Abstract. As a common degenerative disease, osteoporosis (OS) is characterized by reduced bone mass and microarchitectural deterioration of bone tissue. Both genetic and environmental factors are involved in OS development. To date, ~300 genes have been confirmed to be involved in the pathogenesis of OS, a large majority of which have been independently investigated. As OS is a polygenic disease, a comprehensive analysis focusing on the biological functions and interactions of OS-related genes would provide valuable information. In this study, OS related research deposited in PubMed was retrieved and genes related to OS were catalogued. Pathways with an enriched biological function for these genes were extracted, and the crosstalk between the enriched pathways was analyzed. A comprehensive network was constructed, and a minimal network was extracted using the Steiner minimal network algorithm. In this study, a total of 294 genes were retrieved from PubMed. Biological processes found to be enriched included those related to bone metabolism and the immune system. In total, 58 pathways were enriched. Furthermore, the comprehensive network consisting of 3,943 nodes and 7,976 edges was constructed, among which 631 nodes and 2,581 edges contributed to the OS-specific molecular network. In this network, in excess of 300 potential genes associated with OS and two modules were identified. Thus, this study provides a mechanistic insight into OS and suggests more than 300 potential OS-related genes for future research.

Introduction

As a common degenerative disease, osteoporosis (OS) is characterized by reduced bone mass and microarchitectural deterioration of bone tissue (1), which often means patients with OS suffer from back pain and bone fractures (2). It is estimated that ~50% of women and 30% of men >50 years old will suffer a fracture due to osteoporosis globally (3).

The development of OS can be influenced by genetic and environmental factors (4,5). Genetic factors play a greater role than environmental factors. It is estimated that the heritability of vertebral volumetric bone mineral density ranges from 0.27 to 0.51 (6). To date, animal models, gene expression, genome-wide association studies (GWAS) and systems biology approaches have identified hundreds of genes found to contribute to the process of OS (7-9). This suggests that OS is a polygenic disease; that is, numerous genes exert an effect on the development of OS, and multiple genes contribute to the pathophysiology of OS (10). Thus far, the function and role of these genes in the process of OS have only been investigated independently. However, a systematic strategy to collectively analyze the function and connection of these potential OS-related genes is preferable.

Recently, Qin et al (11) applied multiple computational approaches to analyze OS-associated single nucleotide polymorphisms (SNPs) and genes identified in GWAS. This study focused on gene regulatory networks, and retrieved some transcription factors (including NFATC2 and MEF2C) and microRNAs (miRNAs; including miR-3658 and miR-345-5p) that bind to loci of SNPs. However, the use of SNPs alone to identify genes involved in OS could lead to some genes known to be involved in OS being unaccounted for in such an analysis. To overcome this, genes known to be associated with OS were collected from PubMed and biological enrichment analyses was conducted to detect any significant functional themes within these genes. The biochemical pathways associated with these genes were analyzed for interactions among the enriched pathways. Finally, an OS-specific network based on the human protein-protein interaction network was constructed. The current study aimed to promote an understanding of the molecular mechanism of OS and to identify potential OS-related genes for future research.

Materials and methods

Identification of OS-related genes. OS-related gene candidates were curated by retrieving human genetic association studies deposited in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). Referring to published studies (12,13), reports related to OS were queried using the terms 'osteoporosis,
Functional enrichment analysis of OS-related genes. Pathway and process enrichment analysis was carried out using the following ontology sources: Kyoto Encyclopedia of Genes and Genomes (KEGG; Release 87.0; https://www.genome.jp/kegg), Gene Ontology (GO) Biological Processes (Version 2018-08-09; http://geneontology.org/), Reactome Gene Sets (Version 64; https://reactome.org), Molecular Signatures Database (Version 6.2; http://software.broadinstitute.org/gsea/msigdb/index.jsp) and CORUM (Version 3.0; http://mips.gsf.de/genre/proj/corum/index.html) (14). All genes in the genome were used as the enrichment background. In the gene set enrichment analysis, P-values were calculated based on cumulative hypergeometric distribution (15) and Q-values were calculated using the Benjamini-Hochberg procedure for multiple testing (16). The term was regarded as over-represented when $P<0.05$. To reduce redundancy in the ontology terms, the terms with $P<0.01$, a minimum count of 3 and an enrichment factor $>2.0$ were collected and grouped into clusters based on their membership similarities. In the process of hierarchical clustering of the enriched terms, k scores were used as the similarity metric, and sub-trees with similarity $>0.3$ were considered to be a cluster. The most statistically significant term within a cluster was chosen to represent the cluster. All of these results were obtained using the web-based tool, Metascape (Version 3.0, http://metascape.org) (14,17).

Pathway crosstalk analysis. After enriching for biological process, the biochemical pathways of OS-related genes were further enriched using the web-based tool ToppGene (https://toppgene.cchmc.org/). In brief, the UniProt ID (https://www.uniprot.org/) list corresponding to the retrieved genes was uploaded into the server of ToppGene. Then, the list was automatically matched to canonical pathways using the KEGG and BioCarta (www.biocarta.com) pathway databases. A pathway was catalogued when it overlapped with one or more of the genes in the list. The overlap significance between the pathway and the input genes was analyzed using Fisher’s exact test and the Bonferroni correction. The pathways were considered to be significantly enriched when the false discovery rate value was $<0.05$.

Pathway crosstalk was evaluated using the Jaccard coefficient (JC) and overlap coefficient (OC). Both of these describe the overlap between any given pair of pathways, and are defined as follows: $JC = \frac{\min(|A|,|B|)}{|A \cap B|}$ and $OC = \frac{|A \cap B|}{\min(|A|,|B|)}$ where $A$ and $B$ are the number of genes included in the two tested pathways.

As pathways with too few genes may have insufficient biological information, only enriched pathways containing more than three candidate genes were included and pathway pairs with fewer than two overlapping genes were removed. All pathway pairs were ranked according to the average of the JC and OC coefficients. For clarity, crosstalk between pathways was also shown in the Cytoscape software (version 3.7.1; https://cytoscape.org).

Construction of the human interactome and an OS-specific subnetwork. To further investigate the interaction and correlation between these genes, a comprehensive human interactome was constructed based on the InnateDB database (http://www.innatedb.com). In this process, a novel web-based tool, OmicsNet (https://www.omicsnet.ca), was used. OmicsNet allows users to create molecular interaction networks and visually explore them in a three-dimensional space (18). The list of retrieved genes was uploaded to OmicsNet. Under the guidance of OmicsNet, an interactome was constructed based on the InnateDB database and the topological characteristics of potential molecular networks were analyzed using OmicsNet.

Next, an OS-specific network was extracted. The process was similar to the Steiner Tree problem, where the algorithm identifies a minimal sub-network containing all the terminal nodes from the complete network. OmicsNet was used for extraction, which implemented a heuristic approach that provides an approximate answer to this problem in order to reduce computation time.

To assess the non-randomness of the constructed network, 1000 random networks with the same number of vertices and interactions as the OS-specific network were generated using the Erdos-Renyi model (19) in the igraph R package (Version 0.7.1; https://igraph.org/) (20). The arithmetic average values of the shortest path distance and clustering coefficient were calculated. The number of random networks whose average shortest distance was less than that of the OS specific network was calculated and defined as ND. In a similar manner, the number of random networks whose average clustering coefficients was higher than that of the OS specific network was calculated and defined as NC. At last, the empirical P-values were calculated separately using ND/1000 and NC/1000. When both P-values were $<0.05$, the OS specific network was regarded as non-random.

Finally, potential modules were attained using the Infomap algorithm embedded in OmicsNet (https://www.omicsnet.ca) (18). Based on the principles of information theory, the Infomap algorithm transforms the problem of finding the potential modules into the problem of finding a description of minimum information for a random walk on the network graph (21). Using this strategy, the InfoMap algorithm joins neighboring nodes into modules.

Results

Identification of genes reported to be associated with OS. Up to June 1st 2018, a total of 1,335 publications were retrieved for OS. A total of 294 genes were retrieved (Table SI) and curated into a gene set (OS-related genes gene set, OSgset). Referring to a previously published article (9), this gene set can be divided into four subgroups: Calcitropic hormones and receptors (including CASR, CRHR and CTR), cytokines, growth factors and receptors (including BMP2, FGFR1 and
IL-23), bone matrix proteins (including COL1A1, COL1A2 and ITGA1) and miscellaneous (including ADCY10, ALOX15 and ALOX5). The diversity of the genes significantly associated with OS clearly demonstrates that OS is a polygenetic disease.

Biological functions enriched in the OSgset. Functional enrichment analysis can reveal a more specific function of these genes. In total, >2,000 GO terms were significantly enriched in the genes analyzed. Considering the heavy overlap between ontology terms and the need to reduce redundancy, terms were collected with P-values <0.01, a minimum count of 3 and an enrichment factor >2.0, and grouped into clusters based on their membership similarity. The top 20 ranked clusters are shown as a heatmap (Fig. 1 and Table SII). Among these clusters, some biological processes can be discerned: These include the bone forming process (including ‘ossification’, ‘skeletal system development’, ‘tissue remodeling’, ‘tissue morphogenesis’, ‘heart development’, ‘negative regulation of cell differentiation’, ‘vasculature development’ and ‘regulation of secretion’), the bone regulation process (including ‘response to nutrient levels’, ‘response to steroid hormone’, ‘response to peptide’, ‘inflammatory response’, ‘cellular response to organic cyclic compound regulation of animal organ morphogenesis’ and ‘response to growth factor’), and signaling pathways in bone cells (including ‘regulation of MAPK cascade’, ‘GPCR ligand binding’, ‘transmembrane receptor protein tyrosine kinase signaling pathway’, ‘regulation of signaling receptor activity’ and ‘negative regulation of cell proliferation’). These results indicated that the candidate genes collected were reliable for follow-up bioinformatics analysis.

Pathway enrichment analysis in OSgset. Enriching the biochemical pathways in which candidate genes are involved and analyzing their cross-talk can promote our understanding of the molecular mechanisms underlying OS. As shown in Table I (further details are provided in Table SIII), 58 significant enrichment pathways for OS were identified. Among them, the pathway ‘ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins’ was the most significantly enriched (P=3.60x10^-18). In addition, some extracellular matrix-related pathways were also enriched, including ‘ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors’, ‘genes encoding structural ECM glycoproteins’, ‘extracellular matrix organization’ and ‘ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans’. These findings confirmed a role for metabolism of the extracellular matrix in the development of OS.

The second enriched pathway was ‘Wnt signaling pathway’. Moreover, the number of Wnt-related pathways was the highest, including ‘Wnt signaling pathway’, ‘canonical Wnt signaling’, ‘genes related to Wnt-mediated signal transduction’, ‘Wnt signaling network’, ‘negative regulation of TCF-dependent signaling by WNT ligand antagonists’, ‘signaling by Wnt’, ‘signaling by WNT in cancer’, ‘WNT ligand biogenesis and trafficking’, ‘TCF dependent signaling in response to WNT’, ‘Wnt/beta-catenin Pathway’ and ‘RNF mutants show enhanced WNT signaling and proliferation’. These results suggested that Wnt-related pathways play an important role in the development of OS.

The third enriched pathway was ‘breast cancer’, which suggested that OS is closely related to breast cancer. In addition, immune-associated biological processes, including ‘cytokines and inflammatory response’ and ‘cytokine network’, were also significantly enriched, suggesting that the immune system may be involved in the etiology and pathological process of OS.
Table I. Pathways enriched in the OSgset.

| Pathway                                                                | P-value\(^a\)    | \(p_{\text{BH}}\) value\(^b\) |
|-----------------------------------------------------------------------|------------------|-------------------------------|
| Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins | 3.60x10\(^{-18}\) | 6.05x10\(^{-15}\)           |
| Wnt signaling pathway                                                 | 3.71x10\(^{-13}\) | 6.23x10\(^{-10}\)           |
| Breast cancer                                                         | 4.29x10\(^{-13}\) | 7.21x10\(^{-10}\)           |
| Genes encoding secreted soluble proteins                              | 4.63x10\(^{-13}\) | 7.78x10\(^{-10}\)           |
| Cytokine-cytokine receptor interaction                                | 7.86x10\(^{-13}\) | 1.32x10\(^{-9}\)           |
| Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors | 7.91x10\(^{-12}\) | 1.33x10\(^{-8}\)           |
| Canonical Wnt signaling                                              | 8.61x10\(^{-12}\) | 1.45x10\(^{-8}\)           |
| Class B/2 (Secretin family receptors)                                 | 1.10x10\(^{-11}\) | 1.85x10\(^{-8}\)           |
| GPCR ligand binding                                                   | 1.14x10\(^{-11}\) | 1.91x10\(^{-8}\)           |
| Interleukin-4 and 13 signaling                                       | 3.51x10\(^{-11}\) | 5.90x10\(^{-8}\)           |
| Genes related to Wnt-mediated signaling                               | 5.88x10\(^{-11}\) | 9.88x10\(^{-8}\)           |
| Proteoglycans in cancer                                               | 7.02x10\(^{-11}\) | 1.18x10\(^{-7}\)           |
| Cytokine Signaling in Immune system                                   | 1.52x10\(^{-10}\) | 2.56x10\(^{-7}\)           |
| Wnt signaling network                                                 | 1.69x10\(^{-10}\) | 2.83x10\(^{-7}\)           |
| Cytokine Network                                                      | 2.04x10\(^{-10}\) | 3.42x10\(^{-7}\)           |
| Negative regulation of TCF-dependent signaling by WNT ligand antagonists | 2.35x10\(^{-10}\) | 3.95x10\(^{-7}\)           |
| Rheumatoid arthritis                                                  | 7.22x10\(^{-10}\) | 1.21x10\(^{-6}\)           |
| Signaling by Wnt                                                      | 1.07x10\(^{-9}\)  | 1.80x10\(^{-6}\)           |
| Fluid shear stress and atherosclerosis                               | 1.43x10\(^{-9}\)  | 2.40x10\(^{-6}\)           |
| Basal cell carcinoma                                                  | 1.49x10\(^{-9}\)  | 2.51x10\(^{-6}\)           |
| Pathways in cancer                                                    | 1.81x10\(^{-9}\)  | 3.05x10\(^{-6}\)           |
| Alzheimer disease-presenilin pathway                                  | 1.83x10\(^{-8}\)  | 3.08x10\(^{-5}\)           |
| Validated transcription targets of AP1                                 | 2.84x10\(^{-8}\)  | 4.78x10\(^{-5}\)           |
| family members Fra1 and Fra2                                          |                  |                               |
| Signaling pathways regulating pluripotency of stem cells              | 4.81x10\(^{-8}\)  | 8.09x10\(^{-5}\)           |
| ALK in cardiac myocytes                                               | 6.37x10\(^{-8}\)  | 1.07x10\(^{-4}\)           |
| Malaria                                                               | 7.09x10\(^{-8}\)  | 1.19x10\(^{-4}\)           |
| Cytokines and Inflammatory Response                                   | 1.21x10\(^{-7}\)  | 2.03x10\(^{-4}\)           |
| Neuroactive ligand-receptor interaction                               | 1.45x10\(^{-7}\)  | 2.44x10\(^{-4}\)           |
| HTLV-I infection                                                      | 1.54x10\(^{-7}\)  | 2.59x10\(^{-4}\)           |
| Genes encoding structural ECM glycoproteins                          | 2.30x10\(^{-7}\)  | 3.86x10\(^{-4}\)           |
| Metabolism of steroid hormones                                        | 2.79x10\(^{-7}\)  | 4.69x10\(^{-4}\)           |
| Signaling by Interleukins                                             | 4.17x10\(^{-7}\)  | 7.02x10\(^{-4}\)           |
| Extracellular matrix organization                                     | 4.82x10\(^{-7}\)  | 8.11x10\(^{-4}\)           |
| Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans | 5.09x10\(^{-7}\)  | 8.56x10\(^{-4}\)           |
| Signaling by WNT in cancer                                            | 5.91x10\(^{-7}\)  | 9.93x10\(^{-4}\)           |
| Wnt signaling pathway                                                | 7.16x10\(^{-7}\)  | 1.20x10\(^{-3}\)           |
| Angiogenesis                                                          | 8.07x10\(^{-7}\)  | 1.36x10\(^{-3}\)           |
| Ovarian steroidogenesis                                               | 1.00x10\(^{-6}\)  | 1.69x10\(^{-3}\)           |
| Hippo signaling pathway                                              | 1.13x10\(^{-6}\)  | 1.90x10\(^{-3}\)           |
| Inflammatory bowel disease (IBD)                                      | 1.14x10\(^{-6}\)  | 1.92x10\(^{-3}\)           |
| WNT ligand biogenesis and trafficking                                 | 1.22x10\(^{-6}\)  | 2.05x10\(^{-3}\)           |
| AGE-RAGE signaling pathway in diabetic complications                 | 1.36x10\(^{-6}\)  | 2.28x10\(^{-3}\)           |
Crosstalk among significantly enriched pathways. To understand how these enriched pathways interact with each other, we performed a pathway crosstalk analysis for the 58 enriched pathways. All 58 pathways contained three or more members in OSGset. Among the pathways, 706 pairs of pathways met the criterion for crosstalk analysis, that is, each pathway pair shared at least two genes with one or more other pathways.

Table I. Continued.

| Pathway                                             | P-value$^a$   | $p_{BH}$ value$^b$ |
|-----------------------------------------------------|--------------|-------------------|
| G alpha (s) signaling events                        | 3.12x10^{-6} | 5.24x10^{-3}     |
| Disassembly of the destruction complex and recruitment of AXIN to the membrane | 4.21x10^{-6} | 7.07x10^{-3}     |
| Osteoclast differentiation                          | 4.42x10^{-6} | 7.44x10^{-3}     |
| mTOR signaling pathway                              | 4.64x10^{-6} | 7.80x10^{-3}     |
| PI3K-Akt signaling pathway                          | 4.74x10^{-6} | 7.98x10^{-3}     |
| TCF dependent signaling in response to WNT          | 4.92x10^{-6} | 8.27x10^{-3}     |
| Wnt/beta-catenin Pathway                            | 6.48x10^{-6} | 1.09x10^{-2}     |
| Interleukin-10 signaling                            | 8.72x10^{-6} | 1.47x10^{-2}     |
| Glucocorticoid receptor regulatory network          | 9.88x10^{-6} | 1.66x10^{-2}     |
| HIF-1 signaling pathway                             | 1.07x10^{-5} | 1.80x10^{-2}     |
| The IGF-1 Receptor and Longevity                    | 1.12x10^{-5} | 1.88x10^{-2}     |
| IL23-mediated signaling events                      | 1.17x10^{-5} | 1.97x10^{-2}     |
| Detoxification of Reactive Oxygen Species           | 1.41x10^{-5} | 2.37x10^{-2}     |
| Allograft rejection                                 | 1.41x10^{-5} | 2.37x10^{-2}     |
| Circadian Clock                                     | 2.01x10^{-5} | 3.38x10^{-2}     |
| RNF mutants show enhanced WNT signaling and proliferation | 2.43x10^{-5} | 4.08x10^{-2}     |

$^a$P-values were calculated by Fisher's exact test; $^b$p$_{BH}$ values were adjusted by the Benjamini and Hochberg method. OSGset, osteoporosis-related genes gene set.

In Fig. 2, the crosstalk network is arranged based on the closeness centrality of every pathway. The maximum value of 0.84 was for the pathway ‘ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins’ pathway.
Table II. List of genes in different functional groups.

A, Genes in OsGset serving as seeds

| ADRA2A, JAG1, AGTR2, AKR1B1, ALOX5, ALOX15, ALPL, ANXA6, APC, ARHGAP1, ATP5E, AVP, B2M, BDNF, BGLAP, BLVRB, BMP2, BMP7, BMPRI1B, C8A, CALCRL, CAS, CAT, RUNX2, CCNE1, CD40, CD40LG, CFTR, CHD2, CLCN7, CNR1, COL1A1, COL6A1, COL9A1, COMT, CPBR2, CRHR1, CSTA, CTNNB1, CYSL, CYLD, CYPIA1, CYPIB1, CYPI7A1, CYPI9A1, CYPI24A1, DBP, DLX5, DMP1, ESR1, ESR2, ESRMA, MECOM, FABP3, FDPS, FGFR1, FGFR2, FOXC2, FNT1, FRZB, GC, MSTM, GHR, GIP, GIPR, GLA, NR3C1, GSR, GSTM3, GSTP1, HIF1A, HSD11B1, HSD17B2, HTRA2, IBSP, IRF8, ID4, IFNG, IFI1, IFGBP2, IL1A, IL6, IL6R, IL7, IL10, IL15, IL16, IL17A, ITGA1, JUND, LCT, LEPR, LHCGR, LR1, LRP4, LRPS, LTA, LTBP2, MARK3, MEC2, MGF, CIITA, MID1, MMP2, CD200, MXS1, MYO5A, NELL1, NF1, NFATC1, NFKB2, NOS3, NPY, OR2, TNRFSF11B, P2RX1, P2RX4, P2RX7, PCSK6, SERpine1, REG3A, PAPPA, PPNB, PGR, ABCB1, PLAUR, PLOD1, PLS3, PON1, PPARG, PRL, PTH1, PTHLH, PTHR2, PTMA, PTPN, RAC1, RAC2, RBL2, RPL29, SAAl, ATXN1, CCL2, SFRP1, SFRP2, SFRP4, SBHB, SLC6A4, SNA, SOD1, SOD2, SOX4, SOX9, SPARC, SPP2, SPP1, SPTB1, SREBF1, TERT, TIMP1, TIMP2, TLR4, TNRFSF11B, TRPS1, TSHR, TWIST1, TYROBP, VDR, VEGFA, WNT1, WNT7B, WT1, WNTB, WNT10B, PRDM2, CSDE1, CXCR4, SHFM1, HGF5, FZD1, FZD6, SPoP, Pir, TnFSF11, ncoa1, TnFrSF11a, Per3, nG, MaP4K4, aTP6V1G1, TraM2, ZBTB40, nr1i3, Hdac5, GPc6, Slc25a13, SPRY1, ZMPSTe24, aKr1a1, TuBa1B, FaM3c, GPnMB, crTaP, ncoa2, nMu, AKAP13, AKAP11, MMRN1, DKK1, CLEC16A, SATB2, DICER1, DCAF13, FGF20, DKK2, TMEM14A, CNOT7, NOX4, ARHGEF3, SOST, SLMO2, ADIPOR1, WNT16, IL23A, ATP6V1H, SOX6, ITLN1, ADCY10, MEPE, RGMA, SQRDL, TRPV4, SOMCi, IFIHi1, TMEM135, WKORC1, FTO, WLS, CDC170, WNT5B, TARD3NL, NLRCS, WNT3A, SP7, FAM210A, IL23R, VPS13B, ZNF384, THSd7a, ncF1, ccr2 |

B, Genes included in Os-specific network but not in OsGset

| UBC, HNF4A, SP1, ELAVL1, EP300, SUMO2, CXCR4, RELA, CREBBP, SUMO1, STAT3, HSP90AA1, TP53, SMAD3, SRC, HDAC1, APP, SIRT1, JUN, EGR1, COP55, TRAF6, SMAD2, GRB2, PIK3R1, YWHAZ, MARCA4, MYC, HSP90AB1, NFKB1, STAT1, CALM1, HDAC3, HSPA8, HDAC4, E2F1, STUB1, CEPPB, YWHAQ, FOS, FOSL1, CREB1, NCO1, CEPPA, MDMD, KAT2B, GSK3B, SMAD4, CEPEB, IRF1, POIUF1, YY1, F1N1, HSPA4. A TP6, A2M, KPNB1, HNF1A, UBE2I, XRCC5, NOS2, MAPK1, SUM04, NCOA6, ZBTB16, VCP, KIAA0101, NCIK1, TCF3, PLCG1, FYN, H2AFX, HDAC6, EPA51, ACTB, CANDX, HSPA5, EEF1A1, TUBB, TUBA1A, FLNA, SPI1, ETS1, NEDD4, AKT1, CCND1, PRRDC, EZH2, XRCC6, TBP, PPAPC2A, XPO1, BCL6, PIA51, REL, CRK, HNRNPAA1, DNM1T1, UBE3A, CBL, TRAF3, PRKCA, IL7R, TFAP2A, TRAF2, PSMC5, USFI, TNF, PPP1CA, HIST1H4E, MYOD1, STAT5A, VDAC1, PSMC4, MED24, BIRC2, NPM1, STAT5B, USQ11, CDK1, ITGAV, CYCS2, XPC, SOD1, SOD2, SOX9, SPARC, SPP2, SPP1, SPTBN1, SREBF1, TERT, TGFBR3, TIMP1, TIMP2, TLR4, TNRFSF11B, TRPS1, TSHR, TWIST1, TYROBP, VDR, VEGFA, WNT1, WNT7B, WNT10B, PRDM2, CSDE1, CXCR4, SHFM1, HGF5, FZD1, FZD6, SPoP, PIR, TNRFSF11, NCOA1, TNRFSF11A, PER3, NAG, MAP4K4, ATP6V1G1, TRAM2, ZBTB40, NR1I3, HDAC5, G6PC6, SATB2, DICER1, DCAF13, FGF20, DKK2, TMEM14A, CNOT7, NOX4, ARHGEF3, SOST, SLMO2, ADIPOR1, WNT16, IL23A, ATP6V1H, SOX6, ITLN1, ADCY10, MEPE, RGMA, SQRDL, TRPV4, SOMCi, IFIHi1, TMEM135, WKORC1, FTO, WLS, CDC170, WNT5B, TARD3NL, NLRCS, WNT3A, SP7, FAM210A, IL23R, VPS13B, ZNF384, THSd7a, NCFl, CCR2 |

C, Genes in the main module extracted from the Os specific network

| A2M, AHR, AKT1, APC, BIRC2, ARRB1, ATF1, B2M, CCND1, BCL3, BCL6, CALM1, CAV1, RUNX2, CCNE1, CD4, CD40, CDK1, CEPPA, CEPPB, CEPPD, CFTR, CLU, CCR5, CREB1, ATF2, CREBBP, CREM, CNS2A1, CTNNB1, CYP19A1, CYP24A1, DDB1, DHX9, E2F1, EEF1A1, EGFR, EGR1, ELF1, EP300, EPAS1, NR2F6, ESR1, ESR2, ESRRA, |
Table II. Continued.

C, Genes in the main module extracted from the OS specific network

ET51, ET52, MECOM, FLNA, FOS, FTH1, NR5A1, XRCC6, GATA2, NR3C1, GSK3B, HDAC1, HDAC2, HIF1A, HNRNPA1, HSPA4, HSPA5, HSPA8, HSP90AA1, HSP90AB1, IRF8, IFNG, RBPJ, IL6, IL10, IL16, IRF1, JUN, JUND, KPNB1, LRP4, LTA, SMAD2, SMAD3, SMAD4, MARK3, MCM3, MDM2, MECP2, MEF2C, CTITA, MID1, MMP2, CD200, MYC, MYO5A, MYOD1, NF1, NFATC1, NFKB1, NFKB2, NME1, NOS2, NOS3, NPM1, P2RX1, P4HB, SERPINE1, PBX1, PCNA, PGR, PHB, POLR2A, POU2F1, PPAR, PPP2CA, PRKDC, MAPK1, PRL, PSMA6, PSMA7, PSMC3, PSMC4, PSMC5, PSMD1, PSMD11, PTHLH, PTH1R, PTMA, RB1, REL, RELA, RNF4, RXRA, ATXN1, CCL2, CCL7, SFTPD, SGK1, SKP2, SMARCA4, SUMO2, SNCA, SOD1, SOD2, SOX4, SOX9, SP1, SPI1, SPTBN1, SRC, SREBF1, STAT1, STAT5A, STAT6, TBP, TCF3, TERT, TFAP2A, TLR4, TNF, TNFRSF1B, TP53, HSP90B1, TRPS1, TWIST1, UBE2I, UBE3A, SUMO1, USF1, USF2, VCP, VDAC1, VDAC2, VDR, VEGFA, XPO1, XRC5, YY1, YWHAH, YWHAZ, ZBTB16, PRDM2, TUBA1A, CXCXR4, SHFM1, FOSS1, HIST1H4E, PIAS1, NCOA1, E1F3F, RIPK2, HDAC3, KAT2B, SLC9A3R1, NCOA1, HDAC4, MED24, NR113, RBXI, HDAC6, HDAC5, NR1I3, STUB1, TUBA1B, HAX1, NCOA2, CTCF, YWHAQ, COPS5, NCOA6, TAB2, SIRT1, GFG20, TBK1, NOX4, WNT16, SOX6, CAND1, UGTT1, UBQLN4, SQRLDL, WNK4, TUBB

ECM regulators and secreted factors’. The results suggested that these three pathways lie in the central loci of the crosstalk network, and that abnormalities in the extracellular matrix plays an important role in the development of OS.

Among this network, the minimum value was from the pathway of ‘metabolism of steroid hormones’ (including CYP17A1, CYP19A1, SRD5A2, HSD11B1, HSD17B2, AKR1B1, LH8 and STARD3NL). Moreover, this pathway had only one connection with the ‘ovarian steroidogenesis’ pathway, while the other nodes have more edges with other pathway nodes. These findings suggest that steroid hormones affect the formation of OS in an indirect manner.

Network topological characteristics of OSgset. Analyzing the topological properties of nodes and interactions between nodes via protein-protein interaction-based analysis can help to reveal any key biologically related mechanisms associated with the network. The interactome constructed in this present study comprised 3,943 genes/proteins and 7,976 interactions, with a mean degree of 8.23. As specified, 259 of the 294 genes were included in the analysis and accounted for 88.1% of the genes in OSgset, serving as seeds (Table II).

OS-specific molecular network inference. An OS-specific minimal subnetwork was extracted. As shown in Fig. 3, this extracted minimal subnetwork shares 631 nodes and 2,581 edges with the 259 seeds. The average shortest path distance was 3.12, which was significantly smaller than the mean shortest-path distance (5.13; empirical P<0.01). Similarly, the average clustering coefficient of the OS specific network was significantly higher than that of the random networks (0.052 vs. 0.013; empirical P<0.001). These results indicate that the extracted OS-specific network is a non-random network. Therefore, it is proposed that the remaining 372 genes (listed in Table II) are un-identified genes that may be in involved in the OS process.

The functional modules in the minimal subnetwork were also investigated, and two modules were identified. The first shares 219 genes (listed in the Table II). For this module, the first three annotations were for the positive regulation of cellular metabolic processes (P=1.69x10^{-70}), the positive regulation of metabolic processes (P=6.23x10^{-45}) and the positive regulation of secretion from the RNA polymerase II promoter (P=1.51x10^{-45}). The second consisted of 9 genes (AGTR2, GIP, GIPR, HTR2B, SNTA1, TIMP3, NMU, ERBB21P and LNX1). This third gene annotations were for the regulation of secretion (P=5.43x10^{-4}), the positive regulation of secretion (P=6.15x10^{-4}) and regulation of transport (P=9.79x10^{-4}), which indicated that the modules functions in the regulation of secretion.

Discussion

In this study, genes associated with the development of OS were collected and their biochemical pathway functions enriched. Using this, a gene-interacting network was constructed. In doing this, a gene-interacting framework was developed and >300 potential OS-related genes were identified for further research.

Identifying the biochemical pathways that the OS-related genes are involved in will provide an insight to promote the understanding of OS. In the 58 pathways identified, the top three pathways are involved in extracellular matrix metabolism, the cytokine and cytokine receptor network, and in Wnt signaling. Further to this, it was found that all 58 pathways are involved in cross-talk and constitute a module. These findings suggest that pathways function in a concerted manner and not independently, which is supported by other studies (22-24). For instance, Zhang et al (25) reported that lipopolysaccharide activates the mitogen-activated protein kinase/Wnt/NF-κB pathway in cultured chondrocytes, and subsequently induces apoptosis. This process can be rescued by treatment with calcitonin.

When the networks with cross-talk were arranged based on the closeness centrality of every pathway, it was found that the maximum closeness value was from the ‘ensemble of genes encoding extracellular matrix and extracellular
matrix-associated proteins’ pathway. Thus, it is suggested that the aforementioned pathway lies in the central locus of the network. On the contrary, the minimum value was from the ‘metabolism of steroid hormones’ pathway. Thus, it was proposed that abnormalities in the extracellular matrix plays an important role in the development of OS and that steroid hormones affect OS development in an indirect manner.

In the 58 pathways identified, the pathway enriched as ‘breast cancer’ is of interest as it shares 22 genes with the OSgset. It has been noticed that survivors of breast cancer often suffer from OS (26). One explanation for this is that many survivors of breast cancer experience a loss of ovarian function and a drop in estrogen levels due to chemotherapy or surgery, which promotes the development of OS (27). Another explanation is that breast cancer cells can secrete γ-secretase, cyclooxygenase-2 and interleukin-8, promoting the process of osteoclastogenesis (28,29). Based on the results presented here, it is suggested that OS and breast cancer may share a similar genetic background.

Notably, >300 potential genes were found in the OS-specific network presented here. It is predicted that these genes are potential candidates for further research into the molecular mechanism of OS. For instance, the receptor tyrosine-protein kinase ERBB2 interacting protein (ERBB2IP) binds to unphosphorylated ERBB2 protein and regulates ERBB2 function and localization (30). In addition, it can also disrupt the Ras-Raf interaction and affect the Ras signaling pathway (31). As the Ras-Raf signaling pathway participates in bone metabolism (32), it is possible that ERBB2IP is involved in the development of OS. Thus, the extracted network outlines
a portrait of OS, and additionally identified a number of genes for future study.

A module is defined as having more connections than the average number of connections across the entire network, which indicates that the genes involved in a module share the same biological function. In this study, two modules were extracted from the OS-specific network. The first consisted of 219 genes directly involved in bone metabolism. However, the second module, which includes nine genes and is annotated as ‘regulation of secretion’, is more attractive. Among these genes, neuromedin U (NMU), gastric inhibitory polypeptide (GIP), GIPR, AGTR2 and HTR2B have been confirmed to be involved in the process of OS. It has been reported that GIP (a gut hormone) promotes bone anabolism as an entero-osseous hormone by stimulating osteoblast differentiation and increasing osteoblast longevity (33). In addition, GIP attenuates the activity of osteoclastic cells, leading to a net increase in bone deposition and ultimately increasing bone mass (33). NMU is an anorexigenic neuropeptide that acts independently of leptin through poorly defined mechanisms. Physiological and cell-based assays indicate that NMU acts in the central nervous system, rather than directly on bone cells, to regulate bone remodeling (34). It has also been reported that NMU promotes small intestinal transit, and that NMU deficiency results in a lowered intestinal motility rate and diminishes the effect of serotonin-induced defecation and diarrhea (35).

Based on these facts, a ‘neuro-entero-osseous’ model is proposed: The proteins α1-syntrophin, 5-TH2B and ANGII affect the activity of neurons in the gastro-interstitial tissue, alter the secretion of GIP, and influence bone metabolism. Due to a lack of experimental and clinical data supporting the role of other genes in this extracted module, a detailed mode of action cannot yet be given. Therefore, more experimental studies are required in the future.

It is difficult to diagnose or treat OS based on the analysis of all these potential genes. However, it is hypothesized that further studies examining these genes would be of valuable. It is further hypothesized that these genes may serve as a diagnostic test for OS and as drug targets in the future.

At least two studies have applied bioinformatic and computational methods to study OS. Qin et al (11) investigated OS-associated SNPs and genes identified by GWAS, and found a number of SNPs that may influence the binding affinity of transcription factors (NFATC2, MEF2C, SOX9, RUNX2, ESR2, FOXA1 and STAT3) and miRNAs. In 2018, Sheng et al (36) proposed a computational workflow to curate and evaluate OS related genes, and they emphasized the important role of OS-related genes, including TGFBI, IL6, IL1B, TNF, ESR2, IGF1, HIF1A, COL1A1 and IFENG. In comparison with these studies, the work presented here focuses on the relationship between known genes and suggests a number of unknown genes that may be involved in the process of OS.

In summary, this study revealed the pathways of all currently known genes in OS and the cross-talk between these pathways. Based on this analysis, a mechanism of OS for proposed. Additionally, in excess of 300 genes not currently associated with the development of OS have been identified. Therefore, this study increases the understanding of the contribution of genetic factors to OS and identifies genes for further investigation.

Acknowledgements

Not applicable.

Funding

The present study was supported by grants from the National Natural Science Foundation of China (grant nos. 81872847, 81402905 and 81603086).

Availability of data and materials

The datasets generated and/or analyzed during the present study are available upon reasonable request.

Authors’ contributions

YW, LG and DL designed the study. LG and JH performed the experiments. HG and JH analyzed the data, and LG wrote the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Choksi P, Jepsen KJ and Clines GA: The challenges of diagnosing osteoporosis and the limitations of currently available tools. Clin Diabetes Endocrinol 4: 12. 2018.
2. Coughlan T and Dockery F: Osteoporosis and fracture risk in older people. Clin Med (Lond) 14: 187-191. 2014.
3. Jackson RD and Mysiw WJ: Insights into the epidemiology of postmenopausal osteoporosis: The Women's Health Initiative. Semin Reprod Med 32: 454-462. 2014.
4. Suzuki T, Shimoda T, Takahashi N, Tsutsurni K, Samukawa M, Yoshimura S and Ogawara K: Factors Affecting Bone Mineral Density Among Snowy Region Residents in Japan: Analysis Using Multiple Linear Regression and Bayesian Network Model. Interact J Med Res 7: e10, 2018.
5. Del Real A, Riancho-Zarrabeitia L, Lopez-Delgado L, and Riancho JA: Epigenetics of Skeletal Diseases. Curr Osteoporos Rep 16: 246-255, 2018.
6. Liu CT, Karasik D, Zhou Y, Hsu YH, Genant HK, Broe KE, Lang TF, Samelson EJ, Demissie S, Bouxsein ML, et al: Heritability of prevalent vertebral fracture and volumetric bone mineral density and geometry at the lumbar spine in three generations of the Framingham study. J Bone Miner Res 27: 954-958, 2012.
7. Liu YZ, Liu YJ, Recker RR, and Deng HW: Molecular studies of identification of genes for osteoporosis: The 2002 update. J Endocrinol 173: 147-196, 2003.
8. Liu YJ, Shen H, Xiao P, Xiong DH, Li LH, Recker RR and Deng HW: Molecular genetic studies of gene identification for osteoporosis: A 2004 update. J Bone Miner Res 21: 1511-1535, 2006.
9. Xu XH, Dong SS, Guo Y, Yang TL, Lei SF, Papasian CJ, Zhao M, and Deng HW: Molecular genetic studies of gene identification for osteoporosis: The 2009 update. Endocr Rev 31: 447-505, 2010.

10. Liang X, Du Y, Wen Y, Liu L, Li P, Zhao Y, Ding M, Cheng B, Cheng S, Ma M, et al: Assessing the genetic correlations between blood plasma proteins and osteoporosis: A polygenic risk score analysis. Calcif Tissue Int 104: 171-181, 2019.

11. Qin L, Liu Y, Wang Y, Wu G, Chen J, Ye W, Yang J and Huang Q: Computational characterization of osteoporosis associated SNPs and genes identified by genome-wide association studies. PLoS One 11: e0150070, 2016.

12. Hu Y, Pan Z, Hu Y, Zhang L and Wang J: Network and pathway-based analyses of genes associated with parkinson’s disease. Mol Neurobiol 54: 4452-4465, 2017.

13. Hu YS, Jin X, Hu Y, Zhang L and Wang J: Analyzing the genes related to Alzheimer’s disease via a network and pathway-based approach. Alzheimers Res Ther 9: 29, 2017.

14. Zhou Y, Zhou B, Pache l, chang M, Khodabakhshi a H, Zhou G and Xia J: OmicsNet: A web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res 46: w514-w522, 2018.

15. König R, Chiang C Y, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y and Chanda SK: A probability-based approach for the analysis of large-scale RNAi screens. Nat Methods 4: 847-849, 2007.

16. Benjamini and Hochberg: controlling the False discovery rate: A practical and Powerful approach to Multiple Testing. J STaT Soc a STaT Series B (Methodological) 57: 289-300, 1995.

17. Shuaichen L and Guangyi W: Bioinformatic analysis reveals cYP2C9 as a potential prognostic marker for Hcc and liver cancer cell lines suitable for its mechanism study. Cell Mol Biol (Noisy-le-grand) 64: 70-74, 2018.

18. Zhou G and Xia J: OmicsNet: A web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res 46: w514-w522, 2018.

19. Erdos P and Rényi A: On the evolution of random graphs. Publ Math Inst Hungar Acad Sci 5: 17-61, 1960.

20. Sugiyama M, Ghisu ME, Llinares-López F and Borgwardt K: Graphkernels: R and Python packages for graph comparison. Bioinformatics 34: 530-532, 2018.

21. Rosvall M and Bergstrom CT: Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105:1118-1123, 2008.

22. Mansoori Mn, Shukla P and Singh d: Combination of PTH (1-34) with anti-IL17 prevents bone loss by inhibiting IL-17/N-cadherin mediated disruption of PTHR1/LRP-6 interaction. Bone 105: 226-236, 2017.

23. Dar HY, Azam Z, Anupam R, Mondal RK and Srivastava RK: Osteoimmunology: The Nexus between bone and immune system. Front Biosci (Landmark Ed) 23: 464-492, 2018.

24. Ke K, Sul OJ, Chung SW, Suh JH and Choi HS: Lack of NOD2 attenuates ovariectomy-induced bone loss via inhibition of osteoclasts. J Endocrinol 235: 85-96, 2017.

25. Zhang LB, Man ZT, Li W, Zhang W, Wang XQ and Sun S: Calcitonin protects chondrocytes from lipopolysaccharide-induced apoptosis and inflammatory response through MAPK/Wnt/NF-kappaB pathways. Mol Immunol 87: 249-257, 2017.

26. Bruyère O, Bergmann P, Cavaleri E, Gießen E, Goemaere S, Kaufman MJ, Rozenberg S and Body JF: Skeletal health in breast cancer survivors. Maturitas 105: 78-82, 2017.

27. Taxel P, Faircloth E, Idrees S and Van Poznak C: Cancer Treatment-Induced Bone Loss in Women With Breast Cancer and Men With Prostate Cancer. J Endocr Soc 2: 574-588, 2018.

28. Fong JE, Le Nhuovan D and Komarova SV: Tumor-supportive and osteoclastogenic changes induced by breast cancer-derived factors are reversed by inhibition of (gamma)-secretase. J Biol Chem 285: 31427-31434, 2010.

29. Singh B, Berry JA, Vincent LE and Lucci A: Involvement of IL-8 in COX-2-mediated bone metastases from breast cancer. J Surg Res 134: 44-51, 2006.

30. Mei L and Borg JP: ERBB2 oncogenicity: ERBBIN helps to perform the job. Mol Cell Oncol 2: e995033, 2015.

31. Dai P, Xiong WC and Mei L: Erbin inhibits RAF activation by disrupting the sur-8-Ras-Raf complex. J Biol Chem 281: 927-933, 2006.

32. Reel B, Korkmaz CG, Arun MZ, Yildirim G, Ogut D, Kaymak A, Miceli SC and Ergur BU: The Regulation of Matrix Metalloproteinase Expression and the Role of Discoidin Domain Receptor 1/2 Signalling in Zoledronate-treated PC3 Cells. J Cancer 6: 1020-1029, 2015.

33. Ramsey W and Isales CM: Intestinal Incretins and the Regulation of Bone Physiology. Adv Exp Med Biol 1033: 13-33, 2017.

34. Sato S, Hanada R, Kimura A, Abe T, Matsumoto T, Iwasaki M, Inose H, Ida T, Mieda M, Takeuchi Y, et al: Central control of bone remodeling by neuregulin 1. Nat Med 13: 1234-1240, 2007.

35. Nakashima Y, Ida T, Sato T, Nakamura Y, Takahashi T, Mori K, Miyazato M, Kangawa K, Kusukawa J and Kojima M: Neuregulin 1 is necessary for normal gastrointestinal motility and is regulated by serotonin. Ann N Y Acad Sci 1200: 104-110, 2010.

36. Sheng Y, Tang J, Ren K, Manor LC and Cao H: Integrative computational approach to evaluate risk genes for postmeno-pausal osteoporosis. IET Syst Biol 12: 118-122, 2018.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.