OsteoporosAtlas: a human osteoporosis-related gene database

Xun Wang [Equal first author, 1] , **Lihong Diao** [Equal first author, 1] , **Dezhi Sun** [1] , **Dan Wang** [1] , **Jiarun Zhu** [1] , **Yangzhige He** [1] , **Yuan Liu** [1] , **Hao Xu** [1] , **Yi Zhang** [1] , **Jinying Liu** [2] , **Yan Wang** [1] , **Fuchu He** [1] , **Yang Li** [Corresp., 1] , **Dong Li** [Corresp. 1]

[1] Beijing Proteome Research Center, Beijing Institute of Life omics, Beijing, China

[2] School of Traditional Chinese Medicine., Beijing University of Chinese Medicine, Beijing, China

Corresponding Authors: Yang Li, Dong Li
Email address: liyang_bprc@163.com, lidong.bprc@foxmail.com

Background. Osteoporosis is a common, complex disease of bone with a strong heritable component, characterized by low bone mineral density, microarchitectural deterioration of bone tissue and an increased risk of fracture. Due to limited drug selection for osteoporosis and increasing morbidity, mortality of osteoporotic fractures, osteoporosis has become a major health burden in aging societies. Current researches for identifying specific loci or genes involved in osteoporosis contribute to a greater understanding of the pathogenesis of osteoporosis and the development of better diagnosis, prevention and treatment strategies. However, little is known about how most causal genes work and interact to influence osteoporosis. Therefore, it is greatly significant to collect and analyze the studies involved in osteoporosis-related genes. Unfortunately, the information about all these osteoporosis-related genes is scattered in lots of extensive literature. Currently, there is no specialized database for easily accessing relevant information about osteoporosis-related genes and miRNAs.

Methods. We extracted data from literature abstracts in PubMed by text-mining and manual curation. Moreover, a local MySQL database containing all the data was developed with PHP on a Windows server.

Results. OsteoporosAtlas (http://biokb.ncpsb.org/osteoporosis/), the first specialized database for easily accessing relevant information such as osteoporosis-related genes and miRNAs, was constructed and served for researchers. OsteoporosAtlas enables users to retrieve, browse and download osteoporosis-related genes and miRNAs. Gene ontology and pathway analyses were integrated into OsteoporosAtlas. It currently includes 617 human encoding genes, 131 human non-coding miRNAs, and 128 functional roles. We think that OsteoporosAtlas will be an important bioinformatics resource to facilitate a better understanding of the pathogenesis of osteoporosis and developing better diagnosis, prevention and treatment strategies.

OsteoporosAtlas: a human osteoporosis-related gene database

Xun Wang[1], Lihong Diao[1], Dezhi Sun[2], Dan Wang[2], Jiarun Zhu[2], Yangzhige He[2], Yuan Liu[2], Hao Xu[3], Yi Zhang[3], Jinying Liu[3], Yan Wang[3], Fuchu He[4], Yang Li[1*], Dong Li[1*]

[1] State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences, Beijing Institute of Life omics, Beijing 102206, China

[2] School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

Corresponding Author:

Yang Li[1*]

Email address: liyang_bprc@163.com

Dong Li[1*]

Email address: lidong.bprc@foxmail.com

18

Abstract

Background. Osteoporosis is a common, complex disease of bone with a strong heritable component, characterized by low bone mineral density, microarchitectural deterioration of bone tissue and an increased risk of fracture. Due to limited drug selection for osteoporosis and increasing morbidity, mortality of osteoporotic fractures, osteoporosis has become a major health burden in aging societies. Current researches for identifying specific loci or genes involved in osteoporosis contribute to a greater understanding of the pathogenesis of osteoporosis and the development of better diagnosis, prevention and treatment strategies. However, little is known about how most causal genes work and interact to influence osteoporosis. Therefore, it is greatly significant to collect and analyze the studies involved in osteoporosis-related genes. Unfortunately, the information about all these osteoporosis-related genes is scattered in lots of

30 extensive literature. Currently, there is no specialized database for easily accessing relevant
31 information about osteoporosis-related genes and miRNAs.

32 **Methods.** We extracted data from literature abstracts in PubMed by text-mining and manual
33 curation. Moreover, a local MySQL database containing all the data was developed with PHP on
34 a Windows server.

35 **Results.** OsteoporosAtlas (http://biokb.ncpsb.org/osteoporosis/), the first specialized database
36 for easily accessing relevant information such as osteoporosis-related genes and miRNAs, was
37 constructed and served for researchers. OsteoporosAtlas enables users to retrieve, browse and
38 download osteoporosis-related genes and miRNAs. Gene ontology and pathway analyses were
39 integrated into OsteoporosAtlas. It currently includes 617 human encoding genes, 131 human
40 non-coding miRNAs, and 128 functional roles. We think that OsteoporosAtlas will be an
41 important bioinformatics resource to facilitate a better understanding of the pathogenesis of
42 osteoporosis and developing better diagnosis, prevention and treatment strategies.

43

Introduction

45 Osteoporosis is a common skeletal disease under strong genetic control (Ralston et al. 2006),
46 whose characteristics are low bone mineral density (BMD), microarchitectural deterioration of
47 bone tissue and increased risk of fracture (Kanis et al. 1994). Osteoporosis is generally induced
48 by various causes, such as age, sex, genetic factors, a range of hormones and environmental
49 factors. Moreover, genetic factors not only have strong control over osteoporosis and its
50 associated phenotypes, but may also influence sensitivity to hormonal and environmental factors
51 (Eisman 1999). Osteoporosis can be divided into primary, secondary, and idiopathic types.
52 Primary osteoporosis is more common and mostly caused by old age and postmenopause (Ji et al.
53 2015). Secondary osteoporosis is most commonly caused or exacerbated by medication
54 exposures or other disorders (Stein et al. 2003). Idiopathic osteoporosis is a type of rare
55 osteoporosis that often occurs in young, healthy individuals who are not postmenopausal or have
56 other, identifiable secondary causes of osteoporosis (Heshmati & Khosla 1998). The clinical
57 outcomes of osteoporosis are fracture of hip, wrist, spine and other types. Hip fracture, an
58 international barometer of osteoporosis, is the most severe osteoporotic fracture, 10-20% more
59 women with hip fracture die than expected for age within the first year, and the mortality rate in
60 men with hip fracture is even higher (Cummings & Melton 2002). Menopausal estrogen
61 deficiency is one of the most important reasons why women are more susceptible to osteoporosis.
62 In addition, men generally have greater cortical mass and larger bone size than women
63 (Richelson et al. 1984). Alarmingly, osteoporosis affects over 200 million individuals
64 worldwide and the incidence of osteoporotic fractures is expected to rise by 50% over the next
65 decade (Burge et al. 2007). With the aging of the population in recent years, osteoporosis is

66   becoming an increasingly huge public health issue associated with increased mortality and
67   morbidity.

68   Osteoporosis is a polygenic disorder, generally determined by the combined effects of several
69   genes and environmental factors. Osteoporosis rarely occurs as the result of single gene
70   mutations. Twin and family studies have shown that genetic factors are extremely important in
71   the regulation of bone density, skeletal geometry, ultrasound properties of bone, bone turnover,
72   and risk of osteoporosis (Stewart & Ralston 2000). Identification of specific loci or genes
73   determining osteoporosis will contribute to a better understanding of the pathogenesis of
74   osteoporosis and developing better diagnosis, prevention and treatment strategies. The current
75   efforts to identify osteoporosis loci or genes have mainly focused on three approaches: animal
76   models, candidate gene approach and genome-wide scans (Huang et al. 2003). Genome-wide
77   association studies have identified over 300 loci associated with BMD. However, little is known
78   of how most causal genes interact with each other and the mechanisms to cause osteoporosis (Al-
79   Barghouthi & Farber 2018). Recently, some studies have shown that how causal genes contribute
80   to the pathogenesis of osteoporosis. For example, the research has shown that Siglec-15 gene-
81   deficient mice exhibit mild osteoporosis and Siglec-15 gene is involved in osteoclast
82   differentiation induced by estrogen deficiency, which suggests Siglec-15 gene is a promising
83   drug therapy target for postmenopausal osteoporosis and age-related osteoporosis (Kameda et al.
84   2015). C-Abl gene and Atm (ATM Serine/Threonine Kinase) gene result in osteoporosis by
85   positively regulating osteoblast differentiation and bone formation, moreover mice lacking either
86   of them shows osteoporosis, while P53 gene inhibits osteoporosis, which negatively regulates
87   osteoblast differentiation and bone formation, and the knockout mouse shows osteosclerosis
88   (Wang & Li 2007). Sclerostin (SOST gene encoded) is an antagonist of WNT/β-catenin
89   signaling (canonical WNT pathway), which predominantly regulates osteoblast differentiation
90   and plays important roles in regulating bone formation. Inactivating monoclonal antibodies
91   against SOST, a inhibitor of the negative regulation of WNT/β-catenin signaling, has been
92   shown to be a candidate for the prevention and treatment of osteoporosis (Rossini et al. 2013).

93   MicroRNAs (miRNAs), a class of small (~22bp) nucleotides, are single-stranded noncoding
94   RNAs that regulate the expression of target genes by binding to their 3'-untranslated region
95   (Carthew & Sontheimer 2009). miRNAs play critical roles in the regulation of various biological
96   processes by targeting mRNAs, including cellular differentiation and proliferation, apoptosis,
97   and tissue development (Kim et al. 2009). An imbalance between osteoblastic bone formation
98   and osteoclastic bone resorption plays a fundamental role in osteoporosis pathogenesis
99   (Teitelbaum 2000). Recent studies discovered that miRNAs play an important role in the subtle
100   equilibrium between bone formation and bone resorption by targeting various genes to regulate
101   osteoblast and osteoclast differentiation and function (Ji et al. 2016). For example, Zhao et al
102   have found that hsa-mir-21 overexpression aggravates osteoporosis by targeting RECK (Zhao et
103   al. 2015). Hsa-mir-34a is a novel and pivotal suppressor of osteoclastogenesis and bone
104   resorption, which blocks osteoporosis by inhibiting osteoclastogenesis and Tgif2 (Krzeszinski et

105 al. 2014). Hsa-mir-133a was shown to directly target Runx2 gene 3'-UTR when overexpressed
106 in MC3T3, an osteoblast cell line. In addition, hsa-mir-133a was also shown to negatively
107 regulate 3 potential osteoclast-related target genes, CXCL11, SLC39A1 and CXCR3. MiRNA
108 expression analysis in human circulating monocytes shows that hsa-mir-133a is a promising
109 biomarker for postmenopausal osteoporosis (Wang et al. 2012).

110 Currently, therapeutic drugs for osteoporosis include hormone replacement therapy, calcitonin,
111 selective estrogen receptor, and bisphosphonates, etc. However, due to side-effects and high
112 price, drug selection for osteoporosis is still limited (Martin & Sims 2005). Therefore, research
113 for identifying osteoporosis-related genes and their pathogenic mechanisms is greatly significant,
114 which promote a better understanding of the pathogenesis of osteoporosis, identification of novel
115 biomarkers, discovery of therapeutic targets and accurate treatment strategies. However,
116 abundant information of previously confirmed osteoporosis-related genes is scattered in lots of
117 extensive literature. Moreover, there is no specialized knowledge base dedicated to osteoporosis-
118 related genes collection, which makes it difficult to understand the pathogenesis of osteoporosis
119 and develop new drug targets for osteoporosis. In order to address this obstacle, we established
120 OsteoporosAtlas 1.0 (http://biokb.ncpsb.org/osteoporosis/) by literature-mining and manual
121 curation, which currently contains 617 human osteoporosis-related genes, 131 human
122 osteoporosis-regulated miRNAs, 84 biological process analysis, and 128 functional roles. Users
123 can retrieve, browse and download osteoporosis-related genes and relevant information from
124 OsteoporosAtlas. Moreover, gene ontology analysis, pathway analysis, and SNP term analysis
125 were integrated into OsteoporosAtlas. In summary, we believe that OsteoporosAtlas will provide
126 a comprehensive osteoporosis research resource for the pathogenesis of individual cases,
127 discovery of drug target, novel diagnostic biomarker, individual treatment methods and precision
128 medication.

129

Materials & Methods

Text Mining and Manual Curation

132 We developed an ontology-based bio-entity recognizer (see the appendix for details). Abstract
133 data of CRAFT 2.0 (Verspoor et al. 2012) corpus was used as the independent test set to evaluate
134 the performance of our bio-entity recognizer by identifying gene/protein based on Protein
135 Ontology (Natale et al. 2010). After evaluation, the precision (the number of correct entities
136 identified divided by the number of entities identified), recall (the number of correct entities
137 identified divided by the number of entities in the sample), F-measure (the harmonic mean of
138 precision and recall) of our bio-entity recognizer for identifying gene/protein were respectively
139 0.959, 0.802 and 0.874, indicating it is of great performance for entity recognition (both the
140 evaluation program and result was available at
141 http://biokb.ncpsb.org/osteoporosis/Public/file/craft_test.rar). Therefore, we performed a

142 comprehensive search for osteoporosis-related literature abstracts in PubMed to extract
143 osteoporosis-related genes by using the bio-entity recognizer. Specifically, it is divided into three
144 steps (Figure 1A):

145 First, data collection: we collected 79,947 sentences in 7,358 abstracts containing the keywords
146 of "osteoporosis", "osteoporoses", "bone loss, age-related", "age-related bone loss", and
147 "perimenopausal bone loss".

148 Second, candidate data extraction: 1,628 osteoporosis-related candidate genes/microRNAs and
149 13,264 candidate evidence sentences (Supplementary material: TableS1.xlsx) were extracted by
150 using our bio-entity recognizer.

151 Third, manual curation: our experts manually curated all candidate genes and evidence sentences,
152 527 encoding genes and 131 miRNAs were finally confirmed as human osteoporosis-related
153 genes. In addition, we integrated osteoporosis-related genes confirmed in DisGeNET (Piñero et
154 al. 2015) (Figure 1B), an integrated comprehensive platform with information about human
155 disease-associated genes and variants. In summary, our database consists of 617 osteoporosis-
156 related encoding genes, 131 osteoporosis-regulated miRNAs (non-coding RNA).

157 **Gene Annotation**

158 In order to facilitate the deep interpretation of the association between osteoporosis and related
159 genes, we annotated each gene by the basic gene annotation files ("gene2refseq") from NCBI
160 FTP site with the information of gene synonyms symbol, genetic location, gene full name,
161 chromosome, gene type, reference sequence information and chromosomal location. We
162 obtained gene ontology annotation for each gene from the Gene Ontology Annotation database
163 (GOA) (Camon et al. 2004) and gene-pathway mapping relationship from the Reactome database
164 (Fabregat et al. 2015). The database of short genetic variation (dbSNP) (Sherry et al. 2001) was
165 used to map SNPs to gene by the publications' PMIDs (PubMed sole Identifier), which makes it
166 easier to access to genome-wide association study and helps researchers to improve the
167 understanding of the genetic background of osteoporosis. Mapping and annotating are done by
168 using public databases of Entrez gene (Maglott et al. 2005),Ensembl (Flicek et al. 2010), UniProt
169 (Bairoch et al. 2005), Antibodypedia (Björling et al. 2008) and neXtProt (Lane et al. 2011).

170

171 # Results

172 **Database Service and implement**

173 All osteoporosis-related genes and their relevant information were deposited into a local MySQL
174 database. The website of OsteoporosAtlas (http://biokb.ncpsb.org/osteoporosis/) was developed

175   with PHP on a Windows server. All data about OsteoporosAtlas is accessible for all users
176   without registration or login.

Database Navigation and Search

178   In order to provide a user-friendly web interface for searching and browsing, we designed five
179   sections of functions for our database (Figure 2A) and provide three query approaches.

180   For the search by gene name, users can input a gene symbol in the "Gene Name" search box
181   (Figure 2B). After clicking the "search" button, user can get detailed gene annotation for
182   osteoporosis (Figure 2D), including functional role (whether this gene is a biomarker/drug target),
183   supporting literature evidence, SNP information from the database of short genetic variation
184   (dbSNP), gene ontology (GO) terms from Gene Ontology Annotation database (GOA), the
185   protein description from database UniProtKB, pathway terms from Reactome analysis, gene
186   expression information from Expression Atlas, protein expression information from Human
187   Protein Atlas (HPA), and the regulated miRNAs information from miRBase database (Kozomara
188   & Griffiths-Jones 2013). Clicking the number of abstracts as evidence will lead to the detailed
189   info of the evidence abstracts (Figure 2F, 2G), and clicking the number of the involved
190   biological process will lead to the view of the involved biological processes (Figure 2E).

191   The section of 'Browse & Download' provides four different approaches,including genes,
192   miRNAs, biological processes and functional roles. The current version of database contains 617
193   genes, 131 miRNAs, 84 biological processes and 128 functional roles. All the information can be
194   downloaded. In the section of "Feedback", we provide a submitting function. If users found
195   osteoporosis-related genes which are absent from our database, they can submit this gene to our
196   database.

Database Use Case

198   Users can find *SOST* gene by searching gene name, protein sequence or nucleotide sequence in
199   our database. Four functional sections are provided in the search results interface (Figure 2B):
200   Relevant information about gene interface (Figure 2D) provides functional role (whether this
201   gene is a biomarker/drug target) , supporting literature evidence, relevant information about
202   genome-wide association study from the database of short genetic variation (dbSNP), gene
203   ontology (GO) terms from Gene Ontology Annotation database, the protein description from
204   UniProtKB database, pathway terms from Reactome analysis, gene expression information from
205   Expression Atlas, protein expression information from Human Protein Atlas (HPA). Evidence
206   interface (Figure 2F, 2G) provides the literature which supports the association of *SOST* gene
207   with osteoporosis. Biological process interface (Figure 2E) provides gene ontology terms
208   associated with *SOST* gene. Searching the name of the biological process can browse the genes
209   in this biological process.
210   Users can find hsa-mir-100 by searching miRNA name in our database. Three functional sections
211   are provided in the search results interface (Figure 2C). Clicking hsa-mir-100 will jump to

212 miRBase database, an online resource for miRNA sequence and annotation (Figure 2H).
213 Evidence interface provides the literature which supports the association of hsa-mir-100 with
214 osteoporosis (Figure 2I, 2J).
215

Discussion

217 The systematic collection of OsteoporosAtlas 1.0 database provides an overview of human
218 osteoporosis-related genes. Reactome analysis shows that these genes are involved in pathways
219 of signal transduction, immune system, gene expression (transcription), extracellular matrix
220 organization, etc (Figure 3A). Gene ontology analysis using Panther indicates that the most
221 common class of osteoporosis-related genes belongs to signaling molecule, followed by
222 hydrolase/nucleic acid binding, receptor/transcription factors, enzyme modulator, oxidoreductase,
223 transferase, transporter, etc (Figure 3B). The Reactome and gene ontology analysis could help
224 biologists to better understand the functional relevance of these genes and guide the experiment.
225 In addition, the existence of many potential pharmacological targets in these pathways we
226 analyzed makes it attractive for osteoporosis drugs discovery. All the results show the value of
227 our collection, demonstrating that OsteoporosAtlas 1.0 database will greatly benefit exploration
228 of pathogenesis and treatment of osteoporosis.

229 OsteoporosAtlas 1.0 also has community curation feature. All users can log in to provide their
230 feedback, confirm or reject the evidence sentence by clicking the "Yes" or "No" button. We will
231 update the database periodically according to users' feedbacks.

232 We identified 94 drug target genes with definite literature evidences in our database. The
233 existing osteoporosis treatment drugs mainly delay bone loss by inhibiting bone resorption or
234 promoting bone formation. However, due to the high price and side effects, the osteoporosis
235 drugs promoting bone formation without side-effects are still in the basic clinical research stages.
236 So our database will help discover new drugs for the treatment of osteoporosis.

237

Conclusions

239 In conclusion, OsteoporosAtlas 1.0 collects 617 osteoporosis-related encoding genes and 131
240 miRNAs, which is the first database specially to present a comprehensive list of osteoporosis-
241 related genes obtained from published literature. We believe that OsteoporosAtlas 1.0 will be
242 widely used as it can provide facilities for scientists and clinicians in searching the literature on
243 osteoporosis-related genes and their function in diseases.

244

Appendix

246 The detailed description of methods:

247 (1). Construction and processing of vocabulary

248 It is necessary to construct different types of vocabularies for natural language processing and
249 semantic recognition, including stop words, high frequency words, domain words and
250 abbreviations. Stop words are mainly from PubMed, MetaMap (Aronson 2001), and Python
251 NLTK (Loper & Bird 2002). High frequency words are mainly from PIR (Protein Information
252 Resource (Barker et al. 2000)) Common English, General Names, Common English of PIR
253 iProLINK (Hu et al. 2004) Entity Recognition, and Google 1000 English. Domain words are
254 mainly from the gene/protein, compound, enzyme, and other biochemical naming dictionary in
255 PIR. Abbreviations are mainly from the open source software BADREX (Gooch 2012) in Github.
256

257 (2). Collection and processing of literature

258 The PubMed abstracts were downloaded in XML format by NCBI E-Utilities API (Sayers 2010)
259 and parsed by python script. The database currently contains approximately 26 million abstracts.

260 (3). Collection and processing of ontology

261 A high-quality dictionary of human genes/proteins and their synonyms is mainly based on
262 Protein Ontology (Natale et al. 2010) and bioThesaurus (Duek et al. 2011). And the human
263 protein ID of UniProtKB is used as a standardized ID.

264 (4). Entity recognition

265 Indexer (python script) is used to build indexes for the PubMed abstracts. Analyzer (python
266 script) is used to perform synonym expansion on the ontology. Indexes and synonyms were
267 combined to perform span near match and entity recognition.

268

Supplementary material

270 Supplementary data are available at PeerJ Online.

271

Acknowledgements

273 We would like to thank the anonymous reviewers for their useful comments and suggestions on
274 this manuscript.

275

References

Al-Barghouthi BM, and Farber CRJTiG. 2018. Dissecting the Genetics of Osteoporosis using Systems Approaches.

Aronson AR. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proceedings of the AMIA Symposium: American Medical Informatics Association. p 17.

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, and Magrane MJNar. 2005. The universal protein resource (UniProt). 33:D154-D159.

Barker WC, Garavelli JS, Huang H, McGarvey PB, Orcutt BC, Srinivasarao GY, Xiao C, Yeh L-SL, Ledley RS, and Janda JFJNar. 2000. The protein information resource (PIR). 28:41-44.

Björling E, Uhlén MJM, and Proteomics C. 2008. Antibodypedia, a portal for sharing antibody and antigen validation data. 7:2028-2037.

Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson AJJob, and research m. 2007. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005–2025. 22:465-475.

Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, and Apweiler RJNar. 2004. The gene ontology annotation (goa) database: sharing knowledge in uniprot with gene ontology. 32:D262-D266.

Carthew RW, and Sontheimer EJJC. 2009. Origins and mechanisms of miRNAs and siRNAs. 136:642-655.

Cummings SR, and Melton LJJTL. 2002. Epidemiology and outcomes of osteoporotic fractures. 359:1761-1767.

Duek PD, Gleizes A, Zwahlen C, Mottaz A, Bairoch A, and Lane LJB-O. 2011. CALOHA: A new human anatomical ontology as a support for complex queries and tissue expression display in neXtProt.

Eisman JAJEr. 1999. Genetics of osteoporosis. 20:788-804.

Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, and McKay SJNar. 2015. The reactome pathway knowledgebase. 44:D481-D487.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, and Fitzgerald SJNar. 2010. Ensembl 2011. 39:D800-D806.

309   Gooch PJapa. 2012. Badrex: In situ expansion and coreference of biomedical abbreviations
310   using dynamic regular expressions.

311   Heshmati HM, and Khosla S. 1998. Idiopathic osteoporosis: a heterogeneous entity. Annales de
312   medecine interne. p 77-81.

313   Hu Z-Z, Mani I, Hermoso V, Liu H, Wu CHJCb, and chemistry. 2004. iProLINK: an integrated
314   protein resource for literature mining. 28:409-416.

315   Huang Q-Y, Recker RR, and Deng H-WJOI. 2003. Searching for osteoporosis genes in the
316   post-genome era: progress and challenges. 14:701-715.

317   Ji M-X, Yu QJCd, and medicine t. 2015. Primary osteoporosis in postmenopausal women. 1:9.

318   Ji X, Chen X, and Yu XJIjoms. 2016. MicroRNAs in osteoclastogenesis and function: potential
319   therapeutic targets for osteoporosis. 17:349.

320   Kameda Y, Takahata M, Mikuni S, Shimizu T, Hamano H, Angata T, Hatakeyama S, Kinjo M,
321   and Iwasaki NJB. 2015. Siglec-15 is a potential therapeutic target for postmenopausal
322   osteoporosis. 71:217-226.

323   Kanis JA, Melton III LJ, Christiansen C, Johnston CC, Khaltaev NJJob, and research m. 1994.
324   The diagnosis of osteoporosis. 9:1137-1141.

325   Kim VN, Han J, and Siomi MCJNrMcb. 2009. Biogenesis of small RNAs in animals. 10:126.

326   Kozomara A, and Griffiths-Jones SJNar. 2013. miRBase: annotating high confidence
327   microRNAs using deep sequencing data. 42:D68-D73.

328   Krzeszinski JY, Wei W, Huynh H, Jin Z, Wang X, Chang T-C, Xie X-J, He L, Mangala LS, and
329   Lopez-Berestein GJN. 2014. miR-34a blocks osteoporosis and bone metastasis by
330   inhibiting osteoclastogenesis and Tgif2. 512:431.

331   Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, Gateau A, Gaudet P, Gleizes A,
332   and Masselot AJNar. 2011. neXtProt: a knowledge platform for human proteins.
333   40:D76-D83.

334   Loper E, and Bird SJapc. 2002. NLTK: the natural language toolkit.

335   Maglott D, Ostell J, Pruitt KD, and Tatusova TJNar. 2005. Entrez Gene: gene-centered
336   information at NCBI. 33:D54-D58.

337   Martin TJ, and Sims NAJTimm. 2005. Osteoclast-derived activity in the coupling of bone
338   formation to resorption. 11:76-81.

339   Natale DA, Arighi CN, Barker WC, Blake JA, Bult CJ, Caudy M, Drabkin HJ, D'eustachio P,
340   Evsikov AV, and Huang HJNar. 2010. The Protein Ontology: a structured representation
341   of protein forms and complexes. 39:D539-D545.

342   Piñero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, and
343   Furlong LIJD. 2015. DisGeNET: a discovery platform for the dynamical exploration of
344   human diseases and their genes. 2015.

345   Ralston SH, de Crombrugghe BJG, and development. 2006. Genetic regulation of bone mass
346   and susceptibility to osteoporosis. 20:2492-2506.

347   Richelson LS, Wahner HW, Melton III L, and Riggs BLJNEJoM. 1984. Relative contributions of
348   aging and estrogen deficiency to postmenopausal bone loss. 311:1273-1275.

349   Rossini M, Gatti D, and Adami SJCti. 2013. Involvement of WNT/β-catenin signaling in the
350   treatment of osteoporosis. 93:121-132.

351   Sayers EJEPUHBNCfBI. 2010. A General Introduction to the E-utilities.

352   Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, and Sirotkin KJNar. 2001.
353   dbSNP: the NCBI database of genetic variation. 29:308-311.

354   Stein E, Shane EJE, and America mcoN. 2003. Secondary osteoporosis. 32:115-134, vii.

355   Stewart T, and Ralston SJJoE. 2000. Role of genetic factors in the pathogenesis of
356   osteoporosis. 166:235-246.

357   Teitelbaum SLJS. 2000. Bone resorption by osteoclasts. 289:1504-1508.

358   Verspoor K, Cohen KB, Lanfranchi A, Warner C, Johnson HL, Roeder C, Choi JD, Funk C,
359   Malenkiy Y, and Eckert MJBb. 2012. A corpus of full-text journal articles is a robust
360   evaluation tool for revealing differences in performance of biomedical natural language
361   processing tools. 13:207.

362   Wang X, and Li BJIjobsI. 2007. Genetic studies of bone diseases: evidence for involvement of
363   DNA damage response proteins in bone remodeling. 3:217.

364   Wang Y, Li L, Moore BT, Peng X-H, Fang X, Lappe JM, Recker RR, and Xiao PJPo. 2012. MiR-
365   133a in human circulating monocytes: a potential biomarker associated with
366   postmenopausal osteoporosis. 7:e34641.

367   Zhao W, Dong Y, Wu C, Ma Y, Jin Y, Ji YJM, and biochemistry c. 2015. MiR-21 overexpression
368   improves osteoporosis by targeting RECK. 405:125-133.

369

Figure 1

Figure1 The construction workflow of OsteoporosAtlas 1.0 database.

(A) The workflow of OsteoporosAtlas 1.0 database construction; (B) The comparison of osteoporosis-related genes between OsteoporosAtlas 1.0 database and DisGeNET database.
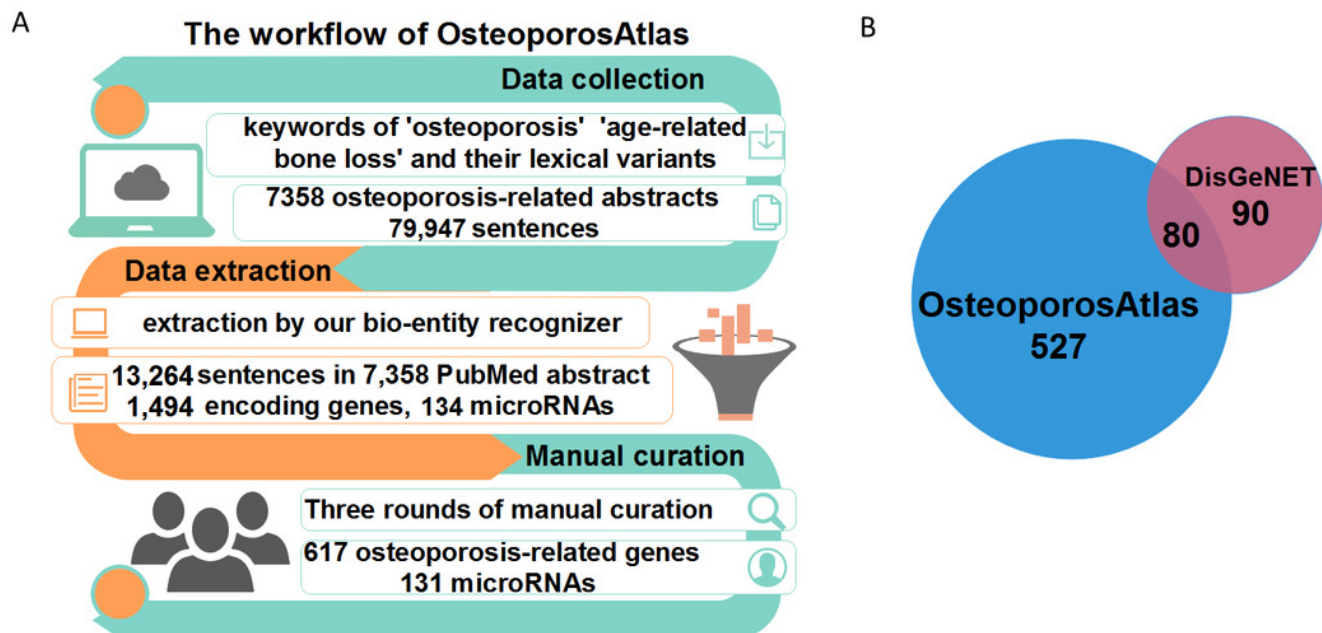
Figure2 The outline for searching in OsteoporosAtlas 1.0 website.

(A) Five functional sections of the database. (B) Users can submit a gene name to the "Gene Name" search box and the search result will be returned, including the information of Gene, related biological processes, functional role, and evidence. (D) After clicking the gene name, users can get more specific information about this gene on the detailed page. (E) After clicking on the number of biological processes, users can scan the biological processes involved by this gene. (F, G) After clicking the number of evidence, the original abstract will be displayed with highlighted matched sentence and matched keywords.(C) Users can submit a microRNA name to the "MicroRNA Name" search box and the search result will be returned, including the information of microRNA, functional role and support evidence. (H) After clicking the microRNA name, users can get more specific information about this microRNA in the detailed page. (I, G) After clicking the number of the evidence, the original abstract will be displayed with highlighted matched sentence and matched keywords.

Figure 3

Figure 3 Bioinformatics analysis of genes associated with osteoporosis.

(A) Biological pathway analysis using Reactome (http://www.reactome.org/). (B) Protein class analysis using PANTHER (http://www.pantherdb.org).