred families suggestive of Autosomal Recessive (AR) inheritance; ii) severe forms of OI showing collagen I biochemical anomalies but no mutations in either collagen I gene; iii) peculiar forms of OI showing neither collagen I anomalies nor mutations in collagen I genes.

The NGS revolution applied to OI, combined with the traditional approaches described above, has unraveled since 2006 its astonishing genetic heterogeneity: fifteen novel disease loci have been discovered in ten years' time; at present eighteen different OI types have been classified, the list will probably expand in the future. A detailed description of each defective gene/protein role in OI pathogenesis would be too cumbersome for an editorial. Instead, a chronologically ordered list of disease genes /proteins with pertinent references, brief info about their physiological role, along with the technical approaches applied for gene hunting, is offered in Table 1. Recently an updated

***Corresponding author:** Monica Mottes, Biomedicine and Movement Sciences Biology and Genetics Section, University of Verona, Italy, Tel: +390458027184; E-mail: monica.mottes@univr.it

Received September 04, 2017; **Accepted** September 06, 2017; **Published** September 10, 2017

Citation: Mottes M, Malerba G (2017) Hunting Novel Human Disease Genes in the Next Generation Sequencing Era: Lessons from Osteogenesis Imperfecta. J Genet Genom 1: e102.

Year [Ref]	Methodological approach	Defective gene/protein	Physiological role	OI type/ Inheritance
1983 [4]	Candidate gene	COL1A1/collagen I	structural/major component of bone ECM	I, II, III, IV/Ad
1984 [11]	Candidate gene	COL1A2/collagen I	structural/major component of bone ECM	I,II,III, IV/Ad
2006 [12,13]	Gwla in inbred families+candidate gene (mouse model)	CRTAP/CRTAP	collagen post-translational modification	VII/Ar
2007 [14]	Candidate gene	LEPRE1/P3H1	collagen post-translational modification	VIII/Ar
2009 [15]	Candidate gene	PPIB/CyPB	collagen post-translational modification	IX/Ar
2010 [16]	Candidate gene	SERPINH1/HSP47	collagen-specific chaperone	X/Ar
2010 [17]	Homozygosity mapping+targeted ngs	FKBP10/FKBP65	chaperone involved in collagen crosslinking	XI/Ar
2010 [18]	Homozygosity mapping+candidate gene (mouse model)	SP7/OSX	master transcription factor for osteogenesis	XII/Ar
2011 [19,20]	Wes in 1 patient; homozygosity mapping+targeted ngs	SERPINF1/ PEDF	anti-angiogenic, pro-osteogenic factor	VI/ Ar
2012 [21,22]	Wes in 1 patient; gwla+targeted ngs	IFITM5/ BRIL	highly expressed in osteoblast; involved in mineralization	V/ Ad
2012 [23]	Homozygosity mapping+candidate gene approach	BMP1/ BMP1	procollagen processing	XIII/Ar
2012 [24]	Autozygosity mapping+wes	TMEM38B/ TRIC-B	regulation of Ca++ flux	XIV/Ar
2012 [25]	Homozygosity mapping+candidate gene	PLOD2/LH2	collagen post-translational modification	unclassified/Ar
2013 [26,27]	Wes; wgla+targeted ngs	WNT1/ WNT1	activates Wnt signaling, which controls bone dev and homeostasis	XV/Ar/Ad
2013 [28]	Candidate gene (mouse model)	CREBL3L1/OASIS	activates transcription of UPR genes	XVI/Ar
2015 [29]	Wes in unrelated patients	SPARC/OSTEONECTIN	protein produced by osteoblasts, binds collagen and other ECM proteins	XVII/Ar
2016 [30]	Gwla + X exome sequencing	MBTPS2/SP2	crucial for RIP of substrates as OASIS, ATF6	XVIII/Xr

Ad: autosomal dominant; Ar: autosomal recessive; RIP: Regulated Intramembrane Proteolisis; UPR: Unfolded Protein Response; Xr: X-linked recessive

Table 1: Flowchart of OI genes identification (1983-2016).

OI clinical classification and nomenclature have been proposed [7]. It is not surprising that seven of the disease genes discovered since 2006, whose defects cause AR forms of OI, code for proteins which are involved in collagen I modifications, processing, folding, cross-linking. Eight additional disease genes, whose defects cause either AR or AD forms of OI, code for proteins involved in various aspects of osteoblast functions and survival. Each of them has brought valuable and sometime unexpected information about its own role in bone biology. Specific epigenetic DNA modifications (i.e., Cytosine methylation) can justify recurrent *de novo* OI causing mutations [8]. New interesting discoveries will certainly come out, as gene hunting in OI and other bone dysplasias goes on. On the practical side, such genotypic and phenotypic variability represents a real challenge for clinical classification and for molecular diagnostics, although it must be kept in mind that >90% OI cases are due to COL1A1/COL1A2 mutations. Most of the AR forms of OI are very rare and were discovered thanks to the analysis of highly inbred families in particular ethnic groups. Accurate pedigree analysis, clinical, biochemical, radiological, bone histology data, may help specialists to address the search for causative mutations in a targeted manner. Moreover, current technological tools, such as NGS platforms designed for simultaneous screening of multiple candidate genes can be employed [9,10].

References

1. McKusick VA (2007) Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet 80: 588-604.
2. Alkuraya FS (2016) Discovery of mutations for Mendelian disorders. Hum Genet 135: 615-623.
3. Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT (2006) Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Hum Mutat 27: 1041-1046.
4. Chu ML, Williams CJ, Pepe G, Hirsch JL, Prockop DJ, et al. (1983) Internal deletion in a collagen gene in a perinatal lethal form of osteogenesis imperfecta. Nature 304: 78-80.
5. Dalgleish R (1997) The human type I collagen mutation database. Nucleic Acids Research 25: 181-187.
6. Dalgleish R (1998) The human collagen mutation database 1998. Nucleic Acids Research 26: 253-255.
7. Van Dijk FS, Sillence DO (2014) Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. Am J Med Genet A 164A:1470-1481.
8. Corradi M, Monti E, Venturi G, Gandini A, Mottes M, et al. (2014) The recurrent causal mutation for osteogenesis imperfecta type V occurs at a highly methylated CpG dinucleotide within the IFITM5 gene. J Pediatr Genet 3: 35-39.
9. Sule G, Campeau PM, Zhang VW, Nagamani SC, Dawson BC, et al. (2013) Next-generation sequencing for disorders of low and high bone mineral density. Osteoporos Int 24: 2253-2259.
10. Rauch F, Lalic L, Glorieux FH, Moffatt P, Roughley P (2014) Targeted sequencing of a pediatric metabolic bone gene panel using a desktop semiconductor next-generation sequencer. Calcif Tissue Int 95: 323-331.
11. Pihlajaniemi T, Dickson LA, Pope FM, Korhonen VR, Nicholls A, et al. (1984). Osteogenesis imperfecta: cloning of a pro-alpha 2(I) collagen gene with a frameshift mutation. J Biol Chem 259: 12941-12944.
12. Barnes AM, Chang W, Morello R, Cabral WA, Weis M, et al. (2006) Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. N Engl J Med 355: 2757-2764.
13. Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, et al. (2006) CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. 127: 291-304.
14. Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, et al. (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet 39: 359-365.
15. Van Dijk FS, Nesbitt IM, Zwikstra EH, Nikkels PG, Piersma SR, et al. (2009) PPIB mutations cause severe osteogenesis imperfecta. Am J Hum Genet 85: 521-527.
16. Christiansen HE, Schwarze U, Pyott SM, AlSwaid A, Al Balwi M, et al. (2010) Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta. Am J Hum Genet 86: 389-398.
17. Alanay Y, Avaygan H, Camacho N, Utine GE, Boduroglu K, et al. (2010)

Mutations in the gene encoding the RER protein FKBP65 cause autosomal-recessive osteogenesis imperfecta. Am J Hum Genet 86: 551-559.

18. Lapunzina P, Aglan M, Temtamy S, Caparrós-Martín JA, Valencia M, et al. (2010) Identification of a frameshift mutation in Osterix in a patient with recessive osteogenesis imperfecta. Am J Hum Genet. 87: 110-114.
19. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, et al. (2011) Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. Am J Hum Genet 88: 362-371.
20. Homan EP, Rauch F, Grafe I, Lietman C, Doll JA, et al. (2011) Mutations in SERPINF1 cause osteogenesis imperfecta type VI. J Bone Miner Res 12: 2798-2803.
21. Semler O, Garbes L, Keupp K, Swan D, Zimmermann K, et al. (2012) A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. Am J Hum Genet 91: 349-357.
22. Cho TJ, Lee KE, Lee SK, Song SJ, Kim KJ, et al. (2012) A single recurrent mutation in the 5'-UTR of IFITM5 causes osteogenesis imperfecta type V. Am J Hum Genet 91: 343-348.
23. Martínez-Glez V, Valencia M, Caparrós-Martín JA, Aglan M, Temtamy S, et al. (2012) Identification of a mutation causing deficient BMP1/mTLD proteolytic activity in autosomal recessive osteogenesis imperfecta. Hum Mutat 33: 343-350.
24. Shaheen R, Alazami AM, Alshammari MJ, Faqeih E, Alhashmi N, et al. (2012) Study of autosomal recessive osteogenesis imperfecta in Arabia reveals a novel locus defined by TMEM38B mutation. J Med Genet 10: 630-635.
25. Puig-Hervás MT, Temtamy S, Aglan M, Valencia M, Martínez-Glez V, et al. (2012) Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome--osteogenesis imperfecta phenotypic spectrum. Hum Mutat 33: 1444-1449.
26. Fahiminiya S, Majewski J, Mort J, Moffatt P, Glorieux FH, et al. (2013) Mutations in WNT1 are a cause of osteogenesis imperfecta. J Med Genet 50: 345-348.
27. Laine CM, Joeng KS, Campeau PM, Kiviranta R, Tarkkonen K, et al. (2013) WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. N Engl J Med 368: 1809-1816.
28. Symoens S, Malfait F, D'hondt S, Callewaert B, Dheedene A, et al. (2013) Deficiency for the ER-stress transducer OASIS causes severe recessive osteogenesis imperfecta in humans. Orphanet J Rare Dis 8: 154.
29. Mendoza-Londono R, Fahiminiya S, Majewski J, Care4Rare Canada Consortium, Tétreault M, et al. (2015) Recessive osteogenesis imperfecta caused by missense mutations in SPARC. Am J Hum Genet 96: 979-985.
30. Lindert U, Cabral WA, Ausavarat S, Tongkobpetch S, Ludin K, et al. (2016) MBTPS2 mutations cause defective regulated intramembrane proteolysis in X-linked osteogenesis imperfecta. Nat Commun 7: 11920.