Computational Drug Discovery in Chemotherapy-induced Alopecia via Text Mining and Biomedical Databases

Nanyang Zhang, PhD; Wenbing Xu, MS; Shijie Wang, MD; Yan Qiao, MS; and Xiaoxiao Zhang, MS

The Affiliated Hospital of Qingdao University, Qingdao, China

ABSTRACT

Purpose: Chemotherapy-induced alopecia (CIA) is a common and often stressful adverse effect associated with chemotherapy. CIA can cause more psychosocial pressure in patients, including effects on sexuality, self-esteem, and social relationships. We analyzed publicly available data to identify drugs formulated for topical use targeting the relevant CIA molecular pathways by using computational tools.

Methods: The genes associated with CIA were determined by text mining, and the gene ontology of the gene set was studied using the Functional Enrichment analysis tool. Protein−protein interaction network analysis was performed using the String database. Enriched gene sets belonging to the identified pathways were queried against the Drug−Gene Interaction database to find drug candidates for topical use in CIA.

Findings: Our analysis identified 427 genes common to CIA text-mining concepts. Gene enrichment analysis and protein−protein interaction analysis yielded 19 genes potentially targetable by a total of 29 drugs that could possibly be formulated for topical application.

Implications: The findings from the present analysis would give a new thought to help discover more effective agents, and present tremendous opportunities to study novel target pharmacology and facilitate drug repositioning efforts in the pharmaceutical industry. (*Clin Ther.* 2019;41:972−980) © 2019 Published by Elsevier Inc.

Key words: chemotherapy-induced alopecia, drugs, text mining.

INTRODUCTION

Hair loss, or *alopecia*, is a common phenomenon, and can be caused by numerous conditions. Some types can be associated with diseases and specific diagnoses, such as androgenic alopecia, alopecia areata, and chemotherapy-induced alopecia (CIA).[1]

As one of the side effects of chemotherapy, CIA generally occurs several weeks after the start of chemotherapy and could be aggravated over 1−2 months. Then the hair regrows after 3−6 months, because the hair follicle (HF) stem cells are not susceptible to chemotherapeutic agents, and have the ability to generate new HF stem cells to promote hair recovery.[2,3] However, some patients experience permanent CIA, and even though chemotherapy may be finished, many patients have color and texture changes in the regrown hair.[4] The effects of CIA may also include psychosocial pressure.[5]

Various chemotherapeutics can cause CIA, such as alkylating agents, cytotoxics, and topoisomerase inhibitors. However, drug type is not the only evocator. The severity of CIA is determined by the ability of chemotherapeutic agents to damage the HF stem cells, which depends on the protocol, dose, and schedule of drug administration, as well as a patient's clinical characteristics.[6] Chemotherapy could injure all rapidly dividing cells, including tumor cells and the hair matrix keratinocytes. At any time, 80%−90% of the HFs are in the anagen phase of the hair cycle, in which mitotically active matrix cells proliferate quickly. Chemotherapeutics may cause apoptosis of these cells, leading to the consequence of hair shedding.

Accepted for publication April 1, 2019
https://doi.org/10.1016/j.clinthera.2019.04.003
0149-2918/$ - see front matter

CIA is common and is a major challenge of cancer treatment.[3] In addition to remedial measures, such as using a wig or hat to hide the scalp after treatment for CIA, many actions have been taken to prevent CIA. Some patients use devices that cool the scalp in order to slow the blood flow and reduce scalp metabolism.[7] This is the most frequent method for patients to use. Patients may also use a scalp tourniquet to reduce the blood supply, aiming to decrease the concentration of chemotherapeutics in the plasma. Drugs have also been developed for CIA treatment. However, the preventive efficacy if these drugs in CIA is still questionable.[8]

The conventional process of drug discovery and development starts with physical experimentation and research on drug compounds. Development may be discontinued because of insufficient efficacy reported in Phase II or III trials. The repositioning of known drugs makes it possible to skip the processes associated with conventional drug development to investigate drug tolerability in healthy individuals, and to move on to a Phase II clinical trial, which would reduce the price and shorten the time of drug development. Following the repurposing paradigm, this study aimed to investigate new purposes for existing drugs as possible modulators of CIA by mining the available published literature combined with biological databases and other analytical tools. Text mining of biomedical literature has been established as an effective way to reveal novel associations between genes and pathologies.[9] Computational prediction of multitargeting drugs has established polypharmacology as a promising alternative approach to tackling some of the daunting complicatedness of drug discovery. Computational prediction of multitargeting drugs will help not only to discover more effective agents, but also to present tremendous opportunities for studying novel targets of pharmacology and to facilitate drug-repositioning efforts in the pharmaceutical industry.[10]

In this study, we made a preliminary list of related genes by mining the literature. Next we validated the association of the identified genes and generated a focused target set by analyzing the signal pathway. With in-depth analysis of the functional enrichments of the genes, we generated a list of high-priority target genes. Candidate drugs were then derived from an analysis of the data on potential drug–gene interactions.

MATERIALS AND METHODS

Gene Collection

The database pubmed2ensembl (http://pubmed2ensembl.ls.manchester.ac.uk was used to perform text mining. We performed queries using the search terms *alopecia, hair loss, baldness,* and *balding* to produce a list of genes. We then extracted all of the unique gene hits from each result. All of these genes were then used in the subsequent analyses.[11]

Gene Ontology and Pathway Analysis

The Functional Enrichment analysis tool (http://www.funrich.org) was used for enrichment analysis of the genes mined from the literature. The pathway analysis was performed with annotations (*chemokine activity, catalytic activity, transcription factor activity, receptor activity, growth factor, receptor binding, cytokine activity*) from the Reactome Pathway Database (https://reactome.org), which contained all of the gene terms we mined and then collected.

Protein–protein Interactions

The String database (http://string-db.org) was used to provide a key assessment of protein–protein functional associations.

Drug-gene Interactions

DGIdb database version 3.0 (www.dgidb.org) was used to search the potential targets for existing associations with drugs or small organic compounds. The results of the search were organized so that information on drug–gene interactions and gene targetability could be obtained.[12] In addition, PubChem (https://www.ncbi.nlm.nih.gov/pccompound/) was used to verify whether the drugs identified in our analysis could target the genes identified.

RESULTS

The investigation strategy of our work is described in Figure 1A. From text-mining searches, 427 unique genes (some genes may have had more than one name) were identified using the search terms *alopecia, hair loss, baldness,* and *balding* (see Supplemental Table I in the online version at https://doi.org/10.1016/j.clinthera.2019.04.003). The gene ontology enrichment analysis of cellular components showed that most of the genes identified are expressed in the

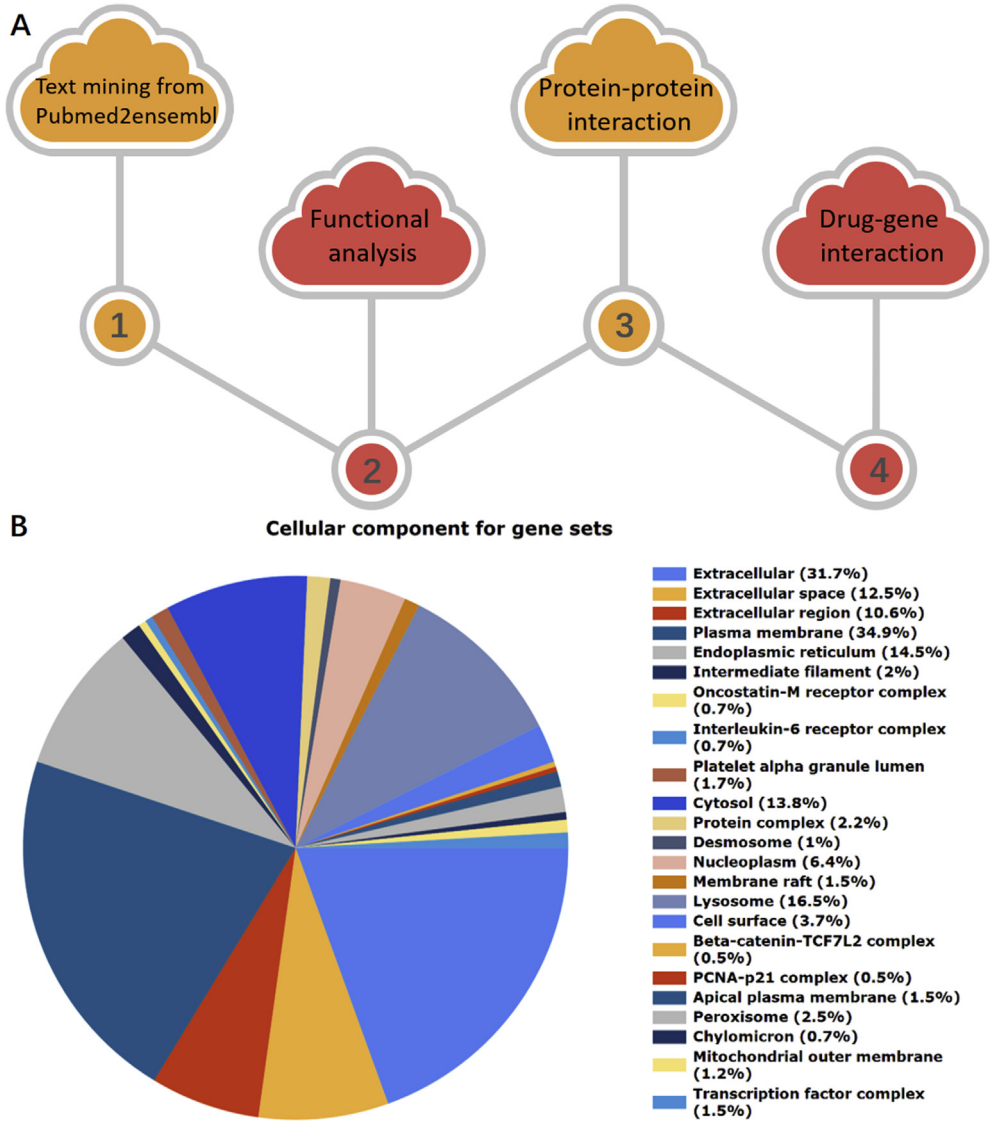

Figure 1. A, The overall strategy of data mining used for the identification of gene sets associated with chemotherapy-associated alopecia. B, The cellular-component annotations of those gene sets.

plasma membrane or are extracellular. Some others mainly are expressed in the endoplasmic reticulum, extracellular space, or extracellular region (Figure 1B).

Molecular-function annotations (*chemokine activity*, *catalytic activity*, *transcription factor activity*, *receptor activity*, *growth factor*, *receptor binding*, and *cytokine activity*) demonstrated that 9.2% of the identified genes had transcription factor activity (Figure 2A). Biological-process annotations revealed that the most highly enriched terms were *cell communication* and *signal transduction*, which are closely related to the pathology of alopecia. The 3 most enriched biological-process annotations—*signal transduction* ($P = 0.022$), *metabolism* ($P < 0.001$), and *energy pathways* ($P < 0.001$)—accounted for 29.4%, 16.4%, and 16.1% of the query set, respectively (Figure 2B). The higher gene representations and relatively low *P*

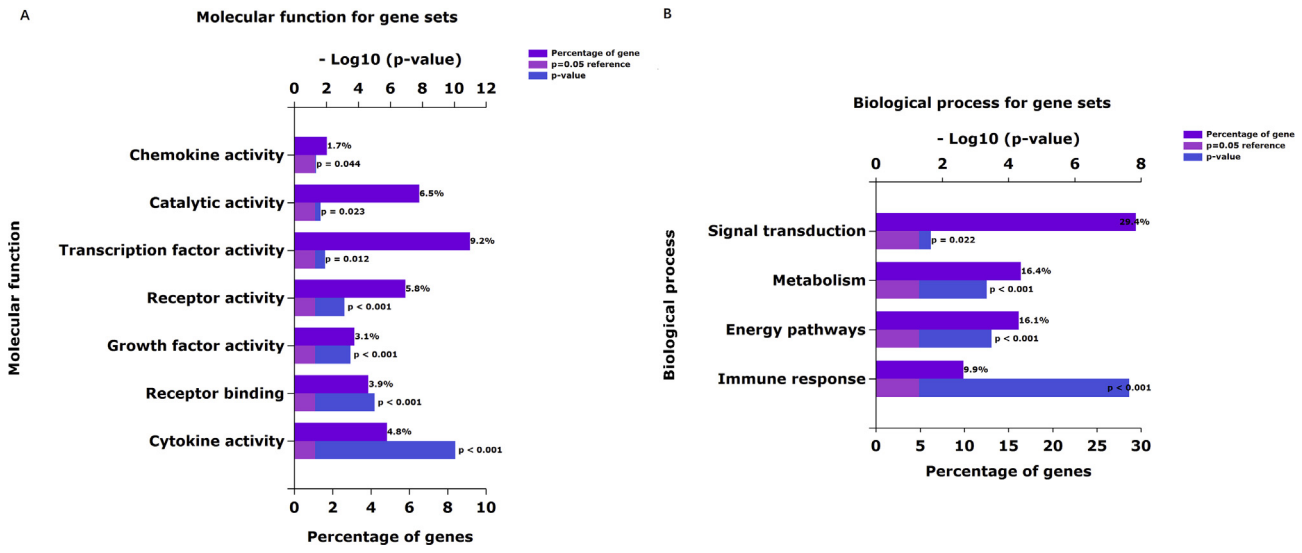

Figure 2. The molecular functions (A) and the biological-process annotations (B) of gene sets associated with chemotherapy-associated alopecia.

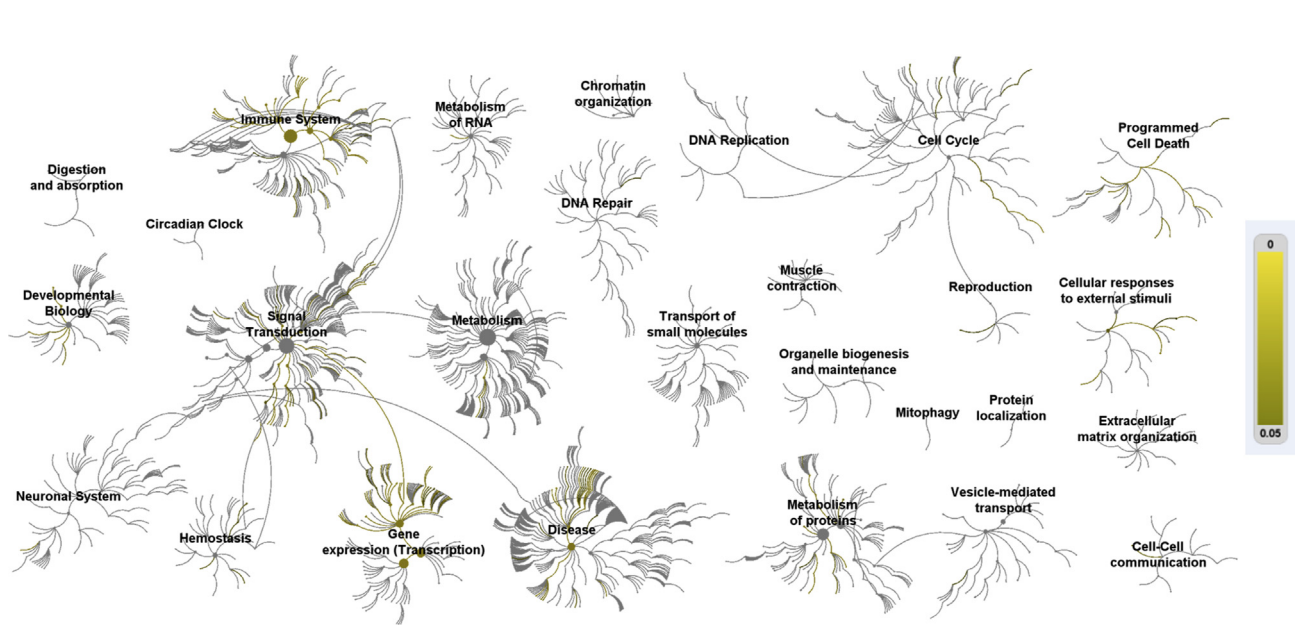

Figure 3. The general pathway molecular function and the biological-process annotations of the gene sets identified as being associated with chemotherapy-induced alopecia. Each step away from the center represents the next-lowest level in the pathway hierarchy. The color code denotes overrepresentation of that pathway in the dataset. Light grey signifies pathways that are not significantly overrepresented.

values suggest that these processes are particularly relevant to alopecia. Other highly enriched biological-process annotations included *immune response*, *regulation of gene expression*, and *epigenetics*. Integrating the molecular functions and biological processes, we selected the molecular-function annotation *transcription factor activity*, and all enriched biological-process annotations resulted in 5 sets of annotations containing a total of 230 unique genes (see Supplemental Table II in the online version at https://doi.org/10.1016/j.clinthera.2019.04.003).

We first analyzed all of the genes mined in the Reactome database. The global pathway analysis included 27 pathways (Figure 3). The enriched pathway annotations of 229 genes were analyzed, from which 10 pathways were selected: (1) interferon γ signaling ($P = 1.11 \times 10^{-16}$), (2) interleukins (IL) 4 and 13 signaling ($P = 1.11 \times 10^{-16}$), (3) cytokine signaling in the immune system ($P = 1.11 \times 10^{-16}$), (4) signaling by IL-6 ($P = 1.11 \times 10^{-16}$), (5) translocation of zinc finger antiviral protein 70 to the immunologic synapse ($P = 5.55 \times 10^{-16}$), (6) IL-10 signaling ($P = 9.99 \times 10^{-16}$), (7) programmed cell death protein 1 signaling ($P = 1.67 \times 10^{-15}$), (8) phosphorylation of CD3 and T-cell receptor ζ chains ($P = 1.67 \times 10^{-15}$), (9) endosomal/vacuolar pathway ($P = 4.10 \times 10^{-14}$), and (10) costimulation by the CD28 family ($P = 9.61 \times 10^{-13}$) (Table).

From the pathway analysis, 32 genes were selected. With these 32 genes, the protein−protein interaction analysis was performed using String (Figure 4 and see Supplemental Table III in the online version at https://doi.org/10.1016/j.clinthera.2019.04.003). Among these, 19 genes met the criteria for the high-confidence protein−protein interaction network. Based on a literature search, these 19 genes included 11 and 8 genes positively associated with higher and lower prevalences and severity of CIA, respectively. Predominant gene functions in the final list included kinases, cytokines, and cell-surface receptors. Pathways

Table. Pathways of the genes involved in chemotherapy-induced alopecia.

Pathway	Entities				Reactions	
	Found	Ratio	*P*	FDR*	Found	Ratio
Interferon γ signaling	44/250	0.018	1.11×10^{-16}	2.01×10^{-14}	12/15	0.001
IL-4 and -13 signaling	35/211	0.015	1.11×10^{-16}	2.01×10^{-14}	22/46	0.004
Cytokine signaling in immune system	123/1055	0.075	1.11×10^{-16}	2.01×10^{-14}	240/639	0.053
Signaling by IL	66/640	0.046	1.11×10^{-16}	2.01×10^{-14}	154/491	0.041
Translocation of ZAP-70 to immunologic synapse	18/42	0.003	5.55×10^{-16}	8.60×10^{-14}	4/4	3.34×10^{-4}
IL-10 signaling	23/86	0.006	9.99×10^{-16}	1.36×10^{-13}	12/15	0.001
PD-1 signaling	18/45	0.003	1.67×10^{-15}	1.82×10^{-13}	4/4	3.34×10^{-4}
Phosphorylation of CD3 and TCR ζ chains	18/45	0.003	1.67×10^{-15}	1.82×10^{-13}	6/7	5.84×10^{-4}
Endosomal/vacuolar pathway	21/82	0.006	4.10×10^{-14}	4.06×10^{-12}	3/4	3.34×10^{-4}
Costimulation by the CD28 family	21/97	0.007	9.61×10^{-13}	8.65×10^{-11}	10/34	0.003

IL = interleukin; PD = programmed cell death protein; TCR = T-cell receptor; ZAP = zinc finger antiviral protein.
The *P* value cutoff for this process was set at 1.00×10^{-10}. Among these highly enriched pathways, those involved in other specific diseases were omitted. Genes involved in one or more of the selected pathways were used for further analysis.
* *False-discovery rate*, calculated using the hypergeometric distribution, which describes the probability of several given types of genes appearing in the query set given the total number of those genes in the reference set (genome). FDR correction is performed to control for the false-positive results expected with a large number of comparisons.

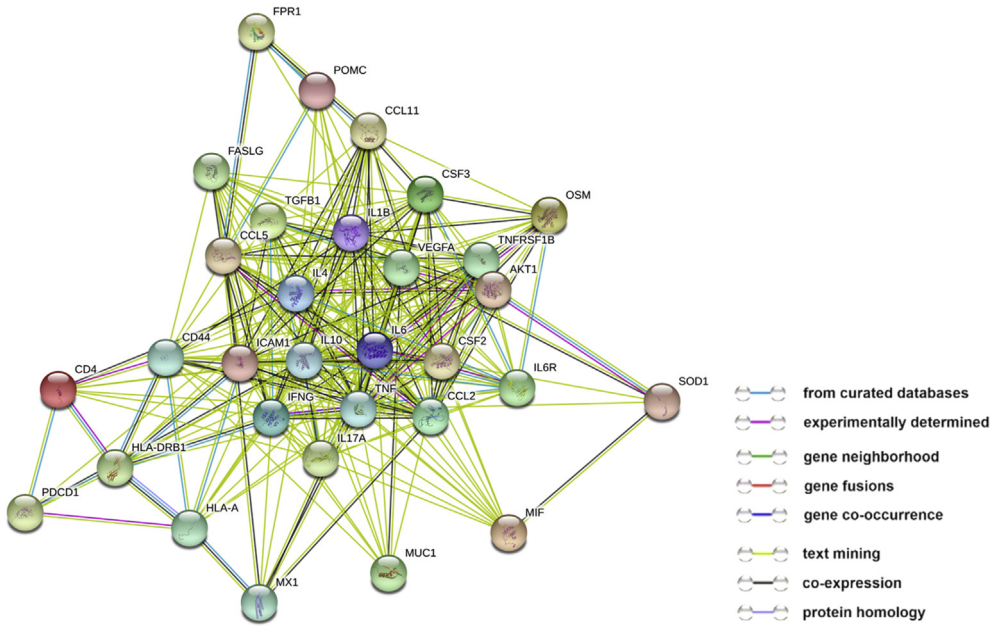

Figure 4. High-confidence protein–protein interaction network of for potential gene targeting in the treatment of chemotherapy-induced alopecia. Connecting line colors indicate the types of information used to infer the interaction with the CI set at 90%. AKT = protein kinase B; CCL = C-C motif chemokine ligand; CSF = colony-stimulating factor; FASLG = tumor necrosis factor superfamily, member 6 (Fas) ligand; FPR = formyl peptide receptor; HLA-DRB = human leukocyte antigen DR B; ICAM = intercellular adhesion molecule; IFNG = interferon-γ; IL = interleukin; MIF = macrophage migration inhibitory factor; MUC = mucin; MX = myxovirus resistance protein; OSM = oncostatin M; PDCD = programed cell death protein; POMC = proopiomelanocortin; SOD = superoxide dismutase; TGFB = transforming growth factor β; TNF = tumor necrosis factor; TNFR = tumor necrosis factor receptor; TNFRSF = tumor necrosis factor receptor superfamily; VEGF = vascular endothelial growth factor.

frequently observed included Toll-like receptor signaling, cytokine–cytokine receptor interaction, and mitogen-activated protein kinase signaling.

The 19 genes in the high-confidence protein–protein interaction network were screened for drug–gene interactions. A list of 29 drugs meeting the standard requirements for CIA therapy was compiled, and included:

Canakinumab
Cladribine
Cyclophosphamide
Cytarabine
Dactinomycin
Everolimus
Gallium nitrate
Gemcitabine
Ifosfamide
Interferon alfa-2b
Irinotecan
Lenalidomide
Mechlorethamine
Melphalan
Mycophenolic acid
Paclitaxel
Pirfenidone
Pomalidomide
Prednisone
Procarbazine
Siltuximab
Sirolimus
Sorafenib

Streptozotocin
Temozolomide
Temsirolimus
Thalidomide
Tretinoin
Vinblastine

Potential gene targets of the drugs in this list are *CSF2* (8 drugs), *AKT1* (7 drugs), *TGFB1* (6 drugs), *IFNG* and *TNF* (5 drugs each), *IL6* (4 drugs), *IL1B* (3 drugs), *IL10* and *VEGFA* (2 drugs each), and *CD44* and *TNFRSF1B* (1 drug each).

Common previously approved uses for these drugs include the treatment of cancer, inflammation, angiocarpy disease, diabetes, Parkinson disease, and Alzheimer disease. A total of 23 drugs on the list are used for anticancer therapy.

Next, we verified whether the drugs identified by our analysis could target pathways associated with CIA. The targeted genes of the potential treatments of CIA were found using the PubChem tool. The analysis revealed that the genes associated with the treatments of CIA were similar to our original pathway enrichment results.

DISCUSSION

Here, we utilized *in silico* text-mining and pathway-analysis tools to identify existing drugs that have the potential for topical administration for CIA treatment in patients receiving chemotherapy for cancer. As HF cells play a core role in CIA therapy, patients have 2 general options to achieve successful CIA treatment: (1) prevention of chemotherapeutics from reaching the HF, and (2) protection of the hair matrix keratinocytes against apoptosis.[13] Since most of these interventions are applied topically to the scalp, their effectiveness in preventing CIA is thought to be limited. Currently, in addition to the scalp-cooling devices mentioned before, several drugs have been used to avoid CIA, of which 2% topical minoxidil may offer the best outcome.[14] Minoxidil is a drug used for the first-line treatment of androgenetic alopecia and second-line treatment of alopecia areata. The use of minoxidil has been approved by the US Food and Drug Administration. In a study of the efficacy of oral minoxidil 2% in preventing CIA in patients with acute myeloid leukemia, patients' hair was significantly regrown after continuous treatment.[15] Although minoxidil was well tolerated, the mechanism of action is not clear. *N*-acetylcysteine is a prostaglandin analogue that may promote hair growth in patients with alopecia areata.[16] The tellurium compound ammonium tri-chloro-(dioxoethylene-*O*,*O*-)-tellurate (AS-101) has shown protective effects against hair loss because of its ability to improve the level of keratinocyte growth factor.[17] The efficacy of topical vitamin D_3 is thought to be related to the general ability of vitamin D_3 to stimulate the differentiation of HF stem cells and to diminish DNA proliferation.[18] Of the 29 drugs screened, 6 drugs are not used for anticancer therapy: canakinumab, mycophenolate, prednisone, siltuximab, sirolimus, and thalidomide.

Chemotherapy against cancer is associated with side effects due to the damage of cellular metabolic processes and the induction of apoptosis in rapidly dividing cells. p53 and its target genes play crucial roles in the pathogenesis of CIA.[19] Genome damage induced by chemotherapy leads to the rapid accumulation of p53 protein in the affected cells, which upregulate the downstream genes of p53, such as *FAS*, *IGFBP3*, and *BAX*.[20] A total of 19 genes associated with the identified target drugs were also associated with p53. Rapid inhibition of p53 activity in HF stem cells via increased mouse double minute 2 homologue expression promotes survival of HF stem cells after DNA damage.[21] HF stem cells more highly express members of the B-cell lymphoma (BCL)-2 family of proteins, which have a negative correlation with the expression of p53.[21] Because BCL-2 has been associated with the increased expression of protein kinase B 1, the encoding gene of which (*AKT1*) appears in our list of target genes, the association of BCL-2 with a lower prevalence and severity of CIA would be expected. However, the role of BCL-2 in chemotherapy-induced alopecia remains unclear. In addition, the proliferative effects of BCL-2 may present a contraindication in patients with cancer, as BCL-2 has been implicated in some cancers.[22] IL-1β, the encoding gene of which (*IL1B*) appears in our target genes list, can induce protection against CIA that occurs with pharmacologically distinct classes of chemotherapeutic agents, and has been used clinically as a target for anticancer treatment.[23] Epidermal growth factor receptor plays a role in the stabilization of p53 during cyclophosphamide-induced alopecia through p38 kinase. The study of CIA treatments has been focused on treating general symptoms or processes using anti-

inflammatory agents or growth factors, which thus far have been of limited success. It is unknown whether this limited success has been in part due to the systemic rather than topical application of drug treatments of CIA. However, one study has reported that treatments targeting epidermal growth factor receptor signaling were associated with reduced inflammation, while also targeting other pathways to promote CIA treatment efficacy, with reduced side effects.[24] Cyclin-dependent kinase (CDK)-2 is a key component in the transition from phase G_1 to late G_2 of cell-cycle regulation, and chemotherapeutic agents may induce G_1 growth arrest in HF cells. In a rat model of CIA, the use of CDK-2 inhibitor was associated with a reduction in hair loss of 33%.[25]

There was a limitation in the present study: The information on the functions or roles of genes could not be verified through experiments but via databases used. The accuracy of the analysis may be increased as the databases are updated. Additionally, without experimentation, we cannot know all of the existing gene interactions of a given drug. Therefore, this analysis may not have included all drugs that potentially work for CIA.

CONCLUSIONS

We have presented a way to survey candidate known drugs that target the genes/pathways relevant to CIA. The list of potential uses for drugs screened by this method could be dynamically changed with the improvement of databases and analytical tools. Thus, in this paper, we identified a list of 29 candidate drugs, of which 7 drugs targeted other CIA-relevant pathways that have also not yet been tested in CIA.

CONFLICT OF INTEREST

The authors have indicated that they have no conflicts of interest with regard to the content of this article.

ACKNOWLEDGMENTS

The sponsor of this work was the Natural Science Foundation, the Affiliated Hospital of Qingdao University. The sponsor had a role in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the article for publication.

N. Zhang contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, and writing, review, and editing of the manuscript. W. Xu contributed to methodology, validation, visualization, and writing of the original draft. S. Wang contributed to visualization and writing of the original draft. Y. Qiao contributed to data analysis and visualization. X. Zhang contributed to visualization.

REFERENCES

1. Santos Z, Avci P, Hamblin MR. Drug discovery for alopecia: gone today, hair tomorrow. *Expert Opin Drug Discov*. 2015;10:269–292.
2. Paus R, Haslam IS, Sharov AA, Botchkarev VA. Pathobiology of chemotherapy-induced hair loss. *Lancet Oncol*. 2013;14:e50–e59.
3. Trueb RM. Chemotherapy-induced alopecia. *Semin Cutan Med Surg*. 2009;28:11–14.
4. Browall M, Gaston-Johansson F, Danielson E. Postmenopausal women with breast cancer: their experiences of the chemotherapy treatment period. *Cancer Nurs*. 2006;29:34–42.
5. Cash TF. The psychology of hair loss and its implications for patient care. *Clin Dermatol*. 2001;19:161–166.
6. Macduff C, Mackenzie T, Hutcheon A, Melville L, Archibald H. The effectiveness of scalp cooling in preventing alopecia for patients receiving epirubicin and docetaxel. *Eur J Cancer Care*. 2003;12:154–161.
7. Katsimbri P, Bamias A, Pavlidis N. Prevention of chemotherapy-induced alopecia using an effective scalp cooling system. *Eur J Cancer*. 2000;36:766–771.
8. Granai CO, Frederickson H, Gajewski W, Goodman A, Goldstein A, Baden H. The use of minoxidil to attempt to prevent alopecia during chemotherapy for gynecologic malignancies. *Eur J Gynaecol Oncol*. 1991;12:129–132.
9. Mosca E, Bertoli G, Piscitelli E, et al. Identification of functionally related genes using data mining and data integration: a breast cancer case study. *BMC Bioinformatics*. 2009;10(suppl):S8.
10. Chaudhari R, Tan Z, Huang B, Zhang S. Computational polypharmacology: a new paradigm for drug discovery. *Expert Opin Drug Discov*. 2017;12:279–291.
11. Kirk J, Shah N, Noll B, et al. Text mining-based in silico drug discovery in oral mucositis caused by high-dose cancer therapy. *Support Care Cancer*. 2018;26:2695–2705.

12. Cotto KC, Wagner AH, Feng Y, et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res*. 2017;4: D1068–D1073.
13. Rugo HS, Klein P, Melin SA, et al. Association between use of a scalp cooling device and alopecia after chemotherapy for breast cancer. *J Am Med Assoc*. 2017;317:606–614.
14. Rossi A, Fortuna MC, Caro G, et al. Chemotherapy-induced alopecia management: clinical experience and practical advice. *J Cosmet Dermatol*. 2017;16:537–541.
15. Yang X, Thai KE. Treatment of permanent chemotherapy-induced alopecia with low dose oral minoxidil. *Australas J Dermatol*. 2016;57:e130–e132.
16. D'Agostini F, Bagnasco M, Giunciuglio D, Albini A, De Flora S. Inhibition by oral *N*-acetylcysteine of doxorubicin-induced clastogenicity and alopecia, and prevention of primary tumors and lung micrometastases in mice. *Int J Oncol*. 1998;13:217–224.
17. Sredni B, Xu RH, Albeck M, et al. The protective role of the immunomodulator AS101 against chemotherapy-induced alopecia studies on human and animal models. *Int J Cancer*. 1996;65:97–103.
18. Yeager CE, Olsen EA. Treatment of chemotherapy-induced alopecia. *Dermatol Ther*. 2011;24:432–442.
19. Lakin ND, Jackson SP. Regulation of p53 response to DNA damage. *Oncogene*. 1999;18:7644–7655.
20. Botchkarev VA, Komarova EA, Siebenhaar F, et al. p53 is essential for chemotherapy-induced hair loss. *Cancer Res*. 2000;60:5002–5006.
21. Lindner G, Botchkarev VA, Botchkareva NV, Ling G, van der Veen C, Paus R. Analysis of apoptosis during hair follicle regression (catagen). *Am J Pathol*. 1997;151:1601–1617.
22. Muller-Rover S, Rossiter H, Paus R, et al. Overexpression of Bcl-2 protects from ultraviolet B-induced apoptosis but promotes hair follicle regression and chemotherapy-induced alopecia. *Am J Pathol*. 2000;156:1395–1405.
23. Stojadinovic O, Wikramanayake TC, Villasante Fricke AC, et al. Wound healing protects against chemotherapy-induced alopecia in young rats via up-regulating interleukin-1β-mediated signaling. *Heliyon*. 2017;3, e00309.
24. Bichsel KJ, Gogia N, Malouff T, et al. Role for the epidermal growth factor receptor in chemotherapy-induced alopecia. *PLoS One*. 2013;8, e69368.
25. Knockaert M, Greengard P, Meijer L. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol Sci*. 2002;23:417–425.

Address correspondence to: Nanyang Zhang, Scientific Research Platform, The Affiliated Hospital of Qingdao University, No.1677, Wutaishan Road, Huangdao District, Qingdao, Shandong 266000, China. E-mail: nanyang.zhang@outlook.com

APPENDIX

Supplementary Table 1.

SERPINE2
HSD17B7
CRP
CNBP
TCF7L2
COL17A1
APOC1
APOE
GABRQ
AC005921.3
CYP2B7P1
MAF
IDE
HSD17B1
BLM
HSD17B3
HLA-DRB1
FTO
HSD17B6
CYP21A2
ABCC8
PTER
PRL
SLC30A8
MYCN
LTA
CYP2D6
APOB
IGFBP3
ZFX
MN1
STATH
SHBG
ABCC9
KCNJ8
FST
MC4R
DSG3
VARS2
CYP51A1
RALBP1
ZFY

Supplementary Table 1. (*Continued*)

GSTM1
TAP2
WNT3A
PARP1
TGFB2
IRF6
HSD11B1
CD34
IL10
PDCD1
REN
PKP1
MLPH
PSMD1
FASLG
SELL
FCGR3B
ABCB6
CYP27A1
CD244
PNKD
FCRL3
NTRK1
NES
ATIC
LMNA
CREB1
MTX1
MUC1
IL6R
LOR
FLG
TCHH
TFPI
CD160
PHGDH
CLDN1
NGF
TP63
LPP
AMPD1
AHSG
PTPN22
C2orf37
LIPH
ABCB11

Supplementary Table 1. (*Continued*)

DPP4
DPYD
F3
ABCD3
CP
CFC1
PARK2
ATR
GLI2
ESR1
IL1B
B3GAT1
MALL
NLRP5
EDAR
SHH
IL11
CD7
NOS3
PTCH2
PLK3
FPR1
MGMT
MKI67
CD200
NPS
BRAF
DPP7
IDS
HTRA1
FOXO3
NOTCH1
C2
PAEP
ZMPSTE24
ATM
SOX9
NTF4
RBM28
LEP
BAX
MMP8
PGR
ORAI1
GH1
GJB3

Supplementary Table 1. (*Continued*)

GJB4
TYR
CTSC
CHDH
NOVA2
DMPK
F8
ZBTB8OS
IKBKG
HDAC1
ACE
RELB
EMD
SCD
CHUK
ABCC2
XRCC1
TRIM37
TUBB3
SERPINE1
ABCD1
EPX
ACHE
EXOSC1
IL17A
COIL
CD79A
EPO
FPGS
MUT
TGFB1
RUNX2
CYP2C19
TXNRD1
VEGFA
REL
IGF1
CD40LG
DLX3
GNPTAB
CYP3A4
NGFR
IL29
FAS
TAC1
OCRL

Supplementary Table 1. (*Continued*)

GGN
KITLG
CIRH1A
CDH1
ABCB1
ABCB4
CDKN1A
PLAU
BRCA1
MEN1
IFNG
TH1L
HMGA2
BAAT
SRD5A2
XDH
CTNNB1
NODAL
SPEN
CYP27B1
COL11A2
ACR
KRT14
FURIN
KRT15
KRT37
CYP24A1
GLA
TYMP
LAT2
DDIT3
EFNA3
AGA
TIMM8A
MT1E
GLI1
RECQL4
PTCH1
FANCC
AF205589.2
TNFRSF1B
CSF3
MMP2
NTRK3
DGAT1
PLCD1
INCENP

Supplementary Table 1. (*Continued*)

ERBB2
CNOT8
MED1
CXCL9
SOX21
PIK3C2A
CYP19A1
CRH
IFNA1
CDKN2A
ABCD2
BCL2L10
ANTXR2
MYO5A
MUC16
DCLRE1C
PCSK1
NDE1
ABCC6
RTEL1
HM13
TPMT
HPSE
ICAM1
AC136618.1
LGR4
BDNF
ARAF
DSPP
VDR
IL6
CCL27
EBP
SHMT1
ASIP
ITCH
CYCS
CAT
CD44
PDSS1
DPP8
CSF2
GSTT1
PROCR
RAG1

(*continued on next page*)

Supplementary Table 1. (*Continued*)

RAG2
NFKB1
IRF1
IL5
IL4
SLC5A11
HLA-A
PPARG
FOXP3
TPO
SMAD3
RAF1
FOXN1
ADI1
EGF
PITX2
TSPYL2
KRT86
ADAM17
KRT5
TRPS1
CDSN
TJP2
CD276
PML
IL2
BTD
POU5F1
NUDT6
NF1
NOTCH3
RAB5A
SERPINA3
TNF
EDA2R
AR
THRB
TCL1A
TOP2B
CCL2
EDA
CCL11
HNF4A
MIF
SMAD1
CTF1
NCOA1

Supplementary Table 1. (*Continued*)

TNFRSF9
CCL5
EIF2C2
CCL4
POMC
H6PD
TF
NCOA4
FBXW7
CDK2
GIF
GLB1
MS4A2
EGFR
PDGFRB
DKK1
AKT1
MTHFR
VCAN
MAX
CALCA
CHD7
DSP
FGF7
NR0B1
RIN2
OSM
AREGB
LIF
AREG
ESR2
LSS
IL8
PLAG1
GATA3
ALB
MBTPS2
GATM
IL2RA
TP53
JAG1
AIRE
MEFV
NHS
GGT1
BMP2
FUT3

Supplementary Table 1. (*Continued*)

KIT
CBS
BMP4
PCNA
OFD1
RB1
DPP9
BCR
MAPK1
MX1
STS
IL6ST
TLR1
KLRC2
AC022075.1
NKX2-1
NFIC
CD69
HLCS
BCL2
OSMR
LIFR
INS
KCNE2
PAFAH1B1
GNRH1
PRLR
ATP8B1
TGM1
STUB1
TXNL1
ALOX5AP
GART
TNFRSF10A
SMAD4
BID
CDX2
CD4
PDZD2
EVC
EVC2
SCG5
MSX1
SMAD2
SOD1
HR
TNFRSF19

Supplementary Table 1. (*Continued*)

SRD5A1
NTF3
MICA
GJB6
GJB2
ZNF396
ZNF24
BTG3
DSG4
DSG1
DSC2
APEX1
TGIF1
TYMS

Supplementary Table 2.

HSD17B7
HSD17B1
HSD17B3
HSD17B6
CYP21A2
VARS2
CYP51A1
GSTM1
HSD11B1
CYP27A1
ATIC
PHGDH
AMPD1
LIPH
DPYD
CP
NOS3
IDS
TYR
CHDH
SCD
ABCD1
ACHE
FPGS
MUT
CYP2C19
TXNRD1
CYP3A4
OCRL
BAAT
SRD5A2
XDH
CYP27B1
CYP24A1
GLA
AGA
TIMM8A
DGAT1
PLCD1
TPMT
HPSE
EBP
SHMT1
CYCS
CAT
PDSS1
GSTT1

Supplementary Table 2. (*Continued*)

TPO
BTD
H6PD
GLB1
MTHFR
LSS
GATM
GGT1
FUT3
CBS
DPP9
STS
HLCS
PAFAH1B1
ATP8B1
TGM1
STUB1
TXNL1
SOD1
SRD5A1
TYMS
PRL
LTA
IGFBP3
MN1
FST
MC4R
DSG3
RALBP1
WNT3A
TGFB2
PDCD1
FASLG
SELL
NTRK1
FLG
NGF
AHSG
PTPN22
ATR
EDAR
SHH
PTCH2
PLK3
FPR1
MKI67
BRAF

Supplementary Table 2. (*Continued*)

NOTCH1
ATM
NTF4
LEP
PGR
GH1
DMPK
IKBKG
CHUK
CD79A
EPO
TGFB1
VEGFA
IGF1
GNPTAB
NGFR
FAS
TAC1
OCRL
KITLG
CIRH1A
CDH1
CTNNB1
NODAL
TYMP
LAT2
EFNA3
GLI1
PTCH1
TNFRSF1B
NTRK3
INCENP
ERBB2
MED1
PIK3C2A
CRH
CDKN2A
MUC16
DCLRE1C
ICAM1
LGR4
BDNF
ARAF
ASIP
CD44
RAF1
EGF

Supplementary Table 2. (*Continued*)

TSPYL2
TJP2
NF1
NOTCH3
RAB5A
TNF
EDA2R
AR
THRB
TCL1A
EDA
CCL11
HNF4A
MIF
CTF1
TNFRSF9
CCL5
POMC
MS4A2
EGFR
PDGFRB
DKK1
AKT1
FGF7
NR0B1
RIN2
OSM
AREG
JAG1
BMP2
KIT
BMP4
BCR
MAPK1
MX1
TLR1
KLRC2
OSMR
LIFR
INS
GNRH1
PRLR
TNFRSF10A

(*continued on next page*)

Supplementary Table 2. (*Continued*)

SMAD2
TNFRSF19
NTF3
BTG3
DSG4
CRP
HLA-DRB1
CD34
IL10
FCGR3B
CD244
FCRL3
MUC1
IL6R
CD160
IL1B
NLRP5
IL11
CD7
CD200
C2
PAEP
EPX
IL17A
CD40LG
IFNG
CSF3
CXCL9
IFNA1
HM13
IL6
CCL27
CSF2
PROCR
IL5
IL4
HLA-A
CD276
IL2
CCL2
CCL4
IL2RA
IL6ST
CD69
ALOX5AP
CD4

Supplementary Table 3.

CD44
HLA-DRB1
HLA-A
ICAM1
AKT1
CD4
PDCD1
FASLG
IL17A
IL6
OSM
TNF
CCL11
IL1B
IL6R
POMC
TNFRSF1B
CCL2
IL10
IL4
MUC1
TGFB1
VEGFA
MX1
CSF2
SOD1
CSF3
MUC1
IFNG
FPR1
MIF
CCL5