An *in silico* drug repurposing pipeline to identify drugs with the potential to inhibit SARS-CoV-2 replication

Méabh MacMahon [a,b,1], Woochang Hwang [a,1], Soorin Yim [c], Eoghan MacMahon [d],
Alexandre Abraham [e], Justin Barton [f], Mukunthan Tharmakulasingam [g], Paul Bilokon [h,i,j],
Vasanthi Priyadarshini Gaddi [k], Namshik Han [a,l,*]

[a] *Milner Therapeutics Institute, University of Cambridge, Cambridge, UK*
[b] *Centre for Therapeutics Discovery, LifeArc, Stevenage, UK*
[c] *Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea*
[d] *UCD School of Chemistry, University College Dublin, Dublin, Ireland*
[e] *Implicity, Paris, France*
[f] *Institute of Structural and Molecular Biology, Birkbeck, University of London, London, UK*
[g] *Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK*
[h] *Department of Computing, Imperial College London, London, UK*
[i] *Department of Mathematics, Imperial College London, London, UK*
[j] *Thalesians Ltd, London, UK*
[k] *Centre for Genetics and Genomics Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK*
[l] *Cambridge Centre for AI in Medicine, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK*

ARTICLE INFO

ABSTRACT

Background: Drug repurposing provides an opportunity to redeploy drugs, which are already tested for safety or approved for use in humans, for the treatment of diseases distinct from their primary indication. For example, the repurposing of dexamethasone and baricitinib has played a crucial role in saving patient lives during the SARS-CoV-2 pandemic. There remains a need to expand therapeutic approaches to prevent life-threatening complications in vulnerable patients with COVID-19.
Method: Using an *in silico* approach based on structural similarity to drugs already in clinical trials for COVID-19, potential drugs were predicted for repurposing. For a subset of identified drugs with different targets to their corresponding COVID-19 clinical trial drug, a mechanism of action analysis based on affected pathways in the context of the first 24 h of infection was applied to establish whether they might have a role in inhibiting the replication of SARS-CoV-2.
Results: Twenty-nine structurally similar drugs with potential for repurposing against COVID-19 were predicted. Two of these with the potential to inhibit SARS-CoV-2 replication were identified using mechanism of action analysis. Triamcinolone is a corticosteroid that is structurally similar to dexamethasone; gallopamil is a calcium channel blocker that is structurally similar to verapamil.
Conclusion: The identification of these drugs as potentially useful for patients with COVID-19 who are at a higher risk of developing severe disease supports the use of *in silico* approaches to facilitate quick and cost-effective identification of drugs for repurposing. Such drugs could expand the number of treatments available to the subset of patients who are not fully protected by vaccination.

1. Introduction

The novel virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes Coronavirus disease 2019 (COVID-19), was first identified in Wuhan in December 2019 and by March 2023 had spread to 192 countries or regions and infected over 670 million people,

---

killing more than 6 million [1]. COVID-19 infection can cause symptoms which fall on a wide spectrum. Those with mild infections may be asymptomatic, have anosmia [2] (the loss of ability to detect smells), or mild respiratory symptoms [3] and need no treatment, whereas in patients that develop severe infections, COVID-19 pathogenesis can progress to acute respiratory distress syndrome (ARDS), cytokine storm [4,5], and in some patients, eventually death. Certain subpopulations are known to be at higher risk of developing severe COVID-19 compared to the whole population. These subpopulations include people over 70 years old [6], those with obesity [7], those with diabetes [7,8] pregnant women [9], and those who are recipients of a solid organ transplant [10]. Despite several highly efficacious vaccines being deployed against SARS-CoV-2 [11–14], COVID-19 continues to spread. Vaccine efficacy wanes with time [15] and this as well as breakthrough infections [16] and unequal vaccine distribution between higher and lower income regions [17] all contribute to the ongoing spread of the virus. Moreover, SARS-CoV-2 strains designated as variants of concern [18] [–] [20] also have the potential to alter infection rates [21,22].

Current treatments for those infected with SARS-CoV-2 and at high risk of developing severe disease [23] are split into early and late therapies. Therapies for early stage disease Molnupiravir [24,25] and nirmatrelvir (with ritonavir) [26,27], which prevent replication of SARS-CoV-2, can be administered in patients with mild to moderate COVID-19 disease to prevent development of severe disease. Being orally administered, these drugs have an advantage over intravenously administered monoclonal antibodies [28], as they can be prescribed for use outside of a clinical setting and are less expensive [29,30]. Hospitalised patients with severe disease can be treated with dexamethasone [31] and in the US, the Food and Drug Administration (FDA) has approved Veklury (remdesivir) for patients who are hospitalised with COVID-19 [32].

Despite these advances in the treatment of mild and severe COVID-19, new drugs are still needed to treat this disease. For example, pregnant women are a high risk population for severe disease and neither nirmatrelvir or molnupiravir are currently approved for use in this population [33,34]. There is also the risk of new variants arising which are resistant to approved drugs, as has occoured in HIV therapeutics, where mutations in HIV protease cause resistance to protease inhibitor treatment [35].

The development of a novel therapeutic can take many years [36] and attrition rates are high, often owing to concerns around safety and toxicity [37]. Drug repurposing aims to use existing approved or investigational drugs for therapeutic uses beyond the scope of the original medical indication [38]. This approach offers de-risked possibilities to identify safe, effective treatments faster and more economically than novel drug development [39]. Computational approaches are now routinely through multiple steps of the drug discovery process [40] [–] [45] and can be applied to repurposing of existing drugs [46,47]. For example, network modelling [48–50], text mining [51] and various machine learning techniques [52–54] have been applied to repurposing drugs for COVID-19.

Chemical structure similarity is another approach that can be used in drug repurposing studies. It is based on the concept that structurally similar molecules often share similar biological function, and is a concept frequently used in drug discovery [55,56]. In this study, this approach has been used to identify drug candidates from publicly available databases [57] that could be potentially be repurposed against COVID-19 based on structural similarity and similar targets to drugs already in clinical trials for COVID-19. A list of COVID-19 clinical trial drugs (C19-CTDs) was obtained from DrugBank [58]. Such C19-CTDs were used as a starting point because they have hypothesis-driven predicted utility against COVID-19 and have passed safety assessment, which is why they have been entered into these clinical trials, and these also includes the few drugs which have been approved for the treatment of COVID-19. The rationale being, that if any C19-CTD is effective in clinical trial, then a structurally similar drug (SSD) which also impacts

the same targets as the C19-CTD from which it was predicted, would be expected to have similar utility and could be rapidly prioritised for investigation. In cases where chemical similarity identified structurally similar drugs that have different targets to the C19-CTD, an additional network analysis was used to predict the probable mechanism of action and to determine whether utility in inhibiting SARS-CoV-2 replication was likely.

2. Results

2.1. Identification of drugs based on chemical structure similarity

To identify drugs that could be safely and rapidly deployed to treat COVID-19, the ChEMBL API was used to computationally predict structurally similar drugs (SSDs) to the C19-CTDs identified in DrugBank. Similarity was based on the Tanimoto coefficient (Tc), a similarity measure widely used in molecular fingerprint comparison [59]. When these analyses were initiated, there were 7,682 compounds listed in ChEMBL [57] as in clinical trials for any indication at Phase II or higher (including approved drugs). Among these 7,682 compounds, 101 compounds were determined to have structural similarity to 153C19-CTDs, with a Tanimoto coefficient >0.7 (Fig. 1).

2.2. Short-listing SSDs by step-by-step triage process

To eliminate any redundancy in the list of SSDs, several stages of filtering were applied (Fig. 2). Initially, any SSDs which were already in the list of C19-CTDs were removed. For example, etripamil is an SSD because of its similarity to C19-CTD verapamil. However, etripamil is already in clinical trials for COVID, so it is already a C19-CTD, for this reason etripamil was removed from the SSD list. Next, any duplicated drugs were removed. For instance, taribavirin and taribavirin hydrochloride were both identified as similar to ribavirin (Tc 0.76). However, since taribavirin hydrochloride is the acid salt of taribavirin, and once solubilized in the body (at approximately pH 7.4), both structures will exist in the same deprotonated form and have the same behaviour, these were considered as effective duplicates and the acid salt was removed from the list. After the removal of these duplicates, 65 unique SSDs remained. This list of 65 SSDs was compared to lists of drugs predicted from 6 other *in silico* repurposing papers which employed different techniques [48–53], 19 SSDs were predicted in other repurposing papers (Supplementary Table 1). These drugs were from 43 different Anatomical Therapeutic Chemical (ATC) classes (Supplementary Table 1), with the top 3 most frequent classes being classes J01-"Antibacterials for systemic use", L01-"Antineoplastic agents" and S01-"Ophtamoligicals. Finally, 5 antineoplastic agents (designated ATC code L01) were excluded from the list of SSDs, this is because these drugs have the potential for severe side effects [60] [–] [62] and therefore would be unsuitable to treat mild COVID-19 infections. This reduced the final list of SSDs to 60.

2.3. Prioritisation of SSDs by their biological function

For the identified SSDs which have similar chemical structure to C19-CTDs, the next stage of the analysis was to map their biological functions based on the proteins they target and the role of these proteins in the context of early SARS-CoV-2 infection. As part of this process, a previously published SARS-CoV-2-induced protein (SIP) network [48] of protein-protein interactions (PPI) that occur at 24-h after SARS-CoV-2 infection was used as the base network to investigate SSDs in the context of COVID-19. This SIP network contains a computationally constructed PPI layer that connects directly interacting proteins (DIPs) and differentially expressed proteins (DEPs) identified by mass spectrometry [48,63,64]. The 24-h SIP network was selected as it contains a high concentration of protein-protein interactions (PPI) in RNA- or viral-replication related pathways [48], which this study wanted to
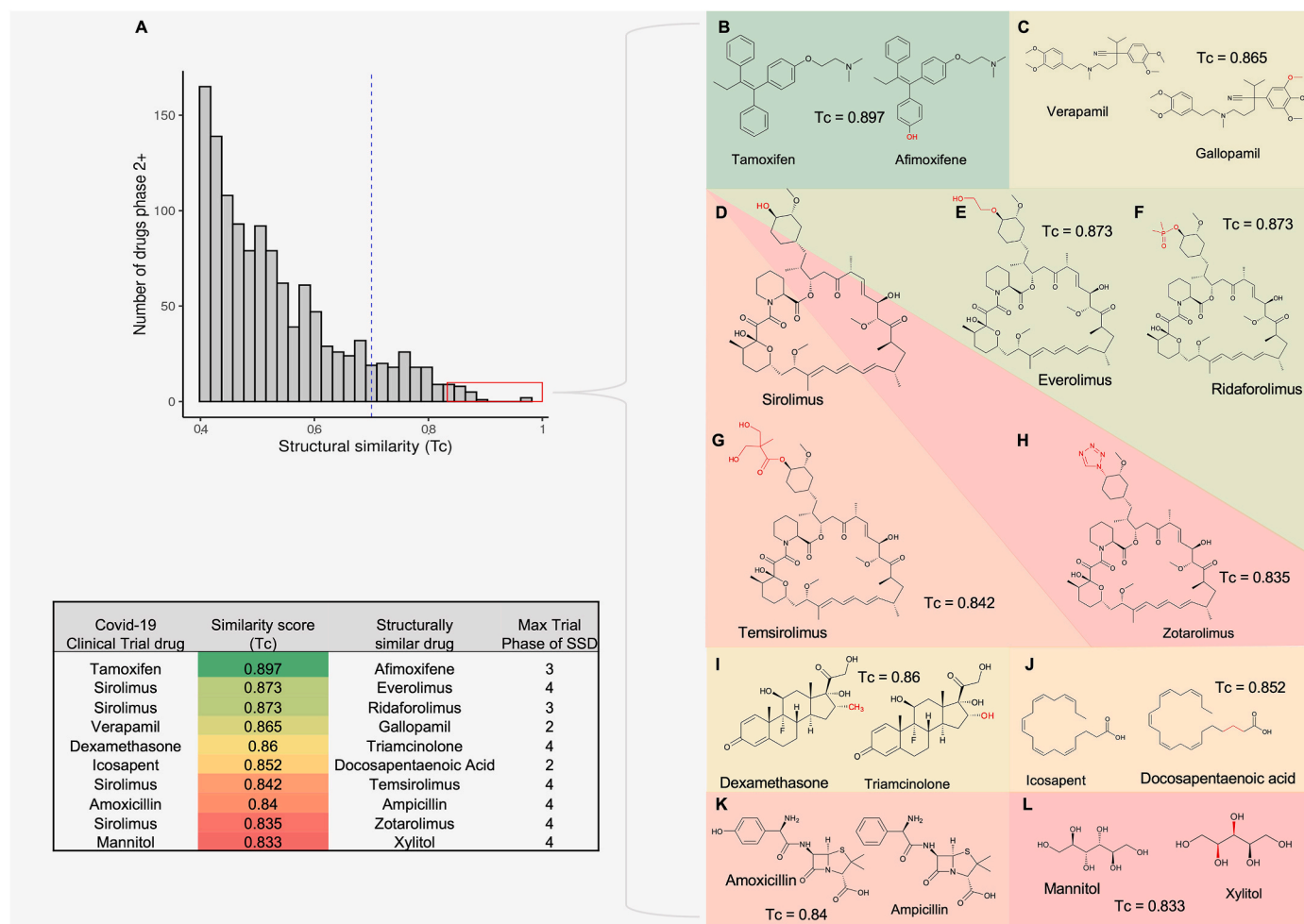
Fig. 1. Drug structural similarity screening (A) Histogram showing the number of drugs in PhaseII + clinical trials with a Tc > 0.4 and <1 that are similar to C19-CTDs. The cut off for similarity of 0.7 used in this study is shown by the blue dashed line. The red box highlights the drug structures shown in panels B–L in the form of a heatmap of similarity. (B) Tamoxifen predicts afimoxifene. (C) Verapamil predicts gallopamil. (D) Sirolimus predicts (E) everolimus, (F) ridaforolimus, (G) temirolimus, and (H) zotarolimus. (I) Dexamethasone predicts triamcinolone. (J) Icosapent predicts docosapentaenoic acid. (K) Amoxicillin predicts ampicillin. (L) Mannitol predicts xylitol. The structural differences are highlighted in red.
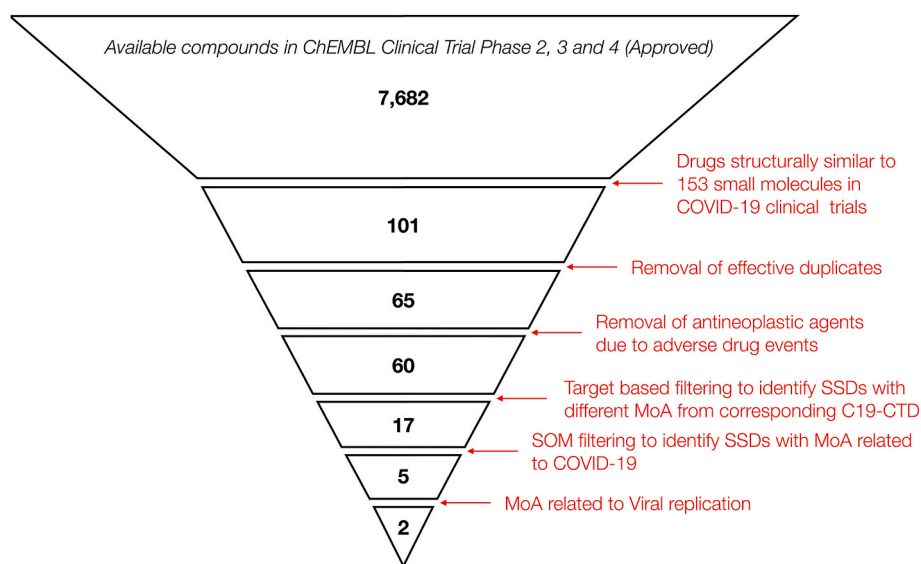

Fig. 2. Overview of SSD prediction and filtering. This figure shows the workflow used to identify SSDs from C19-CTDs and filtering steps used to and identify those with potential for treating early COVID-19 including the number of SSDs retained after each filtering step.

examine in greater detail.

To prioritise SSDs based their biological function, the targets of SSDs and C19-CTDs were compared. The comparison categorised the targets into three groups (Supplementary Table 1). In the first group, 16 SSDs among 60 short-listed SSDs have identical targets to their corresponding C19-CTD and would therefore be expected to have the same biological function. In the second group, 11 SSDs have the same targets as their corresponding C19-CTD in the context of proteins affected in early SARS-CoV-2 infection (ie. proteins within the 24-h SIP network) and might therefore be predicted to have the same biological function in this context. In the third group, 17 SSDs have some different targets from their corresponding C19-CTD, indicating they might have different biological function. Additionally, 16 SSDs were excluded at this stage due to lack of target information meaning no comparison could be carried out.

Since the third group of 17 SSDs target different proteins, they cannot be assumed to have the exact same biological implications as the C19-CTD from which they were predicted, although they could still impact the same pathways. We studied SSDs in this group farther, as although they might have no utility, some may also have an advantage in that SSDs with different biological implication to their C19-CTD could be particularly appropriate as drug repurposing candidates: if any have utility despite having some difference in biological implication to their corresponding C19-CTD then treatment options available to patients could be expanded. For this reason, SSDs in this group underwent further analysis establish their MoA in the context of early SARS-CoV-2 infection.

2.4. Mechanism of action analysis of SSDs in relation to COVID-19

An initial pathway enrichment analysis carried out on the target proteins of all 28 drugs in the third group (17 SSDs and 11 C19-CTDs
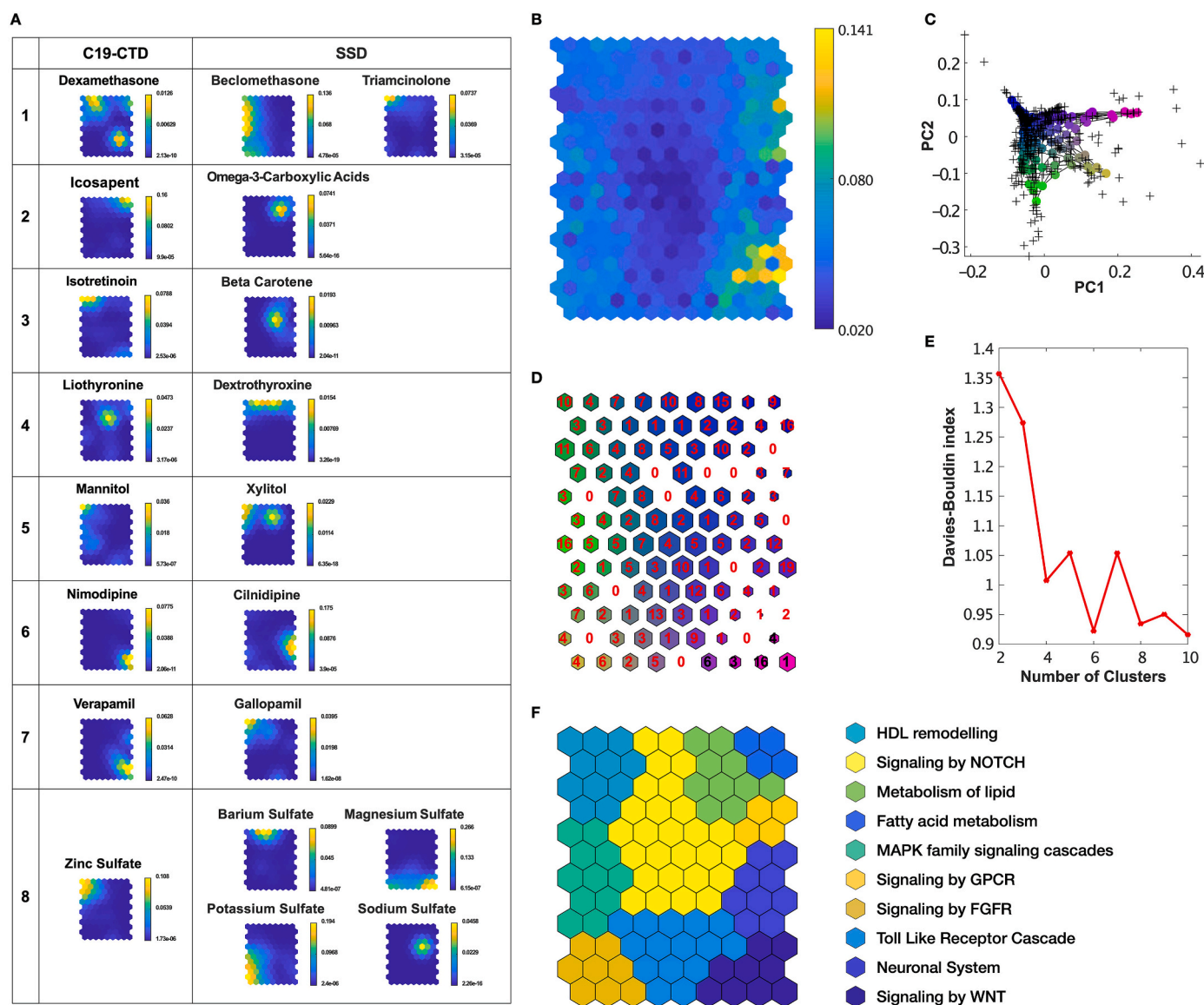

Fig. 3. Drug-pathway association study (A) 20 SOM component plane heatmaps, one for each set of SSDs which were not correlated to their corresponding C19-CTDs. In each plane heatmap, the hexagon in a certain position corresponds to a set of unique pathways which has the same position across all heatmaps (drugs). Each hexagon is coloured according to the distance between corresponding data vectors of neighbour neurons, with low distances areas (bright yellow) indicating high data density. The 488 pathways assigned to hexagons in the SOM heatmap were clustered using a unified distance matrix (U-matrix) (B), that captures the distribution of the trained artificial neurons in the data space (C). The SOM also provided the clustering information based on the U-matrix and number of hits (pathways) per hexagon (neuron) (D). In addition, the Davies–Bouldin index (DBI) was used to establish the optimal number of clusters for *k*-means clustering (E). Using this information, 488 pathways were classified into 10 pathway clusters (F).

from which they were predicted) identified a set of 488 enriched biological pathways. To measure the extent to which the targets of each drug were associated with any of these 488 pathways, an F1 score was calculated for each drug to pathway association. From these, an F1 score matrix was generated, with 488 F1 scores for each of the 28 drugs, scoring their association with each of the 488 enriched pathways.

The unsupervised training of a Self-Organizing Map (SOM) with this F1 score matrix generated 28 SOM component plane heatmaps (Supplementary Fig. 1; Fig. 3). Each heatmap represents one drug, and each hexagon in the heatmap represents unique pathways which have clustered together, and which have the same position across all heatmaps (corresponding to drugs) (Fig. 3B–F). As the position of hexagons remains consistent across all SOM heatmaps, this allows direct visual comparison between pathways affected by each drug. To check how similar drug-pathway associations were between SSDs and their corresponding C19-CTD, the correlation coefficient (r) between SOM heatmap values for each pair was calculated (Supplementary Table 2). This was carried out to narrow the focus on SSDs with different biological implications to their corresponding C19-CTDs. 12 SSDs which were not correlated with their corresponding 8C19-CTDs ($r < 0.5$) were identified (Fig. 3A). These SSDs are triamcinolone and beclomethasone (which are similar to dexamethasone), Omega-3-carboxylic acids (which is similar to icosapent), beta-carotene (which is similar to isotretinoin), dextrothyroxine (which is similar to Liothyronine), xylitol (which is similar to mannitol), Clinidipine (which is similar to nimodipine), gallopamil (which is similar to verapamil) and four SSDs similar to zinc sulfate.

2.5. In silico target validation of SSDs which have potential in treating early COVID-19

As current approved drugs for early stage COVID-19 infection prevent SARS-COV-2 replication [65], and later stage patients with severe disease may develop cytokine storm [4] (an uncontrolled immune response), it was reasoned that drugs which target viral-replication associated proteins/pathways might be more useful for the treatment of early-stage disease, while drugs whose targets are linked with immune-response-associated pathways might be better suited for later-stage COVID-19 infection. To assess whether any of the 12 SSDs would be useful in treating early COVID-19 infection the capacity of these drugs to impact pathways related to viral replication was investigated using the SOM generated data (Fig. 3F). For example, the correlation coefficient between dexamethasone and both SSDs predicted from it (beclomethasone and triamcinolone) was lower than 0.5, making it probable that these drugs have a MoA that is distinct from dexamethasone (Fig. 3A1, Supplementary Table 2). The SOM heatmap for beclomethasone is associated with pathways related to the immune response, including the estrogen receptor (ESR) [66], glycosylation [67], mitogen activated protein kinase (MAPK), activation protein 1 [68] (AP-1) and phospholipase A2 (PLA2) [69] pathways (Fig. 3F, Supplementary Table 3). Interestingly, triamcinolone is associated with the high-density lipoprotein (HDL) remodelling pathway (Fig. 3F), a pathway that is reported to be strongly associated with coronavirus replication [70,71].

Due to their association with viral-replication pathways in the SOM, by the same approach, gallopamil, omega-3-carboxylic acids and beta-carotene were also investigated as potential treatments for COVID-19 at an early stage of infection. The SOM heatmaps for gallopamil and triamcinolone are highly similar (Fig. 3A1 and Fig. 3A7), and unsurprisingly gallopamil is also associated with HDL remodelling (Fig. 3F). Omega-3-carboxylic acids is associated with fatty acid beta-oxidation, which is related to virus replication [72]. Beta-carotene is associated with the retinoic acid (RA) pathway, which is also related to virus replication. However, omega-3-carboxylic acids and beta-carotene are both listed as discontinued in the FDA approved drugs database [73] and so they were not analysed further in relation to early COVID-19. Although beta-carotene appears to be available in other jurisdictions,

it is linked with increased risk of lung cancer [74,75] and possibly pneumonia [76,77] in individuals who smoke. Smoking is associated with increased severity of disease and death in hospitalised COVID-19 patients [78].

As a result of these insights, only triamcinolone and gallopamil were analysed further for a potential impact on viral-replication in patients with early and mild to moderate COVID-19 symptoms. Triamcinolone is a corticosteroid, which is approved for the treatment of a range of inflammatory conditions including arthritis, asthma, and skin conditions. Triamcinolone inhibits nuclear factor kappa-B, which decreases the production of pro-inflammatory signals such as interleukin-6 (IL-6), interleukin-8, and monocyte chemoattractant protein-1 [79]. IL-6 has an important role in cytokine release syndrome [80], which is triggered in patients with severe COVID-19 symptoms. Sustained elevation of IL-6 is also linked to death in acute respiratory distress syndrome [81] (ARDS), the respiratory condition experienced by critically ill patients with COVID-19.

Gallopamil is an L-type calcium channel blocker and an analogue of verapamil. Gallopamil is not yet approved for the treatment of any disease, although it has been tested in a Phase II trial as an orally administered drug for patients with severe asthma [82]. Gallopamil targets membrane metalloendopeptidase (MME), which is a type 2 transmembrane glycoprotein that cleaves angiotensin converting enzyme 1 (ACE) and ACE2. Since ACE2 is a cellular entry point for SARS-CoV-2 into human cells [83], this could be a potential mechanism through which gallopamil impacts SARS-CoV-2 replication. Additionally, ACE directly interacts with the HDL remodelling pathway proteins APOA1, APOE and ALB, as well as kinases that are activated in response to viral infection (AKT1, MAPK3, and MAPK1).

2.6. Triamcinolone and gallopamil targets impact SARS-CoV-2 replication

To establish the potential effect of triamcinolone and gallopamil on the replication of SARS-CoV-2, *in silico* validation was performed on their targets within the SIP network [48]. The proximity of the targets of each drug to a group of kinases that are predicted based on the regulation of their known substrates [84] to be active in the first 24-h post infection with SARS-CoV-2 was measured. Both triamcinolone and gallopamil have 3 targets in the SIP network, and for both drugs, a permutation test showed that their targets are significantly ($P < 0.05$) closer to the active kinases identified by Bouhaddou et al. [84] than 10,000 randomly selected groups of three proteins (Fig. 4A). Indeed, targets of both these drugs directly interact with some of these active kinases identified by Bouhaddou et al. [84] (Fig. 4B).

As a further *in silico* check of the potential efficacy of triamcinolone and gallopamil in inhibiting SARS-CoV-2 replication, the activity levels (expression patterns) of the target genes of these two drugs in data from patients with moderate symptoms of COVID-19 was investigated. Viral replication is known to be active in patients with moderate COVID-19 symptoms but decreased in patients who are severely ill with COVID-19 [85]. Using the SOM results, a list of the proteins and pathways that are targeted by the two drugs was generated (Fig. 3F, Supplementary Fig. 2). To examine pathways being targeted, drugs were mapped to key pathways using SOM component plane heatmap (Fig. 3A) and pathway clusters (Fig. 3F, Supplementary Fig. 2). For example, in the plane of the SOM component of triamcinolone (Fig. 3A), the yellow hexagons are the triamcinolone associated pathways. The positions of the yellow hexagons in the triamcinolone SOM component plane can be mapped to the HDL remodelling pathway cluster, whose location can be seen in Fig. 3F. This cluster is also associated with gallopamil (Fig. 3F), and transcriptome analysis was used to identify upregulated pathways in this cluster. The neighbouring proteins of the drug's target proteins within each pathway were examined; that is, the proteins within each pathway that directly interact with the target proteins in the SIP 24-h network (Fig. 4B). The proteins in the HDL remodelling pathway were
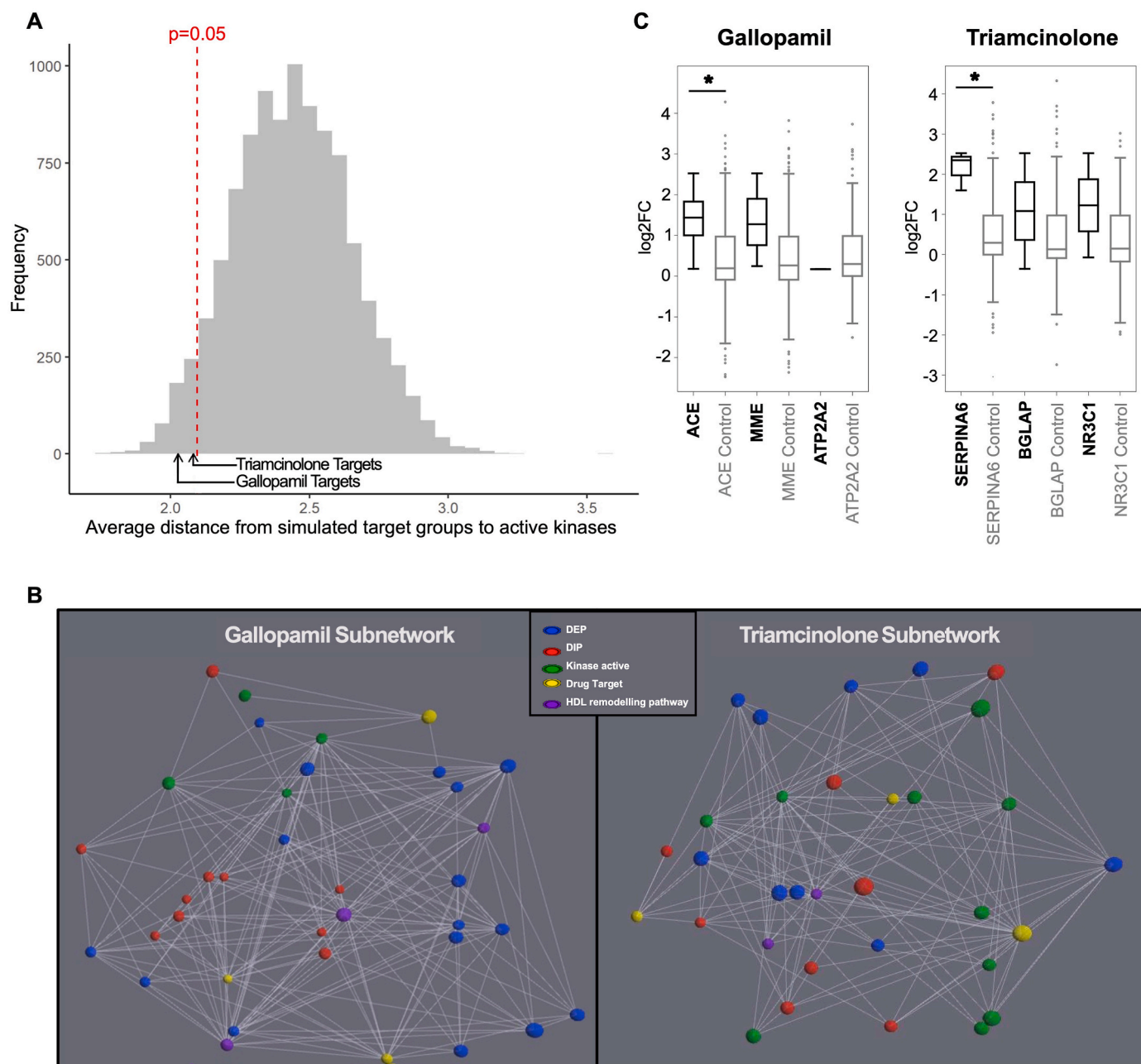
Fig. 4. *In silico* **validation of gallopamil and triamcinolone.** (A) Histogram of the results of a permutation test, the average distances between 10,000 groups of 3 proteins and kinases active at 24 h after infection with SARS-COV-2. The distances of the gallopamil and triamcinolone targets are indicated with arrows, and the red dashed line shows the permutation cut-off for the 5 % of protein groups closest to the kinases (p-value = 0.05) (B) Subnetwork of gallopamil and triamcinolone target proteins and their direct interactors showing how the targets of both drugs interact with proteins, which are proposed to be important in COVID-19 infection. The targets of both drugs (yellow) interact with DIPs (red), DEPs (blue), active kinases (green), and the proteins related to the HDL remodelling pathway (purple). (C) Boxplots showing targets of drugs in HDL remodelling pathway and their neighbouring proteins are up-regulated compared to control in patients with moderate COVID-19 symptoms. The control results in the boxplot show fold-change of random genes with the same number of target proteins and neighbour proteins as the drugs of interest. Significance was tested by the Mann Whitney *U* test (p-value <0.05).

found to be significantly upregulated compared with a random selection of the same number of proteins (the drugs target proteins plus neighbouring proteins of the HDL-remodelling pathway) (Mann Whitney *U* test p-value <0.05, Fig. 4C, Supplementary Table 4). Of the targets of the two drugs, ACE, and serpin peptidase inhibitor clade A member 6 (SERPINA6) directly interacted with APOA1 in the SIP 24-h network. APOA1, which is a major transporter of HDL-cholesterol and is essential for HDL metabolism and remodelling [86], has been linked *in silico* to the severity of COVID-19 [87,88]. The subnetwork for each drug (Fig. 4B) shows that both gallopamil and triamcinolone interacted with

proteins that directly interact with SARS-CoV-2, with kinases proposed to be active during early infection, with the HDL remodelling pathway including APOA1, and with proteins that show expression changes in the first 24 h of COVID-19 infection (Fig. 4B, Supplementary Video 1). As the target proteins of gallopamil and triamcinolone directly interact with these key proteins, these data indicate mechanisms by which these two drugs might impact early COVID-19 infection.

3. Discussion

In this study, an *in silico* approach was taken to identify drugs that have structural similarity to drugs already in clinical trials for the treatment of COVID-19. Of 65 unique drugs predicted, 19 have also been predicted by other *in silico* repurposing studies targeting COVID-19, which used different approaches to this study. This overlap with other repurposing studies suggests lends support to the validity of this approach.

Of the 60 predicted drugs, 29 can be considered from this analysis to have some potential for repurposing in COVID-19. Firstly, 16 have a highly similar structure and the exact same targets, so would be predicted to have the same biological activity, therefore should the drug to which they are similar succeed in clinical trials, these drugs would be expected to be similarly useful in patients with COVID-19 at the same stage of disease pathogenesis as their corresponding C19-CTD. Additionally, 11 other drugs have the same protein targets as their corresponding C19-CTD within an interaction network containing proteins that are predicted to be active in patients 24 h after infection with SARS-CoV-2, and should their corresponding C19-CTD be approved in early or mild-moderate infection, these would be expected to functionally overlap, and could similarly be investigated for repurposing at this stage of infection. This study has further examined drugs that have structural similarity to drugs already in COVID-19 clinical trials but, target different pathways. The analyses from this study indicate that 2 of these, triamcinolone and gallopamil, have a distinct MoA related to virus replication and therefore might be useful in the early treatment of patients infected with SARS-CoV-2 who are at high risk of developing severe COVID-19. This supports the finding in a study by Zhou et al. [50] published before there was available interactome data for SARS-CoV-2, which examined network interactomes of other human coronaviruses including SARS-CoV and MERS-CoV, and with which this study had 3 overlapping drugs, that triamcinolone was found to be significantly associated with the MERS-CoV-host interactome. With regards to gallopamil in the context of repurposing, no other study has proposed this drug. However there are several studies which suggest the use of calcium channel blockers has been associated with decreased mortality in patients hospitalised with COVID-19 [89–91], particularly in those patients with hypertension [90,91], which is a risk factor for severe infection and death with COVID-19 [92,93] and gallopamil is a calcium channel blocker.

This *in silico* approach is based on publicly available data from online databases such as STITCH [94] and ChEMBL [57], and studies published in response to the pandemic [48,64,84,85]. The focus of this study on drugs that are in clinical trials for COVID-19 was a logical starting point given the urgent nature of the COVID-19 pandemic, but this has some obvious limitations. Firstly our understanding of COVID-19 is still developing, and as the controversy surrounding hydroxychloroquine [95] highlights, a drug being in a COVID-19 clinical trial does not necessarily mean it is likely to be effective. A limitation of structural similarity-based approaches such as this is that investigations are limited to the chemical structure data available. Results are limited to similarity to the starting set (in this case C19-CTDs) which is limited by its nature, and are identified from within the chemical space available to screen against, relying on chEMBL as with all externally curated databases limits results within database comprehensiveness and curation [96]. Additionally, this approach relies on the assumption that similar structure has similar biological activity [97,98] whereas small changes in chemical structure may change the way a molecule interacts with a target protein meaning structurally similar compounds, even those with the same targets, may not have similar activity. Instances where two compounds are structurally similar, with the same targets, but different potency levels when interacting with targets, are known as activity cliffs [99]. An additional limitation of this study is that owing to the chemical-structure-based approach the focus was on small molecules only and did not take biologics into account. Polypharmacology is a term describing when drugs interact with multiple targets [100], as many C19-CTDs and SSDs do (Supplementary Table 1). When drugs hit an unknown or unintended biological target this is known as an off-target effect [101]. Off-target effects can cause adverse drug reactions in patients [102]. In this paper we focused on annotated targets of drugs, specifically those active in early COVID-19 infection, this is a limitation as it is unknown how inhibiting other targets (either unknown or outside the network of proteins active in early infection) of these drugs could impact disease progression. The decision to only study SSDs which were in at least a phase 2 clinical trial was taken to ensure drugs identified for repurposing in this study were already tested for safety in humans. Online curated publicly available databases were used to obtain target information for comparing C19-CTDs to SSDs; while these are important and useful resources, they do not take account of incomplete drug information, and especially in relation to inactivity and off-target effects may be skewed towards underreporting [103]. Future *in silico* work could include using chemoinformatic tools to search for proteins with similar binding sites to predict off target effects [104]. Finally, although *in silico* validation has indicated promise for triamcinolone and gallopamil, these results are based on a protein-protein interaction (PPI) network generated from cell-derived mass spectrometry data and a PPI resource [105], which lack directionality for the protein interactions [48]. These limitations highlight the key point that *in silico* work supports experimental work, and wet lab follow-up studies in experimental *in vitro* and *in vivo* models are required to verify the clinical potential of these predictions.

3.1. Conclusions

This *in silico* study has predicted drugs that have potential for repurposing for the treatment of COVID-19. The overlap with other COVID-19 drug repurposing studies lends support to this approach. A further MoA analysis predicts triamcinolone and gallopamil as drugs with potential for targeting the replication of SARS-CoV-2. Of the two, triamcinolone is already approved for other indications, so could be deployed quickly. Gallopamil has passed Phase II trials and could be quickly prioritised into COVID-19 trials. All drugs predicted require further investigation in *in vitro* and *in vivo* models, as well as clinical trials. It is likely that further diseases of zoonotic origin will emerge in the future [106], and the scientific community must be prepared to respond rapidly to future pandemic threats. This kind of *in silico* approach allows for a quick and cost-effective approach to repurposing drugs, which could be deployed in such a situation.

4. Materials and methods

4.1. Obtaining list of structurally similar drugs

A list of 184 COVID-19 clinical trial drugs (C19-CTDs) was obtained from DrugBank [58] (date accessed 25/05/2020). There were 153 small molecules (Supplementary Table 5) and 31 biotech drugs, as this analysis was based on chemical structure, biotech drugs were excluded. Drugbank is an online bioinformatics and chemoinformatics database which contains data on drugs and their targets, for instance the structure of a drug and any clinical trials it has been involved in as well as approval status. ChEMBL is a database containing bioactivity data for drugs and drug-like molecules and tools for mining this information, including by searching for similar chemical structures within the ChEMBL database itself [57]. The ChEMBL application programming interface (API) for chemical structural similarity accepts input in the simplified molecular-input line-entry system (SMILES) format, which is a format for represents chemical structures using ascii strings [107]. For each C19-CTD, the structure in SMILES format was extracted from the DrugBank data (script in code repository). The ChEMBL API library (written in python) was used to obtain a list of structurally similar drugs (SSDs) which were phase II clinical trial or higher from ChEMBL (script

in code repository). SSDs that were in Phase II clinical trials or higher for any condition were retained at this stage because they passed Phase I trials and have therefore been tested for safety in humans. The ChEMBL structural similarity tool used is based on the Tanimoto coefficient (Tc), a similarity measure widely used in molecular fingerprint comparison [59].

4.2. Filtering the list of SSDs

The list of SSDs was refined to remove drugs which fell into several categories (Fig. 2). Tc > 0.7 was used as a similarity cut off. For context, a threshold of 0.7–0.75 is typical for structure based screening [108]. The list of SSDs was cleaned to remove drugs which were already C19-CTDs, drugs which are synonyms of C19-CTDs (using synonym information from Drugbank) and effective duplicates that behave the same in the body despite of a slightly different chemical structure (using salt information from Drugbank).

4.3. Obtaining target information for drugs

Target information for all drugs was collected from target information in DrugBank [58], targets in Therapeutic Target Database (TTD) [109] and targets with interaction confidence score ≥ 0.7 in drug to target interactions from STITCH (v 5.0) [110], a score of 0.7 is considered high confidence by STITCH [111]. TTD is a database containing target-specific drug binding information about known therapeutic targets from literature [109]. STITCH is a database containing scored interactions between proteins and small molecules which has evidence from experimental/biochemical data, association in curated databases and co-mention in abstracts from the scientific literature [94].

4.4. Filtering of drug list based on target information

The SARS-CoV-2-induced protein (SIP) network at 24 h generated by Han et al. [48] was used as the base network to investigate the COVID-19 pathways on which these structurally similar drugs (SSDs) are acting. This SIP network is a network constructed *in silico* of the protein-protein interaction pathways which are activated at 24 h post COVID-19 infection. The methods for construction of this network are described in the Han et al. paper [48]. Briefly, this network contains the directly interacting proteins (DIP) of SARS-CoV-2 identified by mass spectrometry [64], differentially expressed proteins (DEP) after infection by SARS-CoV-2 identified by mass spectrometry [63], and a layer of connections between the two built by obtaining all shortest paths between DIP proteins and DEP proteins using known protein-protein interaction (PPI) data as described in the Han et al. paper [48].

4.5. Distance between targets of SSD and the targets of C19-CTD

The distance dc(S, C) is defined between S, the set of target proteins of a SSD, and C, the set of target proteins of a C19-CTD, as the shortest path length d(s, c) between all pairs of nodes s ∈ S and c ∈ C in the SIP 24-h network [48]. Closest distance measure (equation (1)) was used to calculate the distance between a SSD's target to its corresponding C19-CTD's target in the SIP 24-h network because it showed the best performance in drug-drug relationship prediction in a study by Cheng et al. [112]. If the average distance <1 between the targets of any SSD and its corresponding C19-CTD in the SIP 24-h network [48] the two drugs don't have different targets in the network.

$$d_c(S, C) = \frac{1}{\|C\|} \sum_{c \in C} \min_{s \in S} d(s, c) \tag{1}$$

4.6. Drug-pathway associations

To identify key pathways that are significantly enriched in the proteins that are targeted by 28 SSDs and C19-CTDs, pathway enrichment analysis was conducted using R (v 3.5.2) package gprofiler2 [113] gost function, which uses a hypergeometric test to identify genes over-represented in known pathways for functional enrichment analysis of gene list (FDR-BH < 0.05). Reactome is a database of manually curated and peer reviewed biological interactions, including curated human pathways organised based on biological function [114]. Although there are a number of pathway databases [115,116], Reactome (accessed on 15/05/2020) was chosen for pathway enrichment analysis because it is the most actively updated public database containing human pathways [117]. To understand the MoA of the 28 drugs, pathway enrichment analysis was performed on the targets of each of these drugs. The number of target proteins of the 28 drugs ranges from 2 to 125. The average number of target proteins of the 28 drugs is 14. Drugs targeting fewer than six proteins have too few protein targets to have significantly enriched pathways. To overcome this problem for drugs with fewer than the average number of target proteins (ie. 14), neighbour proteins (ie. proteins with which they directly interact) of the target proteins in the SIP 24-h network were included in their target list for enrichment analysis, up to 14 proteins. Significantly enriched biological pathways of target proteins for each of the 28 drugs were integrated, resulting in 488 key pathways. The Reactome pathway database has a hierarchical structure of biological pathway. For a given function, the lower hierarchy pathways are more specific than the higher hierarchy pathways. Additionally, the parent pathway semantically includes its child pathways. For instance, the parent (highest) hierarchy pathway "Transport of s mall molecules" contains "Plasma lipoprotein assembly, remodelling, and clearance" among its child pathways, which in turn contains "Plasma lipoprotein remodelling" as a child pathway, "HDL remodelling" is contained within this pathway, therefore it is also contained within the higher hierarchy levels. In the process of integrating the enriched pathways per drug, the lowest-possible hierarchy pathways were used to avoid overlapping biological meaning among the hierarchical pathways.

Based on these identifications, a matrix containing F1 scores of the 28 drugs and the 488 key pathways was generated for drug-pathway association. F1 score is the harmonic mean of precision and recall values in a classification and is an accuracy metric used here to calculate enrichment accuracy. The Reactome pathway enrichment analysis for the 28 drugs using gprofiler2 [118] provides enrichment p-values, precision and recall information that were used to produce the F1 scores. The meaning of precision here is that the proportion of a drugs target proteins that are annotated to the pathway. The meaning of recall here is that the proportion of the pathway gene sets that a drugs target proteins are located in. Precision and recall were used to construct a matrix of F1 score (F1 = 2(precision x recall)/(precision + recall) from the pathway enrichment analysis.

4.7. SOM analysis of pathways impacted by targets of drugs

Self-Organizing Map (SOM) [117] was used to analyse the MoA of the candidate drugs based on pathways impacted by their targets in this analysis. SOM, a type of artificial neural network, is an unsupervised machine learning technique used for dimensional reduction and clustering. Due to the map representation feature of SOM, this technique also has an ability to describe and visualise high dimensional data. For this reason, it was useful to directly compare the SOM component heatmaps of the 28 drugs for COVID-19 with one another. SOM also has the advantage of dimensional reduction to allow a clearer clustering result. SOM was used to calculate the low-dimensional abstractions, which are then clustered using k-means. This two-phase approach increases the efficiency of k-means clustering with a relatively small number of samples, a known limitation in hierarchical clustering algorithms including k-means [119]. Another advantage of SOM is noise reduction because the SOM abstractions are less sensitive to random variations than the input data. In addition, SOM offers a systematic

arrangement of the candidate drugs to each neuron (or node) and hence to pathway clusters (Fig. 3C).

The data used in training was the F1 score matrix for Drug-Pathway associations (key pathways to candidate drugs). From the SOM training results a Unified distance matrix is generated which captures the distribution of the trained artificial neurons in the data space, this contains the vector norms between the neighbouring SOM nodes and shows data density in input space. Each sub-unit is coloured according to distance between corresponding data vectors of neighbour units. Low distances areas (dark blue) have high data density (clusters) (Fig. 3B). Davies-Bouldin (DB) index [120], a method for identifying the optimal k number for k-means clustering,was calculated based on the U-matrix to determine the optimal number of clusters. DB index is calculated by obtaining the average similarity of each cluster with the cluster most similar to it. The lower the DB index is, the better separated clusters are. The lowest DB index value occurred at 10 clusters (Fig. 3F). The k-means algorithm was then used to cluster pathways (Fig. 3F). The SOM component maps of key pathways were analysed based on the k-means clustering result (Fig. 3F) and mapped into two MoA categories based on the biological functions (Fig. 3F). The mapping result of key pathways to clusters is available in Supplementary Table 3. The SOM Toolbox package [121] for MatLab was used for this analysis with default settings and parameters.

4.8. In silico mechanism of action analysis of triamcinolone and gallopamil

To identify MoA for triamcinolone and gallopamil in COVID-19 their relationships within the SIP 24-h SARS-CoV-2 network were identified. There are 14,827 proteins and 528,969 interactions in this network. Bouhaddou et al. [84] predicted kinase activities in the first 24h post infection with SARS-CoV-2 based on the regulation of their known substates. From their data all kinases with activity absolute LogFC >1.5 at any time point in the first 24h post infection for verification were used as these are being activated in early SARS-CoV-2 infection. To estimate the proximity of the targets of the drugs of interest to these active kinases in the SIP network, the distance of each target to each kinase was calculated and the average of these distances was taken. To establish significance, a permutation test was used as follows; each drug had 3 targets in the network, so 10,000 groups of 3 proteins present in the network were randomly generated and the average distance of these from the active kinases was calculated (Fig. 4A, Supplementary Video 1). A permutation test is an exact test to establish statistical significance of a null hypothesis by randomly resampling all data points [122], it has been applied in network analysis based drug repurposing [48,50,112]. The closest 5 % distances to the kinases were used as a cut off to establish significance of P < 0.05. To visualise the target proteins interacting with these kinases and other key COVID-10 proteins, subnetworks were plotted using Virtualitics Immersive Platform [123]. An additional movie file shows this in more detail (Supplementary Video 1)

4.9. Expression analysis of the drugs target proteins using COVID-19 patients' data

To indirectly validate the effects of triamcinolone and gallopamil in COVID-19, it was established whether the target proteins of these drugs are significantly changed between patients with moderate and severe COVID-19 as compared to the expression of random proteins. Arunnachalam et al. [85] provide Log2 fold change gene expression values between 4 COVID-19 moderate patients compared to 12 COVID-19 severe patients. In the boxplot shown in Fig. 4C Expression of proteins belonging to pathways in the HDL remodelling cluster (Fig. 3F) which were also among the neighbour proteins of each of the drug's targets in the SIP 24-h network were compared to the same number of randomly selected control proteins, comparisons for the significantly upregulated pathway are shown in Fig. 4C. To ensure the control proteins were unbiased, this random choice was repeated for 100 permutations to create control protein expression values. For example, if five of the target proteins and their neighbour proteins belong to the pathway in the HDL remodelling cluster, 500 proteins were randomly selected and made into control proteins. The Mann Whitney *U* test, a non-parametric version of the independent samples *t*-test [124], was performed to determine whether the expression two groups were significantly different between moderate and severe COVID-19 patients.

Data availability statement

This study uses publicly available data from online databases and other research papers as cited in the text. No new data were generated. Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Namshik Han (nh417@cam.ac.uk). Code is available at the GitHub repository https://github.com/wchwang/COVID-19.HfH.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Namshik Han reports financial support was provided by LifeArc. Meabh MacMahon reports financial support was provided by LifeArc. Woochang Hwang reports financial support was provided by LifeArc. MM is an employee of LifeArc. NH is a cofounder of KURE.ai and CardiaTec Biosciences and an advisor at Biorelate, Promatix, Standigm and Vera-Verse. SY, EM, AA, JB, MT, PB and VPG declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2023.101387.

References

[1] Johns Hopkins Coronavirus Resource Center. COVID-19 map. Johns Hopkins Coronavirus Resour Cent; 2020. https://coronavirus.jhu.edu/map.html [Accessed March 2023].

[2] Walker A, Pottinger G, Scott A, Hopkins C. Anosmia and loss of smell in the era of covid-19. BMJ 2020;370:m2808. https://doi.org/10.1136/bmj.m2808.

[3] Gandhi RT, Lynch JB, del Rio C. Mild or moderate covid-19. N Engl J Med 2020; 383:1757–66. https://doi.org/10.1056/NEJMcp2009249.

[4] Jiang Y, Rubin L, Peng T, Liu L, Xing X, Lazarovici P, et al. Cytokine storm in COVID-19: from viral infection to immune responses, diagnosis and therapy. Int J Biol Sci 2022;18:459–72. https://doi.org/10.7150/ijbs.59272.

[5] Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet Lond Engl 2020;395:1033–4. https://doi.org/10.1016/S0140-6736(20)30628-0.

[6] Pijls BG, Jolani S, Atherley A, Derckx RT, Dijkstra JIR, Franssen GHL, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. BMJ Open 2021;11:e044640. https://doi.org/10.1136/bmjopen-2020-044640.

[7] Zhou Y, Chi J, Lv W, Wang Y. Obesity and diabetes as high-risk factors for severe coronavirus disease 2019 (Covid-19). Diabetes Metab Res Rev 2021;37:e3377. https://doi.org/10.1002/dmrr.3377.

[8] Hartmann-Boyce J, Morris E, Goyder C, Kinton J, Perring J, Nunan D, et al. Diabetes and COVID-19: risks, management, and learnings from other national disasters. Diabetes Care 2020;43:1695–703. https://doi.org/10.2337/dc20-1192.

[9] Boushra MN, Koyfman A, Long B. COVID-19 in pregnancy and the puerperium: a review for emergency physicians. Am J Emerg Med 2021;40:193–8. https://doi.org/10.1016/j.ajem.2020.10.055.

[10] Pereira MR, Mohan S, Cohen DJ, Husain SA, Dube GK, Ratner LE, et al. COVID-19 in solid organ transplant recipients: initial report from the US epicenter. Am J Transplant 2020;20:1800–8. https://doi.org/10.1111/ajt.15941.

[11] Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. N Engl J Med 2020. https://doi.org/10.1056/NEJMoa2034577.

[12] Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet 2021;397:99–111. https://doi.org/10.1016/S0140-6736(20)32661-1.

[13] Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, et al. Safety and efficacy of single-dose Ad26.COV2.S vaccine against covid-19. N Engl J Med 2021;384:2187–201. https://doi.org/10.1056/NEJMoa2101544.

[14] Chagla Z. high-risk adults, the Moderna vaccine had 94% efficacy against COVID-19 ≥14 d after the 2nd dose. Ann Intern Med 2021;174:JC28. https://doi.org/10.7326/ACPJ202103160-028.

[15] Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman L, Haas EJ, et al. Waning immunity after the BNT162b2 vaccine in Israel. N Engl J Med 2021;385: e85. https://doi.org/10.1056/NEJMoa2114228.

[16] Juthani PV, Gupta A, Borges KA, Price CC, Lee AI, Won CH, et al. Hospitalisation among vaccine breakthrough COVID-19 infections. Lancet Infect Dis 2021;21: 1485–6. https://doi.org/10.1016/S1473-3099(21)00558-2.

[17] Rzymski P, Camargo CA, Fal A, Flisiak R, Gwenzi W, Kelishadi R, et al. COVID-19 vaccine boosters: the good, the bad, and the ugly. Vaccines 2021;9:1299. https://doi.org/10.3390/vaccines9111299.

[18] Mishra S, Mindermann S, Sharma M, Whittaker C, Mellan TA, Wilton T, et al. Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England. EClinicalMedicine 2021;39:101064. https://doi.org/10.1016/j.eclinm.2021.101064.

[19] Slavov SN, Patané JSL, Bezerra R dos S, Giovanetti M, Fonseca V, Martins AJ, et al. Genomic monitoring unveil the early detection of the SARS-CoV-2 B.1.351 (beta) variant (20H/501Y.V2) in Brazil. J Med Virol 2021;93:6782–7. https://doi.org/10.1002/jmv.27190.

[20] Vöhringer HS, Sanderson T, Sinnott M, De Maio N, Nguyen T, Goater R, et al. Genomic reconstruction of the SARS-CoV-2 epidemic in England. Nature 2021; 600:506–11. https://doi.org/10.1038/s41586-021-04069-y.

[21] Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. Lancet Infect Dis 2022;22:35–42. https://doi.org/10.1016/S1473-3099(21)00475-8.

[22] Sievers BL, Chakraborty S, Xue Y, Gelbart T, Gonzalez JC, Cassidy AG, et al. Antibodies elicited by SARS-CoV-2 infection or mRNA vaccines have reduced neutralizing activity against Beta and Omicron pseudoviruses. Sci Transl Med 2022. https://doi.org/10.1126/scitranslmed.abn7842.

[23] Cohen MS, Wohl DA, Fischer II WA, Smith DM, Eron JJ. Outpatient treatment of severe acute respiratory syndrome coronavirus 2 infection to prevent coronavirus disease 2019 progression. Clin Infect Dis 2021;73:1717–21. https://doi.org/10.1093/cid/ciab494.

[24] Kabinger F, Stiller C, Schmitzová J, Dienemann C, Kokic G, Hillen HS, et al. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. Nat Struct Mol Biol 2021;28:740–6. https://doi.org/10.1038/s41594-021-00651-0.

[25] Malone B, Campbell EA. Molnupiravir: coding for catastrophe. Nat Struct Mol Biol 2021;28:706–8. https://doi.org/10.1038/s41594-021-00657-8.

[26] Ionescu MI. An overview of the crystallized structures of the SARS-CoV-2. Protein J 2020;39:600–18. https://doi.org/10.1007/s10930-020-09933-w.

[27] Pfizer. Pfizer's novel COVID-19 oral antiviral treatment candidate reduced risk of hospitalization or death by 89% in interim analysis of phase 2/3 EPIC-HR study. https://www.pfizer.com/news/press-release/press-release-detail/pfizers-novel-covid-19-oral-antiviral-treatment-candidate. [Accessed 19 January 2022].

[28] Office of the Commissioner. Emergency use authorization. FDA. https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization. [Accessed 22 March 2021].

[29] Hernandez I, Bott SW, Patel AS, Wolf CG, Hospodar AR, Sampathkumar S, et al. Pricing of monoclonal antibody therapies: higher if used for cancer? Am J Manag Care 2018;24:109–12.

[30] GlobalData. Eli Lilly's COVID-19 mAb bamlanivimab potential cost effectiveness underscores payer support. GlobalData 2020. https://www.globaldata.com/eli-lillys-covid-19-mab-bamlanivimab-potential-cost-effectiveness-underscores-payer-support/. [Accessed 23 March 2021].

[31] WHO. Clinical management of COVID-19. https://www.who.int/publications-detail-redirect/clinical-management-of-covid-19. [Accessed 14 January 2021].

[32] Center for Drug Evaluation and Research. Know your treatment options for COVID-19. FDA 2021. https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19. [Accessed 22 March 2021].

[33] EMA. EMA issues advice on use of Paxlovid (PF-07321332 and ritonavir) for the treatment COVID-19: rolling review starts in parallel. Eur Med Agency 2021. https://www.ema.europa.eu/en/news/ema-issues-advice-use-paxlovid-pf-07321332-ritonavir-treatment-covid-19-rolling-review-starts. [Accessed 31 January 2022].

[34] Commissioner O of the. Coronavirus (COVID-19) update: FDA authorizes additional oral antiviral for treatment of COVID-19 in certain adults. FDA 2021. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-additional-oral-antiviral-treatment-covid-19-certain. [Accessed 31 January 2022].

[35] Agniswamy J, Shen C-H, Aniana A, Sayer JM, Louis JM, Weber IT. HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. Biochemistry 2012;51:2819–28. https://doi.org/10.1021/bi2018317.

[36] Prasad V, Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. JAMA Intern Med 2017;177: 1569–75. https://doi.org/10.1001/jamainternmed.2017.3601.

[37] Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov 2015;14:475–86. https://doi.org/10.1038/nrd4609.

[38] Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 2004;3:673–83. https://doi.org/10.1038/nrd1468.

[39] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov 2019;18:41–58. https://doi.org/10.1038/nrd.2018.168.

[40] Katsila T, Spyroulias GA, Patrinos GP, Matsoukas M-T. Computational approaches in target identification and drug discovery. Comput Struct Biotechnol J 2016;14: 177–84. https://doi.org/10.1016/j.csbj.2016.04.004.

[41] Wu R, Zhou Y, Li L, Shen X, Chen G, Wang X, et al. Computational approaches in preclinical studies on drug discovery and development. Front Chem 2020;8.

[42] Kumar A, Voet A, Zhang KYJ. Fragment based drug design: from experimental to computational approaches. Curr Med Chem 2012;19:5128–47. https://doi.org/10.2174/092986712803530467.

[43] Hoffer L, Muller C, Roche P, Morelli X. Chemistry-driven hit-to-lead optimization guided by structure-based approaches. Mol Inform 2018;37:1800059. https://doi.org/10.1002/minf.201800059.

[44] Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. Comput Biol Med 2021;137:104851. https://doi.org/10.1016/j.compbiomed.2021.104851.

[45] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today 2021;26:80–93. https://doi.org/10.1016/j.drudis.2020.10.010.

[46] Park K. A review of computational drug repurposing. Transl Clin Pharmacol 2019;27:59–63. https://doi.org/10.12793/tcp.2019.27.2.59.

[47] Hwang W, Han N. Identification of potential pan-coronavirus therapies using a computational drug repurposing platform. Methods 2021. https://doi.org/10.1016/j.ymeth.2021.11.002.

[48] Han N, Hwang W, Tzelepis K, Schmerer P, Yankova E, MacMahon M, et al. Identification of SARS-CoV-2–induced pathways reveals drug repurposing strategies. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abh3032. eabh3032.

[49] Sibilio P, Bini S, Fiscon G, Sponziello M, Conte F, Pecce V, et al. In silico drug repurposing in COVID-19: a network-based analysis. Biomed Pharmacother 2021; 142:111954. https://doi.org/10.1016/j.biopha.2021.111954.

[50] Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6: 1–18. https://doi.org/10.1038/s41421-020-0153-3.

[51] Tworowski D, Gorohovski A, Mukherjee S, Carmi G, Levy E, Detroja R, et al. COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics. Nucleic Acids Res 2021;49:D1113–21. https://doi.org/10.1093/nar/gkaa969.

[52] Loucera C, Esteban-Medina M, Rian K, Falco MM, Dopazo J, Peña-Chilet M. Drug repurposing for COVID-19 using machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection. Signal Transduct Targeted Ther 2020;5:1–3. https://doi.org/10.1038/s41392-020-00417-y.

[53] Suvarna K, Biswas D, Pai MGJ, Acharjee A, Bankar R, Palanivel V, et al. Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity with drug repurposing potential. Front Physiol 2021;12.

[54] Han L, Shan G, Chu B, Wang H, Wang Z, Gao S, et al. Accelerating drug repurposing for COVID-19 treatment by modeling mechanisms of action using cell image features and machine learning. Cogn Neurodyn 2021. https://doi.org/10.1007/s11571-021-09727-5.

[55] Kumar A, Zhang KYJ. Advances in the development of shape similarity methods and their application in drug discovery. Front Chem 2018;6. https://doi.org/10.3389/fchem.2018.00315.

[56] Capoci IRG, Faria DR, Sakita KM, Rodrigues-Vendramini FAV, Bonfim-Mendonça P de S, Becker TCA, et al. Repurposing approach identifies new treatment options for invasive fungal disease. Bioorg Chem 2019;84:87–97. https://doi.org/10.1016/j.bioorg.2018.11.019.

[57] Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res 2014;42:D1083–90. https://doi.org/10.1093/nar/gkt1031.

[58] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46: D1074–82. https://doi.org/10.1093/nar/gkx1037.

[59] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminf 2015;7:20. https://doi.org/10.1186/s13321-015-0069-3.

[60] Hamnvik O-PR, Larsen PR, Marqusee E. Thyroid dysfunction from antineoplastic agents. JNCI J Natl Cancer Inst 2011;103:1572–87. https://doi.org/10.1093/jnci/djr373.

[61] Dimopoulou I, Bamias A, Lyberopoulos P, Dimopoulos MA. Pulmonary toxicity from novel antineoplastic agents. Ann Oncol 2006;17:372–9. https://doi.org/10.1093/annonc/mdj057.

[62] Ibañez-Juliá MJ, Berzero G, Reyes-Botero G, Maisonobe T, Lenglet T, Slim M, et al. Antineoplastic agents exacerbating Charcot Marie Tooth disease: red flags to avoid permanent disability. Acta Oncol 2018;57:403–11. https://doi.org/10.1080/0284186X.2017.1415462.

[63] Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature 2020;1–8. https://doi.org/10.1038/s41586-020-2332-7.

[64] 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 2020; 1–13. https://doi.org/10.1038/s41586-020-2286-9.

[65] Petrakis V, Rafailidis P, Trypsianis G, Papazoglou D, Panagopoulos P. The antiviral effect of nirmatrelvir/ritonavir during COVID-19 pandemic real-world data. Viruses 2023;15:976. https://doi.org/10.3390/v15040976.

[66] Cunningham M, Gilkeson G. Estrogen receptors in immunity and autoimmunity. Clin Rev Allergy Immunol 2011;40:66–73. https://doi.org/10.1007/s12016-010-8203-5.

[67] Glycosylation and the Immune System. n.d, https://www.science.org/doi/10.1126/science.291.5512.2370?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub/20/200pubmed. [Accessed 3 February 2022].

[68] Liu Y, Shepherd EG, Nelin LD. MAPK phosphatases — regulating the immune response. Nat Rev Immunol 2007;7:202–12. https://doi.org/10.1038/nri2035.

[69] Nolin JD, Lai Y, Ogden HL, Manicone AM, Murphy RC, An D, et al. Secreted PLA2 group X orchestrates innate and adaptive immune responses to inhaled allergen. JCI Insight 2017;2. https://doi.org/10.1172/jci.insight.94929.

[70] Nainu F, Shiratsuchi A, Nakanishi Y. Induction of apoptosis and subsequent phagocytosis of virus-infected cells as an antiviral mechanism. Front Immunol 2017;8. https://doi.org/10.3389/fimmu.2017.01220.

[71] Tan Y-X, Tan THP, Lee MJ-R, Tham P-Y, Gunalan V, Druce J, et al. Induction of apoptosis by the severe acute respiratory syndrome coronavirus 7a protein is dependent on its interaction with the bcl-XL protein. J Virol 2007;81:6346–55. https://doi.org/10.1128/JVI.00090-07.

[72] Limsuwat N, Boonarkart C, Phakaratsakul S, Suptawiwat O, Auewarakul P. Influence of cellular lipid content on influenza A virus replication. Arch Virol 2020;165:1151–61. https://doi.org/10.1007/s00705-020-04596-5.

[73] FDA. Drugs@FDA: FDA-approved drugs. https://www.accessdata.fda.gov/scripts/cder/daf/. [Accessed 31 May 2021].

[74] Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, et al. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. J Natl Cancer Inst 1996;88:1550–9. https://doi.org/10.1093/jnci/88.21.1550.

[75] Middha P, Weinstein SJ, Männistö S, Albanes D, Mondul AM. β-Carotene supplementation and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: the role of tar and nicotine. Nicotine Tob Res 2019;21:1045–50. https://doi.org/10.1093/ntr/nty115.

[76] Hemilä H, Virtamo J, Albanes D, Kaprio J. Vitamin E and beta-carotene supplementation and hospital-treated pneumonia incidence in male smokers. Chest 2004;125:557–65. https://doi.org/10.1378/chest.125.2.557.

[77] Hemilä H. The effect of β-carotene on the mortality of male smokers is modified by smoking and by vitamins C and E: evidence against a uniform effect of nutrient. J Nutr Sci 2020;9. https://doi.org/10.1017/jns.2020.3.

[78] Smoking and COVID-19 n.d. https://www.who.int/news-room/commentaries/detail/smoking-and-covid-19 (accessed September 7, 2021).

[79] Sidhu G, Preuss CV. Triamcinolone. StatPearls, treasure Island (FL). StatPearls Publishing; 2021.

[80] Zhang C, Wu Z, Li J-W, Zhao H, Wang G-Q. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. Int J Antimicrob Agents 2020;55:105954. https://doi.org/10.1016/j.ijantimicag.2020.105954.

[81] McGonagle D, Sharif K, O'Regan A, Bridgewood C. The role of cytokines including interleukin-6 in COVID-19 induced pneumonia and macrophage activation syndrome-like disease. Autoimmun Rev 2020;19:102537. https://doi.org/10.1016/j.autrev.2020.102537.

[82] Girodet P-O, Dournes G, Thumerel M, Begueret H, Dos Santos P, Ozier A, et al. Calcium channel blocker reduces airway remodeling in severe asthma. A proof-of-concept study. Am J Respir Crit Care Med 2015;191:876–83. https://doi.org/10.1164/rccm.201410-1874OC.

[83] Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci USA 2020;117:11727–34. https://doi.org/10.1073/pnas.2003138117.

[84] Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Correa Marrero M, et al. The global phosphorylation landscape of SARS-CoV-2 infection. Cell 2020;182: 685–712.e19. https://doi.org/10.1016/j.cell.2020.06.034.

[85] Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science 2020;369:1210–20. https://doi.org/10.1126/science.abc6261.

[86] Chistiakov DA, Orekhov AN, Bobryshev YV. ApoA1 and ApoA1-specific self-antibodies in cardiovascular disease. Lab Investig J Tech Methods Pathol 2016; 96:708–18. https://doi.org/10.1038/labinvest.2016.56.

[87] Zamanian Azodi M, Arjmand B, Zali A, Razzaghi M. Introducing APOA1 as a key protein in COVID-19 infection: a bioinformatics approach. Gastroenterol Hepatol Bed Bench 2020;13:367–73.

[88] Ulloque-Badaracco JR, Hernandez-Bustamante EA, Herrera-Añazco P, Benites-Zapata VA. Prognostic value of apolipoproteins in COVID-19 patients: a systematic review and meta-analysis. Trav Med Infect Dis 2021;44:102200. https://doi.org/10.1016/j.tmaid.2021.102200.

[89] Neuraz A, Lerner I, Digan W, Paris N, Tsopra R, Rogier A, et al. Natural Language processing for rapid response to emergent diseases: case study of calcium channel blockers and hypertension in the COVID-19 pandemic. J Med Internet Res 2020; 22:e20773. https://doi.org/10.2196/20773.

[90] Kow CS, Ramachandram DS, Hasan SS. Clinical outcomes of hypertensive patients with COVID-19 receiving calcium channel blockers: a systematic review and meta-analysis. Hypertens Res 2022;45:360–3. https://doi.org/10.1038/s41440-021-00786-z.

[91] PengChi WangHao, GuoYu-Feng, QiGe-Yao, ZhangChen-Xu ChenTing, et al. Calcium channel blockers improve prognosis of patients with coronavirus disease 2019 and hypertension. Chin Med J (Engl) 2021.

[92] Qian Z, Li Z, Peng J, Gao Q, Cai S, Xu X. Association between hypertension and prognosis of patients with COVID-19: a systematic review and meta-analysis. Clin Exp Hypertens 2022;44:451–8. https://doi.org/10.1080/10641963.2022.2071914.

[93] Chen J, Liu Y, Qin J, Ruan C, Zeng X, Xu A, et al. Hypertension as an independent risk factor for severity and mortality in patients with COVID-19: a retrospective study. Postgrad Med J 2022;98:515–22. https://doi.org/10.1136/postgradmedj-2021-140674.

[94] Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008;36:D684–8. https://doi.org/10.1093/nar/gkm795.

[95] Aquino YSJ, Cabrera N. Hydroxychloroquine and COVID-19: critiquing the impact of disease public profile on policy and clinical decision-making. J Med Ethics 2020;46:574–8. https://doi.org/10.1136/medethics-2020-106306.

[96] Papadatos G, Gaulton A, Hersey A, Overington JP. Activity, assay and target data curation and quality in the ChEMBL database. J Comput Aided Mol Des 2015;29: 885–96. https://doi.org/10.1007/s10822-015-9860-5.

[97] Bajorath J. Molecular similarity concepts for informatics applications. In: Keith JM, editor. Bioinforma. Vol. II Struct. Funct. Appl. New York, NY: Springer; 2017. p. 231–45. https://doi.org/10.1007/978-1-4939-6613-4_13.

[98] Terfloth L, Spycher S, Gasteiger J. Drug discovery: an overview. Appl. Chemoinformatics. John Wiley & Sons, Ltd; 2018. p. 165–94. https://doi.org/10.1002/9783527806539.ch6a.

[99] Stumpfe D, Hu H, Bajorath J. Evolving concept of activity cliffs. ACS Omega 2019;4:14360–8. https://doi.org/10.1021/acsomega.9b02221.

[100] Reddy AS, Zhang S. Polypharmacology: drug discovery for the future. Expet Rev Clin Pharmacol 2013;6:41–7. https://doi.org/10.1586/ecp.12.74.

[101] Rao MS, Gupta R, Liguori MJ, Hu M, Huang X, Mantena SR, et al. Novel computational approach to predict off-target interactions for small molecules. Front Big Data 2019;2.

[102] Daly AK. Pharmacogenomics of adverse drug reactions. Genome Med 2013;5:5. https://doi.org/10.1186/gm409.

[103] Smajić A, Rami I, Sosnin S, Ecker GF. Identifying differences in the performance of machine learning models for off-targets trained on publicly available and proprietary data sets. Chem Res Toxicol 2023;36:1300–12. https://doi.org/10.1021/acs.chemrestox.3c00042.

[104] Rasolohery I, Moroy G, Guyon F. PatchSearch: a fast computational method for off-target detection. J Chem Inf Model 2017;57:769–77. https://doi.org/10.1021/acs.jcim.6b00529.

[105] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 2021;49:D605–12. https://doi.org/10.1093/nar/gkaa1074.

[106] Morens DM, Fauci AS. Emerging pandemic diseases: how we got to COVID-19. Cell 2020;182:1077–92. https://doi.org/10.1016/j.cell.2020.08.021.

[107] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28: 31–6.

[108] Vass M, Kooistra AJ, Ritschel T, Leurs R, de Esch IJ, de Graaf C. Molecular interaction fingerprint approaches for GPCR drug discovery. Curr Opin Pharmacol 2016;30:59–68. https://doi.org/10.1016/j.coph.2016.07.007.

[109] Zhou Y, Zhang Y, Lian X, Li F, Wang C, Zhu F, et al. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of

targeted agents. Nucleic Acids Res 2022;50:D1398–407. https://doi.org/10.1093/nar/gkab953.

[110] Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res 2016;44:D380–4. https://doi.org/10.1093/nar/gkv1277.

[111] Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ, et al. STITCH 4: integration of protein–chemical interactions with user data. Nucleic Acids Res 2014;42:D401–7. https://doi.org/10.1093/nar/gkt1207.

[112] Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. Nat Commun 2019;10:1197. https://doi.org/10.1038/s41467-019-09186-x.

[113] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47:W191–8. https://doi.org/10.1093/nar/gkz369.

[114] Fabregat A, Korninger F, Viteri G, Sidiropoulos K, Marin-Garcia P, Ping P, et al. Reactome graph database: efficient access to complex pathway data. PLoS Comput Biol 2018;14:e1005968. https://doi.org/10.1371/journal.pcbi.1005968.

[115] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.

[116] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25:25–9. https://doi.org/10.1038/75556.

[117] Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler,

GSEA, Cytoscape and EnrichmentMap. Nat Protoc 2019;14:482–517. https://doi.org/10.1038/s41596-018-0103-9.

[118] Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler. F1000Research 2020;9:ELIXIR–709. https://doi.org/10.12688/f1000research.24956.2.

[119] Melo Riveros NA, Cardenas Espitia BA, Aparicio Pico LE. Comparison between K-means and Self-Organizing Maps algorithms used for diagnosis spinal column patients. Inform Med Unlocked 2019;16:100206. https://doi.org/10.1016/j.imu.2019.100206.

[120] KOHONEN T. The self-organizing map. Proc IEEE 1990;78:1464–80. https://doi.org/10.1109/5.58325.

[121] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979. https://doi.org/10.1109/TPAMI.1979.4766909. PAMI-1: 224–7.

[122] Phipson B, Smyth GK. Permutation P-values should never Be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 2010;9. https://doi.org/10.2202/1544-6115.1585.

[123] Virtualitics. Make smarter AI-driven decisions. Virtualitics 2021. https://virtualitics.com/. [Accessed 25 June 2021].

[124] McKnight PE, Najab J, Mann-Whitney U. Test. Corsini Encycl. Psychol. John Wiley & Sons, Ltd; 2010. p. 1. https://doi.org/10.1002/9780470479216.corpsy0524. 1.