Journal of Genetic Engineering and Biotechnology (2016) xxx, xxx–xxx

HOSTED BY

ELSEVIER

Academy of Scientific Research & Technology and
National Research Center, Egypt

Journal of Genetic Engineering and Biotechnology

www.elsevier.com/locate/jgeb

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

Catherine Jessica Lai *

Brearley School, New York, NY 10021, USA

Received 18 June 2015; accepted 24 December 2015

KEYWORDS

Bioinformatics;
Drug design;
Virtual screening;
Pharmacophore;
Docking;
Osteoporosis

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 *CH13L1*, 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.

Peer review under responsibility of National Research Center, Egypt.

http://dx.doi.org/10.1016/j.jgeb.2015.12.002

	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