ORIGINAL ARTICLE

Pharmacophore-based screening of differentially-expressed *PGF*, *DDIT4*, *COMP* and *CHI3L1* from hMSC cell lines reveals five novel therapeutic compounds for primary osteoporosis

CrossMark

Catherine Jessica Lai [*]

Brearley School, New York, NY 10021, USA

Abstract   As many societies age, primary osteoporosis (PO) is increasingly a major health problem. Current drug treatments such as alendronate and risedronate have known side effects. We took an agnostic empirical approach to find PO therapeutic compounds. We examined 13,548,960 probe data-points from mesenchymal stromal cell (hMSC) lines and found that *PGF*, *DDIT4*, and *COMP* to be up-regulated, and *CHI3L1*, down-regulated. We then identified their druggable domains. For the up-regulated differentially-expressed genes, we used protein–protein interactions to find residue clusters as binding surfaces. We then employed pharmacophore models to screen 15,407,096 conformations of 22,723,923 compounds, which identified (6R,9R)-6-(2-furyl)-9-(1H-indol-3-yl)-2-(trifluoromethyl)-5,6,7, 9-tetrahydro-4H[1,2,4]triazolo[5,1],(2S)-N1-[2-[2-(methylamino)-2-oxo-ethyl]phenyl]-N2-phenylpyrrolidine-1,2-dicarboxamide, and 2-furyl-(1H-indol-3-yl)-methyl-BLAHone as candidate compounds. For the down-regulated *CHI3L1*, we relied on genome-wide disease signatures to identify (11alpha)-9-fluoro-11,17,21-trihydroxypregn-4-ene-3,20-dione and Genistein as candidate compounds. Our approach differs from previous research as we did not confine our drug targets to hypothesized compounds in the existing literature. Instead, we allowed the full expression profile of PO cell lines to reveal the most desirable targets. Second, our differential gene analysis revealed both up- and down-regulated genes, in contrast to the literature, which has focused on inhibiting only up-regulated genes. Third, our virtual screening universe of 22,723,923 compounds was more than 100 times larger than those in the known literature.

1. Introduction

As many societies age, primary osteoporosis (PO) is increasingly a major health problem. In the U.S., PO incidence has

[*] Corresponding author.
E-mail address: jessicalai98@gmail.com.

	logFC	AveExpr	t	P.Value	adj.P.Val	B
MAB21L2	2.95	6.63	3.64	0.00	0.08	-1.86
XIST	2.85	6.21	2.44	0.03	0.16	-3.86
COL10A1	2.83	7.50	9.39	0.00	0.01	5.26
PGF	2.35	7.61	4.84	0.00	0.04	0.03
IBSP	2.27	7.92	4.57	0.00	0.05	-0.38
DDIT4	2.14	10.35	5.17	0.00	0.04	0.51
NRXN2	2.11	6.04	6.74	0.00	0.02	2.58
MTSS1L	2.08	6.84	9.90	0.00	0.01	5.68
COMP	1.98	8.13	2.05	0.07	0.21	-4.48
PPDPF	1.97	8.63	7.47	0.00	0.01	3.41
RARRES2	1.85	6.49	3.37	0.01	0.09	-2.31
TRIB3	1.84	6.91	4.96	0.00	0.04	0.21
HLA-DRA	1.76	7.62	3.04	0.01	0.11	-2.86
STAT4	1.74	8.62	3.02	0.01	0.11	-2.90
ARHGDIA	1.73	7.54	4.68	0.00	0.05	-0.22
EGR2	1.72	7.21	3.52	0.01	0.08	-2.06
IGFBP2	1.69	9.42	2.37	0.04	0.17	-3.97
SLC27A1	1.68	8.23	5.23	0.00	0.03	0.60
LSP1	1.67	7.14	3.85	0.00	0.07	-1.52
FNDC1	1.65	9.86	1.73	0.11	0.27	-4.96

Figure 1   Up-regulated DEGs using Mendel.

steadily increased from 75 per 100,000 women in the 1950 to 150 per 100,000 women by the 1990s [3]. The most common drug treatments are generic bisphosphonates such as alendronate and risedronate, due to their low cost [15]. However, they have well-known upper gastrointestinal side effects and are associated with atypical fractures of the femur and aseptic necrosis of the mandible [15]. Recently, a number of researchers have sought to identify new therapeutic drugs for PO. For example, Yasuda et al. [20] focused on cathepsins S and K, which have selective expression in the extracellular matrix (ECM). Feder et al. [6] screened a 500-compound library for purple acid phosphatase inhibitors because elevated phosphatases are correlated with osteoporosis. More recently, Vuorinen et al. [19] screened 202,906 compounds for 17B-hydroxysteroid dehydrogenase 2 inhibitors, under the assumption that these inhibitors catalyze the inactivation of estradiol into estrone.

We took a different, novel approach. First, we did not assume or confine our drug targets to hypothesized compounds in the existing literature. Instead, we allowed the full expression profile of PO cell lines to reveal the most desirable targets, subject of course to druggability conditions. Second, our differential gene analysis of the cell lines revealed both up- and down-regulated genes, in sharp contrast to the existing literature, which has focused on inhibiting only up-regulated genes. Third, we used the new ZINC database of compounds [7]. With 15,407,096 conformations of 22,723,923 compounds, this universe is more than 100 times larger than that in Vuorinen et al. [19].

In the Section 2, we describe how we used human mesenchymal stromal cell (hMSC) cell lines for differential expression analysis and used virtual-screening on the ZINC database [7]. In the Section 3, we report four differentially-expressed

genes (DEG): the up-regulated *PGF*, *DDIT4*, and *COMP* and the down-regulated *CHI3L1*. This consideration of a down-regulated gene is a departure from the literature, which has focused on only up-regulated genes. Finally, we report the identity of 5 potent candidate compounds under stringent druggability conditions. We conclude with thoughts on some remaining limitations and suggest further research directions.
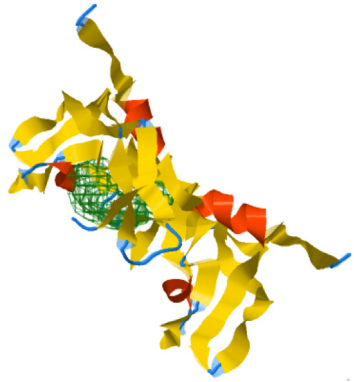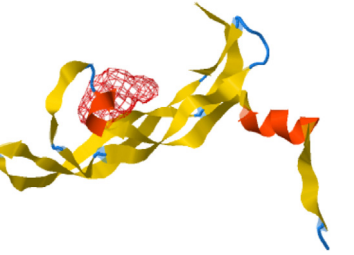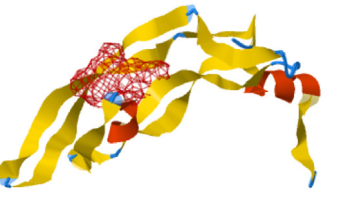
2. Materials and methods

We obtained probe data of hMSC cell lines from the GSE35936 in the NCBI library [1]. The dataset consisted of 1,354,896 million probes for 54,675 genes in each of 10 cell lines, of which 5 are age-matched controls. Data were produced using the Affymetrix U133 Plus 2.0 Array platform.

We undertook differential gene analysis using Mendel [11], which we developed employing the R language. We then preprocessed the probe data with Robust Multichip Average (RMA) analysis. The DEGs were annotated with gene ontology, diseased ontology, and KEGG pathways. All this information allowed us to identify genes that are significantly up- or down-regulated, ranked by log fold changes.

The path from significant DEGs to therapeutic compounds required much filtering. Furthermore, the process for up-regulated DEGs had to be different than that for down-regulated ones. While any ligand docked into an up-regulated DEG domain could be treated as an inhibitor [9], the same could not be said for a down-regulated DEG. Instead, we have to explicitly search for agonists for the latter.
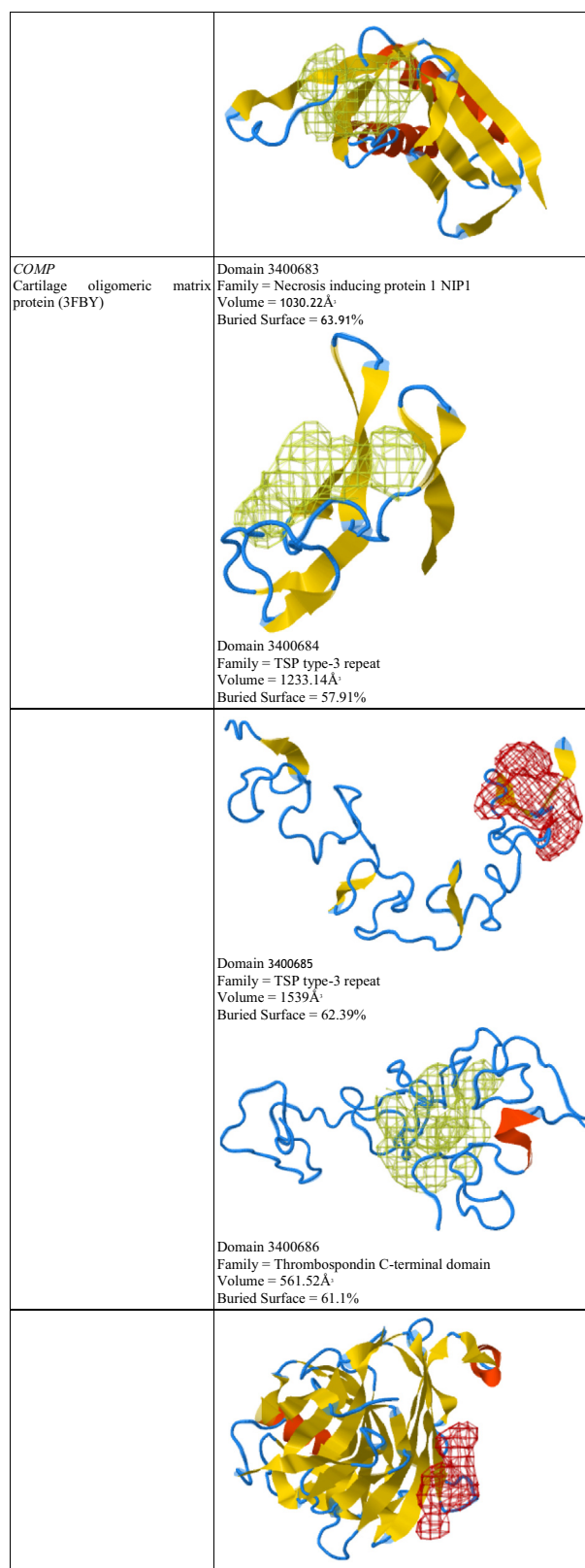
For up-regulated genes, we first used EBI's structured-based engine to identify druggable protein domains, which we in turn used to identify clusters of anchor residues over protein–protein interaction (PPI) surfaces. This was done with

Gene	Druggable Domains (druggable site shown in green mesh)
PGF Human placenta growth factor-1 (1FZV)	Domain 4190116 Family = Platelet-derived growth factor-like (Homo Multimer) Volume = 1714.5Å³ Buried Surface = 66.48%
	Domain 60159 Family = Platelet-derived growth factor-like Volume = 524.39Å³ Buried Surface = 59.63%
	Domain 60160 Family = Platelet-derived growth factor-like Volume = 520.17Å³ Buried Surface = 61.5%
DDIT4 DNA-damage-inducible transcript 4 (3LQ9)	Domain 3530749 Family = PFAM RTP801_C Volume = 2708.86Å³ Buried Surface = 63.57%

Figure 2 Druggable domains of up-regulated DEGs.

|	
COMP Cartilage oligomeric matrix protein (3FBY)	Domain 3400683 Family = Necrosis inducing protein 1 NIP1 Volume = 1030.22Å³ Buried Surface = 63.91%
	Domain 3400684 Family = TSP type-3 repeat Volume = 1233.14Å³ Buried Surface = 57.91%
	Domain 3400685 Family = TSP type-3 repeat Volume = 1539Å³ Buried Surface = 62.39%
	Domain 3400686 Family = Thrombospondin C-terminal domain Volume = 561.52Å³ Buried Surface = 61.1%

Figure 2 (*continued*)

PocketQuery [9] which ranked the clusters with a composite score of solvent-accessible surface area (SASA), free energy (G^{FC}), and sequence conservation, The highest-ranked cluster was used to build pharmacophore models using Pharmer, a high-performance search engine based on geometric hashing, generalized Hough transforms, and Bloom fingerprints [8]. The pharmacophore was then used for virtual screening of over 22 million compounds using ZINCPharmer [10]. We used the stringent criteria of a maximum 1 hit per conformation, maximum 1 hit per molecule, a maximum RMSD of 0.01, and a maximum of 4 rotatable bonds. Finally, each hit was characterized with a set of absorption, distribution, metabolism, excretion, and toxicity (ADMET) rules [17].

Our final step was to undertake molecular simulation to dock the ligands to the protein receptors. To prepare receptors, we use Chimera [16] to remove the protein's original ligand, delete its solvent, replace incomplete side chains with Dunbrack rotamers, add hydrogen and OXT atoms to missing C-termini (with protonation states for histidine), assign partial
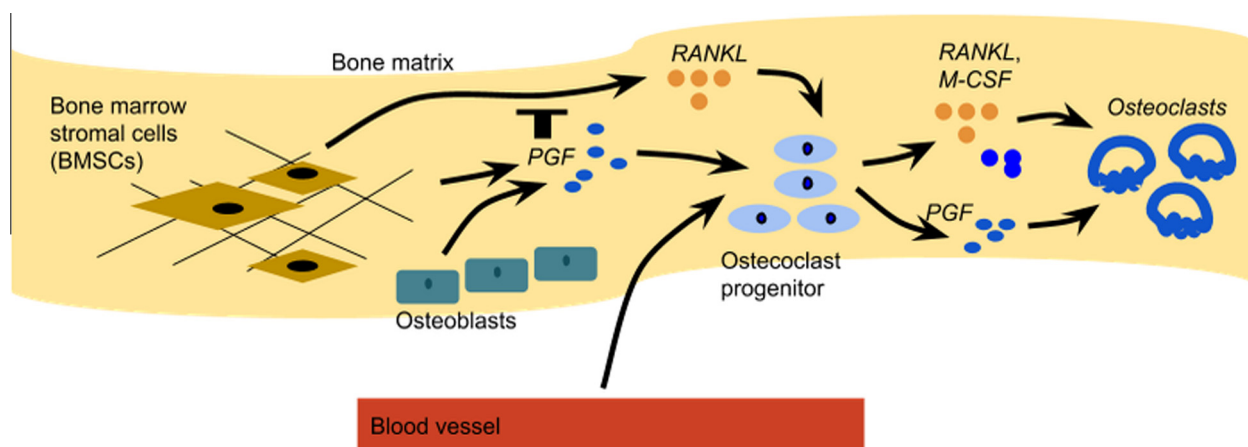
Figure 3   Pathway for *PGF* effect on osteoclastogenesis.

charges from the AMBER ff145B force field with non-monatomic ion residues using semi-empirical (AMI) with bond charge correction (BCC).
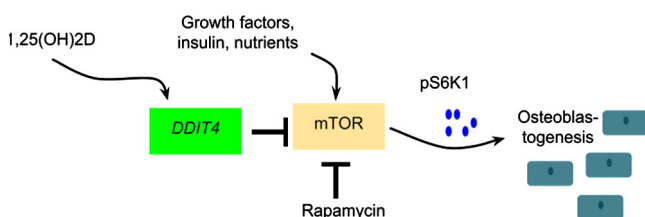
For down-regulated genes, we again confined our attention to those with druggable domains. However, we were not screening for inhibitors, but looking for agonists. We used the functional connections between genetic perturbation and drug action to identify bioactive small molecules [12]. We subject each molecule to the same ADMET rules as for up-regulated DEGs, and dock them with ligands in the same manner.

3. Results and discussion

Fig. 1 reports the up- and down-regulated DEGs from Mendel. LogFC is the value of the contrast in log$_2$ fold-change. AveExpr is the average log$_2$ expression for that gene across all arrays and channels. t and P.Value are the usual statistics, and adj.P.Val adjusts for multiple testing with the Benjamin–Hochberg method to control for false discovery rate. The B is the long-odds statistic, measuring the odds that the gene is differentially expressed.

The most significant up-regulated DEG was *MAB21L2*, on chromosome chr4:151,504,181–151,505,261, that is a repressor of bone morphogenetic protein (BMP)-induced transcription.). However, as Fig. 2 shows, only three up-regulated DEGs have druggable domains: *PGF* (placental growth factor), *DDIT4* (DNA-damage-inducible transcript 4), and *COMP* (cartilage oligomeric matrix protein).

Fig. 3 shows how *PGF* mediates osteoclastogenesis [2]. *PGF* secreted by osteogenic BMSCs stimulates *RANKL* (receptor activator of nuclear factor kappa-B ligand) expression by these

same BMSCs, then working with *RANKL*, *PGF* also mediates the differentiation of osteoclast progenitor cells. *PGF* also signals directly on these precursors in a positive feedback loop, so that its pleiotropic actions lead to bone resorption.

DDIT4 is an inhibitor in mTOR signaling in osteoblasts; see Fig. 4. mTOR is a member of the phosphoinositol kinase family and a regulator of 1,25(OH)2D (vitamin D) action in
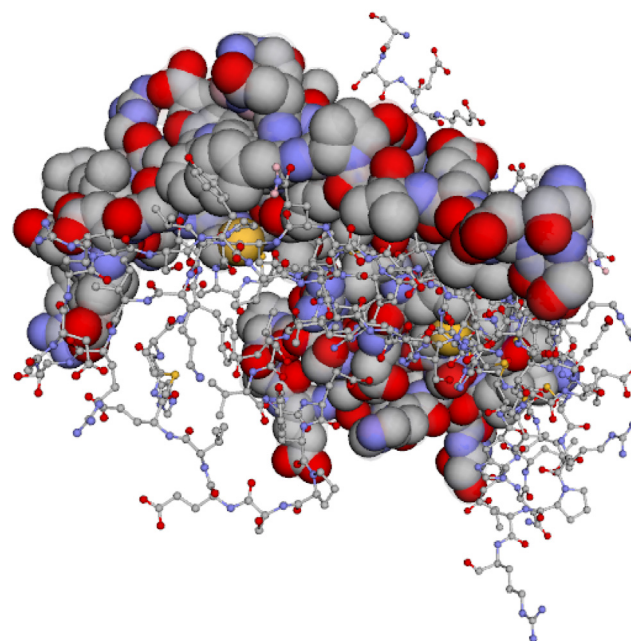
(a) - Hot region of *PGF* (human placenta growth factor-1) defined by residue clusters, at PPI interfaces.

| Residues	Number	ΔG^{FC}	ΔΔG^a	ΔSASA	ΔSASA%	SASA	Conserv.	Evol.Rate
VAL	23	-2.28	1.67	69.72	61	35.19	0.63	0.27
PHE	26	-3.99	3.09	103.29	62.9	36.26	0.9	0.4
VAL	29	-3.28	1.37	69.3	60.7	3.36	0.92	0.72

(b) Residues in hot region of *PGF* (human placenta growth factor-1) defined by residue clusters, at PPI interfaces

Figure 5   Hot region of *PGF* (human placenta growth factor-1) defined by residue clusters, at PPI interfaces.

Figure 4   Pathway for *DDIT4* effect on osteoblast regulation.

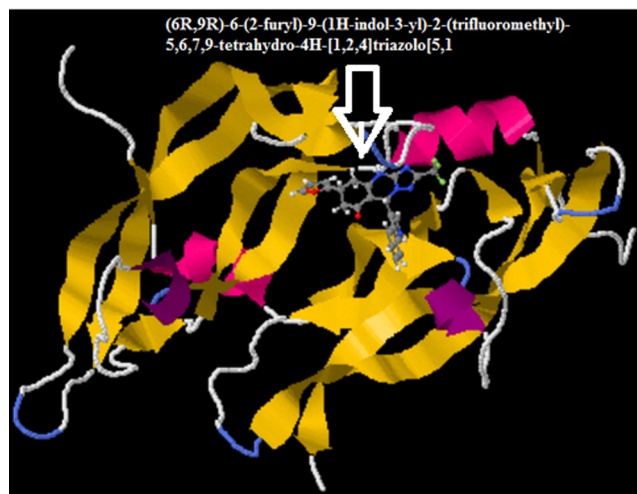| Pharmacophore Class	x	y	z	Radius
HydrogenDonor	9.08	63.66	-11.95	0.5
Hydrophobic	7.72	60.01	-15.38	1
Hydrophobic	9.4	59.59	-9.04	1

(a) Pharmacophore Classes
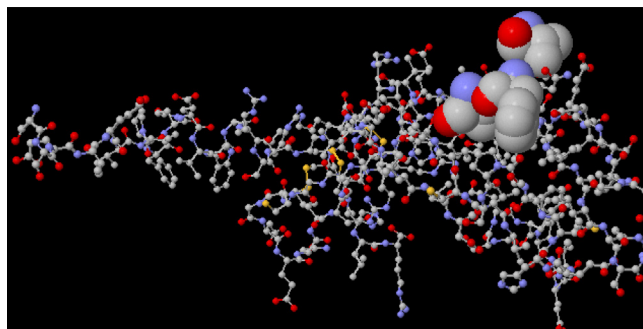
(b) Model with anchors shown as space-filled residues and receptors shown as ball-and-stick residues

Figure 6 Pharmacophore model of *PGF* (human placenta growth factor-1).

(6R,9R)-6-(2-furyl)-9-(1H-indol-3-yl)-2-(trifluoromethyl)-5,6,7,9-tetrahydro-4H-[1,2,4]triazolo[5,1]

Figure 8 *PGF* docked with (2S)-N1-[2-[2-(methylamino)-2-oxo-ethyl]phenyl]-N2-phenyl-pyrrolidine-1,2-dicarboxamide.

bones, via *pS6K1*. *DDIT4* is also an inhibitor of *NFTc1*, a master transcription factor for osteoclastogenesis, via eukaryotic translation initiation factor 2 alpha (eIF2alpha) signaling [18].

COMP expresses a noncollagenous glycoprotein that binds to type I, II, and IX collagen fibers. It has a domain at the N-terminus, a globular domain at the C-terminus, and 4 epidermal growth factor-like (EGF-like) domains, with 8

Compound	Structure	ADMET Results
(6R,9R)-6-(2-furyl)-9-(1H-indol-3-yl)-2-(trifluoromethyl)-5,6,7,9-tetrahydro-4H-[1,2,4]triazolo[5,1] 13568028 Mol. mass = 439 pIC50 = 0.96 (-log M)		Lipinski's rule of five: yes Bioavailability: yes Ghose filter: yes Lead likeness: no Muegge filter: yes Veber filter: yes
(2S)-N1-[2-[2-(methylamino)-2-oxo-ethyl]phenyl]-N2-phenyl-pyrrolidine-1,2-dicarboxamide 58325366 Mol. mass = 380 pIC50 = 3.2 (-log M)		Lipinski's rule of five: yes Bioavailability: yes Ghose filter: yes Lead likeness: yes Muegge filter: yes Veber filter: yes
2-furyl-(1H-indol-3-yl)-methyl-BLAHone 04945118 Mol. mass = 386 pIC50 = 0.96 (-log M)		Lipinski's rule of five: yes Bioavailability: yes Ghose filter: yes Lead likeness: no Muegge filter: yes Veber filter: yes

Figure 7 Candidate compounds for up-regulated DEGs.

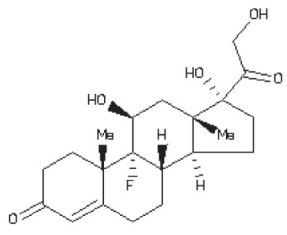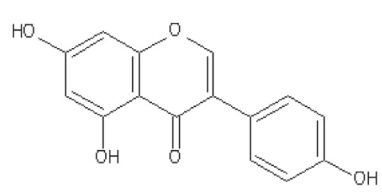|	logFC	AveExpr	t	P.Value	adj.P.Val	B
CXCL6	-3.68	7.89	-3.08	0.01	0.11	-2.80
CMPK2	-3.67	6.95	-4.21	0.00	0.06	-0.94
RGS4	-3.67	8.65	-3.37	0.01	0.09	-2.32
RPS4Y1	-3.65	7.09	-2.55	0.03	0.15	-3.68
CHI3L1	-3.52	9.21	-3.56	0.00	0.08	-2.00
EIF1AY	-3.24	4.16	-2.77	0.02	0.13	-3.32
SCN3A	-2.91	4.53	-2.91	0.01	0.12	-3.08
IFIT1	-2.80	8.95	-3.23	0.01	0.10	-2.55
MX1	-2.78	8.66	-2.95	0.01	0.12	-3.01
MEOX2	-2.74	4.53	-8.40	0.00	0.01	4.36
HAPLN1	-2.71	9.21	-2.30	0.04	0.17	-4.08
RSAD2	-2.64	5.93	-2.48	0.03	0.15	-3.79
CXCL1	-2.63	8.07	-3.58	0.00	0.08	-1.97
THBD	-2.60	5.99	-4.60	0.00	0.05	-0.34
FGD4	-2.57	5.92	-3.31	0.01	0.10	-2.41
TMEM200A	-2.52	7.10	-3.03	0.01	0.11	-2.88
ZNF367	-2.51	5.45	-4.14	0.00	0.06	-1.05
GULP1	-2.48	6.62	-3.19	0.01	0.10	-2.62
IFI27	-2.43	9.04	-2.73	0.02	0.13	-3.39
ZWINT	-2.42	6.95	-3.27	0.01	0.10	-2.48

Figure 9   Down-regulated DEGs using Mendel.

thrombospondin type 3 repeats (Fig. 2). *COMP* stabilizes the collagen fiber network in articular cartilage, tendons, menisci, and synovial tissue. Up-regulation of *COMP*, however, decreases the viability of chondrocytes, which could be a pathway for PO [5].

Next, we report the results of using PPI to discover regions of cluster residues for binding to the three druggable up-regulated *PGF*, *DDIT4*, and *COMP*. Fig. 5 illustrates the region for *PGF*. The residue is shown in space-filled elements and the ligand protein shown with ball-and-sticks. Fig. 6 reports the 3 residues in the cluster model uses the residue cluster at channel A, with 3 residues. G^{FC} is Gibbs free energy and SASA is solvent accessible surface area. Conserv is the conservation score calculated from the EBI Scorecons server. An alternative measure of conservation is Evol.Rate, the evolutionary rate computed from Tel Aviv University's Rate4Site.

With the cluster regions, we successfully built pharmacophore models. Fig. 6 illustrates one such model for *PGF*.

Compound	Structure	ADMET Results
(11alpha)-9-fluoro-11,17,21-trihydroxypregn-4-ene-3,20-dione 4097304 Mol. mass = 380 pIC50 = 1.33457 (-log M)		Lipinski's rule of five: yes Bioavailability: yes Ghose filter: yes Lead likeness: yes Muegge filter: yes Veber filter: yes
Genistein 18825330 Mol. mass = 270 pIC50 = 4.44 (-log M)		Lipinski's rule of five: yes Bioavailability: yes Ghose filter: yes Lead likeness: yes Muegge filter: yes Veber filter: yes

Figure 10   Candidate compounds for down-regulated *CHI3L1*.
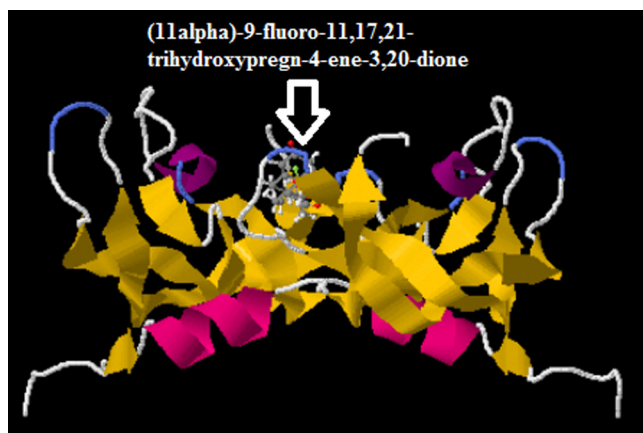
Figure 11 *PGCHI3L1 F* docked with (11alpha)-9-fluoro-11,17,21-trihydroxypregn-4-ene-3,20-dione.

Finally, Fig. 7 shows the candidate compounds based on virtual screening using the pharmacophore models. The screen was able to identify compounds only for *PGF*. We also report a number of ADMET screening results. Although not all compounds pass all ADMET screens, all compounds pass most of the screens.

Of the three compounds, (2S)-N1-[2-[2-(methylamino)-2-oxo-ethyl]phenyl]-N2-phenyl-pyrrolidine-1,2-dicarboxamide has the highest pIC50 (3.2), and therefore can be considered the most potent inhibitor of the up-regulated *PGF*.

Finally, we were able to dock all three compounds with *PGF*. Fig. 8 shows the docking using (2S)-N1-[2-[2-(methylamino)-2-oxo-ethyl]phenyl]-N2-phenyl-pyrrolidine-1,2-dicarboxamide. This has a full fitness of -1167.75 kcal/mol and an estimated ΔG at -7.47 kcal/mol.

We now turn from up-regulated to down-regulated DEGs, shown in Fig. 9. It turns out that only *CHI3L1* (chitinase 3-like 1) is druggable. The role of *CHI3L1* in osteoporosis is still uncertain. On the one hand, *CHI3L1* promotes the proliferation of connective tissue. Its silencing with siRNA is known to lower bone resorption while its transfection decreases the osteoclast pro-differentiative marker *MMP-9* [4]. In this way, PO might be associated actually with *CHI3LI* up-regulation. On the other hand, *CHI3LI* facilitates bacterial adhesion and invasion, especially in the epithelial cells. It does this by activating protein kinase B (AKT) phosphorylation [14]. It is the excessive production of *CHI3L1* that is pathogenic in mucosal tissues. Therefore, it is also plausible that *CHI3L1* down-regulation is associated with PO, as our probe data revealed. In short, the exact pathway in which *CHI3L1* affects PO needs to be further investigated.

The disease-gene signature map identified two agonist compounds, reported in Fig. 10. Both are potent, although Genistein is the more potent, with a pIC50 of 4.44. Genistein is already known to activate peroxisome proliferator-activated receptors (PPARs), especially gamma forms that regulate bone mass [13], so it seems quite natural for it to be repurposed for PO.

Fig. 11 illustrates the docking of (11alpha)-9-fluoro-11,17, 21-trihydroxypregn-4-ene-3,20-dione with *CHI3L1*. The full fitness is associated with -1137.60 kcal/mol, with an estimated ΔG at -7.40 kcal/mol.

4. Conclusion

Our analysis uses an agnostic approach to identify target genes and their inhibitors and activators, so that we were able to obtain a comprehensive list of five candidate therapeutic compounds for PO. Interestingly, many of our targets are consistent with the literature, even if current research has not gone as far as virtual screening of compounds. Furthermore, we could relax the filtering criteria in many ways, from the threshold log fold-change for determining significant DEGs, to the criteria for determining targets, identifying druggable domains, constructing PPI cluster interfaces, modeling pharmacophores, and finally to screening compounds.

Going forward, the candidate compounds will need biological assays and clinical trials, but the lead compounds now identified could produce a new cocktail of possible therapeutic drugs for PO.

References

[1] P. Benisch, S. Tatjana, K. Ludger, S. Frey, L. Seefried, N. Raaijmakers, M. Krug, R.M. Regensburger, S. Zeck, T. Schinke, M. Amiing, R. Ebert, F. Jakob, PLoS One 7 (2012) e45142.

[2] L. Coenegrachts, C. Maes, S. Torrekens, R. Van Looveren, M. Mazzone, T.A. Guise, R. Bouillon, J. Strassen, P. Carmeliet, G. Carmeliet, Cancer Res. 70 (2010) 6537–6547.

[3] C. Cooper, Z.A. Cole, C.R. Holroyd, S.C. Earl, N.C. Harvey, E. M. Dennison, L.J. Melton, S.R. Cummings, J.A. Kanis, Osteoporos. Int. 22 (2011) 1277–1288.

[4] M. Di Rosa, D. Tibullo, M. Vecchio, G. Nunnari, S. Saccone, F. Di Raimondo, L. Malaguarnera, Bone 61 (2014) 55–63.

[5] R. Dinser, F. Zaucke, F. Kreppel, K. Hultenby, S. Kochanek, M. Paulsson, P. Maurer, J. Clin. Invest. 110 (2002) 505–514.

[6] D. Feder, W.M. Hussein, D.J. Clayton, M.W. Kan, G. Schenk, R.P. McGeary, L.W. Guddat, Chem. Biol. Drug Des. 80 (2012) 665–674.

[7] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, J. Chem. Inf. Model. 52 (2012) 1757–1768.

[8] D.R. Koes, C.J. Camacho, J. Chem. Inf. Model. 51 (2011) 1307–1314.

[9] D.R. Koes, C.J. Camacho, Nucleic Acids Res. 40 (2012) 1–66.

[10] D.R. Koes, C.J. Camacho, Nucleic Acids Res. 40 (2012) W409–W414.

[11] C.J. Lai, J. Microbiol. Biotechnol. Res. 5 (2015) 18–30.

[12] J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, T.R. Golub, Science 313 (2006) 1929–1935.

[13] B. Lecka-Czernik, IBMS BoneKEy 7 (2010) 171–181.

[14] I.A. Lee, A. Kamba, D. Low, E. Mizoguchi, World J. Gastroenterol. 20 (2014) 1127–1138.

[15] W.F. Lems, M. den Heijer, Neth. J. Med. 71 (2013) 188–193.

[16] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.

[17] M. Swain, J. Chem. Inf. Model. 52 (2004) 613–615.

[18] N.G. Tanjung, In Vitro and In Silico Analysis of Osteoclastogenesis in Response to Inhibition of De-phosphorylation of EIF2alpha by Salubrinal and Guanabenz Doctoral dissertation, Purdue University, 2013.

[19] A. Vuorinen, R. Engeli, A. Meyer, F. Bachmann, U.J. Griesser, D. Schuster, A. Odermatt, J. Med. Chem. 57 (2014) 5995–6007.

[20] Y. Yasuda, J. Kaleta, D. Brömme, Adv. Drug Deliv. Rev. 57 (2005) 973–993.