

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

Rima Hajjo (✉ r.hajjo@zuj.edu.jo)

Al-Zaytoonah University of Jordan <https://orcid.org/0000-0002-7090-5425>

Alexander Tropsha

UNC Chapel Hill <https://orcid.org/0000-0003-3802-8896>

Research Article

Keywords: BCG vaccine, BCG mimics, COVID-19, innate immunity, SARS-CoV-2, systems biology.

Posted Date: July 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-40730/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Pharmaceutical Research on October 6th, 2020. See the published version at <https://doi.org/10.1007/s11095-020-02930-9>.

Abstract

Purpose: 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 off-target beneficial effects on the immune system, therefore, can be exploited to boost immunity and protect from emerging novel viruses.

Methods: We developed a systems biology workflow capable of identifying small-molecule antiviral drugs and vaccines that can boost immunity and impact a wide variety of viral disease pathways to protect from the fatal consequences of emerging viruses.

Results: We show that BCG affect the production and maturation of naïve T cells, which results in enhanced long-lasting trained innate immune responses to tackle novel viruses. Our workflow identified small-molecule BCG mimics, including antiviral drugs such as raltegravir and lopinavir as high confidence hits. Strikingly, top hits emetine and lopinavir were validated to inhibit the growth of SARS-CoV-2 *in vitro*.

Conclusions: Our results provide systems biology support for using BCG and small-molecule BCG mimics as a protection measure from the lethal consequences of emergent viruses including SARS-CoV-2.

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[1,2] and infectious disease prevention[3–5]. 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[6]. 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[7]. 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[8–10]. 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[11]. 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.

Results

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[12], 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.

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[13]. 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.

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 (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).

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).

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

Key Hub	Molecular Function	Gene Symbol	Predicted Activity†	Correct/Total network predictions‡	Activity Prediction <i>P</i> -value*	Calculation Distance§
HEY1	Transcription factor	HEY1	-	15/15	3.05E-05	3
PHF20	Binding protein	PHF20	+	15/15	3.05E-05	3
DSIPI (GILZ)	Transcription factor	TSC22D3	-	14/14	6.10E-05	3
TAFII70	Transcription factor	TAF6	+	14/14	6.10E-05	3
DSIPI (GILZ)	Transcription factor	TSC22D3	-	13/13	1.22E-04	2
Glutaredoxin 1	Enzyme	GLRX	+	13/13	1.22E-04	3
Jagged1	Receptor ligand	JAG1	-	13/13	1.22E-04	3
RUNX2	Transcription factor	RUNX2	+	13/13	1.22E-04	2
NOTCH1 (NICD)	Transcription factor	NOTCH1	+	16/17	1.37E-04	3
HAND1	Transcription factor	HAND1	-	12/12	2.44E-04	3
PRMT6	Enzyme	PRMT6	+	12/12	2.44E-04	2
miR-129-1-3p	RNA	MIR129-1	-	12/12	2.44E-04	3
SOX10	Transcription factor	SOX10	+	12/12	2.44E-04	3
HAND2	Transcription factor	HAND2	-	12/12	2.44E-04	3
MSK1	Protein kinase	RPS6KA5	+	12/12	2.44E-04	2
USP28	Protease	USP28	+	15/16	2.59E-04	3
c-Fos	Transcription factor	FOS	+	15/16	2.59E-04	3
UBF	Transcription factor	UBTF	+	11/11	4.88E-04	3
miR-520e-3p	RNA	MIR520E	-	11/11	4.88E-04	2
TMEM119	Protein	TMEM119	+	11/11	4.88E-04	3
LRP16	Binding protein	MACROD1	+	11/11	4.88E-04	2
LRP16	Binding protein	MACROD1	+	14/15	4.88E-04	3
CaMK II gamma	Protein kinase	CAMK2G	+	11/11	4.88E-04	2
CaMK II gamma	Protein kinase	CAMK2G	+	14/15	4.88E-04	3
miR-4500	RNA	MIR4500	-	14/15	4.88E-04	3
NOTCH1 (NICD)	Transcription factor	NOTCH1	+	14/15	4.88E-04	2
miR-4516	RNA	MIR4516	-	11/11	4.88E-04	3
NDPK B	Protein kinase	NME2	-	11/11	4.88E-04	3
KLF11 (TIEG2)	Transcription factor	KLF11	-	11/11	4.88E-04	2
miR-320d	RNA	MIR320D1	-	14/15	4.88E-04	3

† 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.

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)[14] 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[15]. 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[16–19]. All top-scoring PKC activators from the CMap, are also known to have antiviral effects or affect T cell activation[20–24].

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[25]. We also found evidence in the biomedical literature indicating that MST-312[26], narciclasine[27] and verrucarin-a[28] 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).

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[29,30]. 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[7].

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[29]. Predicted causal key hubs, SIGMAR1 and GNB1, were also validated to have physical links to SARS-CoV-2[7], and a third key hub PPIA was known as human protein interacting with proteins from corona viruses[29].

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[25]. Other compounds we found to inhibit the growth of corona viruses, or had general antiviral activities (Table 2).

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

Compound	Score†	Description	Validation‡
prostratin	98.65	PKC activator	Antiviral[21]
ingenol	98.52	PKC activator	Antiviral[20]
raltegravir	97.85	HIV integrase inhibitor	Antiviral[54]
emetine	97.25	Protein synthesis inhibitor	SARS-CoV-2[25]
phorbol-12-myristate-13-acetate	96.72	PKC activator	Antiviral[23,24]
mebendazole	95.32	Tubulin inhibitor	Antiviral[55]
lopinavir	95.06	HIV protease inhibitor	SARS-CoV-2[25,26]
MST-312	95.04	Telomerase inhibitor	Antiviral[23,24,26]
narciclasine	94.71	Cofilin signaling pathway activator	Antiviral[27]
verrucarin-a	94.51	Protein synthesis inhibitor	Antiviral[56]
anisomycin	94.40	DNA synthesis inhibitor	Corona viruses[57]
azacitidine	94.29	DNA methyltransferase inhibitor	Antiviral[58]
cytochalasin-b	93.90	Microtubule inhibitor	Antiviral[59]
cephaeline	93.88	Protein synthesis inhibitor	Antiviral[60]
homoharringtonine	93.42	Protein synthesis inhibitor	Antiviral[61]
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
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.

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 boasting their innate immune responses[31]. 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 (Th₀), and triggering their differentiation into the long-lived mature naive T cells (MNTs), such as CD4⁺ and CD8⁺ T cells[32] 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[32].

A recent publication[33] 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 Th₀ 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[34,35] 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[36] 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.[32] 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"[32]. Another 60-year follow-up study, showed that BCG vaccine efficacy persisted for 50 to 60 years after a single dose of BCG.[33] 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 protect from 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[33]. 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.

Methods

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.

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*[37], GSE58636 dataset on NCBI Gene Expression Omnibus (GEO)[14]. 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.

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[38] version 3.8.0 using the STRING[39] protein query application. All retrieved protein-protein interactions (PPIs), including both physical and functional interactions were retrieved from popular databases such as MINT[40], HPRD[41], BIND[42], DIP[43], BioGRID[44], KEGG[45], Reactome[32], EcoCyc[46], NCI-Nature Pathway Interaction Database[47], and Gene Ontology (GO)[48] protein complexes. Network building tools in Cytoscape version 3.7.2 were used to generate PPI networks for BCG-CGS.

Enrichment Analysis. Enrichment analysis was conducted in Cytoscape[38] 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[49]. 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.

The Connectivity Map (CMap). The CMap[14,50] 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[14]; a high-throughput gene expression assay that measures the mRNA transcript abundance of 978 "landmark" genes from human cells.

Causal Reasoning. Causal reasoning[51] 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[52].

R-package gplots. gplots[53] 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.

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

Declarations

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>.

ORCID

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

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

Keywords: BCG vaccine, COVID-19, innate immunity, SARS-CoV-2, systems biology.

References

1. Demaria O, Cornen S, Daëron M, Morel Y, Medzhitov R, Vivier E. Harnessing innate immunity in cancer therapy. *Nature*. Nature Publishing Group; 2019. p. 45–56.

2. Dar TB, Henson RM, Shiao SL. Targeting innate immunity to enhance the efficacy of radiation therapy. *Frontiers in Immunology*. Frontiers Media S.A.; 2019. p. 3077.
3. Pattabhi S, Wilkins CR, Dong R, Knoll ML, Posakony J, Kaiser S, et al. Targeting Innate Immunity for Antiviral Therapy through Small Molecule Agonists of the RLR Pathway. *Journal of Virology*. American Society for Microbiology; 2016;90:2372–87.
4. Seth RB, Sun L, Chen ZJ. Antiviral innate immunity pathways. *Cell Research*. Nature Publishing Group; 2006. p. 141–7.
5. White MR, Doss M, Boland P, Tecle T, Hartshorn KL. Innate immunity to influenza virus: Implications for future therapy. *Expert Review of Clinical Immunology*. NIH Public Access; 2008. p. 497–514.
6. Channappanavar R, Zhao J, Perlman S. T cell-mediated immune response to respiratory coronaviruses [Internet]. *Immunologic Research*. Humana Press Inc.; 2014 [cited 2020 Apr 11]. p. 118–28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24845462>
7. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* [Internet]. Nature Publishing Group; 2020 [cited 2020 Apr 30];1–13. Available from: <http://www.nature.com/articles/s41586-020-2286-9>
8. Covián C, Fernández-Fierro A, Retamal-Díaz A, Díaz FE, Vasquez AE, Lay MK, et al. BCG-Induced Cross-Protection and Development of Trained Immunity: Implication for Vaccine Design. *Frontiers in Immunology*. Frontiers Media S.A.; 2019. p. 2806.
9. Redelman-Sidi G, Glickman MS, Bochner BH. The mechanism of action of BCG therapy for bladder cancer-A current perspective. *Nature Reviews Urology*. Nature Publishing Group; 2014. p. 153–62.
10. Buffen K, Oosting M, Quintin J, Ng A, Kleinnijenhuis J, Kumar V, et al. Autophagy Controls BCG-Induced Trained Immunity and the Response to Intravesical BCG Therapy for Bladder Cancer. Deretic V, editor. *PLoS Pathogens* [Internet]. Public Library of Science; 2014 [cited 2020 May 11];10:e1004485. Available from: <https://dx.plos.org/10.1371/journal.ppat.1004485>
11. BCG vaccines 1 Report on BCG vaccine use for protection against mycobacterial infections including tuberculosis, leprosy, and other nontuberculous mycobacteria (NTM) infections Prepared by the SAGE Working Group on BCG Vaccines and WHO Secretariat. 2017.
12. Hajjo R, Setola V, Roth BL, Tropsha A. Chemocentric informatics approach to drug discovery: Identification and experimental validation of selective estrogen receptor modulators as ligands of 5-hydroxytryptamine-6 receptors and as potential cognition enhancers. *Journal of Medicinal Chemistry*. 2012;55.
13. Matsumiya M, Satti I, Chomka A, Harris SA, Stockdale L, Meyer J, et al. Gene Expression and Cytokine Profile Correlate With Mycobacterial Growth in a Human BCG Challenge Model. 2014;
14. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*. Cell Press; 2017;171:1437-1452.e17.
15. Hollander M, Wolfe D. *Nonparametric Statistical Methods*. New York: Wiley; 1999. p. 178–85.

16. Liao FH, Shui JW, Hsing EW, Hsiao WY, Lin YC, Chan YC, et al. Protein phosphatase 4 is an essential positive regulator for Treg development, function, and protective gut immunity. *Cell and Bioscience*. BioMed Central Ltd; 2014;4:25.
17. Zhan Z, Cao H, Xie X, Yang L, Zhang P, Chen Y, et al. Phosphatase PP4 Negatively Regulates Type I IFN Production and Antiviral Innate Immunity by Dephosphorylating and Deactivating TBK1. *The Journal of Immunology*. The American Association of Immunologists; 2015;195:3849–57.
18. Quevedo M, Meert L, Dekker MR, Dekkers DHW, Brandsma JH, van den Berg DLC, et al. Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nature Communications*. Nature Publishing Group; 2019;10:1–15.
19. Ellmeier W, Seiser C. Histone deacetylase function in CD4+ T cells. *Nature Reviews Immunology*. Nature Publishing Group; 2018. p. 617–34.
20. Kwaa AK, Goldsborough K, Walker-Sperling VE, Pianowski LF, Gama L, Blankson JN. The effect of Ingenol-B on the suppressive capacity of elite suppressor HIV-specific CD8+ T cells. *PLoS ONE*. Public Library of Science; 2017;12.
21. Hezareh M. Prostratin as a new therapeutic agent targeting HIV viral reservoirs. *Drug News and Perspectives*. *Drug News Perspect*; 2005. p. 496–500.
22. Sung TL, Rice AP. Effects of prostratin on Cyclin TI/P-TEFb function and the gene expression profile in primary resting CD4+ T cells. *Retrovirology*. BioMed Central; 2006;3:66.
23. Basta S, Knoetig S, Summerfield A, McCullough KC. Lipopolysaccharide and phorbol 12-myristate 13-acetate both impair monocyte differentiation, relating cellular function to virus susceptibility. *Immunology*. Wiley-Blackwell; 2001;103:488–97.
24. Huang Z, Ye B, Han J, Kong F, Shan P, Lu Z, et al. NACHT, LRR and PYD domains-containing protein 3 inflammasome is activated and inhibited by berberine via toll-like receptor 4/myeloid differentiation primary response gene 88/nuclear factor- κ B pathway, in phorbol 12-myristate 13-acetate-induced macrophage. *Mol Med Rep*. 2018;17:2673–80.
25. Choy KT, Wong AYL, Kaewpreedee P, Sia SF, Chen D, Hui KPY, et al. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Research*. Elsevier B.V.; 2020;178.
26. Pradhan P, Nguyen ML. Herpes simplex virus virucidal activity of MST-312 and epigallocatechin gallate. *Virus Research*. Elsevier B.V.; 2018;249:93–8.
27. Gabrielsen B, Monath TP, Huggins JW, Kefauver DF, Pettit GR, Groszek G, et al. Antiviral (RNA) activity of selected amaryllidaceae isoquinoline constituents and synthesis of related substances. *Journal of Natural Products*. 1992;55:1569–81.
28. Tamura G, Ando K, Takatsuki A, Arima K, Suzuki S. Antiviral activity of brefeldin A and verrucarine A. *Journal of Antibiotics*. 1968. p. 160–1.
29. Guzzi PH, Mercatelli D, Ceraolo C, Giorgi FM. Master Regulator Analysis of the SARS-CoV-2/Human Interactome. *Journal of Clinical Medicine*. MDPI AG; 2020;9:982.
30. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, et al. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv*

- [Internet]. Cold Spring Harbor Laboratory; 2020 [cited 2020 Apr 25];2020.03.22.002386. Available from: <https://www.biorxiv.org/content/10.1101/2020.03.22.002386v1.full>
31. Kleinnijenhuis J, Quintin J, Preijers F, Joosten LAB, Ifrim DC, Saeed S, et al. Bacille Calmette-Guérin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A; 2012;109:17537–42.
 32. Aronson NE, Santosham M, Comstock GW, Howard RS, Moulton LH, Rhoades ER, et al. Long-term Efficacy of BCG Vaccine in American Indians and Alaska Natives: A 60-Year Follow-up Study. *Journal of the American Medical Association*. 2004. p. 2086–91.
 33. Ayoub BM. COVID-19 vaccination clinical trials should consider multiple doses of BCG. *Die Pharmazie* [Internet]. 2020 [cited 2020 Apr 30];75:159. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32295694>
 34. Belizário JE, Brandão W, Rossato C, Peron JP. Thymic and Postthymic Regulation of Naïve CD4+ T-Cell Lineage Fates in Humans and Mice Models. *Mediators of Inflammation*. Hindawi Limited; 2016;2016.
 35. Qi Q, Zhang DW, Weyand CM, Goronzy JJ. Mechanisms shaping the naïve T cell repertoire in the elderly - Thymic involution or peripheral homeostatic proliferation? *Experimental Gerontology*. Elsevier Inc.; 2014;54:71–4.
 36. Mangtani P, Nguipdop-Djomo P, Keogh RH, Sterne JA, Abubakar I, Smith PG, et al. The duration of protection of school-aged BCG vaccination in England: a population-based case-control study. 2017 [cited 2020 May 3]; Available from: <https://academic.oup.com/ije/article-abstract/47/1/193/4098108>
 37. Matsumiya M, Satti I, Chomka A, Harris SA, Stockdale L, Meyer J, et al. Gene Expression and Cytokine Profile Correlate With Mycobacterial Growth in a Human BCG Challenge Model. *The Journal of Infectious Diseases* [Internet]. 2015;211:1499–509. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiu615>
 38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*. Cold Spring Harbor Laboratory Press; 2003;13:2498–504.
 39. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2018;47:D607–13.
 40. Chatr-Aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, et al. MINT: the Molecular INTERaction database. *Nucleic Acids Res*. 2006;35:D572–4.
 41. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, et al. Human protein reference database as a discovery resource for proteomics. *Nucleic acids research* [Internet]. 2004 [cited 2020 Apr 5];32:D497-501. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14681466>

42. Bader GD, Betel D, Hogue CW v. BIND: the Biomolecular Interaction Network Database. [cited 2020 May 22]; Available from: <http://www.ncbi.nlm.nih>.
43. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the Database of Interacting Proteins [Internet]. Nucleic Acids Research. 2000. Available from: <http://dip.doe-mbi.ucla.edu/>
44. Breitkreutz B-J, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, et al. The BioGRID interaction database: 2008 update. Nucleic Acids Res. 2007;36:D637–40.
45. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2007;36:D480–4.
46. Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, et al. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Res. 2005;33:D334–7.
47. Krupa S, Anthony K, Buchoff J, Day M, Hannay T, Schaefer C. The NCI-Nature Pathway Interaction Database: A cell signaling resource. Nature Precedings. Springer Science and Business Media LLC; 2007;1–1.
48. Consortium GO. The Gene Ontology (GO) database and informatics resource. [cited 2020 Apr 5]; Available from: www.geneontology.org/.
49. Nikolsky Y, Kirillov E, Zuev R, Rakhmatulin E, Nikolskaya T. Functional analysis of OMICs data and small molecule compounds in an integrated “knowledge-based” platform. Methods in molecular biology (Clifton, NJ). 2009;563:177–96.
50. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science (New York, NY). 2006;
51. Chindelevitch L, Ziemek D, Enayetallah A, Randhawa R, Sidders B, Brockel C, et al. Causal reasoning on biological networks: interpreting transcriptional changes. 2012 [cited 2020 May 22];28:1114–21. Available from: <http://www.selventa.com>
52. Pollard J, Butte AJ, Hoberman S, Joshi M, Levy J, Pappo J. A computational model to define the molecular causes of type 2 diabetes mellitus. Diabetes Technology and Therapeutics. Diabetes Technol Ther; 2005;7:323–36.
53. gplots package | R Documentation [Internet]. [cited 2020 Jan 24]. Available from: <https://www.rdocumentation.org/packages/gplots/versions/3.0.1.2>
54. Wisskirchen K, Lucifora J, Michler T, Protzer U. New pharmacological strategies to fight enveloped viruses. Trends in Pharmacological Sciences. Elsevier Ltd; 2014. p. 470–8.
55. Barrows NJ, Campos RK, Powell ST, Prasanth KR, Schott-Lerner G, Soto-Acosta R, et al. A Screen of FDA-Approved Drugs for Inhibitors of Zika Virus Infection. Cell Host and Microbe. Cell Press; 2016;20:259–70.
56. Tamura G, Ando K, Takatsuki A, Arima K, Suzuki S. Antiviral activity of brefeldin A and verrucarin A. Journal of Antibiotics. J Antibiot (Tokyo); 1968. p. 160–1.

57. Dyal J, Gross R, Kindrachuk J, Johnson RF, Olinger GG, Hensley LE, et al. Middle East Respiratory Syndrome and Severe Acute Respiratory Syndrome: Current Therapeutic Options and Potential Targets for Novel Therapies. *Drugs*. Springer International Publishing; 2017. p. 1935–66.
58. Ianevski A, Zusinaite E, Kuivanen S, Strand M, Lysvand H, Teppor M, et al. Novel activities of safe-in-human broad-spectrum antiviral agents. *Antiviral Research*. Elsevier B.V.; 2018;154:174–82.
59. Farias KJS, Machado PRL, de Almeida Júnior RF, Lopes da Fonseca BA. Brefeldin A and Cytochalasin B reduce dengue virus replication in cell cultures but do not protect mice against viral challenge. *Access Microbiology*. Microbiology Society; 2019;1:e000041.
60. Yang S, Xu M, Lee EM, Gorshkov K, Shiryaev SA, He S, et al. Emetine inhibits Zika and Ebola virus infections through two molecular mechanisms: Inhibiting viral replication and decreasing viral entry. *Cell Discovery*. Nature Publishing Groups; 2018;4:1–14.
61. Dong HJ, Wang ZH, Meng W, Li CC, Hu YX, Zhou L, et al. The natural compound homoharringtonine presents broad antiviral activity in vitro and in vivo. *Viruses*. MDPI AG; 2018;10.

Supporting Information

Table S1: Consensus gene signature for the BCG vaccine (BCG-CGS).

Table S2: The completer list of ontology terms showing overrepresentation in BCG-CGS.

Table S3: The completer list of ontology terms showing overrepresentation in BCG-CGS expanded gene network in biological pathways.

Table S4: The complete list of regulatory key hubs determined by causal reasoning for BCG-CGS.

Table S5. The complete list of CMap genetic perturbations tested in nine cell lines and ranked according to the similarities of their transcriptional signatures to that of BCG.

Table S6: The complete list of CMap genetic perturbations tested in A549 and ranked according to the similarities of their transcriptional signatures to that of BCG.

Table S7: The complete list of CMap small-molecule perturbagens tested in nine cell lines and ranked according to the similarities of their transcriptional signatures to that of BCG.

Table S8: The complete list of CMap small-molecule perturbagens tested in A549 and ranked according to the similarities of their transcriptional signatures to that of BCG.

Table S9: The complete list of pathways showing statistically significant overrepresentation in positive genetic CMap connections with CMap scores equalto or above 90%.

Table S10: The list of common hits from CMap and causal reasoning.

Figures

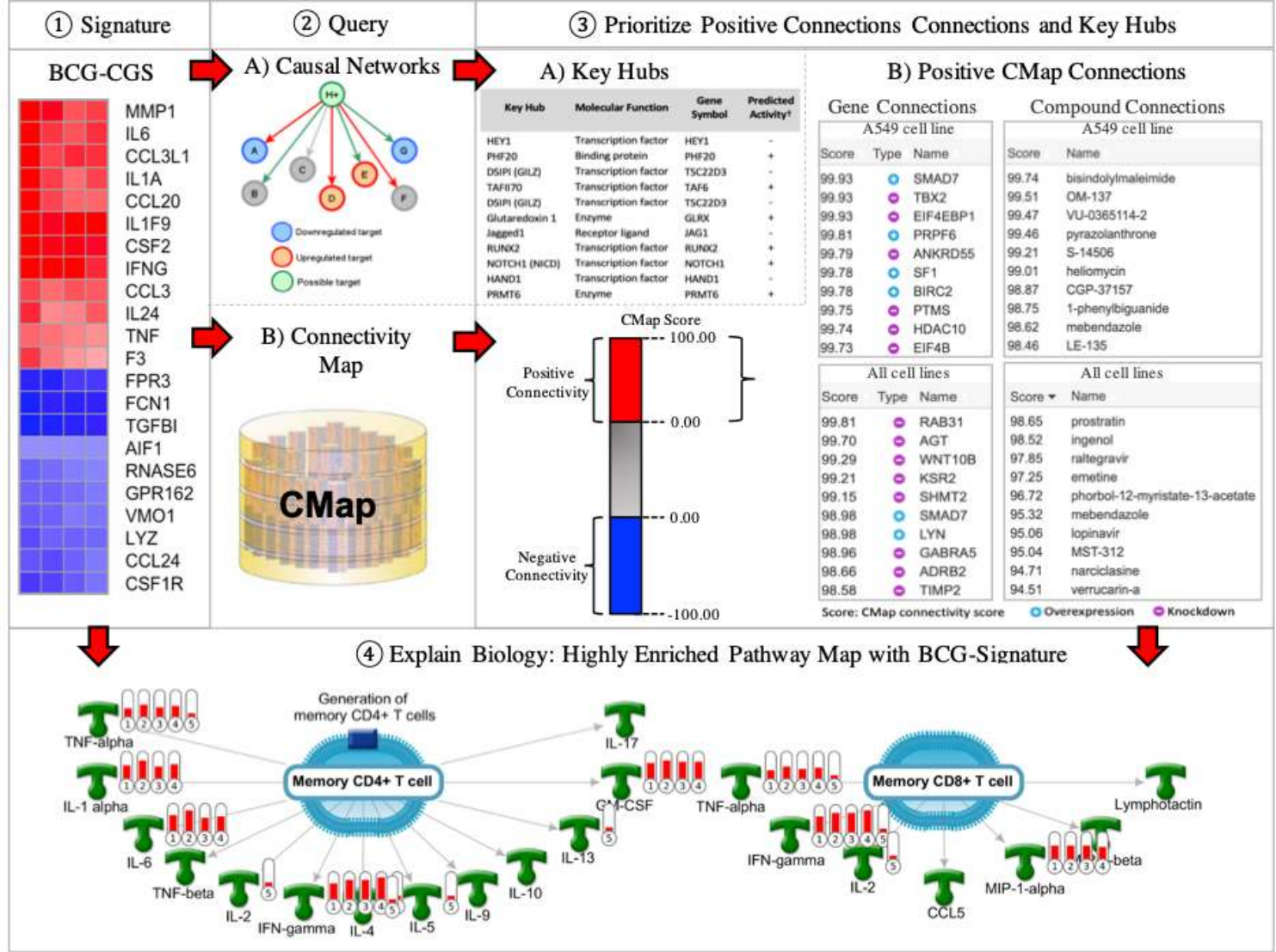


Figure 1

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 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).

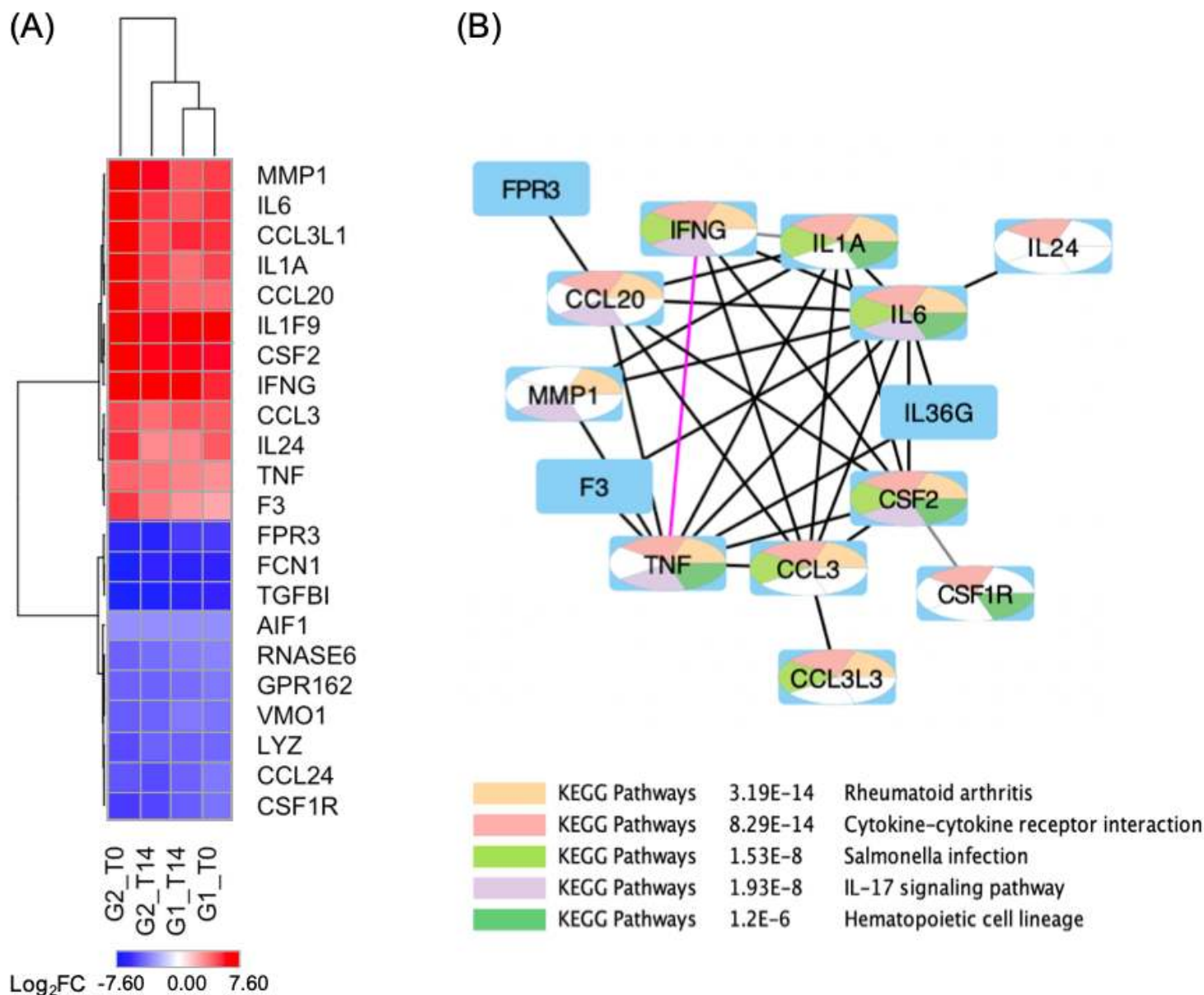


Figure 2

BCG's consensus gene signature (BCG-CGS) and highly enriched pathways. (A) A heat map of the log₂FC of the gene expression for differentially expressed genes representing BCG's consensus gene signature. Upregulated genes have positive log₂FC denoted in red color, and down regulated genes have negative values for log₂FC 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).

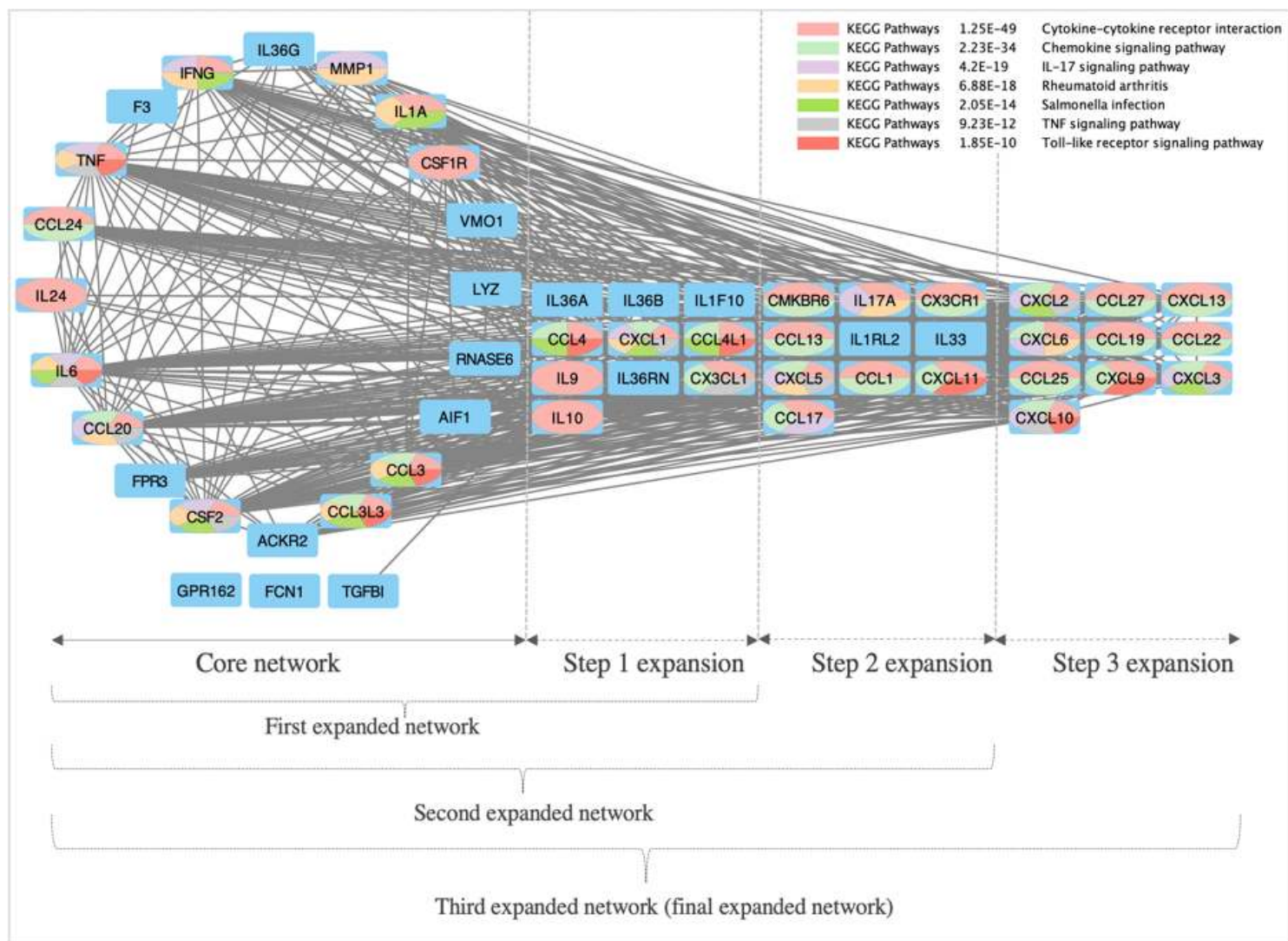


Figure 3

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.

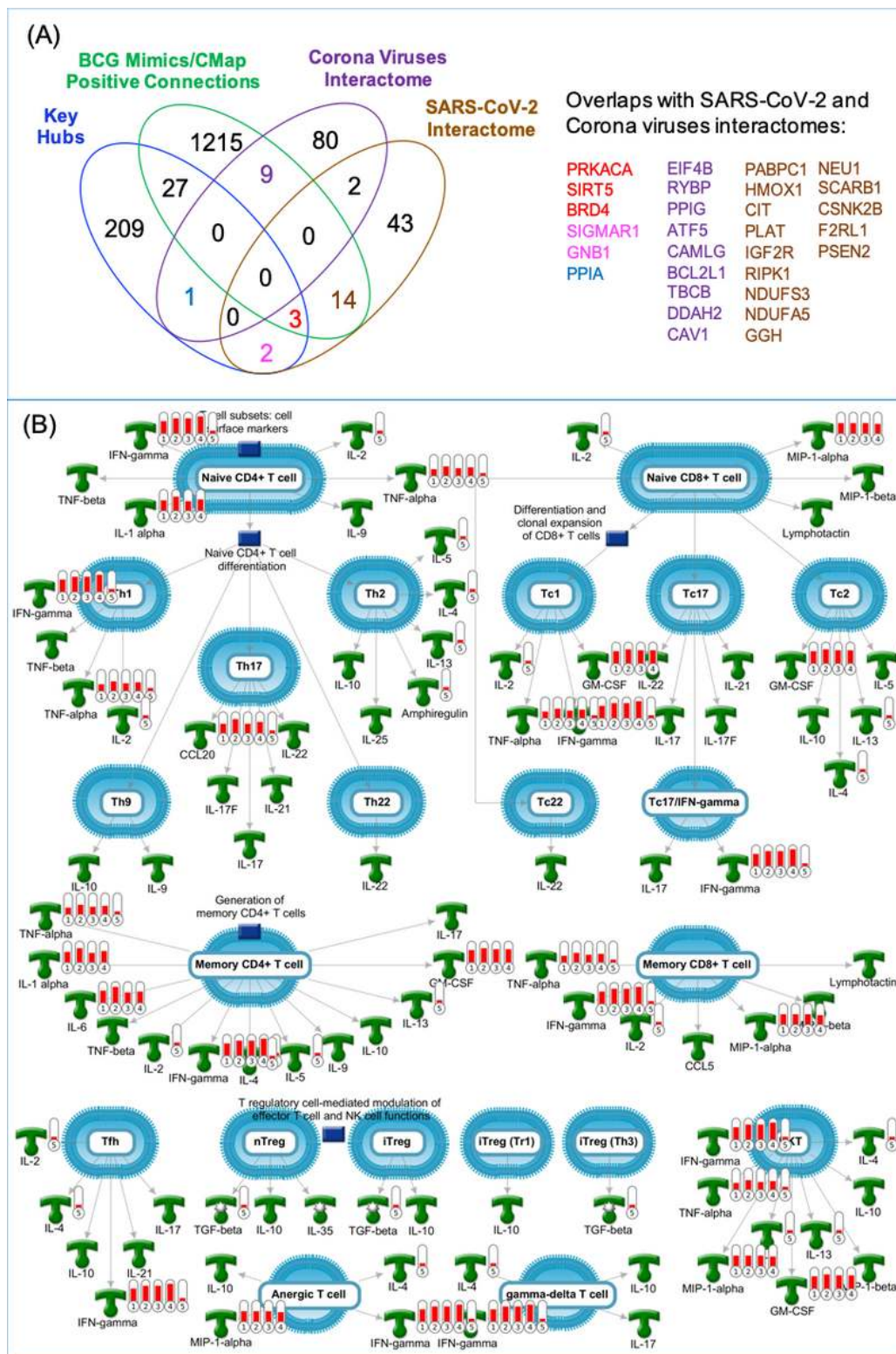


Figure 4

(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., log2 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-naïve 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-naïve 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupportingInformationHajjo.xlsx](#)