HHS Public Access

Author manuscript

Immunity. Author manuscript; available in PMC 2018 October 17.

Published in final edited form as:

Immunity. 2017 October 17; 47(4): 648–663.e8. doi:10.1016/j.immuni.2017.09.006.

Early transcriptional divergence marks virus-specific primary human CD8+ T cells in chronic versus acute infection

David Wolski[1,2,3], **Peter Foote**[1], **Diana Chen**[1], **Lia L Lewis-Ximenez**[4], **Catherine Fauvelle**[2,3], **Jasneet Aneja**[1], **Andreas Walker**[5], **Pierre Tonnerre**[1], **Almudena Torres-Cornejo**[1], **Daniel Kvistad**[1], **Sabrina Imam**[6], **Michael T Waring**[7,8], **Damien C Tully**[7], **Todd M Allen**[7], **Raymond T Chung**[1], **Jörg Timm**[5], **W Nicholas Haining**[6], **Arthur Y Kim**[9], **Thomas F Baumert**[2,3,10], and **Georg M Lauer**[1,11]

[1]Gastrointestinal Unit and Liver Center, Massachusetts General Hospital; Harvard Medical School; Boston, MA, 02114; USA

[2]Inserm, U1110; Institut de Recherche sur les Maladies Virales et Hépatiques; Strasbourg, 67000; France

[3]Université de Strasbourg; Strasbourg, 67081; France

[4]Laboratory of Viral Hepatitis, Oswaldo Cruz Institute; FIOCRUZ, Rio de Janeiro, 21040; Brazil

[5]Institute for Virology, University Hospital Düsseldorf; Heinrich Heine University; Düsseldorf, 40225; Germany

[6]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, 02115; USA

[7]Ragon Institute of MGH, MIT and Harvard; Cambridge, MA, 02139; USA

[8]Howard Hughes Medical Institute; Chevy Chase, MD, 20815; USA

[9]Division of Infectious Disease; Massachusetts General Hospital; Boston, MA, 02114, USA

[10]Institut Hospitalo-Universitaire, Pôle Hépato-digestif; Nouvel Hôpital Civil; Strasbourg, 67000; France

Summary

Distinct molecular pathways govern the differentiation of CD8+ effector T cells into memory or exhausted T cells during acute and chronic viral infection but these are not well-studied in humans. Here, we employed an integrative systems immunology approach to identify transcriptional commonalities and differences between virus-specific CD8+ T cells from patients with persistent and spontaneously resolving hepatitis C virus (HCV) infection during the acute

Correspondence to: Georg M. Lauer, Gastrointestinal Unit, Massachusetts General Hospital and Harvard Medical School, 55 Fruit Street, Boston, MA 02114, glauer@mgh.harvard.edu.
[11]Lead Contact

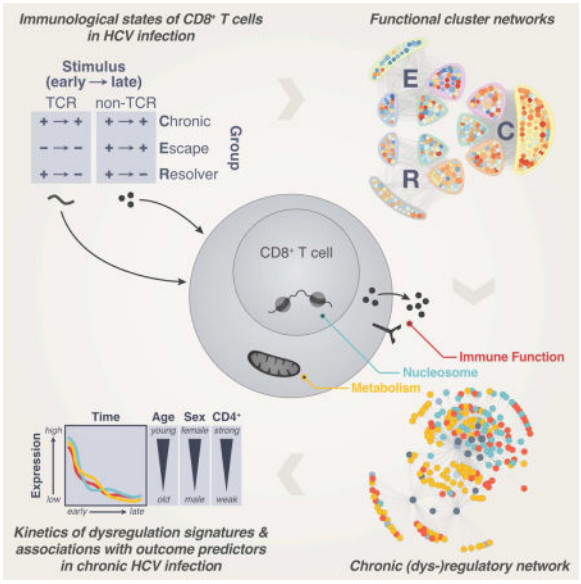
Author Contributions

Conceptualization, Methodology & Writing – Original Draft, D.W. and G.M.L.; Formal Analysis, D.W.; Investigation, D.W., P.K.F., D.C., C.F., A.W., D.K., S.I., M.T.W, and D.C.T.; Resources, L.L.L.X., A.Y.K., T.M.A., R.T.C., J.T., W.N.H., T.F.B., and G.M.L.; Data curation, D.W., J.A., A.Y.K., and G.M.L.; Writing – Review & Editing, D.W., P.T., A.T.C., A.Y.K., T.F.B., and G.M.L.; Supervision, W.N.H., T.F.B., and G.M.L.; Project Administration, G.M.L.; Funding Acquisition, L.L.L.X., A.Y.K., R.T.C., W.N.H, T.F.B., and G.M.L.;

phase. We observed dysregulation of metabolic processes during early persistent infection that were linked to changes in expression of genes related to nucleosomal regulation of transcription, T cell differentiation, and the inflammatory response and correlated with subject age, sex and the presence of HCV-specific $CD4^+$ T cell populations. These early changes in HCV-specific $CD8^+$ T cell transcription preceded the overt establishment of T cell exhaustion, making this signature a prime target in the search for the regulatory origins of T cell dysfunction in chronic viral infection.

eTOC blurb

Wolski et. al show that transcriptional dysregulation of metabolic, nucleosomal, and immune processes in virus-specific $CD8^+$ T cells during early persistent HCV infection is both under tight transcriptional control and associated with differences in predictors of disease outcome, like patient sex, age, and the presence of HCV-specific $CD4^+$ T cells.



Keywords

Hepatitis C Virus; CD8 T cells; CD4 T cell help; T cell dysfunction; metabolism; nucleosome; adaptive immunity; viral escape; network analysis; transcriptional regulation

Introduction

Over the last two decades we have gained important insights into the defining features of T cell responses in controlled versus chronic viral infection. These advances, most notably the differentiation between effective T cell memory and functionally exhausted T cells (Wherry and Kurachi, 2015) have largely been driven by work performed in animal models such as lymphocytic choriomeningitis virus (LCMV) infection in mice, and have had a major impact on our understanding of T cell immunology beyond the scope of viral infections (Pardoll, 2012). Key concepts from these studies have been translated to many human infections, including those with human immunodeficiency virus (HIV), hepatitis B virus (HBV), and HCV (Kahan et al., 2015). Nevertheless, our understanding of how different molecular

mechanisms initiate and govern effective or failing T cell responses remains incomplete, especially for humans, where genetic diversity and variability, differences in age and comorbidities, as well as environmental factors influence both the immune response and the outcome of infection significantly (Brodin and Davis, 2016).

HCV infection presents an exceptional opportunity to study regulation of immune responses, as it is unique among chronic infections with spontaneous and complete resolution of infection in about 20% of subjects within 6–9 months, even after exposure to a viral inoculum that can establish persistent infection in other subjects (Major et al., 2004). Data from humans and the chimpanzee model of infection have demonstrated that the presence of adaptive immune responses by both $CD4^+$ and $CD8^+$ T cells, while indispensable (Grakoui, 2003; Shoukry et al., 2003), is not sufficient for resolution of infection, as such responses are also detected early on in subjects developing chronic infection (Cox et al., 2005a; Schulze zur Wiesch et al., 2012). Neutralizing antibodies might also contribute to viral control, as their induction correlates with viral clearance (Osburn et al., 2014; Pestka et al., 2007). This strong host component to viral control, further evident in associations between HLA expression or sex with infection outcome (Bakr et al., 2006; Kuniholm et al., 2010), separates HCV from LCMV, a classic model of viral infection in inbred mouse strains, where different courses of viral replication are only observed when animals are infected with different viral strains. The analysis of the early anti-HCV T cell response in humans with different clinical outcomes can therefore add additional insights into the complexity of immune response regulation in genetically diverse hosts.

Here, we defined the molecular components of a successful $CD8^+$ T cell response to HCV infection and contrasted it with transcriptional events that govern $CD8^+$ T cells in persisting HCV infection. Taking advantage of two main features of HCV infection, its dichotomous outcome, with chronic viremia or complete clearance of the virus, and the occurrence of HCV viral sequence variation leading to viral escape from the $CD8^+$ T cell response, we were able to not only compare T cell responses that control viremia to those that do not, but also to differentiate between changes mediated by prolonged high level T cell receptor (TCR) stimulation and those driven by a chronic inflammatory environment.

Our results revealed both a preserved co-regulatory structure, referred to as core T cell identity hereafter, as well as distinct outcome-associated patterns of gene regulation indicating early dysregulation of HCV-specific $CD8^+$ T cell responses. Specifically, we identified a signature of major metabolic dysregulation at the transcriptional level during early chronic HCV infection, that was tightly linked to by expression changes of genes related to nucleosomal regulation and compounded by broader patterns of differential expression of genes involved in T cell differentiation and the inflammatory response.

By identifying programs that differentiate T cells from distinct outcomes of HCV infection during the earliest phase of the adaptive immune response, we gainec insights into which processes govern T cells on their way to T cell exhaustion or T cell memory and established a reference framework for more detailed mechanistic studies investigating the determinants of T cell mediated viral control and T cell failure.

Results

Multivariate analysis of adaptive immunity during early HCV infection

To identify regulatory processes controlling the CD8 response during early infection, we isolated 78 HCV-specific $CD8^+$ T cell populations (Figure 1A) from 43 subjects at multiple time points within the first 36 weeks of infection (Figure 1B and Table S1) and analyzed their transcriptional profiles by microarray. To detect potential effects of other branches of the adaptive immune system on $CD8^+$ T cell gene regulation, we assessed the presence of HCV-specific $CD4^+$ T cells targeting a panel of dominant $CD4^+$ epitopes by peptide-MHC (pMHC) multimers and neutralizing antibody titers by HCV pseudoparticle assay in a subset of these samples (Figure 1B and Table S1, STAR Methods). For analysis, we divided subjects and samples into three groups of distinct immunological states based on disease outcome and conservation of the viral epitope recognized by the TCR: 1) Chronic (persisting infection, preserved targeted epitope), 2) Escape (persisting infection, mutated targeted epitope with known diminished signal) and 3) Resolver (resolving infection, preserved targeted epitope), that we refer to as Chronic (C), Escape (E), and Resolver (R) hereafter.

Virus-specific $CD8^+$ T cells display distinct transcriptional profiles of resolution, inflammation and persistent antigen exposure

To identify outcome-associated differences in gene expression levels we analyzed $CD8^+$ T cell transcriptional profiles of Chronic and Resolver groups and found 258 genes that were significantly differentially expressed (|fold change|>1.5, FDR<0.05, Table S2) between Chronic and Resolver during either early (<18 weeks) or late (>18 weeks) phase of infection or between early and late phase of infection in either Chronic or Resolver (Figure 2A). Despite not being used for detection of differentially expressed genes (DEGs), we observed that Escape samples clearly helped to separate genes into distinct expression patterns for each group.

About half of all DEGs were expressed at similar levels between Chronic and Resolver within each phase of infection (Figures A2 and 2B), with higher and lower overall expression levels dominating the early and late phase, respectively. Among these genes were markers of T cell activation, such as *CD74* and *HLA-DRB1*, and T cell homing and proliferation, like *ELF4*, which exhibited much lower expression levels in the later phase compared to the early phase. This apparent dampening of the T cell response, irrespective of outcome, is consistent with the natural course of persistent HCV infection, where inflammation is much less pronounced in the later phase of infection compared to the early acute phase (Bowen and Walker, 2005).

Another group of genes with similar expression levels in Chronic and Resolver included cytokines *TNF* and *TGFB1* and their respective downstream signaling components (*NKFB1, JUNB, KLF10*, and *SMAD7*), as well as the chemokine receptors *CXCR4* and *CCR7*. These functionally related genes were all down-regulated in early Chronic and Resolver samples, but up-regulated in Escape samples throughout both early and late phase, indicating that diminished TCR stimulation evoked a distinct and sustained transcriptional profile in these cells.

While many genes shared expression levels between Chronic and Resolver both in the early and late phase, we also found significant differences between these two groups. Throughout the observation period, Chronic samples were characterized by higher expression levels for important regulators of T cell immune function, proliferation, and survival, such as *TBK1*, *SIRT1*, *BCOR*, and *BCL2L11*, while Resolver samples were marked by higher expression for regulators of T cell differentiation and memory, e.g. *TCF7* and its transcriptional target *LEF1*. Escape samples exhibited a mixed expression profile for most of these genes, either sharing expression levels with Resolver (*TCF7*) or with Chronic (*LEF1*) throughout the infection, or sharing features of Chronic early and Resolver later (*BCL2L11*) or vice versa (*SIRT1*), suggesting that persistent TCR stimulation and chronic inflammation each drive different aspects of the divergent transcriptional programs in $CD8^+$ T cells during the acute phase of HCV infection. These patterns of differential and shared gene expression between $CD8^+$ T cells in different immunological states highlight the complexity of biological processes that dynamically govern the adaptive T cell response to infection.

Modules of tightly co-regulated genes give rise to distinct patterns of transcriptional regulation in different infection outcomes

To address the complexity of the regulatory processes that govern $CD8^+$ T cell responses we applied a weighted gene co-expression network analysis (WGCNA) approach (Langfelder and Horvath, 2008), combining early and late samples for each group and constructed correlation networks for them (Table S4, STAR Methods).

By means of hierarchical clustering, we identified 9, 8, and 8 modules of tightly co-expressed genes (Figure 3A) for Chronic, Escape, and Resolver, respectively. When examining overall network topology (Figure 3B), we observed that genes not only correlated with other genes in the same module, but could also correlate with genes in other modules, indicating a second tier of interconnectedness.

Next, we assessed preservation of modules between outcomes by analyzing the overlap in gene content between individual modules from the different immunological states (Figures 3C and 3D), defining the significance of overlap between a module and another module as core score, and the overlap between a module in one group and the grey pseudo-module of uncorrelated genes in another group as its isolation score (STAR Methods). Most modules in Chronic and Resolver were dominated either by genes that shared co-regulation (Figure 3C, high core score, blue connections) or by genes that were exclusively co-regulated (high isolation score, red connections).

Many genes accounting for a high core score, like *JUNB*, *TNF*, and *KLF10* in modules c1-r5 (Table S4), also exhibited similar differential expression patterns for Chronic and Resolver, but not Escape samples (Figure 2A), supporting preserved regulation in early T cell responses that receive TCR signal. In contrast, other genes within preserved modules were part of the differential expression signature that distinguished Chronic from Resolver, such as *IFNGR1* and *SIRT1* in c2-r5, emphasizing that shared co-regulation did not imply similar directionality of expression.

Escape modules exhibited an overall weaker pattern of co-regulation on the gene and module correlation levels (Figures 3A and 3B), indicating less concerted regulation with abrogation of TCR stimulation. Nonetheless, we identified several modules with significant preservation between Escape and Chronic and/or Resolver, in particular c7-e7-r8, suggesting a core of processes that are regulated independent of TCR signaling.

To formalize patterns of interconnection between modules within and across groups, we combined intra-group module correlations (Figures 3B and Figure 3E, left) and inter-group module overlap (Figure 3C and Figure 3E, middle), identifying module community structures (Figure 3E, right, STAR Methods). We found three core-centric (c7-e7-r8, c2-r3-r4, and c1-e4-e5-r5) and one isolation-centric module community (c3-c5-c6-r1-r2), as well as one with more complex connectivity (c4-c8-c9-e2-r6-r7). While the latter two module communities highlight unique regulatory features of different infection outcomes, the former, core-centric module communities define elements of a core T cell identity, as they comprise genes that are co-regulated irrespective of immunological state. Among these, the community of modules c7-e7-r8 featured a particularly high core score between all three immunological states and contained many important immune genes, including *CXCL8*, *CD14*, *TLR2*, and *LILRB2*, which are parts of important co-stimulatory and inhibitory pathways that help modulate and maintain $CD8^+$ T cell responses (Komai-Koma et al., 2004; Shiroishi et al., 2003). The module community structure revealed by integrating correlations between modules and overlap between experimental groups illustrates how context-specific and universal transcriptional modules in combination shape distinct transcriptional states and highlights the importance of a broadly integrative analytic approach.

Differential expression of network modules is preserved across cohorts of acute HCV infection

To ensure the validity and reproducibility of our observations and to assert robustness of the identified network modules, we confirmed the results in an independent transcriptional data set. We analyzed HCV-specific $CD8^+$ T cells from 10 subjects with acute and chronic courses of infection sampled during the early acute phase of infection (≤ 21 weeks, Table S5). As this data set was too small for WGCNA analysis, we validated network modules by testing differential expression results from both the original and the validation cohort for enrichment of network modules (Figure 4 and Table S6).

We based this approach on the rationale that any enrichment of network modules or lack thereof in the original cohort, should be mirrored in the validation cohort, i.e. that any module that is strongly enriched among differentially expressed genes between Chronic and Resolver patients of the original cohort, should, if modules represent preserved co-regulatory structures, also be enriched in differentially expressed genes in the validation cohort. Similarly, modules for which no enrichment was observed in the original cohort should not be enriched in the validation cohort. We chose the competitive gene set test CAMERA (Wu and Smyth, 2012) (STAR Methods) for this approach, as it accounts for inter-gene correlation of tested gene sets and was therefore particularly suited to this analysis. Testing for enrichment of network module signatures in differential expression results from the original data (early and late) and the validation data (all early), we observed significant

enrichments for only a few modules, such as r2, r5, and r6, in both the validation (Figure 4B) and the early original data set (Figure 4A, left). Most modules exhibited no significant enrichment in the early original or in the validation cohort (Figure 4C). Late samples of the original data set did not share strong enrichment patterns with the validation set, which was limited to early samples. Overall, the large degree of concordance between the early original and validation set (Figure 4C), particularly of the matching strong enrichments of modules r2, r5, and r6, support that our findings are robust and pertain to acute and chronic HCV infection beyond the original cohort.

Outcome-associated clinical traits impact co-regulated modules representing different cellular processes and immune pathways

To identify network modules most likely to be relevant to disease status and progression, we tested modules for correlations with clinical and immunological traits, i.e. subject sex and age, HCV viral load, alanine transaminase levels (ALT), and time from infection, as well as presence of HCV-specific $CD4^+$ T cells targeting a set of dominant viral epitopes (Figure 5A, STAR Methods). To further link modules to biological function we tested them for enrichment of: 1) differential expression signatures of $CD8^+$ T cells in different disease models (HCV (from Figure 2), HIV (Gaiha et al., 2014), LCMV (Doering et al., 2012), and cancer (Singer et al., 2016)) or T cell memory states (Subramanian et al., 2005), 2) module gene signatures of antigen-specific $CD8^+$ T cells in chronic vs acute LCMV infection (Doering et al., 2012), and 3) functional gene sets related to immune pathways (Langfelder and Horvath, 2008), gene ontology, and KEGG pathway databases (Ashburner et al., 2000; Kanehisa, 2000) (Figures 5B, 5C, and 5D, Tables S3 and S7, STAR Methods).

To ease interpretation of functional enrichments in the context of module communities (Figure 3E, right) and immunological states, we grouped enriched terms into clusters of processes related to immune and T cell regulation (Figure 5D, grey panel, red dots), nucleosome associated regulation of transcription (grey panel, blue dots) and metabolism (grey panel, yellow dots), based on functional annotation and co-occurrence in modules across different groups.

We found that about half of all modules correlated significantly with at least one clinical trait (Figure 5A), with ALT, sex, and age showing the most correlations, suggesting a strong impact of liver inflammation and of disease-relevant demographic features. The only two modules, c3 and c5 (Figure 5A), that were (negatively) correlated with time after infection (decreasing expression over time), also exhibited the highest isolation scores (Figure 3C and 3D) and were enriched for metabolic pathways (Figure 5D), indicating unique dynamics of metabolic regulation in $CD8^+$ T cells during early chronic infection. Higher expression levels of genes in c3 and c5 also correlated with female sex and younger age (negative correlation), both of which are predictors of favorable disease outcome, while another module that was enriched for metabolic pathways and apoptosis, c1, exhibited an inverse correlation (Figure 5A), with higher expression levels correlating with higher age and male sex.

c5 was also enriched for a signature associated with long-term viral control in HIV (Gaiha et al., 2014) and correlated with CD4 help status as measured by our targeted CD4 assays.

While not comprehensively assessing all targeted CD4 epitopes in a patient, this assay captures the decline of HCV-specific $CD4^+$ T cell populations in chronic infection (Schulze zur Wiesch et al., 2012). c8 (Figure 5A), displayed an even stronger association with the measured CD4 response and was enriched for multiple processes related to regulation of immune response, cell adhesion and homing, and nucleosomal regulation of transcription (Figure 5D), identifying specific $CD8^+$ T cell programs that rely on sustained T cell help. We did not find correlations with neutralizing antibodies (nAb) as we detected barely any nAb responses in this cohort during the observation period.

While we found little enrichment of differential expression signatures from acute versus chronic LCMV infection (Doering et al., 2012) or cancer (Singer et al., 2016), we observed significant overlap between co-regulation modules from acute vs. chronic LCMV infection (Figure 5C) and the human HCV modules, particularly between immune related modules signatures and the core-centric module community (c7-e7-r8). This demonstrates preservation of central processes across different species, highlights the advantage of employing a co-regulation based network analysis approach and provides further evidence of the validity of our results.

While significant parts of these core modules were preserved both across different infection models and across all immunological states, they also exhibited distinct functional enrichments for immune pathways (Figure 5D), suggesting that co-regulation of core genes with additional genes that are not shared across groups could result in distinct functional output. Such functional differences, coupled with group-specific connectivity patterns that span preserved and non-preserved modules, hold important clues as to how T cell-mediated viral control is orchestrated at the transcriptional level.

Functional immune networks identify patterns of outcome-exclusive transcriptional regulation for T cell metabolic processes in persisting infection

To delineate the commonalities and differences in T cell regulation between Chronic, Escape and Resolver groups, we created functional cluster networks (STAR Methods) from the original gene correlation networks of each group, using the genes that were annotated in the immune, nucleosome, and metabolism clusters (Figure 5D, grey panel, colored dots) and examined them for patterns of shared and exclusive regulation at the gene level. We observed shared components between all three or any combination of two groups in each cluster network, but also group-exclusive components containing genes that were co-regulated in only one immunological state. Components of genes with shared co-regulation between all three groups were most prominent in the immune and nucleosome networks (Figure S2, Table S8). These two networks were further characterized by considerable overlap between Chronic and Resolver samples, indicating that diminished TCR signaling in Escape samples disrupts co-regulation of gene networks otherwise conserved in $CD8^+$ T cells irrespective of viral control. This Chronic-Resolver exclusive component contained more than half of all histone genes annotated in the nucleosome cluster network (Figure S2D-F, Tables S4 and S8).

The metabolism network featured large components that were absent in resolving infection, co-regulated either in Chronic alone (Figures 6A and S1, yellow component) or in Chronic

and Escape samples (pink component) and contained many genes from c3 and c5, modules for which we had identified strong associations with age, sex, presence of HCV-specific $CD4^+$ T cells, and time from infection (Figure 5A). To further pinpoint the functional processes impacted by this dysregulation, we performed a second KEGG pathway enrichment analysis (STAR Methods), revealing a strong enrichment of genes involved in oxidative phosphorylation in the C-exclusive component (Figure 6B, Table S9).

In a recent publication, Bengsch et al. report oxidative phosphorylation (OXPHOS) as one of the key processes involved in metabolic dysregulation of exhausted T cells in persisting clone 13 LCMV infection (Bengsch et al., 2016). When comparing genes driving the enrichment of OXPHOS in the C-exclusive component of the human cluster network to the leading edge signature driving the signature of metabolic dysregulation in LCMV, we found not only that they overlapped highly (12/17 genes, Figure 6C), but also that this overlap was dominated by genes in modules c3 and c5 (and c6), with expression kinetics similar to those observed in chronic LCMV infection, where an early surge in expression was followed by rapid downregulation (Figure 6D). This similarity between HCV and LCMV-specific $CD8^+$ T cells, however, seemed to be limited to enrichment and expression kinetics of OXPHOS-related genes, as enrichment of other metabolic processes found in HCV (Figure 6B), such as biosynthesis of amino acids in the R-exclusive component and cysteine and methionine metabolism in the chronic-exclusive component, have not been described for LCMV. Conversely, genes driving enrichments of fatty acid metabolism and glycolysis in acute LCMV (Figures 6A, 6B and S1, Table S9) did not rank highly in HCV, demonstrating pointed differences between these two infection models.

Together, the results strongly suggest that early dysregulation of metabolic processes, particularly of pathways related to cellular respiration, play a central role on the way towards an exhausted T cell response that is unable to control viral replication. At the same time, our data indicate that specific features of metabolic dysregulation might be highly dependent on infection-and cellular context.

A small group of key transcriptional regulators drive early dysregulation of metabolism in conjunction with nucleosomal and immune processes in persisting HCV infection

As the dysregulation of $CD8^+$ T cell metabolic gene expression in chronic subjects was both associated with time from infection and the presence of HCV-specific $CD4^+$ T cells, and was highly consistent across multiple functionally related modules (Figure 6D), we hypothesized that it might be the result of active transcriptional regulation and that the responsible regulators were likely to share expression patterns with their targets, and should therefore be part of the same module.

To test this hypothesis, we assembled a network of ~2.9 million unique regulatory interactions between 1319 known transcriptional regulators (TRs) and their targets, sourced from publicly available databases, network repositories and published CD8- and lymphocyte-specific regulatory networks (STAR Methods), and intersected it with the correlation networks we constructed (Figure 3) to determine the number of interactions between TRs and targets within their respective modules (TR-target coverage, Table S10, STAR Methods). We then related this information to the intramodular connectivity of each

TR to approximate how likely it was to be involved in active regulation of its respective module. We observed a positive trend of increasing intramodular connectivity (kWithin) with increasing target coverage (Figures 7A, S3A, and S3B), indicating that genes with high kWithin, so-called hub genes, are likely to be of high functional importance.

To identify the TRs that were most likely to be involved in the dysregulation of metabolic processes (Figure 6), we constructed a regulatory network based on the top scoring TRs in c3, c5 and c6, (*STAT5A*, *KDM5B*, *RFX5*, *NELFE*, *CBX3*, *STAT1*, and *TP53*, Figure 7A, Table S10). In addition, and to account for the possibility that the observed changes might not be caused by active regulation, but rather by the lack thereof, we also included the TRs *CREB1*, *IRF3* and *FLI1* from r2, a module from the same community as c3, c5, and c6 (Figure 3E), in which *STAT1* and *TP53* were also co-regulated (Figure S3A).

The resulting regulatory network (Figure 7B), not only mapped many of the metabolic dysregulation signature genes (Figures 6A and 6C, Table S8) to known regulatory interactions for the identified transcriptional regulators, but also contained regulatory interactions for several histone subunits, integrins, class II MHC molecules, chemokines and chemokine receptors (Figure 7B, Table S8) and was only partly preserved in Escape and Resolver groups. Specifically, we identified regulatory interactions between *STAT5A*, *KDM5B*, and *RFX5* and genes of the metabolism cluster (Table S8), that were regulated exclusively in Chronic (Figure S4C, yellow nodes) or Chronic and Escape groups (pink nodes), and which included the hub genes for modules c3, c5, and c6 (*TYMS*, *CMAS*, and *B4GALT3*) as well as important genes annotated to the nucleosome (e.g. multiple histone subunits and integrins, Figure 7B) and immune clusters (e.g. chemokine receptors *CCR5* and *CX3CR1*, and chemokines *CCL4*, and *CCL5*).

The genes in this Chronic regulatory network were, for the most part, marked by the same distinctly temporal expression kinetics we already described for the metabolic dysregulation signature, with strongly elevated levels of expression at the beginning of infection being followed by rapid downregulation (Figures 7C and S3C). What is more, these expression changes occurred during the first 12-18 weeks of infection, before outcome can be predicted clinically and before $CD8^+$ T cells can be differentiated as either exhausted or T cell memory by conventional T cell assays. This makes the identified regulatory interactions prime targets for closer investigation of their role in initiating divergent differentiation programs of $CD8^+$ T cells, leading to either resolved or persistent viral infection.

Discussion

Here we employed a stepwise, integrative systems immunology approach to identify key transcriptional commonalities and differences between virus-specific $CD8^+$ T cells from subjects with persistent and spontaneously resolving hepatitis C virus infection during the critical acute phase. While the inter-individual heterogeneity of human data sets presents challenges for their analysis and interpretation, our work underlines the importance of studying human populations that fully reflect the differences in genetic backgrounds, age, sex, and environmental influences, by adding important insights and offering a unique and

partly contrasting perspective to knowledge gained primarily from disease models in inbred laboratory animals.

To account for human diversity and create a robust data set we generated a large transcriptional data set of human primary virus-specific $CD8^+$ T cells. By comparing cells from three, rather than two distinct immunological states, we simultaneously increased the precision and accuracy of our analysis, since links between multiple groups are less likely occur by chance alone and because it enabled us to distinguish between antigen- and inflammation-mediated effects on adaptive immunity. We tested the validity of our results beyond the original cohort of HCV infection through a constrained validation analysis in a separate gene expression data set from other patients with acute HCV infection that, while limited in scope, confirmed reproducibility of our findings.

The main analysis approaches we used (supervised differential expression analysis and WGCNA), complement each other in that they interrogate both gene expression levels and patterns of co-regulation. Both analyses revealed a host of genes that undergo similar regulation in both Chronic and Resolver samples, but also identified significant differences between the groups that manifested during the earliest stages of infection, indicating that differential regulatory programs are initiated in these T cells well before differences in the quality of the T cell response and the level of viral control become more apparent. Among these differences was a highly distinct co-expression signature of genes related to metabolic processes that were regulated exclusively in Chronic samples, with high initial expression levels followed by rapid downregulation in the very early phase of infection. This rapid onset and marked temporality of gene expression changes make this signature a prime target in the search for the regulatory origins of T cell dysfunction in chronic viral infection, especially since this differential regulation manifests before classical $CD8^+$ T cell studies are able to reveal clear differences in T cell frequency, function or phenotype.

Parts of this metabolic dysregulation signature are shared by a signature recently described for $CD8^+$ T cells in persistent LCMV infection with clone 13 (Bengsch et al., 2016), further highlighting its potential significance for T cell exhaustion and indicating that proper regulation of metabolic processes is likely to present a truly universal feature of a proficient immune response, even across different species. Moreover, one of the driving modules behind this signature was also strongly enriched for genes identified in HIV-specific $CD8^+$ T cells from patients that maintain long-term control of viral replication (Gaiha et al., 2014), lending further support to this notion.

Apart from overlap in metabolic dysregulation signatures, we find other modules that share significant overlap with LCMV infection, as a sign of significant cross-species preservation of T cell regulation. At the same time, we also observed significant differences between HCV and LCMV not just in key genes and pathways comprised in the metabolism signature, but also at the module level of the larger transcriptional networks, with several modules that are entirely exclusive to HCV, suggesting that regulation of metabolism and other T cell functions is, beyond certain immutable properties, also context-specific and influenced by unique characteristics of each infection, on the side of the both the host and the virus.

Our analysis revealed strong correlations between expression changes in the modules containing this metabolic signature and subject age, sex, and the presence of HCV-specific $CD4^+$ T cells, three traits that present the strongest predictors of HCV infection outcome. Identifying these correlations is one of the key strengths of this human study, as diversity in age and sex is rarely reflected in animal models. Here we can, module by module, select the strongest drivers of these correlations as specific targets to investigate the molecular mechanisms by which human host factors impact antiviral immunity and viral control (Giefing Kröll et al., 2015).

Delving deeper into the regulatory interactions that might drive this dysregulation by cross-referencing identified modules with known regulatory interactions, we find that most dysregulation signature genes are regulated by a small group of transcription factors that are not only highly connected within their respective modules, but also annotated for interactions with many genes related to immune function and nucleosomal regulation of transcription in closely connected modules, demonstrating that metabolic dysregulation lies at the heart of more extensive transcriptional changes. This process is highly dynamic, with a very early burst in expression of most genes, followed by rapid downregulation, indicating that this highly activated initial state of the network is not sustained for long.

The transcriptional regulators driving these early changes in T cell regulation include *CREB1*, *FLI1*, *KDM5B*, *RFX5*, *STAT1*, *STAT5A*, and *TP53*, most of which are known regulators of immune function, with *STAT1* and *STAT5A* typically being associated with effective IFN responses and T cell memory differentiation (O'Shea and Plenge, 2012), whose involvement in metabolic regulation has not been elucidated so far. Conversely, the role of *TP53* in metabolism is well documented (Berkers et al., 2013), but its immune-regulatory role has only recently gained increasing attention (Muñoz-Fontela et al., 2016). As such, this regulatory network represents a distinctive feature of the early $CD8^+$ T cell response in persistent HCV infection in humans, highlighting previously unknown connections for a small group of key transcription factors. Outcome-dependent differences in the connectivity of transcriptional regulators have also been described in the chronic phase of LCMV infection, (Doering et al., 2012), albeit for different transcription factors and without dissection of TCR- and inflammation-mediated effects.

Of note in this context, we find that Escape samples only partially shared the signature of metabolic and immune dysregulation we observe in Chronic samples, indicating that this process likely requires strong and sustained TCR signaling. Indeed, there is convincing data supporting that escaped CD8 responses in chronic HCV infection recover to some extent, as these cells adopt a memory phenotype, upregulate CD127 and downregulate PD-1 (Kasprowicz et al., 2010), and respond better to a therapeutic vaccine (Swadling et al., 2014). On the other hand, Escape samples lacked distinct co-regulation of most histones, in contrast to Chronic and Resolver, and exclusively expressed *TNF* and *TGFB1* and their downstream targets at constantly high levels, suggesting that diminished TCR signal in a chronic inflammatory environment does not match spontaneous resolution, where both TCR signal and inflammation disappear together. Whether this distinct transcriptional profile is a result mainly of prolonged exposure to type I interferons, or other inflammatory pathways, and whether it is associated with less overt functional deficits, warrants further investigation.

Together, our data from acute HCV infection identify a set of transcription factors that facilitate a synchronized, early but transient transcriptional burst related to oxidative phosphorylation that is limited to T cells from patients who go on to develop persisting infection. These early changes in HCV-specific $CD8^+$ T cell transcription precede the overt establishment of T cell exhaustion and extend beyond T cell metabolism to regulation of immune function and nucleosomal regulation of transcription. Furthermore, they are associated with strong and prolonged TCR signal and factors of human heterogeneity, like age and sex of the patient as well as with the disappearance of HCV-specific $CD4^+$ T cells. Our data set therefore provides potential links between metabolic and epigenetic regulatory aspects of $CD8^+$ T cell immunity and key clinical determinants of infection outcome, which is highly relevant for pressing question about the impact of sex and age on the adaptive immune response (Giefing Kröll et al., 2015).

In addition to these specific findings, this extensive human data set of virus-specific $CD8^+$ T cells from different infection outcomes and with varying levels of viral control constitutes a valuable resource for the scientific community, that will enable and inspire further inquiry into the regulatory mechanisms that govern adaptive immunity in response to viral infection.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

All requests for reagents and resources can be directed at Georg M. Lauer (glauer@mgh.harvard.edu)

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Human subjects and samples—We collected blood samples from 41 patients at multiple time points during the early and late acute phase (0 to 36 weeks) of infection with hepatitis C virus and extracted PBMC and plasma by Ficoll-Paque (GE Healthcare Life Sciences) density gradient centrifugation. To reduce the impact of sources of variability or bias, we excluded subjects that had tested positive for co-infections such as HIV or HBV as well as subjects that had previously been treated for HCV infection or resolved a previous infection on their own. Furthermore, we excluded subjects for whom the natural course of infection could not be determined accurately, due to lack of follow-up or treatment within the acute phase of infection. All included patients were followed for at least 1 year post the estimated infection date in order to obtain reliable data on the outcome of infection. Informed consent was given by all subjects under two protocols approved by the Partners Healthcare Human Research Committee and the IRB of the Oswaldo Cruz Institute in Rio de Janeiro, Brazil.

To capture both cross-sectional and longitudinal changes, we sampled subjects during the entire phase of acute infection, capturing 1–3 time points per patient, for a total of 78 samples, 41 from chronic and 37 from resolving patients. For analysis purposes, we distinguished between $CD8^+$ T cells from chronic infection targeting conserved vs escaped viral epitopes as determined by viral sequencing (described below), resulting in three groups of samples that we refer to as Chronic (C. chronic with conserved epitope, $n_{chronic} = 27$),

Escape (E, chronic with escaped epitope, n_{escape} = 14), and Resolver (R, resolver with conserved epitope, $n_{resolver}$ = 37) (Figure 1B, Table S1).

For the validation cohort (Table S5), we sampled 7 chronic (C) and 3 resolving (R) patients at the earliest time point each (≤ 21 weeks) using the same criteria as outlined above.

METHOD DETAILS

Flow Cytometry and Cell Sorting—We detected, quantified, and sorted HCV-specific $CD8^+$ T cells from human PBMC by multi-color flow cytometry and fluorescence assisted cell sorting using class I MHC multimers (Figure 1A). Subject-specific response profiles were established by screening using IFN-gamma secretion assays and available HLA-matched peptide-MHC multimers. Positive multimer responses were used for sorting of specific cells. HCV-specific $CD8^+$ T cells were defined as cells positive for CD8/pMHC-multimer (inclusion of specific $CD8^+$ T cells) and negative for CD4/CD14/CD19/CD16/CD56 (exclusion of $CD4^+$ T cells, natural killer T cells, natural killer cells, monocytes, and B cells) as well as negative for Live/Dead viability dye and Annexin V (exclusion of dead and apoptotic cells). All steps were performed at 4°C, unless specified otherwise by the manufacturer. In brief, cells were thawed to a temperature of 4°C washed in FBS (1x) and R10 (1x), and resuspended in phosphate buffered saline (PBS) for LiveDead viability staining (Thermo Fisher Scientific) at room temperature for 30min. Cells were washed in FACS buffer, stained for TCR using peptide-MHC multimers and washed again (FACS buffer, 1x) prior to surface antigen staining (KEY RESOURCES TABLE). Following surface antigen staining, cells were washed (FACS buffer, 2x) and stained for Annexin V (Thermo Fisher Scientific) per manufacturer instructions and cells were sorted in the Ragon Institute Imaging Core Facility using a BD FACSAria II cell sorter (BD Biosciences) at 70 psi with a 70um nozzle in a Bioprotect IV biosafety cabinet (The Baker Company) into PBS in eppendorf tubes maintained at 4 °C in the collection tube holder, subsequently pelleted by centrifugation, and resuspended in cold TRIzol (Invitrogen). Samples were then stored at −80°C until further processing. Details on used antibodies and reagents are listed in the KEY RESOURCES TABLE.

Data on Other Adaptive Immune Effectors

Magnitude of $CD4^+$ T cell response: In a subset of the samples used for CD8 gene expression analysis we measured HCV-specific $CD4^+$ T cell responses targeting a set of dominant HCV CD4 epitopes (Schulze zur Wiesch et al., 2012; 2005) by multi-color flow cytometry using pooled HLA-matched pMHC class II multimers, with the same experimental protocol described for $CD8^+$ T cells above. We used this data as a basic indicator for the presence of HCV-specific $CD4^+$ T cells that could provide help to $CD8^+$ T cells (n = 37/78, Figure 1, Table S1). Different levels of HCV-specific $CD4^+$ T cell presence (referred to as CD4 Status in Figure 1) were categorized as follows-Samples with no detectable $CD4^+$ T cells (≤ 0.001% of total T cells (CD3+)) at the time point in question or within 2 weeks of it and which had detectable $CD4^+$ T cells at other time points (including time points outside of this study) were classified as “none”, samples with detectable $CD4^+$ T cells (>0.001% and < 0.01% of CD3+) at the time point in question or within 2 weeks of it were classified as “low”, and samples with detectable $CD4^+$ T cells (e 0.01% of CD3+) at

the time point in question or within 2 weeks of it were classified as “high”. Furthermore, samples from subjects with “high” CD4 status at a later time point (> 2 weeks) were classified as “high” and samples from subjects with CD4 status “none” at an earlier time point (> 2 weeks) were also classified as “none”.

Neutralizing antibodies: HCV-specific neutralizing antibodies were quantified in serum of a subset of patients (n=32/78, Figure 1B, Table S1) using a neutralization assay based on the infection of Huh7 cells with retroviral pseudoparticles (HCVpp) that express genotype-specific HCV envelope glycoproteins E1 and E2. Neutralizing antibodies were quantified using endpoint titration (Fafi-Kremer et al., 2010; Pestka et al., 2007). For analysis purposes, titers were classified as non-neutralizing (<1:100) and neutralizing (1:100). Neutralizing antibodies could not be used for trait correlations with module genes (Figure 5A), since only 4 subjects, distributed over all three groups, displayed significant levels of neutralization.

Viral Genome Sequencing and Testing of Variant Epitopes: We sequenced the Hepatitis C viral genome on a per patient and time point basis for samples where a sufficiently high viral load was detectable. We surveyed the relevant viral epitopes used in the flow cytometry/cell-sorting step for sequence conservation. Samples with conserved epitopes were grouped into Chronic and Resolver samples based on infection outcome. If the circulating virus contained epitope variants, we generated short-term T cell lines by stimulating the patient’s PBMC with the prototype peptide under addition of IL-2 and subsequently tested recognition of both prototype and variant epitope by intracellular cytokine staining. We did not test again the mutation in the HLA B*0801 epitope (HSKKKCDEL to HSKKKCDEF), which leads to altered peptide processing and presentation (Timm et al., 2004). The effect of epitope sequence variations were categorized into three types of 1) complete absence of T cell stimulation 2) significantly diminished T cell stimulation (more than one log higher peptide concentration required for same level of T cell stimulation with variant peptide) and 3) minimal effect on T cell stimulation (Cox et al., 2005b; Kasprowicz et al., 2010). For our analysis, we considered samples of type 1 and 2 to be Escape samples and samples of type 3 to be Chronic samples.

RNA Processing: We extracted total RNA from samples with >500 sorted cells using the RNAdvance Tissue Isolation Kit (Agencourt), reverse transcribed RNA into cDNA and amplified it using the WT-Ovation Pico Amplification System (Nugen). Following that, we tested for cDNA yield by Nanodrop (Thermo Fisher Scientific) and purity by 2100 Bioanalyzer (Agilent Technologies). We then hybridized concentration-adjusted amounts of cDNA to GeneChip HG U133A 2.0 (Affymetrix) microarrays to measure gene expression. Details on used reagents are listed in the KEY RESOURCES TABLE.

QUANTIFICATION AND STATISTICAL ANALYSIS

Microarray Data Preprocessing—Raw CEL files, containing probe level expression data, were imported into R using *affy* (Gautier et al., 2004), background-corrected, normalized, summarized to probe set level, and finally log transformed by robust multi-array average (RMA) using a custom HGU133A2_Hs_ENTREZG v20.0.0 chip definition file (CDF) file obtained from http://brainarray.mbni.med.umich.edu (Dai et al., 2005). Prior to

correction for known technical batch effects (sample source and array batch), we ensured that variables of interest (disease outcome, viral escape) had a reasonably balanced distribution among batches and estimated the number of latent variables by *sva* (Leek and Storey, 2007). Probe set level data were batch effect corrected using the risk-conscious PCA-based adjustment method *Harman* (Oytam et al., 2016), while controlling for variables of interest. We reran *sva* to confirm successful batch correction and confirm that no significant latent variables remained in the data.

To remove lowly expressed genes and reduce the impact of noise on analysis of the data, probe sets were filtered based on mean expression and variance, keeping only those with a mean log2 expression value >4.25 and a variance >0.3 across all samples. Finally, probe sets were collapsed to the gene level using *genefilter* (Kauffmann et al., 2009), removing duplicate entries and manufacturer quality control probe sets, resulting in a set of 5246 genes.

Data were split into three sample sets by disease outcome and status of viral escape. Obvious outlier samples were identified by hierarchical clustering and excluded from further analysis.

In order to improve speed and sensitivity of our analysis pipeline we calculated the nearest neighbor connectivity for each of the 5246 genes (per group) using *WGCNA* (Langfelder and Horvath, 2008) by set (nneighbors=50) and kept only the union of the top 45% most connected genes from each group, resulting in a final set of 3850 genes used for all downstream analysis.

For the validation experiment (Figure 4, Tables S5 and S6), we processed data in the same manner as outlined above using a custom CDF file (HTHGU133A_Hs_ENTREZG v20.0.0) and used the same subset of genes that were used in the original data set for subsequent analysis.

Differential Expression Analysis—To improve group homology for detection of differentially expressed genes, sample groups were split into early (≤18 weeks post infection) and late (>18 and <36 weeks post infection) subgroups (Figure 1B). We assessed differential gene expression between early (differential gene expression and late (>18 and <36 weeks post infection) Chronic and Resolver samples (Figure 2A, Table S2), using an empirical Bayes moderated ANOVA after fitting an observation level mean-variance model to data subsets to adjust for heteroscedasticity as implemented in *limma* (Ritchie et al., 2015), returning an F test statistic and p value analogous to those of regular ANOVA. P-values were corrected for multiple hypothesis testing by controlling the false discovery rate (FDR). Genes were called differentially expressed with an FDR <0.05 and an absolute fold change > 1.5, corresponding to a 50% difference in expression between Chronic and Resolver groups during early or late acute phase of infection or between early and late phase of acute infection in Chronic or Resolver groups, resulting in a set 4 contrasts between groups and timeframes across which we detected 258 differentially expressed genes (Figure 2, Table S2). We expressly based detection of differentially expressed genes on Chronic and Resolver samples only, as the differences between these samples were the focus of this

study, but also plotted the mean scaled expression scores for the early and late Escape samples (Figure 2A), to identify whether Escape samples shared part of their transcriptional signature with Chronic or Resolver groups.

Based on shared expression profiles of differentially expressed genes between Chronic, Escape and Resolver groups (Figure 2A) we defined a set of 8 gene expression signatures for both the early and late acute phase of infection (Figure 2B, Table S3). Each set was defined as 7 signatures of upregulated genes (Resolver, Escape/Resolver, Chronic/Resolver, Escape, Chronic/Escape, Chronic, and All) and one signature of downregulated genes (All).

The same parameters as described above were used for differential expression analysis in the validation cohort (Figure 4, Tables S5 and S6).

Network Construction and Module Detection—We used *WGCNA* (Langfelder and Horvath, 2008) to construct a weighted gene co-expression network for each sample set. *WGCNA* assumes that co-expression of genes, measured by their correlation across samples within an experimental group also implies their co-regulation. In this semi-supervised way, *WGCNA* allows for detection of modules of highly correlated and often functionally related genes that can be tested for their degree of preservation and difference in connectivity between individual networks. By applying weights to correlations by raising them to a power beta, giving preference to strong correlations over weak ones, *WGCNA* achieves both a reduction of noise in the data and a scale-free network topology. This scale-freeness is considered to reflect an intrinsic and desirable property of most biological networks, as it is strongly correlated with a network's robustness to random perturbation (e.g. random mutation) (Barabási and Oltvai, 2004). Accordingly, a signed network adjacency matrix was calculated from the biweight midcorrelation between all gene pairs of each group (Chronic, Escape, and Resolver) using soft thresholding powers (see *WGCNA* parameter table). Network adjacencies were then transformed into topological overlap dissimilarity measures and subjected to average distance hierarchical clustering. The resulting clustering dendrograms (Figure 3A), that visualize distances between genes in terms of their co-expression profiles, with lower hanging branches indicating clusters of highly correlated genes, were then used to isolate 9 Chronic, 8 Escape, and 8 Resolver gene modules using a dynamic tree cutting algorithm (Langfelder and Horvath, 2008).

Highly similar modules were merged based on a minimum dissimilarity threshold. Since WGCNA allows a host of different parameter choices, and to ease reproduction of our results, we have itemized details of parameter choices below.

Parameter choices for construction of weighted gene co-expression networks

Category	Function	Parameter	Value
Network Construction	adjacency	corFnc	"bicor"
		corOptions	"use = 'p', maxPOutliers = 0.05"
		type	"signed hybrid"
		power	4/7/3 (Chronic /Escape/Resolver)

Category	Function	Parameter	Value
	TOMsimilarity	TOMType	“signed”
		TOMDenom	“min”
	TOM scaling[a] (scale TOM matrices for comparability)	percentile	0.95
		sample size	1/(1-percentile) * 1000
		reference set	“Resolver”
	Gene clustering[a] (cluster genes based on scaled TOM)	dist	1-scaledTOM
		Clustering method	“average”
Module Detection	cutreeDynamic	method	“hybrid”
		minClusterSize	60
		maxCoreScatter	0.75
		minGap	(1-maxCoreScatter)*3/4
		pamStage	FALSE
	mergeCloseModules	useAbs	TRUE
		cutHeight	0.25

Identification of Module Communities—We calculated pairwise biweight midcorrelations between first principal components of individual modules (module eigengenes) within groups and determined correlation significance using Fisher’s exact test. We also determined overlap in gene content between modules between different groups and calculated the significance of this overlap using a one-tailed Fisher’s exact test. Both correlation and overlap p-values were corrected for multiple comparisons by control of FDR. Furthermore, for Figures 3C and 3D we defined the significance of each overlap as its core or isolation score, depending on whether the overlap was between two modules in two groups (core score) or a module and unconnected genes (grey pseudo module) (isolation score). Next, we created correlation and overlap module networks (Figure 3E, left and middle), where module correlations or overlaps with FDR <0.01 were defined as connections between individual modules. We then used the strength, defined as -log10(FDR), and number of these connections to identify communities of related modules (Figure 3E) in a merged network of correlations and overlaps (Figure 3E, right). Modules communities were detected using a multi-level algorithm (Blondel et al., 2008) as implemented in *igraph* (Csardi and Nepusz, 2006).

Trait Correlations—To correlate identified modules with clinical traits, we calculated the first principal component (eigengene) of each module and correlated it with several clinical and immunological traits (Figure 5A), including HCV viral load (log2 transformed), alanine aminotransferase (log2 transformed), time from infection (weeks), subject sex and age, as well as status of CD4 help (measured as number of detectable HCV-specific $CD4^+$ T cells, see above) or neutralizing antibody status (measured as titer threshold, see above). Biweight midcorrelation was used for all continuous variables (age, viral load, ALT, time from infection) and Pearson correlation was used for all ordinal variables (sex, CD4 help). Ordinal variables were coded as follows-sex, female = 1, male = 2; CD4 help, none = 0, low = 1, high = 2.

Gene Set Enrichment

Analysis Assembly of external gene signatures: We assembled a set 6 of HIV-specific $CD8^+$ T cell signatures from based on genes that were differentially expressed (FDR ≤ 0.05, |log2(fold change)| ≥ 1.5) in HIV-specific $CD8^+$ T cells from elite or viremic controllers and chronic progressors (Gaiha et al., 2014) based on comparisons between peptide-stimulated controllers and progressors and between stimulated and unstimulated cells in controller and progressor groups (Table S3).

For mouse signatures of T cell dysfunction vs activation, we used a set of 4 signatures (Table S3) recently published by Singer et al., that define genes related to T cell dysfunction, T cell activation, T cell activation/dysfunction and naïve/memory like T cells based on their correlation with the first two principal components that separate tumor-infiltrating lymphocytes according to their functionality and activation status based on expression of inhibitory receptors Tim-3 and PD-1 and deficiency of metallothioneins MT1 and MT2 (Singer et al., 2016).

We assembled other mouse signatures of naïve, memory and exhausted T cells from the literature as detailed by Singer et al. (Singer et al., 2016), to ensure direct comparability with the novel signatures of T cell dysfunction and activation they describe. In brief, we compiled consensus signatures of naïve and memory T cells from multiple MSigDB v5.0 gene signatures (Subramanian et al., 2005) (naïve-28 genes upregulated in > 10/26, memory-23 genes upregulated in > 6 out of 13 signatures, Table S3). Furthermore, we compiled a T cell exhaustion signature from genes that were differentially expressed (FDR < 0.05 and |fold change| ≥ 2) in LCMV-specific $CD8^+$ T cells between chronic and acute disease on day 15 and 30 of LCMV infection (Doering et al., 2012) (73 genes, Table S3). In addition to the late exhaustion signature, we also assembled an early exhaustion signature (31 genes) based on differential expression (FDR < 0.05 and |fold change| ≥ 2) at day 8 of LCMV infection as well as resolution signatures at day 8 (4 genes, Table S3) and day 15/30 (7 genes, Table S3).

For external module signatures, we used *WGCNA* based gene modules from $CD8^+$ T cells in chronic and acute LCMV infection (Doering et al., 2012).

To account for mismatches in gene annotation between organisms, we transformed all mouse signature into human signatures by mapping from MGI symbols to ENSEMBL identifiers and from ENSEMBL identifiers back to HGNC symbols using biomaRt (Durinck et al., 2009).

HCV module signature enrichments for module validation: For validation of detected network modules we used CAMERA (Wu and Smyth, 2012) to test both the original data set (Table S1) and a separate data set of microarray data from HCV-specific $CD8^+$ T cells from patients with persisting and resolving infection within 20 weeks of infection ("validation" set, Table S5) for enrichment of the modules we derived from the original data set. Similar to GSEA (Subramanian et al., 2005), CAMERA allows ranked list of genes as input (here, we rank genes by differential expression between Chronic and Resolver within the different data sets). The idea behind CAMERA is based on the fact that gene set tests that use gene-label permutation are often affected by inflated false discovery rates, and that thee

alternative, sample permutation, is not applicable for experiments with small numbers of replicates. Therefore, CAMERA instead incorporates an estimated variance inflation factor that is associated with inter-gene correlation into the test procedure to correctly control the type I error rate. Since this approach will specifically penalize sets of co-regulated genes which are often the most biologically relevant, CAMERA also allows for a preset inter-gene correlation to regain biological relevance at a slight cost of error rate control. Since we were interested in sets of genes that were defined by their high degree of correlation, we set the inter gene correlation to be at a low, fixed level of 0.01, deliberately allowing our module gene sets with high inter-gene correlation to achieve high enrichment scores, which were then corrected for multiple hypothesis testing within each data set (early, late, and validation) by controlling the FDR and setting a strict significance threshold of FDR <0.001 to offset the slight loss of type I error control resulting from a preset inter-gene correlation value. Figure 4 shows examples of enrichment plots for modules r2, r5, r6, and c5 for both the early and late group of the original data set (Figure 4A) and the validation data set (Figure 4B). Full enrichment results are listed in Table S6.

Differential Expression and Module Signature Enrichments: We performed a variety of gene set enrichment analyses per module per experimental group (Chronic, Escape, Resolver). For the differential expression signature enrichment analysis (Figure 5B), we used 30 $CD8^+$ T cell gene expression signatures related to different T cell states, including the 16 gene signatures defined from differential expression profiles in our data set of HCV-specific $CD8^+$ T cells (Figure 2, Tables S2 and S3) as well as 14 gene sets assembled from the literature which were related to $CD8^+$ T cell dysfunction, exhaustion and memory state in murine models (Doering et al., 2012; Singer et al., 2016; Subramanian et al., 2005) or derived from peptide stimulated $CD8^+$ T cells in subjects with controlled or progressing HIV infection (Gaiha et al., 2014) (Table S3).

For a module signature enrichment analysis (Figure 5C), we used 19 $CD8^+$ T cell co-expression module signatures that were identified using *WGCNA* in longitudinally sampled, antigen-specific $CD8^+$ T cells isolated from mice infected with acute (13 modules) or chronically persisting (6 modules) strains of lymphocytic choriomeningitis virus (LCMV) (Doering et al., 2012) (Table S3). Details for the assembly of the used gene signatures are explained earlier in STAR Methods.

Both expression and module signature enrichments (Figures 5B and 5C) were calculated using a standard hypergeometric test as implemented in *WGCNA* using the entire set of 3850 genes included in the analysis as background. Gene signature enrichment scores in the form of p-values were corrected for multiple hypothesis testing within experimental groups and modules by control of FDR. All enrichments with FDR <0.05 are shown in Figures 4B and 4C. A comprehensive list of signature enrichment results for all groups and modules can be found in Table S7.

Functional Enrichments: For the Immune Pathway enrichment (Figure 5D, top) we used a manually curated list of 54 immune pathways as contained in the *WGCNA* package (assembled from Ingenuity, WikiPathways, and literature (Langfelder and Horvath, 2008) (Table S3). Immune pathway enrichments were calculated using a standard hypergeometric

test as implemented in *WGCNA*. Enrichment scores in the form of p-values were corrected for multiple hypothesis testing within experimental groups and modules by control of FDR. The top 5 enrichments (by score and size of overlap) per group and modules that were also enriched above a significance threshold of FDR < 0.05 are shown in Figure 5D.

For the functional database enrichment (Figure 5D, bottom) we used the publicly available gene ontology and KEGG pathway databases (Ashburner et al., 2000; Kanehisa, 2000). We used *topGO* (Alexa et al., 2006) to test the identified network modules for enrichment of gene ontology terms related to biological processes, cellular components and molecular function with scoring based on *topGO's* mixed elimination and weight algorithm and Fisher's exact test. For analysis of KEGG pathway enrichments (Figures 5D) we used the corresponding function implemented in *limma* (Ritchie et al., 2015). Rather than controlling the FDR, we opted to impose a more stringent cutoff for significance thresholding of GO and KEGG enrichment analyses ($p < 1e-10$), since these generally produced very strong p-values. Furthermore, *topGO* utilizes the underlying graph structure of the GO topology to return enrichment scores which can be considered to not be affected by multiple testing. A comprehensive list of functional enrichment results for all groups and modules can be found in Table S7.

For the KEGG enrichment analyses of the metabolism cluster network (Figure 6B, Table S8) we corrected enrichment scores (p-values) for multiple hypothesis testing by controlling the FDR. For this enrichment, all significant enrichments with ≥4 genes and FDR <0.05 are shown.

All functional enrichments were calculated using the entire set of 3850 genes included in the analysis as background.

Identification and visualization of cluster networks: Based on detected module communities (Figure 3E, right) and results from enrichment analysis of gene ontologies, pathways and gene signatures performed (Figure 5) we selected gene sets of common functional annotation to construct networks of co-regulated gene clusters related to metabolism, nucleosome and immune signaling (Table S8). We visualized these networks with *igraph* (Csardi and Nepusz, 2006) using a fixed network structure to allow for easy identification of functionally preserved components and connections within subnetworks (Figures 6A, S1 and S2).

Regulatory Network Analysis

Regulatory interaction network assembly: We integrated information on experimentally confirmed and predicted interactions of regulatory genes and their targets from multiple publicly available repositories to create a large regulatory interaction network, consisting of 1319 regulators and ~2.9 million unique regulatory interactions. We combined data from CHEA, ENCODE (Dunham et al., 2012), TRANSFAC (Matys, 2006), and JASPAR (Mathelier et al., 2014) data sets obtained from the Harmonizome (Rouillard et al., 2016) repository and combined it with information on interactions from the TRRUST (Han et al., 2015) and RegNetwork (Liu et al., 2015) repositories, which also included information on interaction type and confidence. Links to the individual resources are given in the KEY

RESOURCES TABLE. Furthermore, we included interactions from two weighted, cell type-specific regulatory networks for $CD8^+$ T cells and lymphocytes inferred recently published by Marbach et al. (Marbach et al., 2016). For purposes of this study, only interactions with an interaction weight >0.005 were included from these two networks.

Regulatory molecules were annotated for their class (transcription factor, transcription co-factor, and chromatin remodeling gene) in accordance with their annotation in the animal transcription factor database (Zhang et al., 2015).

Identification of High Confidence Transcriptional Regulators: We intersected the assembled regulatory interaction network with the constructed gene co-expression networks for Chronic, Escape, and Resolver groups that were pruned to exclude weak interactions (only interactions with weight >0.025 were included), resulting in networks of high-confidence regulatory interactions present in the HCV data.

To identify the most important within-module regulatory interactions for each regulator annotated for a module, we counted the number of regulatory interactions between the regulator and its targets within the same module and compared it to its intramodular connectivity (Figures 7A and S3, Table S10). To account for varying module sizes, we assessed regulatory target coverage both by total number of connections and number of connections relative to module size (percentage).

Construction of HCV-specific $CD8^+$ T cell regulatory networks: To identify and highlight high-confidence regulatory interactions between important transcriptional regulators (Figure 7A) and genes of interest from the functional cluster networks (Figures 6A, S1, and S2), we subset regulatory interaction networks for Chronic, Escape, and Resolver groups to important transcriptional regulators of interest (Figures 7A and S3, Table S10) and genes with intramodular connectivity >0.25 contained in metabolism, immune, and immune/nucleosome cluster networks. We allowed for connections of regulatory genes to targets across modules that had been omitted in the transcription factor identification step, since the existence of such connections is likely to better reflect the underlying biology as investigated modules are highly co-regulated. In contrast, their inclusion in the identification step would have inevitably lead to overestimation of the importance of individual transcription factors (as some TFs are annotated for >20,000 interactions) and likely obscure the highly combinatorial design, by which transcriptional regulation is organized (Ernst and Smale, 1995).

We created regulatory networks based on the interactions of the transcriptional regulators *STAT5A*, *KDM5B*, *CBX3*, *RFX5*, *NELFE*, *STAT1*, *TP53*, *FLI1*, *IRF3*, *CREB1* within Chronic, Escape and Resolver groups in our data set and visualized them by functional annotation (Figure 7B) as well early and late gene expression levels (see Figure 2), module color (see Figure 3) and shared co-regulation among groups (see Figures 6A, S1 and S2) using *igraph* (FigureS4). To reduce visual redundancies, regulatory interactions for annotated as transcription co-factors and chromatin remodeling genes were subset to those targeting transcription factors.

DATA AND SOFTWARE AVAILABILITY

Data Resources—The raw and processed microarray data generated for this study are deposited in the gene expression omnibus (GEO) and can be found under the accession numbers GSE93712 for the original data set and GSE93711 for the validation data set.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
FITC mouse Anti-Human CD4, mAb, clone 13B8.2	*Beckman Coulter*	*Cat# IM0448U*
APC-Cy7 mouse Anti-Human CD8, mAb, clone SK1	*BD Biosciences*	*Cat# 557834*
FITC mouse Anti-Human CD14, mAb, clone RMO52	*Beckman Coulter*	*Cat# IM0645U*
PE-Cy7 mouse Anti-Human CD16, mAb, clone 3G8	*BD Biosciences*	*Cat# 557744*
FITC mouse Anti-Human CD19, mAb, clone J3-119	*Beckman Coulter*	*Cat# IM1284U*
PE-Cy7 mouse Anti-Human CD56, mAb, clone B159	*BD Biosciences*	*Cat# 557747*
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Unlabeled peptide-MHC Multimers (class I)	*ProImmune*	*Cat# Fxxx-80A-D*
Critical Commercial Assays		
Annexin V, Pacific Blue conjugate	*Thermo Fisher Scientific*	*Cat# A35122*
Annexin Binding Buffer (5X)	*Thermo Fisher Scientific*	*Cat# V13246*
Live/Dead Fixable Yellow Dead Cell Stain Kit	*Thermo Fisher Scientific*	*Cat# L34959*
APC Fluorotag	*ProImmune*	*Cat# K4A-D*
PE Fluorotag	*ProImmune*	*Cat# K2A-D*
Agencourt RNAdvance Tissue Kit	*Beckman Coulter*	*Cat# A32649*
Trizol	*Invitrogen*	*Cat# 15596026*
Ovation® Pico WTA System V2	*Nugen*	*Cat# 3302-12*
MinElute PCR Purification Kit	*Quiagen*	*Cat# 28004*
Phase Lock Gel Heavy	*5 Prime*	*Cat# 2302810*
Human Genome U133A 2.0 Array	*Affymetrix*	*Cat# 900471*
GeneChip HT Human Genome U133A 96-Array Plate	*Affymetrix*	*Cat# 900751*
Deposited Data		
HCV-Specific CD8 T cells, microarray, original data set	*This paper*	*GEO: GSE93712*
HCV-Specific CD8 T cells, microarray, validation data set	*This paper*	*GEO: GSE93711*
HIV-specific CD8 T cells, microarray, differential expression signature	Gaiha et al., 2014	*GEO: GSE56775; GEO: GSE56971*
LCMV-specific CD8 T cells, microarray, differential expression and module signatures	Doering et al., 2012	*GEO: GSE41867*
CD8+ tumor infiltrating lymphocytes, high-throughput sequencing, activation/dysfunction score signature	Singer et al., 2016	*GEO: GSE86042*

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LCMV-specific CD8 T cells, microarray, leading edge signature	Bengsch et al., 2016	*GEO: GSE41867*
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
Recombinant DNA		
Sequence-Based Reagents		
Software and Algorithms		
affy (v1.52.0)	Gautier et al., 2004	https://www.bioconductor.org/packages/release/bioc/html/affy.html
biomaRt (v2.30.0)	Durinck et al., 2009	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
Harman (v1.2.0)	Oytam et al., 2016	https://bioconductor.org/packages/release/bioc/html/Harman.html
igraph (v1.0.1)	Csardi and Nepusz, 2006	http://igraph.org/r/
limma (v3.30.0)	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
sva (v3.22.0)	Leek and Storey, 2007	https://www.bioconductor.org/packages/release/bioc/html/sva.html
topGO (v2.26.0)	Alexa et al., 2006	https://bioconductor.org/packages/release/bioc/html/topGO.html
WGCNA (v1.51)	Langfelder and Horvath, 2008	https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/
Other		
Custom microarray annotation hgu133a2hsentrezg.db (v20.0.0)	*Psychiatry/MBNI Microarray Lab*	http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/20.0.0/entrezg.asp
Custom microarray annotation chip definition file hgu133a2hsentrezgcdf (v20.0.0)	*Psychiatry/MBNI Microarray Lab*	http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/20.0.0/entrezg.asp
Custom microarray annotation hthgu133ahsentrezg.db (v20.0.0)	*Psychiatry/MBNI Microarray Lab*	http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/20.0.0/entrezg.asp
Custom microarray annotation chip definition file hthgu133ahsentrezgcdf (v20.0.0)	*Psychiatry/MBNI Microarray Lab*	http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/20.0.0/entrezg.asp
Harmonizome database, transcriptional regulatory interactions	Rouillard et al., 2016	http://amp.pharm.mssm.edu/Harmonizome/
RegNetwork database, transcriptional regulatory interactions	Liu et al., 2015	http://www.regnetworkweb.org
TRRUST database, transcriptional regulatory interactions	Han et al., 2015	http://www.grnpedia.org/trrust/
Cell type and tissue-specific regulatory networks, Lymphocytes and CD8 T cells, transcriptional regulatory interactions	Marbach et al., 2016	http://regulatorycircuits.org/download.html

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Brad Rosenberg, Naglaa Shoukry, Paul Klenerman, and Kate Jeffrey for constructive discussions.

This work was funded by the NIH through U19 AI066345/U01 AI13131401 (G.M.L., L.L.L.X., A.Y.K.), U19 AI082630 (G.M.L., A.Y.K., R.T.C., W.N.H.), R01 AI105035/R01 DA016017 (G.M.L.) and by ARC, Paris, IHU Strasbourg/TheraHCC IHUARC IHU201301187 and the European Union/ERC-AdG-2014-671231-HEPCIR, EU H2020-667273-HEPCAR, FP7 HEPAMAB GAN 305600, ANR/LABEX ANR-10-LABX-0028_HEPSYS (all to T.F.B.).

References

Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006; 22:1600–1607. [PubMed: 16606683]

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology-tool for the unification of biology. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]

Bakr I, Rekacewicz C, Hosseiny El M, Ismail S, Daly El M, El-Kafrawy S, Esmat G, Hamid MA, Mohamed MK, Fontanet A. Higher clearance of hepatitis C virus infection in females compared with males. Gut. 2006; 55:1183–1187. [PubMed: 16434426]

Barabási AL, Oltvai ZN. Network biology-understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101–113. [PubMed: 14735121]

Bengsch B, Johnson AL, Kurachi M, Odorizzi PM, Pauken KE, Attanasio J, Stelekati E, McLane LM, Paley MA, Delgoffe GM, et al. Bioenergetic Insufficiencies Due to Metabolic Alterations Regulated by the Inhibitory Receptor PD-1 Are an Early Driver of CD8(+) T Cell Exhaustion. Immunity. 2016; 45:358–373. [PubMed: 27496729]

Berkers CR, Maddocks ODK, Cheung EC, Mor I, Vousden KH. Metabolic Regulation by p53 Family Members. Cell Metab. 2013; 18:617–633. [PubMed: 23954639]

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech: Theory Exp. 2008; 2008:P10008.

Bowen DG, Walker CM. Adaptive immune responses in acute and chronic hepatitis C virus infection. Nature. 2005; 436:946–952. [PubMed: 16107834]

Brodin P, Davis MM. Human immune system variation. Nat Rev Immunol. 2016; 17:21–29. [PubMed: 27916977]

Cox AL, Mosbruger T, Lauer GM, Pardoll D, Thomas DL, Ray SC. Comprehensive analyses of CD8+ T cell responses during longitudinal study of acute human hepatitis C. Hepatology. 2005a; 42:104–112. [PubMed: 15962289]

Cox AL, Mosbruger T, Mao Q, Liu Z, Wang XH, Yang HC, Sidney J, Sette A, Pardoll D, Thomas DL, et al. Cellular immune selection with hepatitis C virus persistence in humans. J Exp Med. 2005b; 201:1741–1752. [PubMed: 15939790]

Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006:713–716.

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005; 33:e175–e175. [PubMed: 16284200]

Doering TA, Crawford A, Angelosanto JM, Paley MA, Ziegler CG, Wherry EJ. Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8+ T Cell Exhaustion versus Memory. Immunity. 2012; 37:1130–1144. [PubMed: 23159438]

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Kaul R, Khatun J, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4:1184–1191. [PubMed: 19617889]

Ernst P, Smale ST. Combinatorial regulation of transcription I: General aspects of transcriptional control. Immunity. 1995; 2:311–319. [PubMed: 7719936]

Fafi-Kremer S, Fofana I, Soulier E, Carolla P, Meuleman P, Leroux-Roels G, Patel AH, Cosset FL, Pessaux P, Doffoël M, et al. Viral entry and escape from antibody-mediated neutralization influence hepatitis C virus reinfection in liver transplantation. J Exp Med. 2010; 207:2019–2031. [PubMed: 20713596]

Gaiha GD, McKim KJ, Woods M, Pertel T, Rohrbach J, Barteneva N, Chin CR, Liu D, Soghoian DZ, Cesa K, et al. Dysfunctional HIV-Specific CD8+ T Cell Proliferation Is Associated with Increased Caspase-8 Activity and Mediated by Necroptosis. Immunity. 2014; 41:1001–1012. [PubMed: 25526311]

Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20:307–315. [PubMed: 14960456]

Giefing Kröll C, Berger P, Lepperdinger G, Grubeck-Loebenstein B. How sex and age affect immune responses, susceptibility to infections, and response to vaccination. Aging Cell. 2015; 14:309–321. [PubMed: 25720438]

Grakoui A. HCV Persistence and Immune Evasion in the Absence of Memory T Cell Help. Science. 2003; 302:659–662. [PubMed: 14576438]

Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, et al. TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep. 2015; 5:11432. [PubMed: 26066708]

Kahan SM, Wherry EJ, Zajac AJ. T cell exhaustion during persistent viral infections. Virology. 2015; 479–480:180–193.

Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000; 28:27–30. [PubMed: 10592173]

Kasprowicz V, Kang YH, Lucas M, Schulze zur Wiesch J, Kuntzen T, Fleming V, Nolan BE, Longworth S, Berical A, Bengsch B, et al. Hepatitis C Virus (HCV) Sequence Variation Induces an HCV-Specific T-Cell Phenotype Analogous to Spontaneous Resolution. J Virol. 2010; 84:1656–1663. [PubMed: 19906915]

Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics–a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009; 25:415–416. [PubMed: 19106121]

Komai-Koma M, Jones L, Ogg GS, Xu D, Liew FY. TLR2 is expressed on activated T cells as a costimulatory receptor. Proc Natl Acad Sci USA. 2004; 101:3029–3034. [PubMed: 14981245]

Kuniholm MH, Kovacs A, Gao X, Xue X, Marti D, Thio CL, Peters MG, Terrault NA, Greenblatt RM, Goedert JJ, et al. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. Hepatology. 2010; 51:1514–1522. [PubMed: 20169624]

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008; 9:559.

Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. 2007; 3:e161.

Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database. 2015; 2015:bav095. [PubMed: 26424082]

Major ME, Dahari H, Mihalik K, Puig M, Rice CM, Neumann AU, Feinstone SM. Hepatitis C virus kinetics and host responses associated with disease and outcome of infection in chimpanzees. Hepatology. 2004; 39:1709–1720. [PubMed: 15185313]

Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods. 2016; 13:366–370. [PubMed: 26950747]

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 2014; 42:D142–D147. [PubMed: 24194598]

Matys V. TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–D110. [PubMed: 16381825]

Muñoz-Fontela C, Mandinova A, Aaronson SA, Lee SW. Emerging roles of p53 and other tumour-suppressor genes in immune regulation. Nat Rev Immunol. 2016; 16:741–750. [PubMed: 27667712]

Osburn WO, Snider AE, Wells BL, Latanich R, Bailey JR, Thomas DL, Cox AL, Ray SC. Clearance of hepatitis C infection is associated with the early appearance of broad neutralizing antibody responses. Hepatology. 2014; 59:2140–2151. [PubMed: 24425349]

Oytam Y, Sobhanmanesh F, Duesing K, Bowden JC, Osmond-McLeod M, Ross J. Risk-conscious correction of batch effects: maximising information extraction from high-throughput genomic datasets. BMC Bioinf. 2016; 17:733.

O'Shea JJ, Plenge R. JAK and STAT Signaling Molecules in Immunoregulation and Immune-Mediated Disease. Immunity. 2012; 36:542–550. [PubMed: 22520847]

Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012; 12:252–264. [PubMed: 22437870]

Pestka JM, Zeisel MB, Blaser E, Schurmann P, Bartosch B, Cosset FL, Patel AH, Meisel H, Baumert J, Viazov S, et al. Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. Proc Natl Acad Sci USA. 2007; 104:6025–6030. [PubMed: 17392433]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47–e47. [PubMed: 25605792]

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database. 2016; 2016:baw100. [PubMed: 27374120]

Schulze zur Wiesch J, Ciuffreda D, Lewis-Ximenez L, Kasprowicz V, Nolan BE, Streeck H, Aneja J, Reyor LL, Allen TM, Lohse AW, et al. Broadly directed virus-specific CD4+ T cell responses are primed during acute hepatitis C infection, but rapidly disappear from human blood with viral persistence. J Exp Med. 2012; 209:61–75. [PubMed: 22213804]

Schulze zur Wiesch J, Lauer GM, Day CL, Kim AY, Ouchi K, Duncan JE, Wurcel AG, Timm J, Jones AM, Mothe B, et al. Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. J Immunol. 2005; 175:3603–3613. [PubMed: 16148104]

Shiroishi M, Tsumoto K, Amano K, Shirakihara Y, Colonna M, Braud VM, Allan DSJ, Makadzange A, Rowland-Jones S, Willcox B, et al. Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. Proc Natl Acad Sci USA. 2003; 100:8856–8861. [PubMed: 12853576]

Shoukry NH, Grakoui A, Houghton M, Chien DY, Ghrayeb J, Reimann KA, Walker CM. Memory CD8 +T Cells Are Required for Protection from Persistent Hepatitis C Virus Infection. J Exp Med. 2003; 197:1645–1655. [PubMed: 12810686]

Singer M, Wang C, Cong L, Marjanovic ND, Kowalczyk MS, Zhang H, Nyman J, Sakuishi K, Kurtulus S, Gennert D, et al. A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell. 2016; 166:1500–1511.e1509. [PubMed: 27610572]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–15550. [PubMed: 16199517]

Swadling L, Capone S, Antrobus RD, Brown A, Richardson R, Newell EW, Halliday J, Kelly C, Bowen D, Fergusson J, et al. A human vaccine strategy based on chimpanzee adenoviral and MVA

vectors that primes, boosts, and sustains functional HCV-specific T cell memory. Sci Transl Med. 2014; 6:261ra153–261ra153.

Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, Pillay T, Ouchi K, Reyor LL, Wiesch zur JS, et al. CD8 Epitope Escape and Reversion in Acute HCV Infection. J Exp Med. 2004; 200:1593–1604. [PubMed: 15611288]

Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015; 15:486–499. [PubMed: 26205583]

Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res. 2012; 40:e133–e133. [PubMed: 22638577]

Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo AY. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Res. 2015; 43:D76–D81. [PubMed: 25262351]

Highlights

- Disease outcome and viral escape are linked to transcriptional differences in T cells
- T cells from different outcomes share a core of co-regulated T cell identity genes
- Metabolic, nucleosome, and immune genes are dysregulated early in chronic infection
- Dysregulation correlates with sex, age, and presence of HCV-specific CD4 T cells

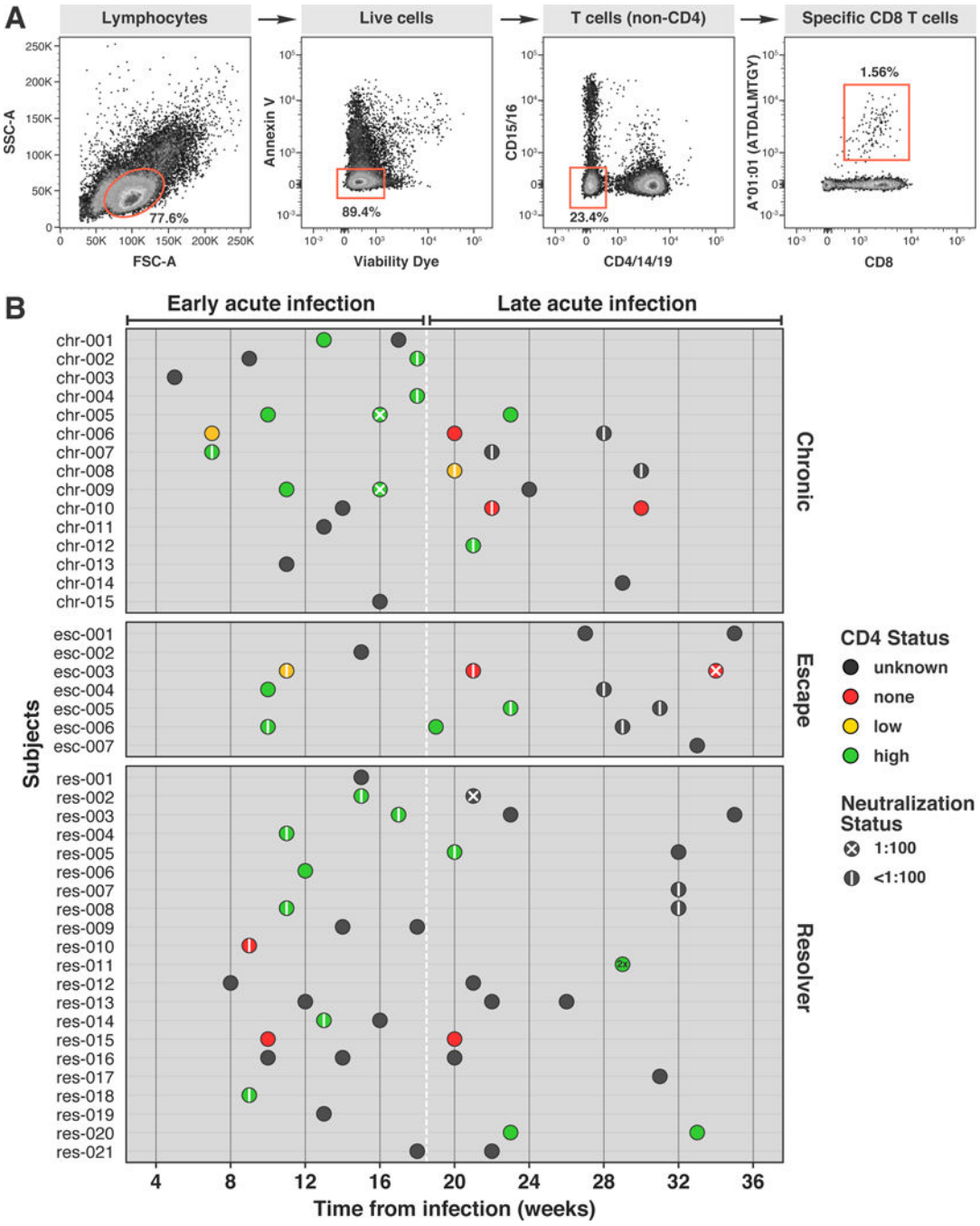


Figure 1. Multivariate analysis of adaptive immunity during early HCV infection

(A) Gating strategy for sorting of HCV-specific $CD8^+$ T cells, defined as CD8 and pMHC multimer positive lymphocytes negative for CD4, CD14, CD19, CD16, CD56, Annexin V, and LiveDead viability stain.

(B) Sampling and assay overview. HCV-specific $CD8^+$ T cells from 43 subjects with different outcomes and varying levels of antigen recognition (Chronic (C), Escape (E), and Resolver (R)) were sampled at multiple time points (dots) within the first 36 weeks of infection (n_{total} = 78, STAR Methods). $CD4^+$ T cell status (magnitude of HCV-specific

$CD4^+$ T cell response, dot color) and neutralizing antibodies (titers from HCV pseudoparticle neutralization assay, white bars) were assessed for a subset of these samples (n_{help} = 37 and $n_{neutralization}$=32).
See also Table S1.

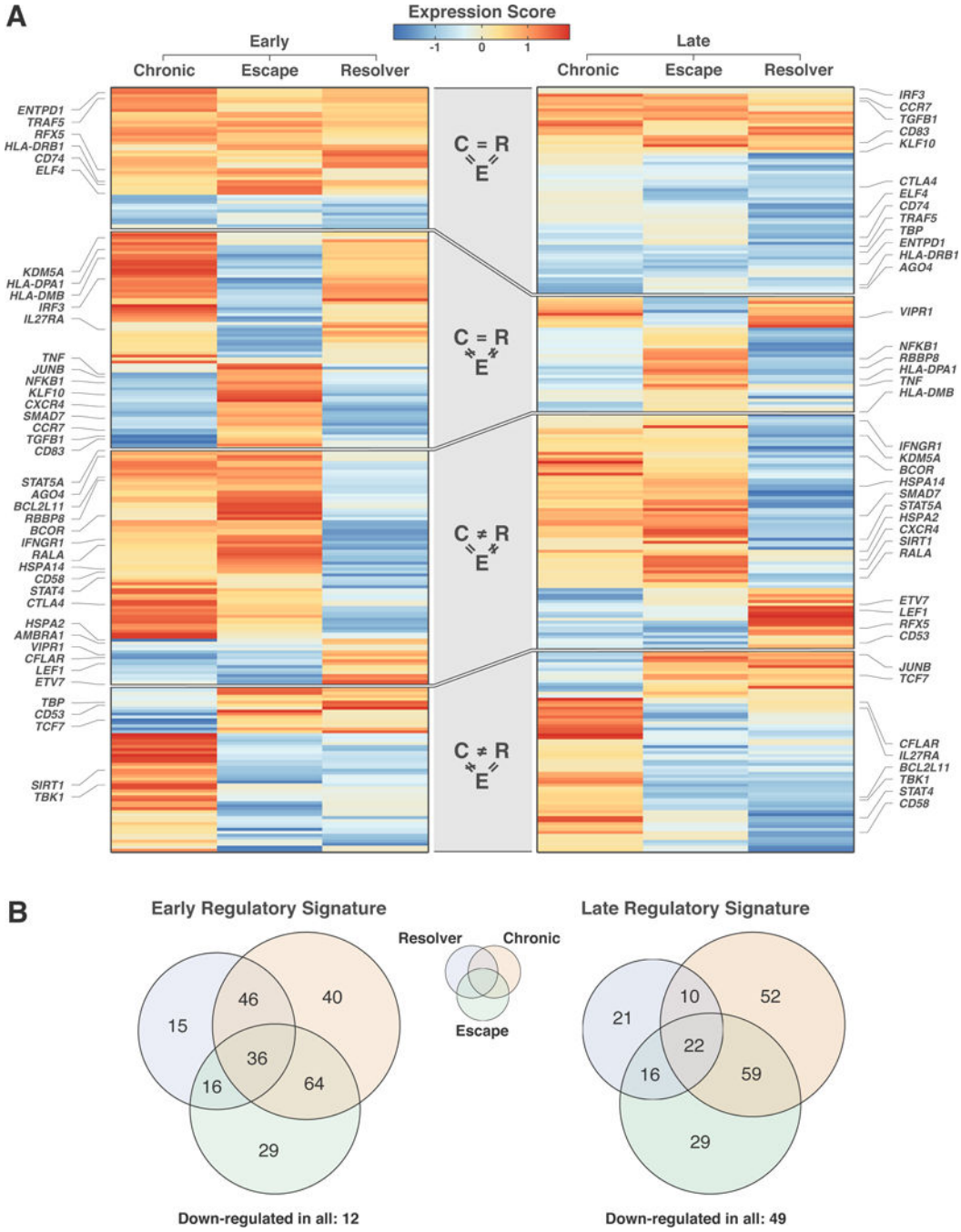


Figure 2. Virus-specific CD8$^+$ T cells display distinct transcriptional profiles of resolution, inflammation and persistent antigen exposure

(A) Heatmap of 258 differentially expressed genes across C, E, and R groups during early and late acute phase of HCV infection (<18 and >18 weeks post infection). Transcriptional profiles of HCV-specific CD8$^+$ T cells were analyzed for differential gene expression between C and R in early or late acute infection or between early and late infection within C or R. Shown are means of scaled expression levels (expression scores) by group. Expression levels for E samples are shown but not used for detection of differential expression.

(B) Venn diagrams of distinct transcriptional signatures defined by patterns of shared differential expression between C, E, and R groups and in early and late acute phase of infection.

See also Tables S1, S2, and S3.

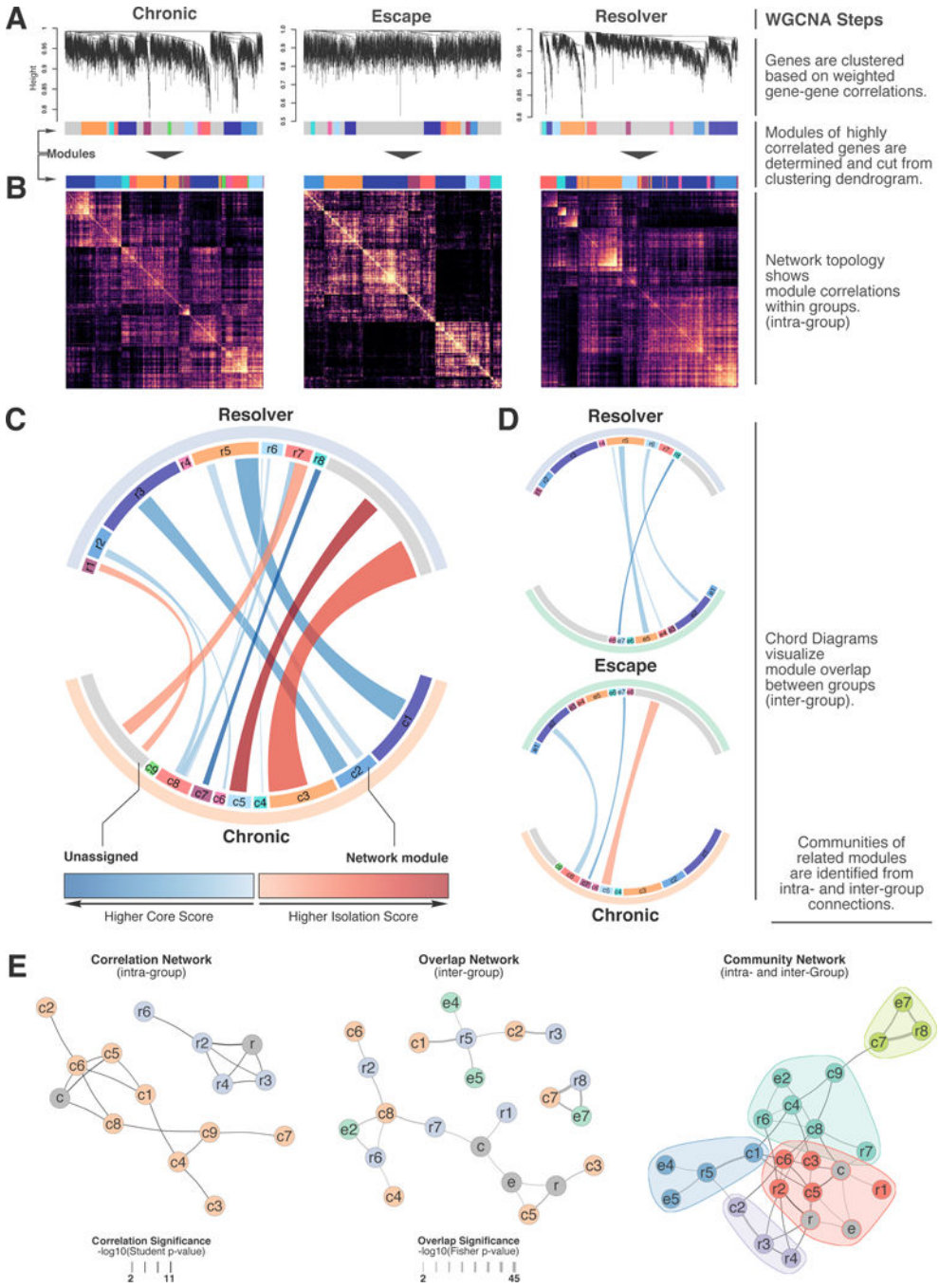


Figure 3. Modules of tightly co-regulated genes give rise to distinct patterns of transcriptional regulation in different infection outcomes

(A) Detection of gene co-expression modules in C, E, and R groups. Weighted gene co-expression networks were constructed using WGCNA and modules of highly correlated genes were detected by hierarchical clustering. Colored bars indicate detected modules and their relative size, similar colors across groups do not imply overlap.

(B) Heatmap of scaled topological overlap matrices for isolated network modules from C, E, and R groups. Lighter colors indicate stronger positive correlation between genes within and across modules.

(C and D) Chord diagrams of gene module correspondence between C, E, and R groups. Diagrams show significance of gene overlap between two modules (core score, blue connection) or a module and unconnected genes (isolation score, red connections) for individual group pairs of (C) C/R or (D) C/E and E/R. Shown are scores that passed a significance threshold of FDR <0.01.

(E) Correlation, overlap and community networks for modules detected in C, E, and R groups. Correlation and overlap networks were constructed using within-group correlation significance (FDR <0.01) and across-group overlap significance (FDR <0.01) and merged for detection of module communities (indicated by color) (STAR Methods). Significance of correlation and overlap were used for network layout and shown as width of the edges connecting modules within networks.

See also Table S4.

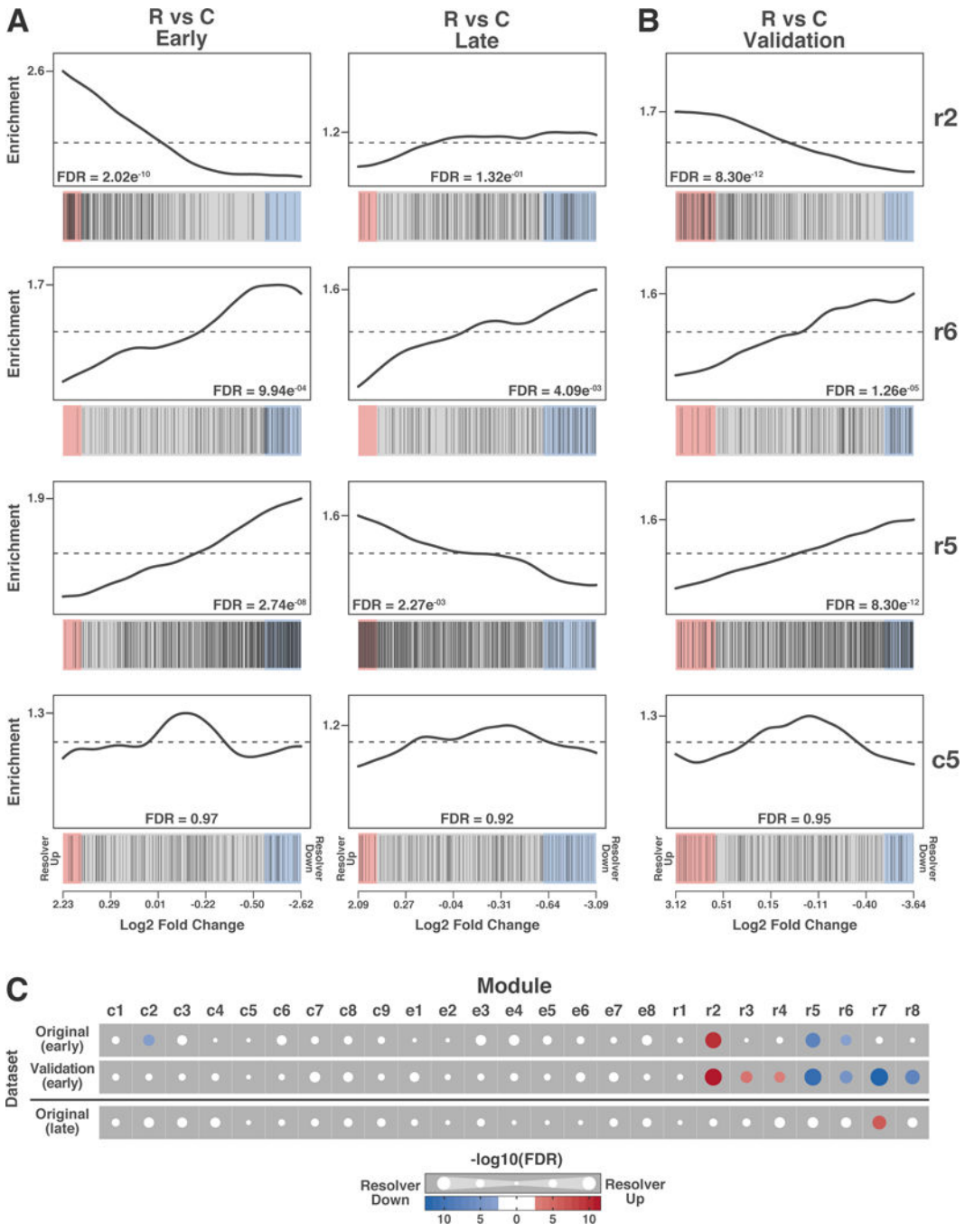


Figure 4. Differential expression of network modules is preserved across cohorts of acute HCV infection

(A and B) Barcode plots of network module signature enrichments for HCV-specific $CD8^+$ T cells from (A) C and R patients during the early and late phase of acute infection and (B) a separate validation cohort of chronic and resolving patients sampled during the early acute phase of infection (≤ 21 weeks post infection).

Ranked lists of log2 fold changes from analysis of differential expression between C and R groups were tested for enrichment of network modules (STAR Methods). Shown are representative examples of modules that were significantly enriched in early samples of the

original data set (FDR < 0.001) (r2, r6 and r5) as well as a universally non-enriched module (c5). (C) Overview of enrichment results for all three sample subsets and all modules. Dot size indicates FDR corrected p-values of enrichment. Statistically significant enrichments (FDR < 0.001) are highlighted in color according to directionality of enrichment in respect to up-regulation (red) or down-regulation (blue) in the R group.
See also Tables S4, S5, and S6.

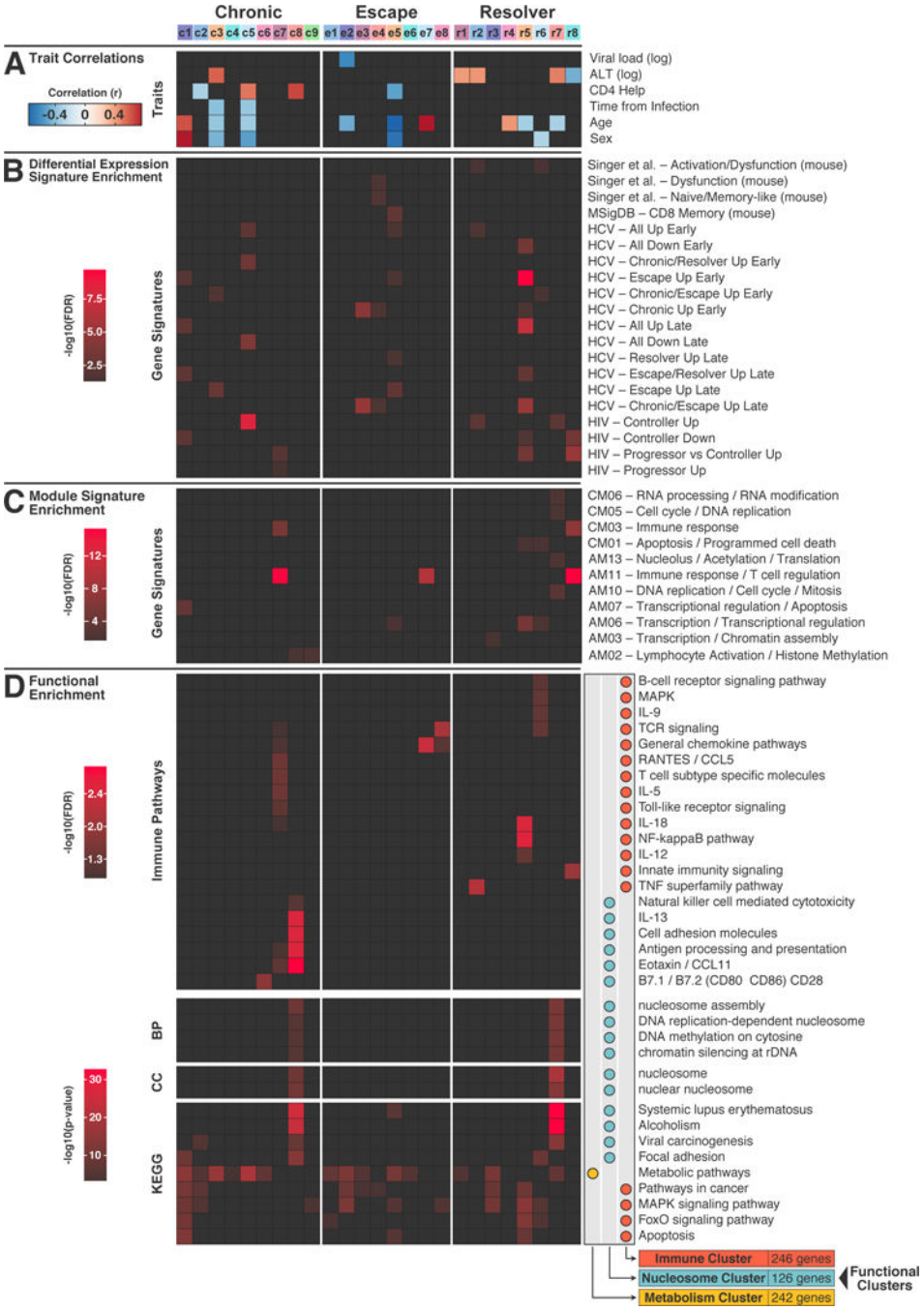


Figure 5. Correlation and enrichment maps link gene co-expression modules from different immunological states to clinical traits, immune signatures and biological function

(A) Trait correlation map for C, E and R groups in acute HCV infection. Eigengenes (first principal components) of network modules were correlated with clinical traits (colors indicate r value and sign of correlations (red = positive, blue = negative). Shown are correlations with FDR < 0.1.

(B and C) Enrichment maps for C, E, and R groups, showing enrichments scores for all significant enrichments (FDR < 0.05). (B) Modules were tested for enrichment of $CD8^+$ T cell differential expression signatures from HCV (Figure 2), HIV (Gaiha et al., 2014), and

mouse models of T cell memory and exhaustion (Doering et al., 2012; Singer et al., 2016; Subramanian et al., 2005) (C) Modules were tested for enrichment of $CD8^+$ T cell WGCNA module signatures from acute (AM) and chronic (CM) LCMV infection (Doering et al., 2012).

(D) Functional enrichment maps for C, E, and R groups. Detected modules were tested for enrichment of immune pathways (Langfelder and Horvath, 2008), GO terms, and KEGG pathways. Shown are enrichment scores for top-scoring immune pathways (top 5 per group per module by enrichment score and number of enrichment-driving genes) with FDR <0.05 and KEGG pathway/GO term enrichments below a significance threshold of FDR <1e-10. Enriched processes were grouped into clusters of functionally related and co-regulated terms related to metabolism, immune, and nucleosome clusters as indicated by colored dots in grey panels to the right of enrichment maps.

See also Tables S3, S7 and S8.

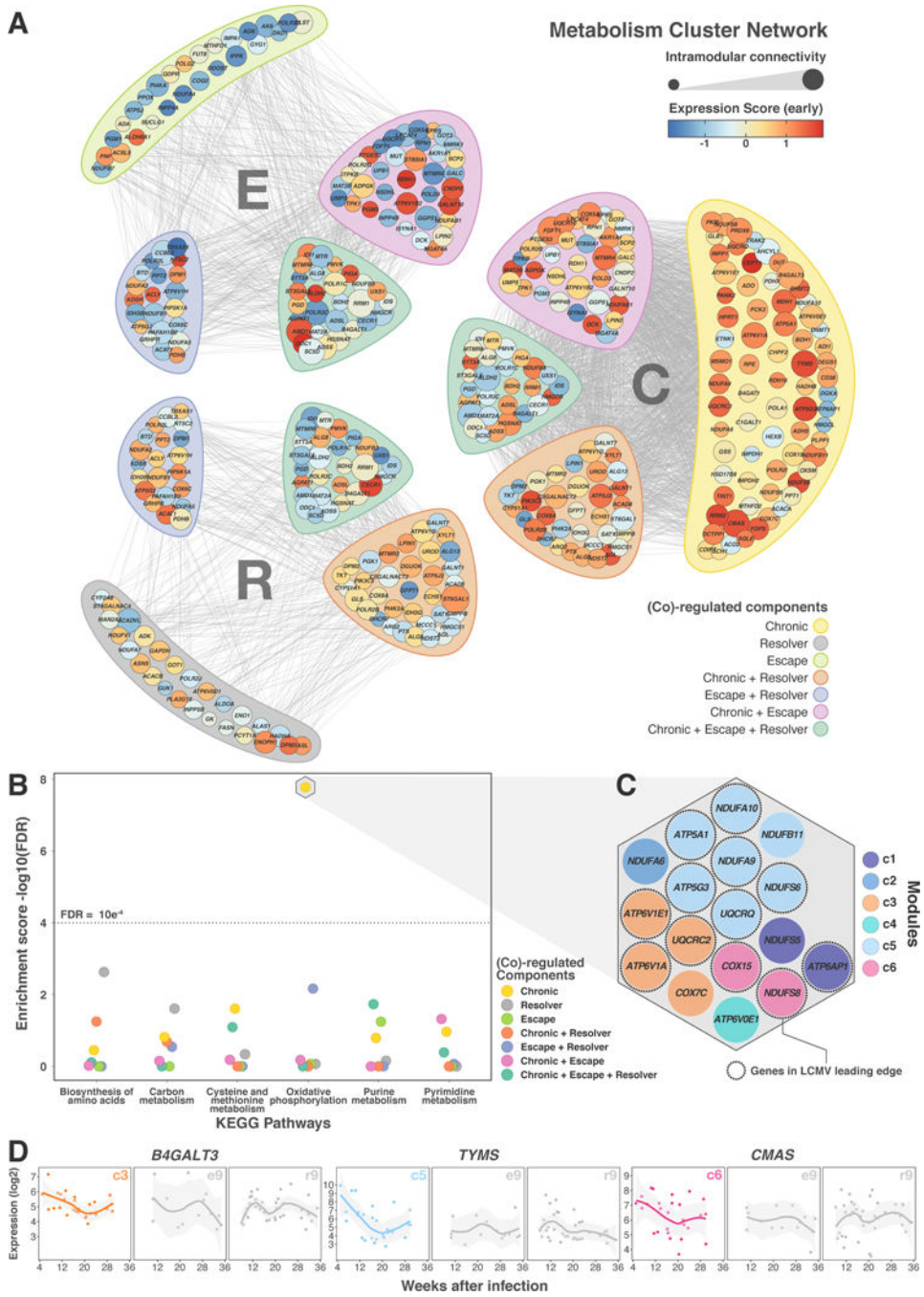


Figure 6. A distinct temporal pattern of metabolic regulation marks chronic HCV infection

(A) Metabolism Cluster Network for C, E, and R, groups. Co-expression networks for each group were subset to genes annotated to the metabolism cluster and visualized as networks with components of genes grouped by exclusive or shared regulation between groups (colored bubbles). Nodes are sized by intramodular connectivity and colored by group expression score during the early phase of infection.

(B) KEGG pathway enrichment in metabolism cluster network. Individual components from (A) were tested for enrichment of KEGG pathways. All pathways that were significantly enriched (FDR <0.05) in at least one component are shown.

(C) Chronic-exclusively regulated genes annotated for OXPHOS. Genes (shown as nodes), colored by their module membership in the C group. Genes that overlap with a leading-edge signature of metabolic processes associated with T cell exhaustion (Bengsch et al., 2016) are encircled by a dashed line (12/17 genes).

(D) Expression kinetics of hub genes for modules c3, c5, c6. Log2-transformed gene expression levels for the most highly connected gene of each module are plotted over time for each group (dots = individual samples, color = module membership in C). Longitudinal trends in expression are visualized using a local regression based (LOESS) smoother (grey shading = 95% confidence interval).

See also Figures S1 and S2, and Tables S4, S8, and S9.

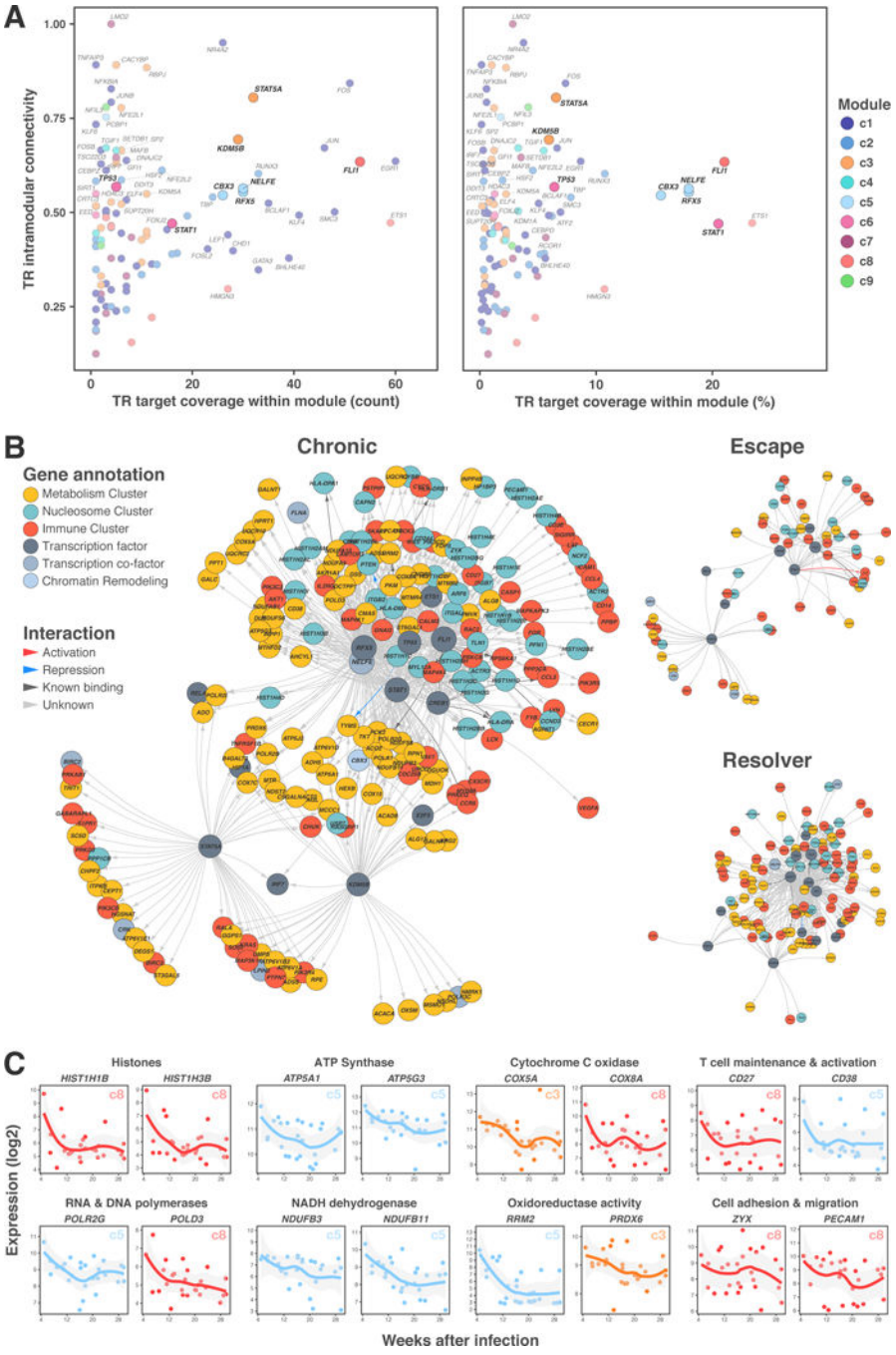


Figure 7. Distinct regulatory interactions link dysregulation of metabolic, nucleosomal, and immune processes in CD8$^+$ T cells from early chronic HCV infection

(A) TR-target coverage for CD8$^+$ T cells in chronic HCV infection. A large network of regulatory interactions was assembled from the literature (STAR Methods) and subset to TR-target interactions with matching module assignments. TRs are plotted by number (left) or percentage (right) of within-module regulatory interactions and intramodular connectivity. Transcription factors with intramodular connectivity > 0.5 and/or target coverage count > 20 or percentage >5% are labeled with their respective gene symbols. TR module membership is indicated by color.

(B) Regulatory networks for select TRs from modules c3, c5, c6, and r2. Weighted gene correlation networks were restricted to strong interactions (weight > 0.025) between selected TRs and known targets (k_{Within} >0.25). For transcription co-factors and chromatin remodeling genes, only interactions targeting transcription factors are shown. TRs and targets are labeled with gene symbols and colored by regulatory function or annotation in metabolism, nucleosome, or immune cluster networks. Interaction types are indicated by color of edges connecting TRs and targets.

(C) Representative expression kinetics of Chronic regulatory network genes. Shown are log2-transformed expression levels (dots represent individual samples) of genes annotated for regulation of transcription (Histones, RNA & DNA polymerases), metabolism (ATP Synthase, NADH dehydrogenase, Cytochrome C oxidase, and Oxidoreductase activity), and T cell function and homing (T cell maintenance & activation, Cell adhesion & migration). Color and number indicate C group module membership. Longitudinal trends in expression are visualized using a local regression based (LOESS) smoother (grey shading = 95% confidence interval).

See also Figures S3 and S4, and Tables S8 and S10.