

<https://doi.org/10.1038/s42003-024-05897-0>

Priority index for critical Covid-19 identifies clinically actionable targets and drugs

Check for updates

Zhiqiang Zhang^{1,2,13}, Shan Wang^{1,13}, Lulu Jiang^{3,13}, Jianwen Wei^{4,13}, Chang Lu^{5,13}, Shengli Li⁶, Yizhu Diao⁷, Zhongcheng Fang^{1,2}, Shuo He⁸, Tingting Tan¹, Yisheng Yang¹, Kexin Zou^{1,2}, Jiantao Shi⁹, James Lin⁴, Liye Chen¹⁰✉, Chaohui Bao^{1,11}✉, Jian Fei^{11,12}✉ & Hai Fang¹✉

While genome-wide studies have identified genomic loci in hosts associated with life-threatening Covid-19 (critical Covid-19), the challenge of resolving these loci hinders further identification of clinically actionable targets and drugs. Building upon our previous success, we here present a priority index solution designed to address this challenge, generating the target and drug resource that consists of two indexes: the target index and the drug index. The primary purpose of the target index is to identify clinically actionable targets by prioritising genes associated with Covid-19. We illustrate the validity of the target index by demonstrating its ability to identify pre-existing Covid-19 phase-III drug targets, with the majority of these targets being found at the leading prioritisation (leading targets). These leading targets have their evolutionary origins in Amniota ('four-leg vertebrates') and are predominantly involved in cytokine-cytokine receptor interactions and JAK-STAT signaling. The drug index highlights opportunities for repurposing clinically approved JAK-STAT inhibitors, either individually or in combination. This proposed strategic focus on the JAK-STAT pathway is supported by the active pursuit of therapeutic agents targeting this pathway in ongoing phase-II/III clinical trials for Covid-19.

Host genetics, encompassing human genetic contributions to infectious diseases, holds the potential to unveil genetically informed mechanisms for disease prevention and drug therapy. Naturally occurring genetic variations within the host genome have catalysed strategies in preventing infectious diseases, often referred to as 'natural immunity'. Successful examples include the discovery of HIV-1 infection natural resistance attributed to the *CCR5* gene¹, the identification of the malaria resistance locus at the *HBB* gene², and the protective effects conferred by the *WT1* locus against tuberculosis³. In addition to providing evidence of disease prevention mechanisms, these host variations can offer insights into therapeutic targeting by mimicking the on-target effects of pharmacological interventions, known as 'clinical trials by nature'⁴. Moreover, genetic variations can exert regulatory influence on genes ('effector genes'), which encode protein targets of drugs — both those approved/licensed and those in clinical phase development — thus providing opportunities for repurposing pre-existing drugs into new indications^{5,6}.

Genetic targets are defined as early-stage genetically informed and validated therapeutic candidates⁷. The endorsement of genetic targets

enhances the chance of approval along the drug development pipeline, as compared to drug-target pairs without such genetic backing^{8,9}. The proposition of genetic targets is timely, particularly amid the pandemic crisis brought about by coronavirus disease 2019 (Covid-19), attributed to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)¹⁰. Its severity exhibits strong correlations with factors such as older age, being male, lower socio-economic status, non-European demographic ancestry, and pre-existing clinical comorbidities (including diabetes)¹¹. Nonetheless, these non-genetic factors do not fully explain disease severity, as instances of severe cases among young, otherwise healthy individuals, often run in families¹² and persistently hint at genetic contributions to disease severity¹³.

In a fast-moving landscape of Covid-19 field^{14,15}, there emerges a need for the timed translation of host genetic findings into precise therapeutics targeted to treat critically ill patients with life-threatening Covid-19 (referred to herein as 'critical Covid-19'). Individuals suffering from critical Covid-19 are likely to develop acute respiratory distress syndrome, accompanied by a hyperinflammation phenotype characterised by the release of excessive pro-inflammatory cytokines, resembling immune-related disorder

A full list of affiliations appears at the end of the paper. ✉ e-mail: liye.chen@ndorms.ox.ac.uk; chaohuibao@sjtu.edu.cn; feijian@hotmail.com; fh12355@rjh.com.cn

manifestations¹⁶. With the continual expansion of sample sizes and a clearer definition of disease severity, genome-wide studies^{17–24} offer a wealth of genetic targets pertinent to critical Covid-19. Nonetheless, the task of identifying genetic targets is intricate, primarily due to the prevalence of disease severity loci within the non-coding human genome. This challenge is further exacerbated by the regulatory nature of non-coding loci on effector genes, often involving long-range and cell-type-specific mechanisms²⁵.

Addressing the challenge mentioned above necessitates a paradigm shift in strategies. To this end, we have devised a principled strategy^{26,27} that links non-coding loci to effector genes, eventually cascading down to drug targets. This strategy has found diverse applications for various diseases^{28–33}. Building upon our previous accomplishments and aligning with the transition from Covid-19 host genetic findings to translational applications, in this study we report a genetically powered drug-target discovery engine, culminating in the introduction of a resource, namely PIC², made accessible and reproducible via a publicly available web portal (www.genetictargets.com/PIC2) (Fig. 1). The engine features a genetics-led target prioritisation (an approach PIT generating the target index PIC²Target), and a crosstalk-based drug repurposing (an approach PID generating the drug index PIC²Drug). We demonstrate the validity of the target index in successfully recovering known therapeutics. The drug index highlights opportunities for repurposing clinically approved JAK-STAT inhibitors, either individually or in combination. The introduction of this dual-indexes strategy sets this work apart from our earlier studies^{7,34,35}. This strategy is especially distinctive in its capacity to streamline computational translational medicine for Covid-19, seamlessly transitioning from genetic target prioritisation to rational drug repurposing.

Results

The target index

As an extension to our previous genetics-led approach³⁴, PIT converts host genetic findings (derived from the latest critical Covid-19 GWAS summary-level data²¹) into an index for genetically informed therapeutic targets (Fig. 2a). The target-index generation/prioritisation process involved two key steps (see Materials and Methods): (1) predictor preparation, which harnessed the value of multi-modal regulatory genomic datasets on proximity, quantitative trait locus (QTL), and promoter capture Hi-C (PCHi-C) (constituting genomic evidence), while simultaneously leveraging the knowledge of high-quality protein interactions (sourced from the STRING database³⁶) through the random walk with restart (RWR) algorithm (constituting network evidence); and (2) predictor combination, accomplished by benchmarking different target-index generation schemes, including meta-analysis-like schemes (logistic and Fisher's) and conventional schemes (sum and max). In evaluating how to best combine predictors for generating the target index (Supplementary Fig. 1), it became evident that meta-analysis-like schemes, particularly the logistic scheme, exhibited superior performance over conventional schemes (sum and max), under the same parameter setting for RWR (optimised at the restarting probability of 0.5). We also introduced the Naive scheme as a baseline for evaluation and additionally explored another alternative employing network evidence sourced from the BioGRID database³⁷, albeit both alternatives demonstrated inferior performance to PIT (i.e., the logistic scheme anchored by STRING) (Fig. 2b and Supplementary Data 1).

Applying PIT to critical Covid-19 led to the generation of the target index, which involved approximately 14,000 genes, ranked according to their priority rating (Fig. 2c and Supplementary Data 2). To outline this process in more detail, we first used critical Covid-19 GWAS SNPs to define core genes based on evidence drawn from proximity, QTL, and PCHi-C criteria. After establishing core genes, we proceeded to incorporate network evidence obtained from STRING to define peripheral genes. Following this, we computed affinity scores for both core and peripheral genes through RWR, with core genes serving as seed nodes and capitalising on network connectivity. These steps resulted in the preparation of a gene-predictor matrix containing affinity scores for both core and peripheral genes. Next, we used the logistic method to combine the predictors, ensuring the incorporation of both genetic and network evidence. This method, widely employed for combining evidence (*P*-values) from various studies³⁸, accommodates the interdependence/non-independence of *P*-values during the aggregation process. By embracing this approach, we combined different predictors grounded in proximity, QTL, and PCHi-C criteria to formulate the target index. The specific steps involved in this process are as follows: (1) for each predictor, the affinity scores resulting from RWR were converted into *P*-like values; (2) these converted *P*-like values were collectively combined across predictors using the logistic method for each gene; and (3) the combined *P*-values were subsequently rescaled to yield a single, unified priority rating (i.e., target index). This unified priority rating ranged from 0 to 5 for each gene (further details provided in Materials and Methods).

In the uppermost (top) 1% of prioritised genes within the target index, approximately three-fourths were genes that did not belong to the core genes (Fig. 2d, left). This highlights a substantial contribution of non-core peripheral genes to the top-ranked genes within our target index, further emphasising the value of incorporating network evidence. To assess the contribution of core genes based on different criteria (proximity, QTL, and PCHi-C), we calculated the proportion contributed by each criterion, with QTL being the most prominent (Fig. 2d, right). Consistent with this, QTL exhibited the most substantial impact on performance when used as the single criterion for prioritisation (Supplementary Fig. 2).

To elucidate the disease relevance of the target index, with a specific focus on the top 1% of prioritised genes, we conducted enrichment analysis using two distinct gene lists (Fig. 2e): (1) a list encompassing human genes/proteins that directly interact with SARS-CoV-2 viral proteins (considering virus-host interactions identified by two or more independent studies^{39–45}), and (2) another list comprising Covid-19 human host pathway genes (sourced from KEGG⁴⁶). Amongst the top 1% of prioritised genes, we did not observe any significant enrichment (*P*-value = 0.95 based on Fisher's exact test) for human proteins that directly interact with SARS-CoV-2 viral proteins (Supplementary Fig. 3). Instead, we found enrichment for Covid-19 host pathway genes. Notably, a substantial proportion of these host pathway genes were present within the top 1% of the target index (odds ratio = 11.1; 95% confidence intervals (CI) = [6.4, 18.4]; *P*-value = 8.0×10^{-14}). This supports the capacity of the target index in illuminating disease relevance within the context of the human host pathway (Fig. 2f).

In addition to its disease relevance, the target index also successfully recovered 23 pre-existing phase-III drug targets for Covid-19

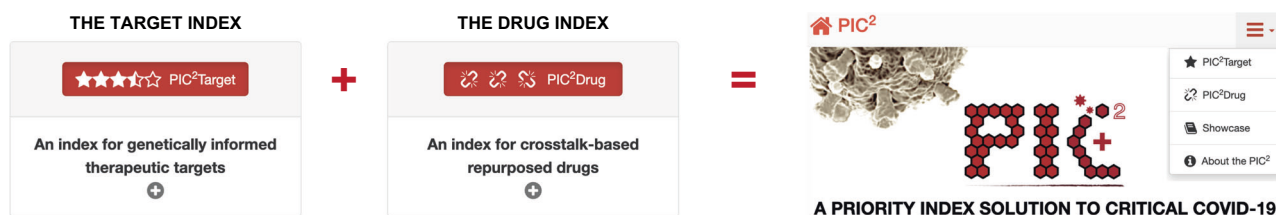


Fig. 1 | A priority index solution to critical Covid-19. The PIC² resource comprises two indexes: the target index (PIC²Target) for genetically informed therapeutic targets, and the drug index (PIC²Drug) for crosstalk-based repurposed drugs.

Incorporating 'PI' (emphasising priority index) and 'C2' (symbolising critical Covid-19), the PIC² logo also bears symbols ('red cross' and 'spreading coronavirus') to signify different interpretations of the letter 'C'.

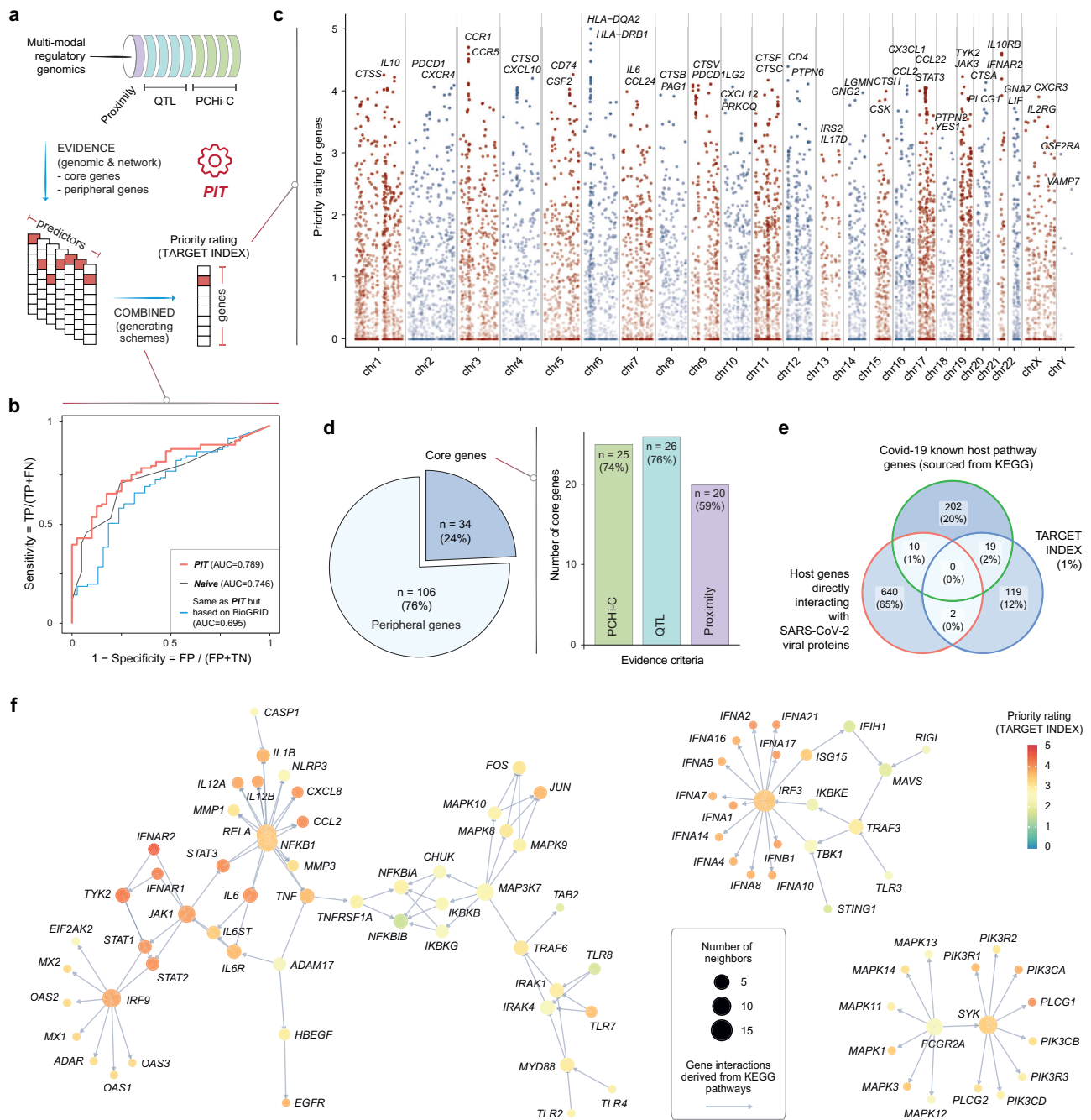
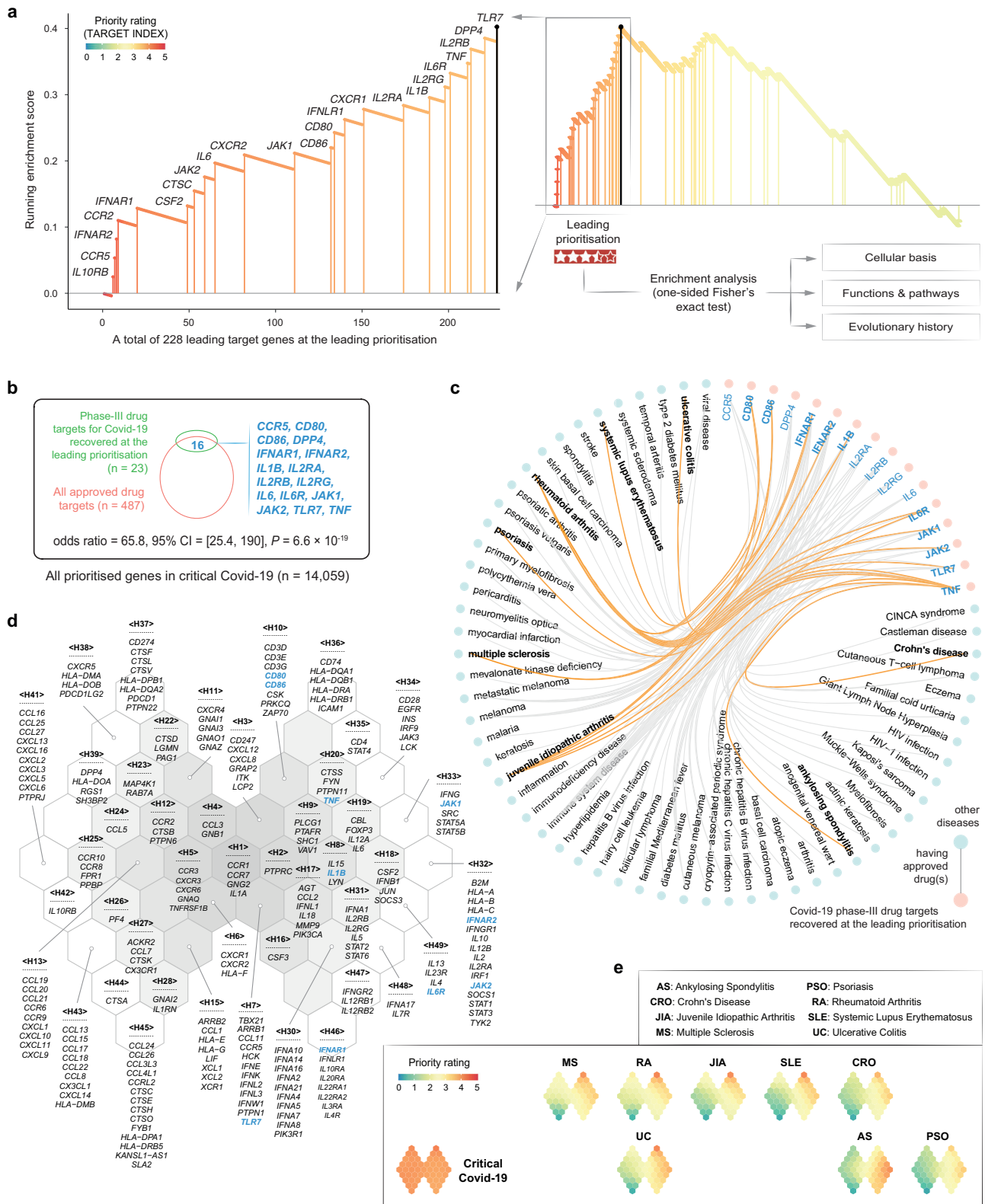


Fig. 2 | The target index (PIC²Target) for critical Covid-19. a Target prioritisation workflow. **b** Performance benchmarking. Performance is measured by AUC, distinguishing Covid-19 phase-III drug targets from simulated negative targets. Several target-index generation schemes are compared: PIT (using a logistic meta-analysis combined method and employing network evidence sourced from the STRING database), Naive (frequency of therapeutic targeting by existing licensed/approved drugs, motivated by drug repurposing), and an alternative method employing network evidence sourced from the BioGRID database. Abbreviations include AUC (area under the ROC curve), FN (false negatives), FP (false positives), TN (true negatives), and TP (true positives). **c** Manhattan plot. Priority rating (y-axis) for ~14,000 target genes is illustrated across genomic locations, with the top two prioritised genes per chromosome labelled on the x-axis. **d** Pie plot illustrating the composition of core and peripheral genes amongst the top 1% of the target index. Right panel: bar plot for core genes identified through each indicated evidence of criteria. **e** Venn diagram illustrating the overlap between the top 1% of the target index, the Covid-19 host pathway genes (KEGG), and host genes directly interacting with SARS-CoV-2 viral proteins. **f** Network visualisation of the Covid-19 host pathway (sourced from KEGG), with gene nodes color-coded by priority rating and sized by node degree (i.e., the number of neighbours).

(P -value = 5.0×10^{-5} based on leading prioritisation analysis; Fig. 3a). These Covid-19 phase-III targets included:

- IL10RB* (6th) and *IFNL1* (140th), both targeted by peginterferon lambda-1a (an interferon λ receptor agonist);
- CCR5* (7th), the target of ceniciviroc and leronlimab (two C-C chemokine receptor Type 5 antagonists);

- IFNAR2* (8th) and *IFNAR1* (20th), both targeted by interferon beta-1a and peginterferon beta-1a (two interferon α/β receptor agonists);
- CCR2* (9th), the target of ceniciviroc (a C-C chemokine receptor Type 2 antagonist);
- CSF2* (49th), targeted by lenzilumab (a GM-CSF inhibitor);



CTSC (53rd), the target of brensocatib (a dipeptidyl peptidase I inhibitor);
 JAK2 (59th), the target of baricitinib and pacitinib (two JAK2 inhibitors);
 IL6 (65th), targeted by siltuximab (an interleukin-6 inhibitor);
 CXCR2 (82nd), the target of reparixin (an interleukin-8 receptor B modulator);

JAK1 (111th), the target of baricitinib (a JAK1 inhibitor);
 CD86 (132nd) and CD80 (134th), both targeted by abatacept (a T-lymphocyte activation antigen CD86/CD80 inhibitor);
 CXCR1 (151st), the target of reparixin (an interleukin-8 receptor A modulator);
 IL2RA (174th), IL2RG (198th), and IL2RB (213th), all targeted by aldesleukin (an interleukin-2 receptor agonist);

Fig. 3 | Leading prioritisations for critical Covid-19. **a** Leading prioritisation plot for Covid-19 phase-III drug targets. Those Covid-19 phase-III drug targets ($n = 23$) recovered at the leading prioritisation (zoomed in left panel) are denoted by gene symbols and indicated in vertical lines (also color-coded by priority rating). The leading prioritisation represents the core subset of prioritised target genes that accounts for the enrichment signal, visualised as the left-most region of the peak in the running enrichment plot (middle panel). All genes at the leading prioritisation, referred to as ‘leading target genes’ ($n = 228$), were subjected to enrichment analysis in terms of cellular basis (cell types; see Fig. 4a), functions and pathways (see Fig. 4b), and evolutionary origins (ancestors when they were first created; see Fig. 4c). **b** Venn diagram illustrating the intersection between Covid-19 phase-III drug targets recovered at the leading prioritisation and the entire pool of approved drug targets in any disease indications. The statistical significance (P -value), odds ratio, and its 95%

confidence intervals (CI; represented by lines) were calculated using one-sided Fisher’s exact test. **c** Hierarchical edge bundling for targets and drugs. Covid-19 phase-III drug targets recovered at the leading prioritisation (in pink) that are already targeted (connected by edges) by approved drugs in other diseases are connected by edges. Those edges involving immune-mediated diseases are highlighted in orange. **d, e** Cross-disease prioritisation map between critical Covid-19 and immune-mediated diseases. The map with a 2D butterfly-like topology was trained using a self-organising algorithm for leading target genes in critical Covid-19. **d** The architecture of the map, consisting of 49 hexagons (H1-H49) with genes listed within per hexagon (if any). **e** Disease-specific illustration of target gene prioritisation, with locations in a 2D square lattice capturing inter-disease relationships. This square map was also trained using self-organising algorithm but for diseases.

IL1B (189th), the target of canakinumab (an interleukin-1 β inhibitor); *IL6R* (201st), targeted by levilimab, sarilumab, and tocilizumab (three interleukin-6 receptor α subunit inhibitors); *TNF* (211th), targeted by adalimumab and infliximab (two TNF- α inhibitors); *DPP4* (221st), the target of linagliptin (a dipeptidyl peptidase IV inhibitor); and *TLR7* (228th), the target of hydroxychloroquine sulfate (a Toll-like receptor 7 antagonist).

Notably, baricitinib, a selective oral JAK1/2 inhibitor, has been shown to improve the use of remdesivir, the only FDA-approved antiviral drug for the treatment of Covid-19 (ref. ⁴⁷). Baricitinib not only reduces recovery time but also accelerates improvement in clinical status⁴⁸. The REMAP-CAP randomised clinical trial revealed that critical Covid-19 patients receiving IL6 receptor inhibitors (tocilizumab or sarilumab) exhibited an improved 180-day mortality rate⁴⁹. Collectively, genes identified at the leading prioritisation (referred to as ‘leading target genes’ hereinafter; listed in Supplementary Data 3) tended to be therapeutically considered for targeting to treat Covid-19. Approximately 70% (16/23) of these genes have already been targeted by approved drugs in other diseases (odds ratio = 65.8; 95% CI = 25.4, 190); P -value = 6.6×10^{-19} ; see Fig. 3b), especially in immune-mediated diseases. These immune diseases include ankylosing spondylitis, Crohn’s disease, juvenile idiopathic arthritis, multiple sclerosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, and ulcerative colitis (Fig. 3c). Moreover, when incorporating our previous findings³⁴, we found that these approved drug targets were also highly prioritised in their respective immune diseases, including *CD80/86*, *IFNAR1/2*, *IL1B*, *IL6R*, *JAK1/2*, *TLR7*, and *TNF* (Fig. 3d, e and Supplementary Data 4).

Based on leading target genes, we delved deeper into: (1) the cellular basis, revealing the involvement of multiple immune and inflammatory lineages (Fig. 4a and Supplementary Fig. 4); (2) molecular functions, biological processes, and environmental information processing pathways (Fig. 4b and Supplementary Data 5); and (3) their evolutionary origins, particularly emerging at the time of Amniota (the most recent common ancestor of all living reptiles, birds, and mammals) (Fig. 4c and Supplementary Data 6). Amniotes represent the clade of tetrapod (‘four-leg’) vertebrates characterised by protective extra-embryonic membranes. Remarkably, genes originating at Amniota were mostly involved in cytokine-cytokine receptor interactions, as well as the JAK-STAT signaling (i.e., *IFNAR2*, *IFNG*, *IFNGR1/2*, *IFNK*, *IL10RA*, *IL22RA1*, *IL2RA*, *IL4R*, *IL5/6*, and *LIF*). The knowledge of molecular interactions within the JAK-STAT signaling was illustrated, with genes color-coded based on their priority rating (Fig. 5).

The drug index

One of the distinguishing features of PID lies in its capacity to perform crosstalk-based drug repurposing and effect-by-removal analysis to determine the impact of node removal on pathway crosstalk, generating an index for crosstalk-based repurposed drugs (refer to Fig. 1c). Central to this capability is the identification of genes that mediate crosstalk between

pathways. The resulting 50-gene network for pathway crosstalk (P -value = 1.41×10^{-112} based on the permutation test) all contained highly prioritised genes in critical Covid-19 (Fig. 6a and Supplementary Data 7). Remarkably, 8 genes (i.e., *CSF2*, *CXCR2*, *IFNAR1/2*, *IL10RB*, *IL6*, and *JAK1/2*) targeted by drugs currently undergoing phase-III clinical trials for Covid-19 were present within this network (Fig. 6b, c). By focusing on pathways significantly over-represented in this crosstalk network (Fig. 6d and Supplementary Data 8), we could represent the crosstalk at the pathway level, with edges estimated based on the extent of genes being shared between two endpoint pathways (Fig. 6e).

We proceeded to explore the evidence in support of drug repurposing by asking whether pathway crosstalk genes could be targeted by approved or phased drugs in diseases other than Covid-19. Our analysis yielded substantial clinical evidence (approved drugs) with false discovery rate (FDR) of 1.9×10^{-8} , identifying 13 genes already targeted by approved drugs used to treat other diseases, that is, licensed medications in clinical use for diseases other than Covid-19 (Supplementary Fig. 5). Notably, out of these 13 approved drug targets, 12 genes (*IFNAR1/2*, *IFNG*, *IFNGR1/2*, *IL6/13*, *JAK1/2/3*, *PIK3CA*, and *TYK2*) were integral to the JAK-STAT signaling pathway, highlighting repurposing opportunities of targeting this pathway. Additionally, we identified four phase-III drug targets (*CSF2*, *CXCR2*, *IL10RA*, and *IL10RB*; FDR = 5.9×10^{-3}).

Finally, we delved into drug repurposing opportunities by implementing crosstalk-based effect-by-removal analysis. Effect-by-removal analysis gauged the vulnerability of the pathway crosstalk to node removal, either individually or in combination; removing critical network nodes would result in a larger proportion of disconnected nodes in the crosstalk (Fig. 7 and Supplementary Data 9). This analysis generated an index for repurposed drugs, that is, the drug index measured as the fraction of disconnected nodes (ranged 0–1). The drug index underscored repurposing opportunities of clinically approved inhibitors targeting the JAK-STAT signaling. Notably, removing the gene *PIK3CA* alone resulted in a maximum of 36.4% node disconnection (i.e., the drug index of 0.34). This gene is intricately linked to the PI3K/AKT/mTOR axis and presents a potential Covid-19 pharmacological target⁵⁰. The disconnection fraction increased to 40.0% when further removing *JAK2* (i.e., *PIK3CA* + *JAK2*) or *IFNB1* (i.e., *PIK3CA* + *IFNB1*), representing the maximal/optimal effect achievable by removing any two nodes in combination (i.e., the drug index of 0.38). Further disconnection, reaching a maximum of 42% disconnections (i.e., the drug index of 0.42), was observed upon removing *PIK3CA* + *JAK2* + *IFNB1*. In summary, drug repurposing coupled with effect-by-removal analysis provided evidence for targeting key components (i.e., *PIK3CA*, *JAK2*, and *IFNB1*) of the JAK-STAT signaling pathway, either individually or in combination (potentially for the poly-therapeutics discovery in critical Covid-19)⁵¹. The ongoing focus on targeting this pathway is reinforced by the presence of therapeutic agents now in active phase-II/III clinical trials in Covid-19. This includes therapeutic agents targeting the gene *IFNB1*, which are currently under phase-II clinical trials for Covid-19 (Fig. 7, top), while agents targeting the gene *JAK2* are undergoing phase-III trials for Covid-19 (also see Fig. 6c).

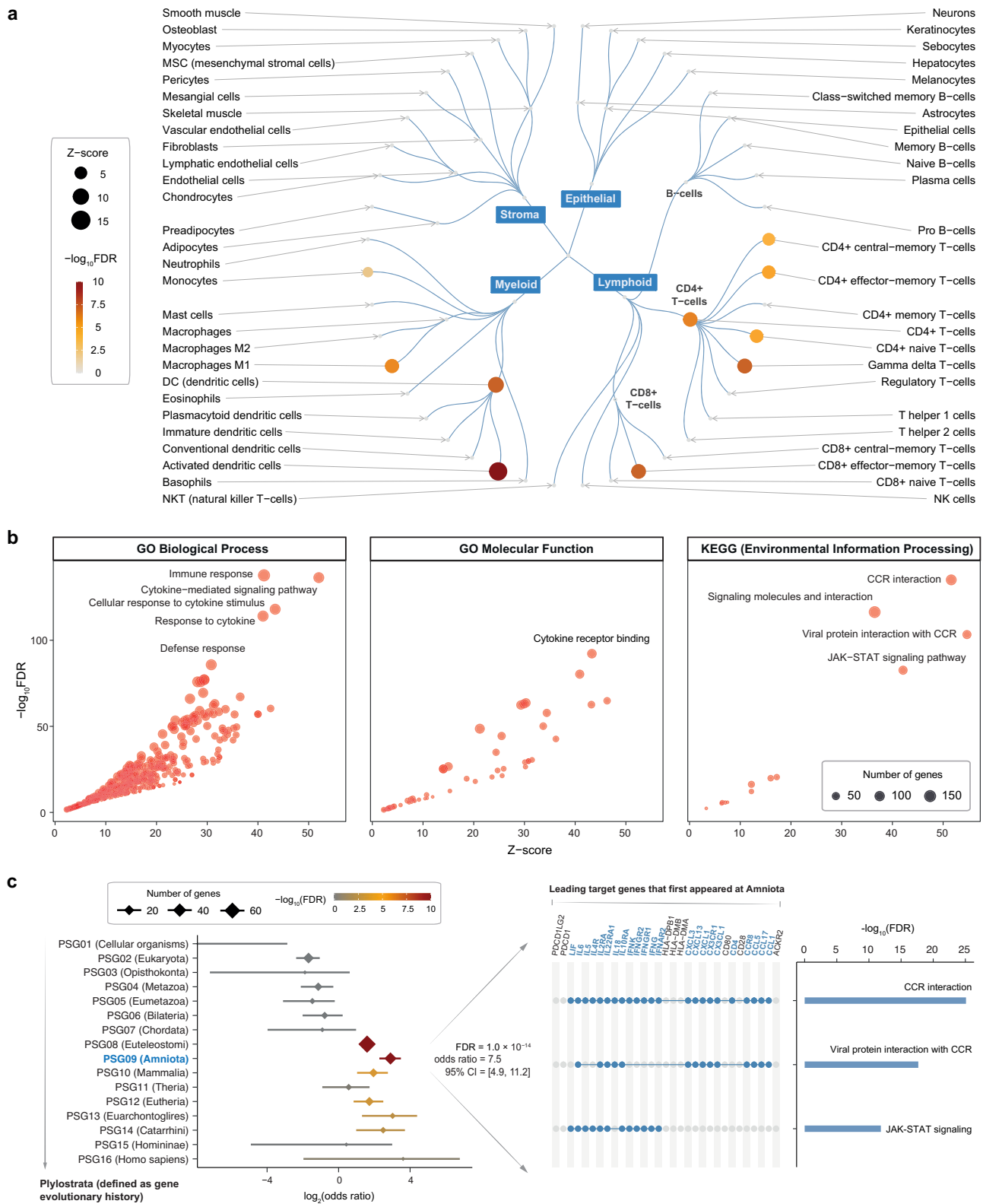
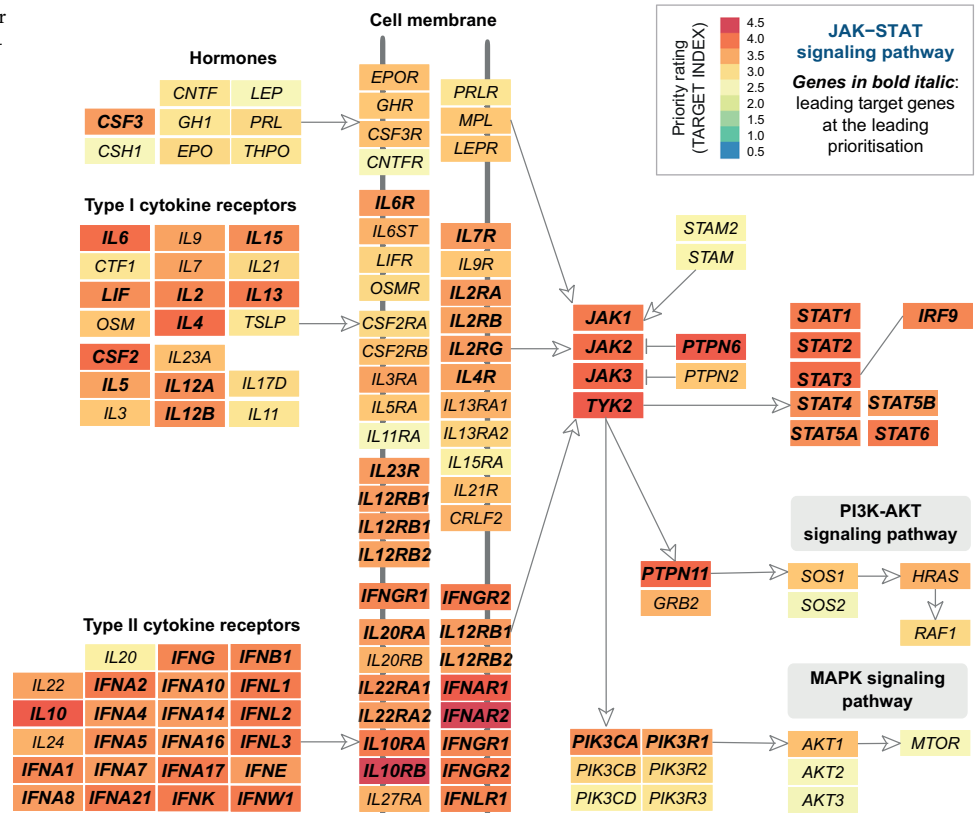


Fig. 4 | Enrichment analysis for leading target genes. a Circular illustration of cell type enrichments, with nodes sized by Z-score and color-coded by FDR. **b** Scatter plot illustrating enriched functions and pathways, with the 10 most significant terms/pathways labelled. Functions are based on Gene Ontology (GO) Biological Process and Molecular Function terms, while pathways based on KEGG (Environmental Information Processing pathways). Each dot corresponds to an individual

term/pathway, with the size proportional to the number of its member genes. **c** Forest plot illustrating enriched phylostrata (ancestors) ordered by evolutionary history. Genes first created at the ancestor 'Amniota' are indicated in the right panel, along with their pathway enrichments. Notably, the significance level (FDR), odds ratio, and its 95% confidence intervals (CIs; represented by lines) were calculated using one-sided Fisher's exact test.

Fig. 5 | JAK-STAT signaling pathway. Its member genes are color-coded by priority rating, with leading target genes highlighted in bold.



Discussion

In deepening our long-standing efforts in therapeutic target prediction and validation, our work has culminated in the creation of the PIC² solution, dedicated to expediting Covid-19 translational medicine research. It streamlines the running from genetic target prioritisation to rational drug repurposing, thereby contributing to informed decisions about potential therapeutic candidates for next preclinical validations and, ultimately, clinical trials.

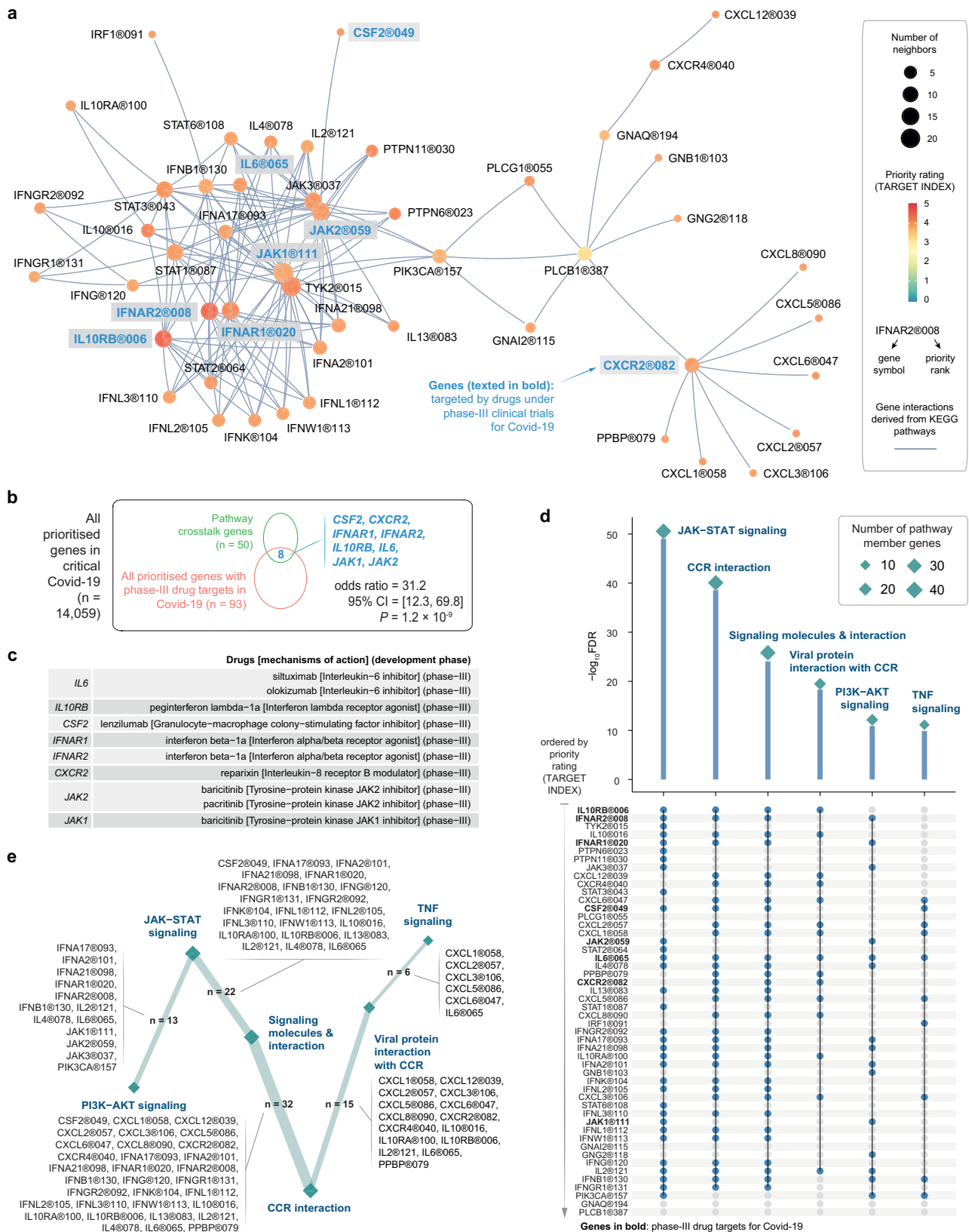
An outstanding feature of PIC² lies in its capacity in seamlessly integrating GWAS summary-level data with a wealth of multi-modal regulatory genomic datasets and high-quality protein interactions. Comprehensively incorporating both genetic and network information, this integration culminates in the generation of the target index, demonstrated to be informative in recovering pre-existing Covid-19 phase-III drug targets. Another notable feature offered by PIC² is its capacity in discerning crosstalk between pathways, which represents an advancement compared to our previous studies^{34,35}. This crosstalk-based approach enhances the identification of genes for potential drug repurposing. The effect-by-removal analysis is an addition that provides insights into the impact of removing specific nodes (genes) from the network, allowing for the evaluation of the consequences of targeting individual genes or their combinations, particularly in the context of pathway crosstalk, which is essential for drug repurposing. This crosstalk-based drug repurposing, when coupled with effect-by-removal analysis, is a component that enhances the methodology's depth and applicability. As exemplified in this study, it offers an avenue for repurposing clinically approved JAK-STAT inhibitors, whether as single agents or in combination, for potential application in the treatment of severe Covid-19 cases.

Acknowledging the limitations of PIC² is crucial, particularly in its reliance on the availability and quality of regulatory genomic datasets. The quality and abundance of these datasets play a pivotal role in establishing links between non-coding loci and the core genes that underlie host genetic associations. As QTL and PChI-C datasets continue to expand across diverse cell types, states, and tissues, improvements in the accuracy and

comprehensiveness of target prioritisation and drug repurposing are anticipated. Exploring alternative rating schemes, such as Brown's method⁵² and the Cauchy combination test⁵³, offers a potential avenue to explicitly address the challenge of interdependence/non-independence between predictors. Another limitation arises from the lack of extensive experimental or clinical validations, even though we have demonstrated the validity and utility of the leading prioritisation and drug repurposing. The dual indexes for targets and drugs form the foundation for validation and repurposing within the wider research community. However, the need for substantiated therapeutic efficacy remains paramount and necessitates further evidence, potentially derived from human disease-relevant assays.

Our study, primarily oriented toward the prioritisation of potential therapeutic targets for critical Covid-19, does not explicitly determine the specific directionality (i.e., agonism or antagonism) of the prioritised genes. Instead, the prioritisation stems from the perceived relevance of these genes within the disease context. Additional experimental and functional investigations are required to ascertain the precise directionality of therapeutic intervention (agonism or antagonism) for prioritised genes. Such studies encompass *in vitro* and *in vivo* assays, preclinical models (particularly human organoids^{54,55}), and clinical trials, enabling a comprehensive understanding of the effects resulting from the modulation of the activity of these genes. Our approach yields a prioritised list of potential therapeutic targets, mandating further empirical validation to define the most optimal therapeutic intervention strategy.

The development of the PIC² solution represents a big step forward in the realm of critical Covid-19 research. However, it is just the beginning of an exciting journey toward more accurate, comprehensive, and context-specific target prioritisation and validation. We are committed to embracing the potential future directions (detailed below) and welcome collaborative efforts to realise these objectives. By integrating omics data, utilising functional annotations, and considering context-specific interactions, we aim to enhance our understanding of Covid-19 mechanisms and improve the precision of target prioritisation, ultimately



contributing to the development of effective therapeutic strategies for critical Covid-19 patients.

- Integration of omics data. One of the most promising areas for future exploration involves the integration of diverse omics data, including transcriptomic, epigenomic, and proteomic information, which can enhance the

comprehensiveness of the target index. While our study primarily focused on genetic and network evidence (which is less context-specific), we recognise the significance of including these additional layers of omics data information to further refine and validate the target index. The inclusion of cell-type-specific expression information, for instance, could facilitate the

Fig. 6 | Pathway crosstalk for critical Covid-19. It was identified from pathway-derived gene interactions but constrained by target gene prioritisation information (i.e., target index). **a** Gene-centric representation of the crosstalk, with nodes labelled by 'gene symbol @ priority rank' and color-coded by priority rating (target index). The crosstalk involves highly prioritised and interconnected genes. Genes targeted by drugs in phase-III clinical trials for Covid-19 are highlighted in bold. **b** Venn diagram illustrating the overlap between crosstalk genes and Covid-19 phase-III drug targets. The statistical significance (*P*-value), odds ratio, and its 95% CI were calculated using one-sided Fisher's exact test. **c** Tabular display of crosstalk genes

also being Covid-19 phase-III drug targets, along with information on drug candidates and mechanisms of action. **d** Kite-like plot for KEGG pathways enriched in crosstalk genes. Enrichment significance (FDR) was calculated using one-sided Fisher's exact test. Each kite is sized by the number of member genes (indicated by blue dots beneath). Abbreviations include CCR (cytokine-cytokine receptor) and TNF (tumor necrosis factor). **e** Pathway-centric representation of the crosstalk, with node size indicating the number of pathway member genes, and edge thickness proportional to the number of shared genes between two-endpoint pathways.

selection of prioritised targets in specific cell types for experimental validation. This context-specific data is vital for refining the prioritisation of genes for therapeutic targeting. Additionally, the integration of epigenomic and proteomic data can provide insights into post-transcriptional and post-translational modifications, offering a more comprehensive understanding of target functionality. By combining these data types, future research can achieve a deeper comprehension of critical Covid-19 mechanisms and enhance therapeutic target prioritisation.

- Utilisation of functional annotations. The incorporation of functional annotations, such as gene ontology terms and curated pathway information, is another avenue for improving target prioritisation. By associating target genes with specific biological processes or pathways relevant to Covid-19, this approach can enhance the contextualisation of the relationship between gene function and disease progression. Curated pathways from reputable databases, like KEGG, can serve as reliable sources to guide the functional annotation process. This additional layer of data can refine the selection of prioritised targets in specific biological contexts, thus increasing the precision of target prioritisation.

- Context-specific interactions. The importance of considering context-specific interactions, especially tissue-specific or cell-type-specific interactions, cannot be overstated. Accurately representing host-virus interactions necessitates leveraging context-specific networks to obtain a more precise understanding of the disease process within the context of critical Covid-19. These context-specific interactions are invaluable for identifying critical nodes within specific cell types or tissues that play pivotal roles in the disease process. As the importance of context-specific information grows in phenomics^{56,57}, our future endeavors will delve deeper into this aspect.

In summary, our PIC² solution, tailored for critical Covid-19, offers a platform for in silico and timed translation of host genetic findings into drug repurposing. The accessibility, reproducibility, and the potential for scalability make it a robust solution. Looking beyond the scope of this study, the dual-indexes strategy holds promise for scalability, positioning it as an adaptable approach for addressing a wide range of disease domains. Within these domains, computational translational medicine emerges as an indispensable tool, swiftly bridging the gap between host genetic insights and their effective applications in translational therapeutics.

Materials and methods

Generating an index for genetically informed therapeutic targets for critical Covid-19 (PIC²Target) through a genetics-driven and network-based approach (PIT)

The creation of the index involved the implementation of our previously established multi-step integrative prioritisation strategy, which integrates a genetics-driven and network-based approach^{26,34,35}. To elucidate, we used critical Covid-19 GWAS summary-level data²¹ as inputs and harnessed both genetic evidence (from multi-modal regulatory genomic datasets) and network evidence (from high-quality protein interactions), to output a comprehensive list of approximately 14,000 target genes, ranked by their priority ratings on a scale of 0–5. This intricate prioritisation procedure comprises the following key components.

- GWAS SNPs. We input GWAS-detected SNPs with a significant threshold of *P*-value < 5 × 10⁻⁸. In addition, SNPs in linkage disequilibrium (LD; R² ≥ 0.8) were considered according to the European population⁵⁸. The

scoring formula for SNPs accounted for various aspects of disease genetic associations, including *p*-values, the significance threshold, and R² values.

- Core genes. SNPs scored above were used to define core genes through evidence of genomic proximity (nearby genes), e/pQTL (QTL genes)^{59–65}, and PCHi-C (conformation genes)^{66–71}. In addition to SNP scores, the scoring for core genes also encompassed the following factors: (1) the distance-to-SNP window for genomic proximity; (2) the significance level of genetic associations with gene expression for eQTL datasets or protein abundance for pQTL datasets; and (3) the strength of gene promoters physically interacting with SNP-harboring genomic regions for PCHi-C datasets. The empirical cumulative distribution function (eCDF) was estimated based on all SNP-gene pairs to ensure scaling within the 0–1 range.

- Peripheral genes. Using core genes as seeds, we implemented the RWR algorithm to identify (non-seed) peripheral genes under network influence through exploiting knowledge of protein interactions. The RWR algorithm is a propagation-based method that simulates the information flow within a network to estimate connectivity (affinity) between two nodes. This method assigned affinity scores to genes based on their connectivity to a set of 'seed' genes (in this context, core genes identified through genetic evidence). We established a network of protein interactions primarily from the STRING database³⁶, incorporating interactions with an interaction score no less than 0.7 (i.e., high confidence) and labelled as 'experiments' and 'databases' (i.e., manually curated). This interaction network encompassed approximately 14,000 nodes/genes (and around 201,000 interactions/edges). There were two types of nodes in this gene interaction network: nodes labelled as seed genes (along with their corresponding scores as weights), and the remaining nodes as non-seed genes. The algorithm started a random walk from each seed gene and iteratively traversed neighboring genes along network edges. At each iteration, the walker faced a choice: either moving to a randomly chosen neighbor or jumping back to the seed node. This process continued until a steady state was reached, characterised by the probability distribution of each gene stabilising. The resulting steady probability values, ranging from 0 to 1, stored affinity scores for each gene relative to seed core genes. For each of datasets (i.e., proximity, QTL, and PCHi-C), the process above yielded a predictor containing core and peripheral genes, accompanied by affinity scores that quantified their network connectivity to seed core genes. Notably, genes with higher network connectivity to seed core genes were assigned higher affinity scores. Generally, seed genes were more likely to receive higher affinity scores than non-seed peripheral genes. Nonetheless, a peripheral gene might receive a high affinity score if it exhibited high connectivity to most (if not all) seed genes. Thus, non-seed peripheral genes with higher affinity scores denoted genes highly influenced by the network.

- Gene-predictor matrix. Employing the abovementioned multi-step scoring procedures (GWAS/LD SNPs – core genes – peripheral genes), a gene-predictor matrix was constructed. This matrix featured rows for genes and columns for predictors. Within this matrix, affinity scores were amalgamated, collectively incorporating genetic evidence and network evidence.

- Performance benchmarking for target-index generation/prioritisation schemes. Leveraging the gene-predictor matrix constructed above, various generation schemes were evaluated for combining predictors. Performance comparisons were conducted between (1) meta-analysis-like schemes, including logistic and Fisher's combined methods (detailed in the

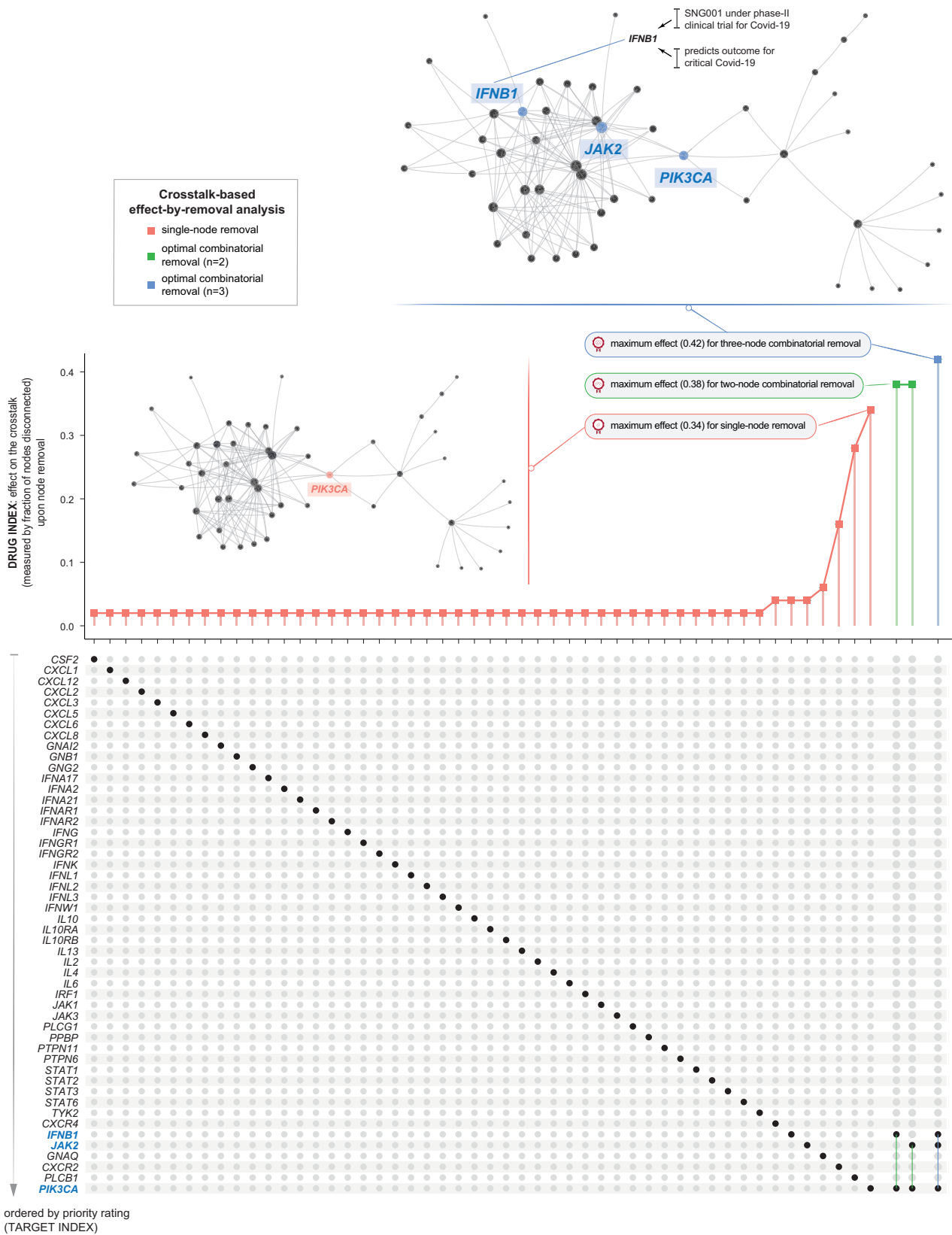


Fig. 7 | The drug index (PIC²Drug) for critical Covid-19. Effect of node removal, either individually or in combination, on the crosstalk. The fraction of disconnected nodes (i.e., the drug index) on the y-axis is plotted against node removal (indicated by solid circles beneath) on the x-axis. Notably, only the optimal removal with the

largest effect is illustrated for two- or three-node combinatorial removal. Inserted is visualisation of the crosstalk, with the same layout as Fig. 6a, but labelled only for genes with optimal removals.

subsequent section ‘Gene-level prioritisation’), and (2) conventional schemes, comprising sum (additively aggregating affinity scores across predictors) and max (selecting the maximum one across predictors). Performance was also compared with a Naive scheme, wherein genes were prioritised based on the frequency of being therapeutically targeted by licensed/approved drugs. This Naive prioritisation, motivated by drug repurposing (though not applicable for prioritising new targets), served as the *status quo* baseline for target-index prioritisation. Performance assessment relied on the area under the ROC curve (AUC) for distinguishing Covid-19 phase-III drug targets from simulated negative target controls. Information on phase-III drug targets in Covid-19 was sourced from the ChEMBL database⁷², encompassing drug candidates and mechanisms of action explaining drug efficacy. Simulated negative target controls for Covid-19 were established based on known therapeutics from ChEMBL, involving three steps: (1) defining the druggable landscape as all known target genes across various diseases and drug development phases; (2) extracting Covid-19 drug targets (regardless of drug development phase) from ChEMBL, along with their interacting neighbours determined by interaction information from databases (STRING³⁶ and Pathway Commons⁷³); and (3) simulating negative targets as genes from the druggable landscape after excluding Covid-19 drug targets and their interacting neighbours.

- Target-index generation/prioritisation. For comprehensive mathematical explanations regarding Fisher’s combined meta-analysis method, please consult our previous publication³⁴. Building upon the performance evaluation elucidated in the previous section, in this study, PIT used a logistic combined meta-analysis method to consolidate predictors within the gene-predictor matrix prepared earlier. Within the gene-predictor matrix, affinity scores for a given predictor were first converted into *P*-like values using empirical cumulative density function (eCDF) estimated from affinity scores of all genes within that predictor (Eq. 1).

$$P_i^j = eCDF(AF_i^j), \quad (1)$$

Here, AF_i^j denotes the affinity score for the *i*th gene concerning the *j*th predictor, P_i^j is the corresponding converted *P*-value, and *eCDF* is estimated based on all genes.

Subsequently, converted *P*-values for a gene were combined across predictors via a logistic combined method (Eqs. 2–4); see the publication³⁸ for a more detailed explanation. Ultimately, the combined *P*-value underwent rescaling to yield priority rating (continuously from 0–5 and corresponding to priority rank from 1 to ~14,000). This output is referred to as an index for target genes (denoted as PIT; Eq. 5).

$$x_i = -2 \sum_j^J \left[P_i^j / (1 - P_i^j) \right], \quad (2)$$

$$x_i \sim St(5J + 4), \quad (3)$$

$$CP_i = CDF(x_i), \quad (4)$$

$$PIT_i = 5 \times \frac{-\log CP_i - \text{MIN}_k^K (-\log CP_k)}{\text{MAX}_k^K (-\log CP_k) - \text{MIN}_k^K (-\log CP_k)}, \quad (5)$$

In these equations, P_i^j denotes the converted *P*-value for the *i*th gene within the *j*th predictor, *J* is the number of predictors, *St*(5*J*+4) denotes Student’s *t*-distribution with 5*J* + 4 degrees of freedom, CP_i symbolises the combined *P*-value for the *i*th gene (that is, CDF valued at x_i), and *PIT_i* represents the priority rating for the *i*th gene (among *K* genes, where *K* is ~14,000).

- Analysis using human genes/proteins that directly interact with SARS-CoV-2 viral proteins. Interactions between SARS-CoV-2 viral proteins and human proteins were obtained from the BioGRID database (version 4.4.214)³⁷. A total of 2637 virus-host interactions, encompassing 29 SARS-CoV-2 viral proteins and 652 human proteins, were independently

identified by two or more studies^{39–45}. Fisher’s exact test was employed to estimate the significance of the overlap between virus-interacting human proteins and the top 1% of prioritised genes. This test was also performed for human proteins interacting with each of the 29 SARS-CoV-2 viral proteins.

- Analysis using known Covid-19 human host pathway. The Covid-19 human host gene interaction pathway, expertly curated and extensively documented in www.genome.jp/pathway/hsa05171 as part of the KEGG database⁴⁶, focuses notably on the downstream effects of the SARS-Cov-2 virus on the host. These effects involve not merely direct interactions with host proteins, but also encompass various downstream outcomes such as the activation of the NF-κB pathway, IL-6 production, systemic inflammation, and many others. Retrieving this information from KEGG involved utilising the KEGGgraph package (version 1.56.0) to download the KGML-formatted file, which was subsequently processed and converted into an *igraph* R object. Fisher’s exact test was used to assess the significance of the overlap between the top 1% prioritised genes and the genes associated with the host pathway. The stress majorisation algorithm was applied to visualise the network layout of gene interactions within the host pathway.

- Leading prioritisation. The concept of leading prioritisation pertains to the core subset of all prioritised genes, conveniently referred to as ‘leading target genes’, which contribute to the enrichment of phase-III drug targets in Covid-19. This enrichment was visually presented as the leftmost region of the peak within the running enrichment plot generated by target set enrichment analysis. The *dnet* package (version 1.1.7)⁷⁴ facilitated this analysis by quantifying the degree to which Covid-19 phase-III drug targets were enriched at the leading prioritisation. Essentially, Covid-19 phase-III drug targets found at the leading prioritisation were assigned high priority. The significance (*P*-value) of the enrichment was ascertained through permutation tests conducted 50,000 times.

- Cross-disease prioritisation map. We employed a self-organising algorithm implemented in the *supraHex* package (version 1.40.0)^{75–77} to construct a cross-disease prioritisation map for leading target genes in critical Covid-19. A butterfly-shaped map, consisting of *N* = 49 hexagons, was trained using the input prioritisation matrix (containing priority rating) across *M* = 9 diseases. These diseases included critical Covid-19 (this study) and 8 immune-mediated diseases³⁴: ankylosing spondylitis, Crohn’s disease, juvenile idiopathic arthritis, multiple sclerosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, and ulcerative colitis. The codebook matrix associated with the trained map was utilised to provide a disease-specific view of target gene prioritisation. To further visualise inter-disease relationships, a 2D square map lattice was used to self-organise diseases in a manner that geometric locations within this square lattice delineated the relationships between diseases.

- Enrichment analysis for leading target genes. *OpenXGR*^{78,79} was used for conducting enrichment analysis on leading target genes to identify over-representation (i.e., enrichments) in terms of cellular basis (cell types), functions (GO terms), pathways (KEGG pathways), and evolutionary history (phylostrata). Cell-type-specific gene signatures were extracted from *xCell*⁸⁰. GO Biological Process and Molecular Function terms, alongside their annotated genes, were obtained from NCBI⁸¹. Environmental Information Processing (EIP) pathways and their member genes were sourced from KEGG⁴⁶, which covers a diverse range of molecular pathways, including EIP and six other categories. Each category captures its unique knowledge domain of molecular interactions with relative completeness. Genes first created at each of 16 phylostrata were extracted from this study⁸² to delve into the evolutionary history of leading genes. Enrichment analysis was performed using one-sided Fisher’s exact test, producing FDR, *Z*-scores, odds ratio and its 95% CI. A bipartite graph was crafted to illustrate the relationship between enriched cell types and their member genes.

Generating an index for repurposed drugs in critical Covid-19 (PIC²Drug) through pathway crosstalk-based effect-by-removal analysis (PID)

- Identification of pathway crosstalk. This identification focused on genes with high ratings and strong interconnectedness. The search involved

selecting a subset of gene interactions that were merged from KEGG EIP pathways. The intention of choosing EIP pathways aimed to emphasise molecular interactions involved in signal transductions that are less organism, cell type, and disease-specific. These EIP pathways encompassed 2158 genes, collectively forming a network of 15,375 interactions between these genes. Each interaction in this gene network was present in at least one EIP pathway(s). The process of identifying pathway crosstalk utilised a heuristic solution for solving the prize-collecting Steiner tree problem^{74,78}. To assess the statistical significance (P -value) of the identified crosstalk, a degree-preserving node permutation test was conducted with 100 iterations. Additionally, the analysis allowed for the specification of a desired number (e.g., 50) of nodes or genes in the resulting crosstalk could be specified, and this desired output was obtained through a well-established iterative search procedure. For comprehensive details, please refer to our previous publications^{74,78,79}. The identified crosstalk, apart from its portrayal as a gene network, was also visualised as a pathway-centric network, in which pathways were depicted as nodes and their inferred connections as network edges. Only pathways that exhibited significant over-representation in crosstalk genes, as determined by a one-sided Fisher's exact test, were included as nodes. The edges were initially inferred based on the shared member genes between pathways, and then filtered by identifying the minimum spanning tree using the *igraph* package (version 1.6.0). This process retained only the edges present within the resulting tree, with the thickness of edges adjusted proportionally according to the number of shared member genes between the two endpoint pathways.

- Crosstalk-based drug repurposing. The drug repurposing analysis relied on information extracted from ChEMBL⁷², which aggregated therapeutic data on current phase-III and approved therapeutics (including drugs, development phases, target genes and disease indications). For a disease indication, drugs reaching the maximum phase of development were selected for a target gene, considering that selected target genes had well-defined mechanisms of action and could explain the efficacy of drugs in treating the disease. These selected target genes were categorised into two distinct groups: one encompassing approved drug targets (i.e., genes targeted by any approved drugs), and the other comprising phased drug targets (i.e., genes targeted by non-approved phased drugs that are in developmental phases, not by any approved drugs). One-sided Fisher's exact test was used to evaluate the statistical significance of crosstalk genes enriched for two drug target groups (i.e., approved drug targets and phased drug targets).

- Crosstalk-based effect-by-removal analysis. The removal analysis was designed to evaluate the effect of nodes on the crosstalk. This analysis comprised two dimensions: (i) the individual removal of nodes (i.e., single-node removal), and (ii) the simultaneous removal of nodes in combination (i.e., combinatorial removal). Single-node removal was undertaken to define an index for repurposed drugs targeting the node being removed. In instances where a specific node earmarked for removal was pivotal for the network, its removal would result in the disconnection of a substantial fraction of nodes from the largest network component. This fraction, embodying the number of disconnected nodes, was used to quantify the drug index. Combinatorial removal, on the other hand, sought to select optimal combinations for targeting, with the objective of maximising the effect resulting from the simultaneous removal of specific node combinations (e.g., removing two nodes at once). In other words, the largest fraction of disconnected nodes was sought for removing a specific node combination. The effect of node removal was visually captured through an upset plot using the *ggupset* package (version 0.3.0). This visualisation method allowed us to clearly depict the outcomes of both single-node and combinatorial node removal.

Statistics and reproducibility

All statistical analyses were performed using R (version 4.3.0), with one-sided Fisher's exact test used for performing enrichment analysis, reporting the statistical significance, odds ratio, and its 95% confidence intervals. In the spirit of supporting reproducibility, we offer showcases (<http://www.genetictargets.com/PIC2/showcase>) that enable users to reproduce all

findings presented in this study. These showcases include input data, line-by-line codes, as well as tabular and graphical outputs, all embedded into a single self-contained HTML file. Within this file, users can find comprehensive, step-by-step instructions guiding them through showcases and providing a preview of anticipated results.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data for graphs within this paper are provided in the Supplementary Data files. More specifically, source data in Supplementary Data 1 are used to generate Fig. 2b, source data in Supplementary Data 2 for generating Fig. 2c, source data in Supplementary Data 3 for generating Fig. 3a, source data in Supplementary Data 4 for generating Fig. 3d, e, source data in Supplementary Data 5 for generating Fig. 4b, source data in Supplementary Data 6 for generating Fig. 4c, source data in Supplementary Data 7 for generating Fig. 6a, source data in Supplementary Data 8 for generating Fig. 6d, and source data in Supplementary Data 9 for generating Fig. 7. Together with these source data, data used and generated during the current study are also accessible through a web-based open-access portal at <http://www.genetictargets.com/PIC2>. This portal empowers users to interactively explore two indexes comprising genetic targets and repurposed drugs for critical Covid-19.

Code availability

Software codes have been encapsulated within an R package, accessible at <https://hfang-bristol.github.io/PIC2>. Additionally, codes have been deposited into a public data repository at Figshare (<https://doi.org/10.6084/m9.figshare.24967797>).

Received: 22 April 2023; Accepted: 7 February 2024;

Published online: 16 February 2024

References

- Samson, M. et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
- Timmann, C. et al. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446 (2012).
- Thye, T. et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet.* **44**, 257–259 (2012).
- Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
- Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug repurposing. *Nat. Rev. Genet.* **22**, 658–671 (2021).
- Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, 1–16 (2017).
- Fang, H. PIER: web-based facilities tailored for genetic target prioritisation harnessing human disease genetics, functional genomics and protein interactions. *Nucleic Acids Res* **50**, W583–W592 (2022).
- Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
- Waring, M. J. et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486 (2015).
- Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
- Williamson, E. J. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).

12. Van Der Made, C. I. et al. Presence of Genetic Variants among Young Men with Severe COVID-19. *JAMA* **324**, 663–673 (2020).
13. Niemi, M. E., Daly, M. J. & Ganna, A. The human genetic epidemiology of COVID-19. *Nat. Rev. Genet.* **23**, 533–546 (2022).
14. Nicholson, J. K. Molecular phenomic approaches to deconvolving the systemic effects of SARS-CoV-2 infection and post-acute COVID-19 syndrome. *Phenomics* **1**, 143–150 (2021).
15. Eskandarzade, N. et al. Network for network concept offers new insights into host- SARS-CoV-2 protein interactions and potential novel targets for developing antiviral drugs. *Comput Biol. Med* **146**, 105575 (2022).
16. Ramos-Casals, M., Brito-Zerón, P. & Mariette, X. Systemic and organ-specific immune-related manifestations of COVID-19. *Nat. Rev. Rheumatol.* **17**, 315–332 (2021).
17. Ellinghaus, D. et al. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
18. Shelton, J. F. et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).
19. Niemi, M. E. K. et al. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
20. Pairo-Castineira, E. et al. Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
21. Odhams, A. et al. Whole genome sequencing reveals host factors underlying critical Covid-19. *Nature* **607**, 97–103 (2022).
22. Horowitz, J. E. et al. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* **54**, 382–392 (2022).
23. Huffman, J. E. et al. Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nat. Genet* **54**, 125–127 (2022).
24. Pairo-Castineira, E. et al. GWAS and meta-analysis identifies multiple new genetic mechanisms underlying severe Covid-19. *Nature* **617**, 764–768 (2023).
25. Feng, Z., Ren, X., Duren, Z. & Wang, Y. Human genetic variants associated with COVID-19 severity are enriched in immune and epithelium regulatory. *Netw. Phenom.* **2**, 389–403 (2022).
26. Fang, H., Chen, L. & Knight, J. C. From genome-wide association studies to rational drug target prioritisation in inflammatory arthritis. *Lancet Rheumatol.* **2**, e50–e62 (2020).
27. Plenge, R. M. Priority index for human genetics and drug discovery. *Nat. Genet.* **51**, 1073–1075 (2019).
28. Robertson, C. C. et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for Type 1 diabetes. *Nat. Genet* **53**, 962–971 (2021).
29. Schwartzenuber, J. et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet* **53**, 392–402 (2021).
30. Chia, R. et al. Identification of genetic risk loci and prioritization of genes and pathways for myasthenia gravis: a genome-wide association study. *Proc. Natl Acad. Sci.* **119**, e2108672119 (2022).
31. Bao, C. et al. Priority index for asthma (PIA): In silico discovery of shared and distinct drug targets for adult- and childhood-onset disease. *Comput Biol. Med* **162**, 107095 (2023).
32. Bao, C., Wang, H. & Fang, H. Genomic evidence supports the recognition of endometriosis as an inflammatory systemic disease and reveals disease-specific therapeutic potentials of targeting neutrophil degranulation. *Front Immunol.* **13**, 758440 (2022).
33. Fang, H. & Jiang, L. Genetic prioritization, therapeutic repositioning and cross-disease comparisons reveal inflammatory targets tractable for kidney stone disease. *Front Immunol.* **12**, 687291 (2021).
34. Fang, H. et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet* **51**, 1082–1091 (2019).
35. Fang, H. & Knight, J. C. Priority index: database of genetic targets in immune-mediated disease. *Nucleic Acids Res.* **50**, D1358–D1367 (2022).
36. Szklarczyk, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
37. Oughtred, R. et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541 (2019).
38. Loughin, T. M. A systematic comparison of methods for combining *p*-values from independent tests. *Comput Stat. Data Anal.* **47**, 467–485 (2004).
39. Banerjee, A. K. et al. SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell* **183**, 1325–1339.e21 (2020).
40. Gordon, D. E. et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eaabe9403 (2020).
41. Stukalov, A. et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**, 246–252 (2021).
42. Li, J. et al. Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* **2**, 99–112.e7 (2021).
43. Lee, J. G. et al. Characterization of SARS-CoV-2 proteins reveals Orf6 pathogenicity, subcellular localization, host interactions and attenuation by Selnexor. *Cell Biosci.* **11**, 1–12 (2021).
44. Liu, X. et al. SARS-CoV-2–host proteome interactions for antiviral drug discovery. *Mol. Syst. Biol.* **17**, 1–26 (2021).
45. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
46. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
47. Beigel, J. H. et al. Remdesivir for the treatment of Covid-19 — Final Report. *N. Engl. J. Med.* **383**, 1813–1826 (2020).
48. Kalil, A. C. et al. Baricitinib plus remdesivir for hospitalized adults with Covid-19. *N. Engl. J. Med.* **384**, 795–807 (2021).
49. Writing Committee for the REMAP-CAP Investigators. Long-term (180-Day) outcomes in critically ill patients with COVID-19 in the REMAP-CAP randomized clinical trial. *JAMA* **329**, 39–51 (2023).
50. Basile, M. S. et al. The PI3K/Akt/mTOR pathway: a potential pharmacological target in COVID-19. *Drug Discov. Today* **27**, 848–856 (2022).
51. Hopkins, C. et al. Platforms for personalized polytherapeutics discovery in COVID-19. *J. Mol. Biol.* **433**, 166495 (2021).
52. Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B. & Knijnenburg, T. A. Combining dependent *P*-values with an empirical adaptation of Brown's method. *Bioinformatics* **32**, i430–i436 (2016).
53. Liu, Y. & Xie, J. Cauchy combination test: a powerful test with analytic *p*-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* **115**, 393–402 (2020).
54. Lv, T. et al. Defense of COVID-19 by human organoids. *Phenomics* **1**, 113–128 (2021).
55. Basu, A., Pamreddy, A., Singh, P. & Sharma, K. An adverse outcomes approach to study the effects of SARS-CoV-2 in 3D organoid models. *J. Mol. Biol.* **434**, 167213 (2022).
56. Ying, W. Phenomic studies on diseases: potential and challenges. *Phenomics* **3**, 285–299 (2023).
57. Jin, L. Welcome to the phenomics journal. *Phenomics* **1**, 1–2 (2021).
58. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. *Nature* **491**, 56–65 (2012).
59. Leland Taylor, D. et al. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc. Natl Acad. Sci. USA* **166**, 10883–10888 (2019).

60. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
61. Momozawa, Y. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
62. Quach, H. et al. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016).
63. Chen, L., Ge, B., Casale, F. P., Downes, K. & Pastinen, T. Genetic drivers of epigenetic and transcriptional variation in human immune. *Cells Cell* **167**, 1398–1414 (2016).
64. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1118 (2014).
65. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
66. Jung, I. et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* **51**, 1442–1449 (2019).
67. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
68. Choy, M. et al. Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. *Nat. Commun.* **9**, 2526 (2018).
69. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
70. Miguel-Escalada, I. et al. Human pancreatic islet 3D chromatin architecture provides insights into the genetics of Type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).
71. Song, M. et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).
72. Mendez, D. et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
73. Rodchenkov, I. et al. Pathway commons 2019 update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* **48**, D489–D497 (2020).
74. Fang, H. & Gough, J. The ‘dnet’ approach promotes emerging research on cancer patient survival. *Genome Med.* **6**, 64 (2014).
75. Fang, H. & Gough, J. supraHex: An R/Bioconductor package for tabular omics data analysis using a supra-hexagonal map. *Biochem. Biophys. Res. Commun.* **443**, 285–289 (2014).
76. Fang, H. & Wang, K. Regulatory genomic data cubism. *iScience* **3**, 217–225 (2018).
77. Tan, Y., Jiang, L., Wang, K. & Fang, H. I3: a self-organising learning workflow for intuitive integrative interpretation of complex genetic data. *Genom. Proteom. Bioinforma.* **17**, 503–510 (2019).
78. Fang, H., Knezevic, B., Burnham, K. L. & Knight, J. C. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med.* **8**, 1–20 (2016).
79. Bao, C. et al. OpenXGR: a web-server update for genomic summary data interpretation. *Nucleic Acids Res.* **51**, W387–W396 (2023).
80. Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
81. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **39**, D52–D57 (2011).
82. Trigos, A. S., Pearson, R. B., Papenfuss, A. T. & Goode, D. L. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proc. Natl Acad. Sci.* **114**, 6406–6411 (2017).

Acknowledgements

This work is funded by National Natural Science Foundation of China (32170663 to H.F. and 82300263 to C.B.), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (awarded to H.F.), and Innovative Research Team of High-Level Local Universities in Shanghai.

Author contributions

H.F.: Conceptualisation, Methodology, Software, Investigation, Resources, Data Curation, Visualisation, Supervision, Project Administration, Funding Acquisition, and Writing – original draft.; J.F.: Supervision, Funding Acquisition, and Writing – Review & Editing.; C.B.: Methodology, Data Curation, Visualisation, Funding Acquisition, and Writing – Review & Editing.; L.C.: Supervision, and Writing – Review & Editing.; Z.Z.: Investigation, and Writing – Review & Editing.; S.W.: Methodology, and Writing – Review & Editing.; L.J.: Investigation, Visualisation, and Writing – Review & Editing.; J.W.: Investigation, and Writing – Review & Editing.; C.L.: Methodology, and Writing – Review & Editing.; S.L.: Writing – Review & Editing.; Y.D.: Writing – Review & Editing.; Z.F.: Writing – Review & Editing.; S.H.: Writing – Review & Editing.; T.T.: Writing – Review & Editing.; Y.Y.: Writing – Review & Editing.; K.Z.: Writing – Review & Editing.; J.S.: Writing – Review & Editing.; J.L.: Writing – Review & Editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains Supplementary Material available at <https://doi.org/10.1038/s42003-024-05897-0>.

Correspondence and requests for materials should be addressed to Liye Chen, Chaohui Bao, Jian Fei or Hai Fang.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Chien-Yu Chen and David Favero.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

¹Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ²School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. ³Translational Health Sciences, University of Bristol, Bristol BS1 3NY, UK. ⁴Network and Information Center, Shanghai Jiao Tong University, Shanghai 200240, China. ⁵MRC London Institute of Medical Sciences, Imperial College London, London W12 0HS, UK. ⁶Precision Research Center for Refractory Diseases, Institute for Clinical Research, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 201620, China. ⁷College of Finance and Statistics, Hunan University, Changsha 410079 Hunan, China. ⁸College of Health Science and Technology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ⁹Key Laboratory of RNA Science and Engineering, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China. ¹⁰Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK. ¹¹Department of General Surgery, Ruijin Hospital Luwan Branch, Shanghai Jiao Tong University School of Medicine, Shanghai 200020, China. ¹²Department of General Surgery, Pancreatic Disease Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ¹³These authors contributed equally: Zhiqiang Zhang, Shan Wang, Lulu Jiang, Jianwen Wei, Chang Lu. ✉ e-mail: liye.chen@ndorms.ox.ac.uk; chaohuibao@sjtu.edu.cn; feijian@hotmail.com; fh12355@rjh.com.cn