Methods

High throughput data acquisition and analysis

We collected transcriptomics data from the GEO[41] database, and proteomics data from relevant publication[67]. A total of 159 transcriptomics and 20 proteomics samples relevant to the CD4+ T cells were selected (Supplementary Data 1). Transcriptomics data analysis was performed using the affy[68] and limma[69] R packages. Because we aimed to characterize gene activities instead of gene expression levels, the processed transcriptomics data were discretized (active = 1; inactive = 0) and samples for each cell type were combined together. Next, if a gene was active in more than 50% of the samples in which the probe was present were considered as active (*see Supplementary Note 2*). Similarly, proteins expressed in more than 50% samples in the proteomics dataset were considered as active. In the proteomics datasets, protein IDs were mapped to gene IDs. Next, we integrated activities from transcriptomics and proteomics datasets. First, biological entities that were overlapped in both types of data were selected. Second, the non-overlapped high confidence genes were added to the overlapped genes if exclusively present in the transcriptomics data, and expressed in at least 90% of samples. Third, gene IDs of non overlapping high-confidence proteins were selected if exclusively present in the proteomics dataset, and ranked in the top 25% (fourth quartile) based on their abundance. These genes were added to the previous list of overlapped genes and high confidence genes exclusively active in transcriptomics data and resulted in the final gene activity matrices (*Supplementary Note 2; Supplementary Data 7*). We used these cutoffs to include high-confidence genes while not limiting the number of false positives by selecting a loose cutoff.

Cell type specific genome-scale metabolic model reconstruction

We used GIMME[29] method (in COBRA toolbox) to construct the metabolic models of different CD4+ T cells (Naïve, Th1, Th2, Th17, iTreg). The input for GIMME were generic human Recon 3D (as a template) and gene activities based on integrated multi-omics data (for naïve, Th1, Th2 and Th17) and transcriptomics data (for iTregs). The template Recon 3D[26] was modified prior to construct CD4+ T cell specific metabolic models. These modifications include GPR rules, media conditions, and reaction directionality. In addition, new reactions involved in the biomass objective function were added, and some reactions were removed (*Supplementary Note 3* Supplementary Data 8). The used transcriptomics or proteomics data have information about genes/proteins instead of transcript variants. Therefore, we updated transcript IDs provided in Recon 3D to Entrez gene IDs. A total of 1,892 genes were included in the Recon 3D model. Furthermore, because different CD4+ T cells have different nutrient uptake preferences, we used three types of media conditions (one for each effector, naïve, and inducible regulatory T cells). For all cell subtypes, in addition to the basal metabolites (i.e. h2o, o2, H, o2s, co2, pi, h2o2, hco3, h2co3, and co), glucose, glutamine and other amino acids were set as open (un-constrained) for uptake. Effector T cells are highly glycolytic and do not uptake fatty acids from environment whereas naïve T cells and iTregs sense FAs. Thus, the fatty acid exchange reactions were opened for construction of naïve and iTreg cells whereas these were closed for effector T cells. Furthermore, the directionality of 39 reactions was updated based on the Recon 2.2.05 model[70]. Because of the lack of T cell specific data, the biomass objective function was adopted from macrophage model iAB-AMØ-1410 and added to the Recon3D. Additionally, to investigate the effect of biomass objective function on constructed models, we built two models using biomass objective functions from (1) Recon 3D and (2) iAB-AMØ-1410 models. The reactions in output models generated based on each biomass function were compared. The

models based on the two objective functions were not significantly different, therefore models that are constructed based on biomass reaction adopted from iAB-AMØ-1410 were used in next analyses. Models were further reduced by removing the dead end reactions. The models were investigated to perform basic properties using leak test, gene deletion and further refined in iterative manner. Refined models were then subjected to 460 metabolic tasks that were used with Recon 3D model and included in *Test4HumanFctExt* function in COBRA (Supplementary Data 9). The constructed models were simulated using the Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA).

Model validation:

Models were validated based on literature knowledge related to active pathways and based on the accuracy of predicted essential genes. CD4+ T cell specific metabolic functions were searched in the literature using PubMed. The flux distribution of metabolic pathways was obtained under wild type conditions. We searched the flux through pathways active in T cells such as aerobic glycolysis (lactate production), fatty acid synthesis or fatty acid oxidation in all the models. For gene deletion based validation, flux distribution was obtained under knockout conditions using Flux Balance Analysis (FBA) and Minimization of Metabolic Adjustment (MoMA)[71] in COBRA toolbox[72]. Gene essentiality analysis was performed using *singleGeneDeletion* in the COBRA toolbox using the MoMA method[71]. The data for experimentally tested essential and nonessential genes for human were obtained from the OGEE database[44]. In this database the essentiality data for human was compiled using 18 experiments across various cell lines that include RNAi based inhibition, CRISPR, and CRISPR-CAS9 systems. To calculate the predictive power of the model, predicted essential genes were compared with experimentally observed essential and conditionally essential genes

reported in the OGEE database. Essential and conditionally essential genes were merged together.

Drugs target map

The drugs and their annotations including targets were exported from The Drug Repurposing Hub[40] in theConnectivityMap (CMap) databse[25]. All withdrawn drugs were first removed from the list. In this list, the gene symbols of target genes of drugs were converted to Entrez IDs. Next, we searched Entrez IDs from CMap data in the genes of metabolic models. For each mapped gene in the model, the drugs were listed to create a drug target map.

Differentially expressed metabolic genes in autoimmune diseases

Datasets GSE56649[73], GSE43591[74], and GSE93170[75] were obtained from the GEO database. These datasets were focused on case control studies and collected from peripheral blood mononuclear cells (PBMC). Raw data files were processed using the affy and limma packages[68,69] in Bioconductor/R. The limma package was used to identify differentially expressed genes between patients with a confirmed disease and healthy control. For significant differential expressed, a fold change cutoff of 1.5 or 2 (for GSE56649) was used with adjusted P-value < 0.05. Two fold cutoff was selected when we obtained a high number of differentially expressed genes, whereas 1.5 fold cutoff chosen when the two fold was too strict for that datasets.

Perturbation of metabolism and perturbation effect score (PES)

In naive, Th1, Th2, and Th17 models, the knockout of genes that are targets of existing drugs was performed in the COBRA toolbox using MoMA[71]. For each knockout we investigated change in fluxes of disease associated genes identified by differential expression analysis. The change in fluxes was computed using flux ratio of perturbed fluxes/ WT flux. Next, using flux

ratios we calculated all fluxes that are affected by each perturbation. We summed up all up-regulated genes that are down-regulated after perturbation (N1), and up-regulated genes that are up-regulated after perturbation (N2). The effects on up-regulated genes were calculated using E1. The effect size of downregulated genes that were upregulated upon knock out were also calculated. The knockout effect on the downregulated genes was calculated using E2. For each perturbed gene, a perturbation effect score (PES) was calculated using the sum of Eq 1 and Eq 2.

$$Effect\ on\ upregulated\ disease\ genes(E1) = \frac{(N1 - N2)}{Total\ number\ of\ affected\ fluxes}$$

$$Effect\ on\ down\ regulated\ disease\ genes(E2) = \frac{(N1 - N2)}{Total\ number\ of\ affected\ fluxes}$$

$$PES = E1 + E2$$

Furthermore, to prioritize drug targets utilizing all the ranks across all four metabolic models, we computed Z-score ($\frac{x-\mu}{\sigma}$) for each CD4+ T cell using ranks of PES scores. For each disease type, all Z-scores were summed up to calculate an aggregated Z-score.

Pathway enrichment analysis

For biological processes enrichment analysis, we used DAVID[76], and STRING databases[77]. Gene Ontology biological processes[78], KEGG pathways[79], and reactome pathways[80] were used for enrichment. A cutoff of 1% False Discovery Rate (FDR) was used for significant enrichment.