

A Systems Biology Workflow for Drug and Vaccine Repurposing: Identifying Small-molecule BCG Mimics to Prevent COVID-19 Mortality

Rima Hajjo¹ and Alexander Tropsha²

¹Al-Zaytoonah University of Jordan

²University of North Carolina at Chapel Hill

June 1, 2020

Abstract

Background. Coronavirus disease 2019 (COVID-19) is expected to continue to cause worldwide fatalities until the World population develops ‘herd immunity’, or until a vaccine is developed and used as a prevention. However, the vaccine may prove ineffective due to rapid changes in viral antigenic determinants. Bacillus Calmette–Guérin (BCG) vaccine has been recognized for its beneficial effects on the immune system, and it is currently in being tested in clinical trials for COVID-19. However, BCG shortages may affect clinical decisions regarding the prioritization of BCG to protect from viral infections, hence, small-molecule BCG-mimics will be valuable alternatives. **Methods.** We developed and applied a systems biology workflow capable of identifying antiviral drugs and vaccines that can boost immunity and impact viral disease pathways to prevent the fatal consequences of COVID-19. **Results.** Our results indicate that BCG and small-molecule BCG-mimics affect the production and maturation of naïve T cells, which results in enhanced long-lasting innate immune responses to tackle novel viruses. Our workflow identified several antiviral drugs including raltegravir and lopinavir as high confidence BCG mimics. Top hits including emetine and lopinavir were validated to inhibit the growth of novel coronavirus SARS-CoV-2 in vitro. **Conclusions.** Herein, we provide systems biology support for using BCG as a protection measure from the lethal consequences of emergent viruses including SARS-CoV-2. We also provide systems biology evidence that certain small molecule drugs could mimic the effects of BCG and serve as alternatives to BCG.

Abbreviations:

BCG: Bacillus Calmette–Guérin

BIND: Biomolecular Interaction Network Database

BioGRID: Biological General Repository for Interaction Datasets

CMap: Connectivity Map

COVID-19: Coronavirus Disease 2019

DIP: Database of Interacting Proteins

FDA: The Food and Drug Administration

GEO : Gene Expression Omnibus

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

Mtb: *Mycobacterium tuberculosis*

RCTs: Randomized Clinical Trials

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

WHO: World Health Organization

1 INTRODUCTION

Few months after the declaration of COVID-19 pandemic by the World Health Organization (WHO), the disease-causing virus is still sweeping the globe, causing more fatalities, failing health care systems, and resulting in severe economic losses. Currently there are no approved drugs to treat COVID-19, and new vaccine development is expected to take at least 12-18 months, with growing fears of possible failure associated with rapid changes in viral antigenic determinants. Additionally, the highly specific virus-neutralizing antibodies in recovered patients may be short lived and ineffective in preventing the disease caused by the emerging variable strains of the virus. With these uncertainties regarding an eminent specific SARS-CoV-2 vaccine, we should start thinking about alternatives, such as exploiting the unique capabilities of our innate immune system.

Recent immuno-oncology success stories indicate that the best cancer-fighting strategies results from unleashing the patients' immune power. And there is an increased awareness that harnessing innate immune responses, opens up new possibilities for long-term, multifaceted tumor control and infectious disease prevention. Therefore, next generation antiviral vaccines should be capable of boosting innate immune responses to tackle a wide range of novel pathogens very early after exposure, as single treatments or adjuvants to traditional vaccines targeting the adaptive immune system.

Accumulating evidence from the biomedical literature indicates that SARS-CoV-mediated pathology, a very similar pathology to SARS-CoV-2, was mainly caused by ineffective innate immune responses, associated with a severe reduction in the number of T cells in the blood. Recent evidence indicated that SARS-CoV-2 and *Mycobacterium tuberculosis* (*Mtb*) share unique similarities in terms of host protein interaction partners, and both pathogens infect lung tissues. On the other hand, old 'polypharmacological vaccines', such as the BCG vaccine for tuberculosis (TB), has shown promising therapeutic effects on a wide range of infectious and non-infectious diseases including bladder cancer. Studies showed that BCG's polypharmacological effects were not limited to memory T cell immunity, but promoted strong, beneficial, and long-lasting effects on innate immunity. The WHO has also recognized these beneficial 'off-target' effects of BCG, calling for a further investigation to repurpose for other orphan life-threatening diseases. There are 35 clinical trials reported on clinicaltrials.gov, testing BCG for non-TB conditions including COVID-19. Additionally, few recent peer-reviewed reports have pointed to an epidemiological relationship between BCG and COVID-19 without providing substantial evidence. Therefore, the results of the randomized clinical trials (RCTs) will provide more clues soon.

Herein, we describe a unique drug and vaccine repurposing workflow, and list high confidence proteins and pharmacological classes of compounds, that work as BCG mimics on the systems level by inducing beneficial long lasting trained immune response. Thus, BCG mimics can be used as alternatives to BCG in protecting from COVID-19 and other emergent infectious diseases, or as treatments for bladder cancer and other tumors.

2 METHODS

2.1 Integrative Systems Biology Workflow. We developed and applied a systems biology workflow to study BCG network pharmacology and prioritize small-molecule BCG mimics and antivirals. Our workflow (Figure1) incorporates three major components: (1) a module for mining and prioritizing gene signatures representative of a condition or a biological state; (2) a network-mining module to identify genetic perturbations that induce gene expression profiles that are highly enriched with the genes constituting the condition gene signature; and (3) a pathway enrichment module to understand the biological processes involved in the mechanism of action of BCG and highly correlated genetic perturbagens.

2.2 BCG consensus gene signature. A consensus gene signature for BCG vaccine was derived from gene expression profiles in peripheral blood mononuclear cells (PBMCs) in response to a BCG challenge test reported by Matsumiya *et al* , GSE58636 dataset on NCBI Gene Expression Omnibus (GEO). All whole blood samples were collected from healthy human subjects enrolled in phase 1 trial (clinical trials registration: NCT01194180). For the purposes of this study we used the gene expression profiles generated from two human subject groups included in the above trial: group 1 (BCG naive), and group 2 (BCG vaccinated; median time since vaccination, 10 years). The consensus gene signature we prepared to study network pharmacology and query the connectivity map consisted of the genes that showed significant differential gene expression in response to a BCG challenge test (stimulated) in comparison with controls (unstimulated) on days 0 and 14 in both groups 1 and 2.

2.3 Network Building. A systematic search, for nearest neighbor (NN) genes/proteins of the upregulated and downregulated genes in BCG’s gene signature, was conducted in Cytoscape version 3.8.0 using the STRING protein query application. All retrieved protein-protein interactions (PPIs), including both physical and functional interactions were retrieved from popular databases such as MINT, HPRD, BIND, DIP, BioGRID, KEGG, Reactome, EcoCyc, NCI-Nature Pathway Interaction Database, and Gene Ontology (GO) protein complexes. Network building tools in Cytoscape version 3.7.2 were used to generate PPI networks for BCG-CGS.

2.4 Enrichment Analysis. Enrichment analysis was conducted in Cytoscape and MetaCore to identify pathways and biological processes associated with BCG-CGS and CMap genetic connections. The significance of the enrichment was determined by the hypergeometric test. All terms from the ontology are ranked based on their calculated p-values. Ontology terms with p-values less than the p-value threshold 0.05 are defined as statistically significant and therefore relevant to the studied list of genes. All terms from the ontology are ranked according to their calculated p-values.

2.5 The Connectivity Map (CMap). The CMap is a chemogenomics database that catalogs 1.3 Million profiles of transcriptional responses of human cells to chemical and genetic perturbations. Currently, there are 27,927 perturbagens (19,911 small molecules, and 7,494 genetic perturbagens) producing 476,251 expression signatures in 9 human cell lines: PC3, VCAP, A375, A459, HA1E, HCC515, HT29, MCF7, HEPG2. This database of cellular signatures has been produced using the L1000 platform; a high-throughput gene expression assay that measures the mRNA transcript abundance of 978 "landmark" genes from human cells.

2.6 Causal Reasoning. Causal reasoning analysis identify genes and proteins of a ‘topological significance’ in order to make decisions whether these genes/proteins are eligible for targeting in the studied phenotype. In this study we applied causal reasoning to identify molecular regulators that most likely directly cause the observed expression changes in transcriptional profiles in response to BCG. In this approach, changes in gene expression, both direction and effect of edges in the network are taken into account. For each node (i.e., gene) in causal reasoning network, observed changes in expression are matched with the expected changes inferred from network structure given the hypothesis that the observed gene expression is decreased or increased due to its activity. Each node has outgoing activation or inhibition effects on other objects in the knowledge database, and a key hub with a predicted increase in activity shows increased expression for those genes that the hub is known to activate, and it shows decreased expression for genes it is known to inhibit. Each predicted key hub has a prediction P-value which is produced as a result of a binomial test used to assess the probability of making a given number of supportive data out of all defined differentially expressed genes (DEGs) in examined data. It is noteworthy that causal reasoning examines both direct neighbors of differentially expressed genes, and remote (several steps away) regulators. All causal reasoning predictions were performed in Key Pathway Advisor from Clarivate Analytics, using the Pollard method.

2.7 R-package gplots. gplots v3.0.1.2 was used as for plotting enhanced heatmaps for transcriptional data (e.g., heatmap representing BCG-CGS in Figure 2). Heat maps were generated using the heatmap.2 function included in this package.

3 RESULTS

3.1 Drug and vaccine repurposing workflow

To study BCG’s polypharmacology and potential beneficial effects of this vaccine in preventing the fatal consequences of COVID-19, we have devised and implemented a ‘network biology’ workflow (Figure 1) to interrogate the hypothesis that BCG vaccination may protect from COVID-19 fatalities. This workflow is based on our drug repurposing chemocentric informatics workflow, which has been validated previously for small-molecule drug repurposing. The current workflow is tweaked towards vaccine repurposing by employing novel bioinformatic approaches to computationally model and connect molecular networks in an effort to understand the underlying ‘network’ biology of vaccines, and pinpoint the regulatory genes and proteins responsible for causing the observed beneficial multitherapeutic effects. Although we are not the first group to use network biology approaches to study the transcriptional changes of vaccines, to our knowledge, this is the first study that uses these approaches to support vaccine repurposing, specifically for COVID-19.

3.3 BCG consensus gene signature

Our workflow starts with the prioritization of a gene signature to study BCG’s network pharmacology. First, we derived a consensus gene signature (CGS) for BCG based on GEO’s dataset GSE58636. Details on BCG-CGS signature are found in table S1 (Supporting Information). Twenty-two differentially-expressed genes across all 4 experiments (2 Groups x 2 time points discussed in Methods) formed BCG’s consensus gene signature (BCG-CGS) shown in Figure 2A.

3.4 BCG protein-protein interactions (PPIs)

All 22 genes in BCG-CGS were used as seed nodes to build a protein-protein interaction network for signature genes (Figure 2B). Interactions were extracted from STRING database and included high confidence interactions including: physical interactions (e.g., binding), functional interactions (e.g., activation, inhibition, catalysis), or gene co-expression. Two types of networks were generated: 1) high-confidence ‘core’ network restricted to BCG signature genes as network nodes and high confidence (≥ 0.70) interactions as network edges, and 2) medium-confidence interaction network obtained from expanding the core network by 20 additional nodes (Figure 3).

Enrichment analysis results performed in Cytoscape, using STRING’s protein-protein interactions, indicated that BCG-CGS is enriched in inflammatory cytokines and immune response modulators (Figure 2B). Some signature genes are also involved in the negative control of important viral processes (e.g., (FCN1, TNF and CCL3), and others are involved in the response to viral infections (e.g., IFNG, RNASE6, IL6 and TNF). The complete lists of enriched pathways are included in tables S2 and S3 (Supporting Information).

3.5 Identification of key hubs

We identified 291 key hubs using the causal reasoning method which seeks to identify molecular regulators that will directly cause the observed transcriptional changes in response to BCG vaccination. Key regulators can be transcriptional factors and proteins with potentially altered activity that explains the transcriptional changes. Top five statistically significant inhibited key hubs were: HEY1, DSIPI (GILZ), Jagged1, HAND1 and miR-129-1-3p. And top five statistically significant activated key hubs were: PHF20, TAFII70, Glutaredoxin, RUNX2 and NOTCH1 (NICD). Top 30 causal key hubs are shown in table 1 and all identified 291 key hubs are included in table S4 (Supporting Information).

3.6 Identifying BCG ‘mimics’

In order to identify experimentally validated upstream regulators that cause transcriptional changes similar to those induced by BCG, we queried the Connectivity Map (CMap) database of the Broad Institute with BCG-CGS and identified proteins and small-molecule drugs that have strong connectivity scores with BCG (Figure 1). The CMap approach enabled us to compare BCG-CGS with ‘experimentally’ predefined signatures of therapeutic compounds and genetic perturbations (i.e., over expression or knockdown) included in the CMap and ranked according to a connectivity scores (ranging from +100 to -100), representing relative similarity to BCG-CGS. The connectivity score itself is derived using a nonparametric, rank-based, pattern-matching

strategy based on the Kolmogorov-Smirnov statistic. All instances in the database are then ranked according to their connectivity scores with BCG-CGS; those at the top (+) are most strongly correlated to the query signature and looked at as BCG mimics, and those at the bottom (-) are most strongly anticorrelated and can reverse BCG's gene signature.

Our analysis identified three highly enriched classes of genetic knockdown (KD) perturbagens and one pharmacological class of drugs that have positive connectivity scores in alveolar A549 cells (i.e., caused similar transcriptional changes to those induced by BCG in alveolar A549 cells). These hits can be considered as BCG mimics capable of inducing transcriptional changes similar to those caused by BCG vaccine. Therefore, we suggest that BCG mimics can be used as alternatives to BCG vaccination to promote long-lasting beneficial effects on immune cells. The three enriched protein classes are: protein phosphatases (with best positive connection for PPP4C KD), histone deacetylases (with best positive connection for HDAC10 KD followed by HDAC11 KD), and mediator complex proteins (with best positive connection for MED6 KD followed by MED7 KD). Additionally, protein kinase C (PKC) activators were enriched as a drug class; and top 3 PKC activators with highest CMap connectivity scores to BCG-CGS prostratin, phorbol-12-myristate-13-acetate, and ingenol. It is evident that all of the above 4 classes of proteins share one common feature: they participate in the transcriptional and metabolic regulation of immune cells in response to environmental cues including responses to pathogens. All top-scoring PKC activators from the CMap, are also known to have antiviral effects or affect T cell activation.

Remarkably, analyzing top ten CMap positive connections with BCG-CGS obtained from nine cell lines indicated that two compounds are approved antiviral drugs: raltegravir (top 3rd positive connection, an HIV integrase inhibitor) and lopinavir (top 6th positive connection, an HIV protease inhibitor). More interestingly, emetine (top 4th positive connection) and lopinavir were recently validated to inhibit SARS-CoV-2 replication in vitro. We also found evidence in the biomedical literature indicating that MST-312, narciclasine and verrucarin-a possess antiviral activities. All CMap hits are provided in tables S5 and S8 (Supporting Information).

In order to prioritize high confidence BCG genetic mimics, we integrated hypotheses derived independently from the CMap with those predicted by causal reasoning, and accepted common hits only (i.e., CMap positive connections with BCG-CGS that are also predicted as beneficial drug targets by causal reasoning). This analysis resulted 30 high confidence common hits reported in table S9 (Supporting Information).

3.8 Any validation for functional connections with SARS-CoV-2?

We tested whether BCG-CGS, CMap positive connections, or predicted key hubs will have any impact on COVID-19 by identifying overlaps with SARS-CoV-2 interactome, i.e., human proteins that were experimentally validated to interact with SARS-CoV-2 and extracted from two recent reports. This analysis (Figure 4A) validated 3 proteins hits to have physical links to SARS-CoV-2. The three proteins are transcribed by: BRD4, PRKACA and SIRT5; they all were positive connections from the CMap, predicted as statistically significant key hubs, and were also validated SARS-CoV-2 interacting proteins.

Additionally, 14 high-confidence CMap positive connections, were validated to make physical interactions SARS-CoV-2 proteins. These proteins are: PSEN2, PABPC1, HMOX1, CIT, PLAT, IGF2R, RIPK1, NDUFS3, NDUFA5, GGH, NEU1, SCARB1, CSNK2B, F2RL1. And two positive connections, MARK2 and MARK3, were reported to have interactions with corona viruses. Predicted causal key hubs, SIGMAR1 and GNB1, were also validated to have physical links to SARS-CoV-2, and a third key hub PPIA was known as human protein interacting with proteins from corona viruses.

Additionally, we mined the biomedical literature to identify evidence for linking BCG small molecule mimics with SARS-CoV-2, corona viruses or viral infections in general. We found that two out of ten top positive compound connections (emetine and lopinavir), were recently validated to inhibit SARS-CoV-2 replication in vitro. Other compounds we found to inhibit the growth of corona viruses, or had general antiviral activities (Table 2).

4 DISCUSSION

Our results indicate that BCG-CGS, key regulatory hubs and BCG-mimics identified from the CMap enrich common biological pathways important for key viral processes such as viral RNA synthesis and processing, virus-host interactions, positive regulation of viral genome replication, and are also important for the immune response mounted against the virus. Evidence from the biomedical literature confirms that BCG has many beneficial ‘off-target’ effects that can protect humans from emerging novel pathogens by boosting their innate immune responses. Our studies suggest that BCG can potentially reverse or prevent some of the detrimental consequences, caused by SARS-CoV-2 on vital regulatory processes, by promoting wide-range transcriptional and metabolic changes that are capable of producing a balanced immune response against SARS-CoV-2. BCG can accomplish that mainly by increasing the production of thymus-generated short-lived undifferentiated CD4+ cells known as naive T cells (Th_0), and triggering their differentiation into the long-lived mature naive T cells (MNTs), such as CD4+ and CD8+ T cells. These conclusions are supported by the enrichment results produced using the ‘Compare Experiment’ algorithm in MetaCore from Clarivate Analytics, which looks for significant coordinated gene expression effects across all experiments to test whether the pathway is being up- or down-regulated in a manner that is unlikely to be accounted for by random chance. The top enriched pathway map, with upregulated genes in response to BCG, is ‘Immune response T cell subsets: secreted signals’ (Figure 4B). As a validation, a recent study showed that SARS-CoV-2 reshapes central cellular pathways, such as translation, splicing, carbon metabolism and nucleic acid metabolism.

A recent publication in Lancet has questioned whether BCG’s effects can last for a long time. Our top enriched pathway map (Figure 4B) indicates that BCG’s effects can be long-lasting if the effects were exerted on thymus-generated Th_0 cells, which can occur to a greater extent very early in life before reaching thymic involution by puberty. This pathway map indicates that BCG is capable of affecting both the numbers and the types of produced innate immune cells, as well as their maturation to long-lived memory T cells (i.e., what is known as trained immunity). This is very significant in the context of BCG’s protective effects from SARS-CoV-2 and other emergent novel viruses; where the individual’s ability to eradicate such viruses is dictated by the number and diversity of naive T cell reservoir. And this is a clear indication that BCG protects individuals from lethality by novel pathogens by priming their trained immunity to fight such pathogens, including SARS-CoV-2.

Supporting evidence for this hypothesis is found in the literature indicating that the protective effects of the BCG against TB, can last from 15 to 60 years after vaccination, with longer lasting effects when the vaccine is administered during the first year of life. A recent study indicated that “school-aged BCG vaccination offered moderate protection against tuberculosis for at least 20 years, which is much longer than previously thought”. Another 60-year follow-up study, showed that BCG vaccine efficacy persisted for 50 to 60 years after a single dose of BCG. These studies serve as additional evidence from the literature supporting our claim that a single dose of an ‘effective’ BCG vaccination to infants can have a very long duration of protection against pathogens including SARS-CoV-2.

Our findings provided systems biology support for using BCG to prevent the fatal consequences of COVID-19. BCG is currently on WHO’s List of Essential Medicines; it is considered one of the safest and most effective medicines needed in a health system. Therefore, we propose BCG administration to all newborns will act as a protection measure from SARS-CoV-2 and other emerging pathogens. BCG can be given to newborns according to the regulations known for TB prevention. We also recommend that multiple doses of the vaccine are necessary to protect adults from COVID-19 since the protective effects of BCG are weaker if the vaccine is given after the first year of life and especially after puberty. Since this is an approved vaccine for TB, it can directly enter Phase III testing for the protection from COVID-19 caused fatalities. However, we caution that running these experiments during an active COVID-19 outbreak, might expose participants to aggravated immune responses if they contract COVID-19 during the study. We also advise that clinical study design takes into account several factors that are known to affect the performance of BCG vaccine, such as: the age of the participants, geographies, ethnicities, route of administration and the

mycobacterium strain used in the vaccine. It is equally important to run experimental validation studies to evaluate the effects of BCG mimics, in preventing COVID-19 or for treating urological cancers. BCG mimics can solve problems associated with potential supply shortages of BCG, or even address some of the problems associated with the use of attenuated live vaccines.

ACKNOWLEDGEMENTS

RH acknowledges support from the Deanship of Scientific Research at Al-Zaytoonah University of Jordan grant 2020-2019/17/03. AT acknowledges partial support from NIH grant OT2TR003441. We acknowledge Clarivate Analytics for providing access to MetaCore; a specialized pathway and functional genomics analysis product. Reference to commercial products or services does not constitute endorsement.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

R. Hajjo generated the idea, designed the workflow, generated content, performed data analysis and wrote the manuscript. A. Tropsha provided critical discussion, recommended some studies and provided feedback on manuscript.

DATA AVAILABILITY

BCG dataset, GSE58636 is available on the Gene Expression Omnibus GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58636>.

All generated data and R-scripts will be deposited on github: <https://github.com/rhajjo/>

ORCID

Rima Hajjo: <https://orcid.org/0000-0002-7090-5425>

Alexander Tropsha: <https://orcid.org/0000-0003-3802-8896>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

REFERENCES

Table 1. Top twenty key hubs predicted by causal reasoning.

Key Hub	Molecular Function	Gene Symbol	Predicted Activity+	Correct/Total network predic
HEY1	Transcription factor	HEY1	-	15/15
PHF20	Binding protein	PHF20	+	15/15
DSIP1 (GILZ)	Transcription factor	TSC22D3	-	14/14
TAFII70	Transcription factor	TAF6	+	14/14
DSIP1 (GILZ)	Transcription factor	TSC22D3	-	13/13
Glutaredoxin 1	Enzyme	GLRX	+	13/13
Jagged1	Receptor ligand	JAG1	-	13/13
RUNX2	Transcription factor	RUNX2	+	13/13
NOTCH1 (NICD)	Transcription factor	NOTCH1	+	16/17
HAND1	Transcription factor	HAND1	-	12/12
PRMT6	Enzyme	PRMT6	+	12/12
miR-129-1-3p	RNA	MIR129-1	-	12/12
SOX10	Transcription factor	SOX10	+	12/12
HAND2	Transcription factor	HAND2	-	12/12
MSK1	Protein kinase	RPS6KA5	+	12/12

Key Hub	Molecular Function	Gene Symbol	Predicted Activity+	Correct/Total network predic
USP28	Protease	USP28	+	15/16
c-Fos	Transcription factor	FOS	+	15/16
UBF	Transcription factor	UBTF	+	11/11
miR-520e-3p	RNA	MIR520E	-	11/11
TMEM119	Protein	TMEM119	+	11/11
LRP16	Binding protein	MACROD1	+	11/11
LRP16	Binding protein	MACROD1	+	14/15
CaMK II gamma	Protein kinase	CAMK2G	+	11/11
CaMK II gamma	Protein kinase	CAMK2G	+	14/15
miR-4500	RNA	MIR4500	-	14/15
NOTCH1 (NICD)	Transcription factor	NOTCH1	+	14/15
miR-4516	RNA	MIR4516	-	11/11
NDPK B	Protein kinase	NME2	-	11/11
KLF11 (TIEG2)	Transcription factor	KLF11	-	11/11
miR-320d	RNA	MIR320D1	-	14/15

+ Predicted activity of the key hub by causal reasoning is denoted by - if the hub is inhibited, and denoted by + if the hub is activated.

++ Correct/total network predictions: correct for the genes in the dataset predicted correctly; total for the total number of genes in the causal reasoning network.

§ Calculation distance: Using causal reasoning one-step key hubs are defined as statistically significant transcriptional factors that are associated with experimental differential expressed genes regulation. Two-step and three-step key hubs are distant key hubs that regulate one-step transcriptional factors.

*P -value calculated for the polynomial test.

Table 2. Small-molecule BCG mimics with potential antiviral effects.

Compound	Score+	Description	Validation++
prostratin	98.65	PKC activator	Antiviral
ingenol	98.52	PKC activator	Antiviral
raltegravir	97.85	HIV integrase inhibitor	Antiviral
emetine	97.25	Protein synthesis inhibitor	SARS-CoV-2
phorbol-12-myristate-13-acetate	96.72	PKC activator	Antiviral
mebendazole	95.32	Tubulin inhibitor	Antiviral
lopinavir	95.06	HIV protease inhibitor	SARS-CoV-2
MST-312	95.04	Telomerase inhibitor	Antiviral
narciclasine	94.71	Cofilin signaling pathway activator	Antiviral
verrucarin-a	94.51	Protein synthesis inhibitor	Antiviral
anisomycin	94.40	DNA synthesis inhibitor	Corona viruses
azacitidine	94.29	DNA methyltransferase inhibitor	Antiviral
cytochalasin-b	93.90	Microtubule inhibitor	Antiviral
cephaeline	93.88	Protein synthesis inhibitor	Antiviral
homoharringtonine	93.42	Protein synthesis inhibitor	Antiviral
ruxolitinib	92.81	JAK inhibitor	COVID-19 CT§
HU-211	92.64	Glutamate receptor antagonist	Unknown
vinblastine	92.36	Microtubule inhibitor	Unknown
RO-28-1675	92.12	Glucokinase activator	Unknown

Compound	Score+	Description	Validation++
vincristine	91.61	Tubulin inhibitor	Unknown

+Score refer to the CMap score. It represents the level of similarity between transcriptional effects induced by BCG and each of the compounds.

++ Validation refers to the presence of any supporting evidence from the biomedical literature that the predicted BCG mimics have any antiviral activities. Antiviral means there is evidence that the compound is used as or has antiviral activity; SARS-CoV-2 means that the compound should antiviral activity against SARS-CoV-2; Corona viruses means that the compound showed antiviral activity against corona viruses other than SARS-CoV-2.

§ COVID-19 CT: there is evidence that the compound is being tested in clinical trials for COVID-19. There are 12 Studies found for Ruxolitinib in COVID-19 on clinicaltrials.gov.

Figure 1. Workflow for drug and vaccine repurposing. A gene signature is identified/derived and a consensus gene signature is prioritized, all genes are nomenclated according to HUGO Gene Nomenclature Committee (HGNC). The consensus gene signature is used to query the CMap to identify positive connections capable of producing gene signatures similar to BCG-CGS. Prioritize genes and compounds that induce transcriptional changes similar to those induced by BCG: A) key hubs predicted by causal reasoning; B) Positive genes and compound connections from the connectivity map. Prioritize top enriched pathway map explaining the biological effects of BCG (gene annotations on the pathway map are assigned by Clarivate Analytics, IFN-gamma is an alias for INFG, GM-CSF is an alias for CSF2, MIP-1-alpha is an alias for CCL3).

Figure 2. BCG’s consensus gene signature (BCG-CGS) and highly enriched pathways. (A) A heat map of the $\log_2FC$ of the gene expression for differentially expressed genes representing BCG’s consensus gene signature. Upregulated genes have positive $\log_2FC$ denoted in red color, and down regulated genes have negative values for $\log_2FC$ denoted in blue color. (B) Core network for BCG-CGS showing highly connected genes in BCG-CGS, deleting all singleton genes. Nodes are color-coded using a split pie chart coloring scheme indicating pathway/gene set contribution to each node from the top 5 most enriched pathways/gene lists. All details about pathway/gene set ID are found in Table S2 (Supporting Information).

Figure 3. High-confidence expanded network for BCG-CGS. Nodes are color-coded using a split pie chart coloring scheme indicating pathway/gene set contribution to each node from the top 5 most enriched pathways/gene lists. Core network is composed of genes in the BCG-CGS that are not singletons. Step 1 expansion, added 10 additional nodes (i.e., genes) to the core network. Step 2 expansion, added another 10 nodes for the first expansion. Step 3 expansion, added another 10 nodes to the second expansion. Expansions were performed to see which pathways remained most statistically significant, and therefore are considered high confidence pathways.

Figure 4. (A) A venn diagram showing overlaps between BCG genetic mimics and key hubs with SARS-CoV-2 and Corona viruses interactomes. (B) Top “pathway map” with the highest level of enrichment by genes in BCG-CGS. This map is generated using MetaCore from Clarivate Analytics. Red thermometers indicate genes overexpressed in response to BCG treatment, and the height of the red bars is representative of the differential gene expression level (i.e., $\log_2$ values of the fold change). The numbers under the thermometers 1- 5 refer to the experiment number: 1) gene expression on day 1 in response to BCG vaccination to a BCG-naive population on day 1; 2) gene expression on day 1 in response to BCG re-vaccination to a previously vaccinated population; 3) gene expression on day 1 in response to BCG vaccination to a BCG-naive population on day 14; 4) gene expression on day 14 in response to BCG re-vaccination to a previously vaccinated population, and 5) positive connections from the connectivity map, and the red bar in the thermometer number 5 represents presence of the gene only.





