The Individuality of (virus-specific) CD8⁺ T cells

van Aalderen, M.C.

Citation for published version (APA):
LABEL-FREE ANALYSIS OF CD8+ T CELL SUBSET PROTEOMES SUPPORTS A PROGRESSIVE DIFFERENTIATION MODEL OF VIRUS-SPECIFIC T CELLS IN HUMANS

M.C. van Aalderen,1,2# M. van den Biggelaar,3#
E. B. M. Remmerswaal,1,2 F. P.J. van Alphen,3 A. B. Meijer,3
I. J. M. ten Berge,1,2 R. A. W. van Lier3

1 Department of Experimental Immunology, Academic Medical Center,
   Amsterdam, the Netherlands
2 Renal Transplant Unit, Academic Medical Center,
   Amsterdam, the Netherlands
3 Sanquin Research and Landsteiner laboratory,
   Amsterdam, the Netherlands.
# Authors contributed equally.
ABSTRACT

Different pathogens trigger naïve T cells to express distinct sets of effector proteins. To better understand the molecular mechanisms that drive this functional specification, we used high resolution mass spectrometry and label-free protein quantification to measure proteomic differences between the seven largest circulating human CD8+ T cell subsets. Hierarchical clustering of the proteomes placed naïve and CD45RA-expressing effector-type T cells at the extremes of the spectrum with central-memory and other effector-memory stages located in between. Prominent differences between the subsets included expression of various granzymes, signalling proteins and molecules involved in metabolic regulation and cell adhesion. Remarkably, whereas most of the proteomic changes between the subsets were gradual, a small proportion of proteins were regulated only in discrete subsets. The data obtained from this proteome analysis correspond best to a progressive differentiation model in which specific stable traits are gradually acquired during pathogen-specific development.
INTRODUCTION

Functionally distinct CD8+ memory T cell subsets have been implicated in immunological control of different viruses in humans (reviewed in 1). Differences concern the expression of transcription factors, cytokine- and chemokine receptors, co-stimulatory receptors, proteases, and the production of cytokines (reviewed in 1-4). For example, whereas Epstein-Barr virus (EBV)-specific cells highly express the transcription factor eomesodermin (Eomes) and the serine protease granzyme K, human cytomegalovirus (hCMV)-specific cells predominantly express the transcription factors T-bet and Hobit, and granzyme B 5,6. Importantly, the distinct functional properties of different types of antigen-specific CD8+ T cells have been implicated not only in the immunological control of viruses, but also in the immune defence against tumours and in tumour treatment outcomes 7-9. These notions corroborate the importance of specific differentiation pathways, where different types of CD8+ memory T cell responses have adapted to optimally combat different types of pathogens.

It is unsettled if distinct CD8+ memory T cell subsets develop as a consequence of progressive and linear differentiation, where the persistence of viral proteins drives the acquisition of specific traits, or rather, if distinctive adaptations of CD8+ T cells occurs in response to unique pathogens (i.e. branched differentiation 1). One approach to differentiate between these two, non-mutually exclusive, paradigms is the unbiased analyses of genes and proteins that are expressed by CD8+ T cells that belong to well-demarcated subpopulations. Transcriptome analyses revealed that both memory-type and effector-type CD8+ memory T cells differ from naïve cells in the expression of many genes while at the same time also emphasizing major differences between antigen-primed subsets 10. Notably however, these analyses fall short in providing insight in linear versus branched differentiation models since only two sets of phenotypically distinct primed subsets were analysed 11.

Next to the aforementioned need for unbiased analyses of intermediate T cell subsets it is important to consider that the fast action of memory/effector CD8+ T cells to reinfection or viral reactivation is likely to be dictated primarily by expressed proteins rather than by the transcriptome. Therefore, to enhance our understanding of CD8+ T cell differentiation in humans, we utilized high resolution mass spectrometry (MS) combined with label-free protein quantification (LFQ) to determine the proteomes of the seven largest circulating human CD8+ T cell populations defined by distinctions in CD45RA, CCR7, CD28 and CD27 expression 5,11. We detected a total of 6114 proteins and quantified the expression of 4816, but interestingly, only 286 proteins were differentially expressed between the subsets. Within the set of proteins with differential expression, various serine proteases (granzymes), signalling molecules, proteins involved in metabolic regulation, and proteins involved in (regulation of) cell-cell adhesion were present. Unsupervised hierarchical clustering ordered the CD8+ T cell subsets along a line where the expression of the majority of the proteins gradually changed from the naïve (T_N), to the central-memory (T_CM), effector-memory (T_EM) and RA+ T_EM (effector-type, T_EM_RA) differentiation states in a seemingly linear fashion. Importantly, a small proportion of the proteome displayed non-
PROTEOMIC TRANSITIONS BETWEEN HUMAN CD8+ T CELL SUBSETS

linear expression profiles, peaking or dipping in early-differentiated and effector-memory subsets, respectively. We infer that the apparently linear acquisition and loss of expression of particular proteins by circulating CD8+ T cell subsets would fit best with a progressive differentiation model.

RESULTS AND DISCUSSION

Quantitative proteomics reveals a hierarchical order in CD8+ T cell proteome transitions

Flowcytometry studies identified seven functionally distinct CD8+ T cell subsets in the circulation of healthy adults, as defined by a differential expression of CD45RA, CCR7, CD28 and CD27.5,11-14. These subsets comprise the CD45RA+CCR7+CD28+CD27+ naïve population (T_N), the CD45RA−CCR7−CD28+CD27+ early-differentiated effector-memory population (T_EM1), CD45RA−CCR7+CD28+CD27+ early-like (T_EM2), the CD45RA−CCR7−CD28−CD27+ intermediately-differentiated (T_EM3), the CD45RA−CCR7−CD28−CD27+RA− effector-type (T_EM4), and the CD45RA−CCR7−CD28−CD27− T_EM_RA populations (see reference 4 for a review on the nomenclature).

We first isolated these populations through cell sorting (Supplemental Figure 1). To define the protein expression profiles of these subsets, proteomic analysis of the purified subset isolates was performed (Figure 1A). High resolution MS data were analyzed using the MaxQuant computational platform, by which we identified a total of 6114 proteins (Supplemental Table I). Of these, 4816 were quantified in at least 2 samples in one of the subsets (Supplemental Table II). The proteomic ruler methodology (Wisniewski et al MCP 2014) was applied to derive absolute copy number values for each of the quantified proteins (Supplemental Table III). For each of the subsets, proteins showed a wide range of expression (Supplemental Figure 2A) and a similar cumulative protein abundance (Supplemental Figure 2B). When comparing our data on human CD8+ T cells to that reported for murine CD8+ T cells by Hukelmann et al.15, we find about equal expression of proteins like profilin-1, coflin-1 and vimentin (Supplemental Table III). To assess the biological variation between and within CD8+ subsets, Pearson correlation testing was performed on filtered MaxLFQ intensities omitting missing values (NaN). This showed a high correlation between all CD8+ T cell subsets (r > 0.8, Figure 1B). Unsupervised hierarchical clustering revealed a similar ordering of the subsets as was found previously.5 T_N formed a cluster of its own, being entirely different from all the antigen-experienced subsets. TCM and T_EM1 cells together formed another separate cluster, after which the spectrum continued with the T_EM2, the T_EM3, the T_EM4 and the T_EM_RA subsets, as third cluster. This third cluster could then be further subdivided into a T_EM2, T_EM3 and T_EM4 cluster, and a separate fourth cluster containing only the T_EM_RA (effector-type) subset (Figure 1B).

To identify the proteins with the most prominent differences in expression profiles within the different CD8+ T cell subsets, we performed imputation of missing values with a normal distribution (width = 0.3; shift = 1.8) and used the built-in analysis-of-variance functions in PERSEUS using an FDR of 5% and S0 of 0.4 (the S0 parameter sets a threshold for minimum fold change). This identified a total of 286 differentially expressed proteins...
PROTEOMIC TRANSITIONS BETWEEN HUMAN CD8+ T CELL SUBSETS

Figure 1. Subset analyses using MS: associations between the populations. A. PBMCs were enriched for CD8+ T cells using negative selection with magnetic beads. Thereafter, T_N, T_CM, T_EM1-4 and T_EM RA cells were isolated from the enriched cell fraction using multichannel flowcytometry (see also Supplemental Figure 1). The isolated subsets were lysed into peptides and were then analysed using high resolution mass spectrometry. B. Unsupervised hierarchical clustering analysis and Pearson correlation testing was performed on filtered MaxLFQ intensities omitting missing values (NaN) to determine the subset ordering and the Pearson correlation coefficients between the CD8+ T cell subsets. C. Principle component analysis was done to further assess how the subset proteomes correspond to each other. Symbol and colour corresponds to the respective subsets.

(supplemental Table II). Principle component analysis performed on these proteins further emphasized that differentiated CD8+ T cell subsets can be distinguished based on their protein expression profile (Figure 1C). Enrichment analysis of the differentially expressed proteins indicated that they are associated with various biological processes such as T cell activation (e.g. ‘lymphocyte activation’ and ‘signal transduction’), and cytotoxic T
cell differentiation (e.g. ‘cytolysis’, ‘lytic vacuole’ and ‘(cytoskeletal anchoring at) plasma membrane’) (Figure 2 and Supplemental Table IV).

As such, the proteomes of the T_N and the T_EM RA subsets differed the strongest from one another with the T_CM and the four types of T_EM located in between. Furthermore, these differences relied only on a small proportion of the entire proteome expressed by the overall pool of CD8^+ T cells circulating in healthy adults. These subset-specific proteins indeed fit previous observations that distinguish naïve CD8^+ T cells from cells with typical memory and effector traits.\textsuperscript{5,11-14}

Figure 2. Enrichment analysis of the differentially expressed proteins. Gene ontology enrichment analysis of the 286 significantly different genes. Enrichment analysis of the 286 differentially expressed proteins was performed using the Cytoscape (version 3.2.1) plug-in BiNGO (version 3.0.3) with an FDR threshold of 0.05 done to determine how these associate to distinct biological processes. The color of the node represents the adjusted p-value. White nodes are not significantly over-represented, the other ones are, with a color scale ranging from yellow (adjusted p-value = 0.05) to dark orange (adjusted p-value = 6 orders of magnitude smaller than significance level, e.g. 10^-6 * 0.05). The size of the node is proportional to the number of genes within the node. Node size and color correspond to the number and importance of the interactions with other biological processes.
Proteome changes occur mostly gradually between the antigen-experienced CD8+ T cell subsets

To gain more insight into the overall protein expression profiles within the different CD8+ T cells, as well as the relations between the proteomes of the subsets, we performed a hierarchical clustering, pre-processed with k-means, of the 286 differentially expressed proteins. This approach identified clusters of proteins of which the expression levels changed in a linear fashion, viz. clusters with an increasing expression (clusters 1-3) or a decreasing expression (cluster 4); and clusters in which the expression levels changed in a nonlinear fashion, viz. a cluster with an increasing expression profile only in the T_{CM}, T_{EM1} and T_{EM2} subsets (cluster 5) and a cluster with a decreased expression profile only in the T_{EM2}, T_{EM3} and T_{EM4} subsets (cluster 6) (Figure 3A).

STRING analysis revealed a highly connected network of known physical and functional interactions between the differentially expressed proteins, which partly cluster together according to their expression profile (Figure 3B). The major characteristics of the distinct clusters will be summarized in separate paragraphs below.

The granzyme K cluster

All the proteins in the granzyme (GZM) K cluster, which comprises 42 proteins in total, were expressed in an increased manner by all the subsets when compared to the T_{N} subset (Figure 3A). Enrichment analysis revealed the abundance of proteins related to cellular activation, including ‘calcium ion binding’ (supplemental Table V). The majority of the enriched biological process were linked to the annexins ANXA1 and ANXA2/ANXA2P2 (Supplemental Table V). Notably, Wherry et al. previously observed a similar expression of ANXA1 and ANXA2 at the mRNA level by lymphocytic choriomeningitis virus-specific CD8+ T cells in mice, while our group also reported increased expression of ANXA4 and ANXA5 at the mRNA level by human hCMV-specific CD8+ T cells^{10,16}. These and other proteins, such as the transcription factor EOMES (eomesodermin or Eomes) and the Rab family member RAB27A in this same cluster (Supplemental Figure 3), have been associated with biological processes like ‘tissue development’, ‘system development’, ‘calcium ion binding’, ‘calcium-dependent phospholipid binding’, ‘lipid transport’, ‘regulation of vesicle fusion’ and ‘T cell activation in immune response’, all referring to a specific regulation of cellular activation and cytotoxic T cell development (Supplemental Table V). Interestingly, we recently identified a strong association in EBV-specific CD8+ memory T cells between granzyme K and Eomes expression, which indeed cluster together in the present analysis^{5}. Finally, when matching the proteins contained in this cluster to the STRING database, ITGB1 (integrin β-1 or CD29), an integrin that both mediates cell-cell adhesion and serves as a co-stimulatory receptor for T cells^{17,18}, was found to be strongly integrated in the network of functional associations with various other proteins (Figure 3B).

Taken together, this cluster contains molecules that function in all antigen-experienced CD8+ T cells and points at the development of several functional, but also metabolic traits that are not acting in naïve CD8+ T cells.
The granzyme A/granzyme M cluster

The GZMA and GZMM cluster (cluster 2), contains 64 proteins that were expressed in a gradually increasing manner from the naïve to the T_{EM}^{RA} subset, suggestive of a progressive gain of functional specialization by the subsets (Figure 3A). Enrichment analysis revealed HEXB (hexosaminidase B), NAGA (N-acetylgalactosaminidase-α), GGH (Gamma-Glutamyl Hydrolase), GNS (glucosamine (N-Acetyl)-6-Sulfatase), GLB1 (galactosidase-β1), IDS (iduronate 2-Sulfatase), ACP2 (acid Phosphatase 2) and MAN2B1 (Mannosidase-α-2B1), proteins involved in processes such as ‘lysosome’ and ‘lytic vacuole’ (Supplemental Table V, Supplemental Figure 4). These processes were also enriched in cluster 3, but were then linked to other molecules, hinting at a stepwise maturation of vesicles during CD8^+ T cell development (Supplemental Table V). Concomitantly, the cells within this cluster also had expression of SERPINA1 (α-1 antitrypsin) and SERPINA3 (α-3 antitrypsin), which act as serine protease inhibitors and may thus function to counteract the effects of among others elastases and granzymes. This expression may be important for protecting T cells against hostile (inflammatory) environments and possibly also against their own acquired cytotoxic capability. Further enrichments entailed processes like ‘response to stimulus’, ‘response to wounding’, ‘defence response’ and ‘inflammatory response’. This originated from the expression of specific proteins like PRF1 (perforin) and KLRG1 (killer cell lectin-like receptor G1), known to function as a cytotoxic pore-forming effector molecule and an inhibitory receptor, respectively (Supplemental Table V)\(^{19,20}\). Another protein in this cluster was ITGAL (CD11a), which together with CD18 forms the LFA-1 complex that mediates firm adhesion to antigen-presenting cells and the subsequent formation of an immunological synapse \(^{21,22}\). This may be especially relevant when considering the co-expression of CD11a with perforin and EZR (ezrin), and TLN1 (talin 1) in this cluster, proteins that have been linked to ‘cytolysis’ and ‘cytoskeletal anchoring to plasma membrane’ processes, respectively (Supplemental Table V).

As such, not only do these cells specialize towards the ability to form an immunological synapse with possible target cells, they also appear to guard themselves against possible hostile inflammatory environments.

The granzyme B/granzyme H cluster

The degree of (cytotoxic) specialization intensifies in the GZMB and GZMH cluster, which comprises 44 other proteins that were expressed predominantly by the T_{EM}^{4} and T_{EM}^{RA} subsets (Figure 3A). Not only do granzyme B and granzyme H cluster together, these two granzymes are also very similar to each other with respect to amino acid sequence \(^{23}\). Whereas not much is known about the functions of granzyme H, granzyme B is considered to be the main apoptosis-inducing molecule employed by CD8^+ T cells and NK cells to kill pathological cells. TBX21 (the transcription factor T-bet) is known to directly induce the expression of granzyme B in murine T cells, and likely also in human T cells, and was located within the same cluster (Supplemental Table II and Supplemental Figure 5)\(^{5,24-26}\). T-bet protein expression accorded to the high T-bet mRNA levels found in hCMV-specific CD8^+ T cells, which indeed also abundantly express granzyme B \(^{10,27}\).
Enrichment analysis of the proteins in this cluster emphasized the cytotoxic effector nature of the T_{EM}^3 and T_{EM}^4 and the T_{EM} RA cells, which was reflected by the association with several processes involved in ‘adhesion’, indicating an acquisition of functions with respect to better allowing for cell-cell interactions (Supplemental Table V). Important players in these processes comprised CX3CR1 (CX3CR1), ICAM3 (intercellular adhesion molecule 3), and ITGAM (integrin-α-M), molecules that may mediate interaction with target cells through binding of fractalkine, LFA-1, or molecular patterns found on bacterial surface, respectively \(^{28}\) (Supplemental Table V, Supplemental Figure 5). Interestingly, biological processes like ‘response to bacterium’ and ‘cellular defence response’ were also enriched and these were mainly related to the presence of GRNLY (granulysin) (Supplemental Table V, Supplemental Figure 5). Key players identified through STRING analysis, apart from GZMB, were the tyrosine kinase LYN and PLCG2 (an enzyme of the phospholipase C family), both key modulators of intracellular signalling cascades (Figure 3B, Supplemental Figure 5).

In conclusion, the cells that are expressing this cytolytic set of proteins have developed into immediate killers, that despite being quiescent cells are capable not only of interacting with target cells but also have the ability to instantaneously secrete highly cytotoxic substances during this encounter.

The protein kinase C cluster

The fourth cluster holds 70 proteins of which the expression decreased from the T_N to the T_{EM} RA differentiation state (Figure 3A). These proteins may therefore be important in regulating the cellular processes necessary to acquire the specific effector functions of the T_{EM} and the T_{EM} RA populations. Recent findings have emphasized the importance of metabolic changes in CD8\(^+\) T cells to obtain specific effector or memory functions \(^{29-31}\). Interestingly, ADK (adenosine kinase), PAICS (phosphoribosylaminomimidazole carboxylase), CAD (carbamoyl-Phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) and PFAS (phosphoribosylformylglycinamidine synthase), all proteins involved in nucleotide biosynthesis, were found within this cluster (Supplemental Figure 6).

A central protein in this cluster was PRKCA (protein kinase C), a protein that is involved in various intracellular signalling cascades. Other processes were associated to the antioxidant enzymes GPX1 (gluthathione peroxidase 1) and PRDX2 (peroxiredoxin-2) and EPHX2 (epoxide hydrolase 2), which are known to act as antioxidants, while also conferring protection against different harmful substances (Supplemental Table V, Supplemental Figure 6). This may be particularly interesting when considering that these proteins are mainly expressed by subsets that are known for their stem-cell-like traits, amongst which a greater resistance to cell death \(^{32-34}\). STRING analysis showed that protein kinase C was pivotal in the known functional interactions within this cluster (Figure 3B). Other focal proteins in this cluster comprised ACTN1 (α-actinin-1), which is important for cell-cell interactions, and also ALDOC (fructose-bisphosphate aldolase) and LDHB (lactate dehydrogenase B) that are both implicated in glycolysis and gluconeogenesis (Figure 3B, Supplemental Figure 6).
Taken together, protein kinase C is functionally associated with many proteins involved in nucleotide, xenobiotic and glucose metabolism, the expression levels of which decrease in a linear fashion from the $T_N$ to the $T_{EM RA}$ subsets.

Non-linear transitions in protein expression

Interestingly, whereas clusters 1 to 4 show that a large proportion of the variation in protein expression by different CD8$^+$ T cell subsets concerns gradual, rather than sudden changes, the proteins found in clusters 5 and 6 are expressed only by particular subsets (Figure 3A). Cluster 5 is the smallest cluster, comprising only 11 proteins that are expressed mainly by the $T_{CM}$, $T_{EM 1}$ and $T_{EM 2}$ populations (Supplemental Figure 7). Cluster 6 (53 proteins) comprises proteins of which the expression levels markedly drop in the $T_{EM 2}$, $T_{EM 3}$ and $T_{EM 4}$ subsets. None of the proteins in cluster 5 could be associated to discrete physiological processes based on enrichment analysis. Nevertheless, DPP4 (CD26) and SLAMF1 (signalling lymphocytic activation molecule 1) are well known as co-stimulatory signalling molecules modulating T cell receptor (TCR)-mediated activation $^{35,36}$, the expression of which appears to be downregulated by all the CD28$^-$ subsets (Supplemental Figure 7).

Notably, cluster 6 was enriched for the ‘(mitochondrial) proton transporting complex’ and ‘purine nucleoside metabolic’ processes (Supplemental Table 4). Both processes revolve around the same proteins, being ATP5EP2/ATP5E and ATP5D. The ‘dipping’ expression also included proteins involved in processes such as ‘generation of precursor metabolites’ and ‘oxidative phosphorylation’, pointing at a significantly altered metabolic state of these three $T_{EM}$ populations (Supplemental Table V). The central proteins in this cluster, as revealed by STRING analysis, were the transcription factors HRNPD (heterogeneous nuclear ribonucleoprotein D0) and NFKB1 (NF-$\kappa$B1) (Figure 3B and Supplemental Figure 8).

In contrast to the other clusters, which all contain proteins of which the expression increases or decreases along the subset spectrum in a seemingly linear fashion, clusters 5 and 6 hold specific TCR-modulating proteins, as well as proteins involved in the mitochondrial electron transfer chain. Instead, the expression of these proteins is lost rather abruptly only from specific subsets in what appears to be a non-linear manner.

Validation of label-free proteomic analysis

We first aimed to validate the MS results by using multichannel flowcytometry. Overall we found that the expression levels of the majority of the both cell surface as well as intracellular molecules determined by flowcytometry closely resemble those identified by MS (Figure 4A,B). For instance, a close resemblance was found for the expression levels of CD26, which displayed the rather abrupt drop in expression after the $T_{EM 2}$ state using both techniques (Figure 4B) and the distinct granzymes followed the patterns identified by MS (Figure 4B).

Still, it is important to note that not all the cell surface proteins can be quantified in a reliable manner by the MS approach. For instance, we were unable to retrace the levels of CCR7 in all the subsets (Supplemental Figure 9). The reason for this is not known. Possibly, the antibodies used to stain these molecules somehow hinder the peptide (and
Figure 3. Distinct CD8+ T cell protein expression patterns correspond to functional specialization. A. Heat map and hierarchical clustering (with k-means) of the 286 significantly different proteins revealed distinct expression patterns of the proteins from the TN to the TEMRA subsets. The 6 clusters discriminate between upregulated (clusters 1-3), downregulated (cluster 4) and non-linear regulated proteins (clusters 5 and 6). Heat map colors are based on the Z-scored (log2) LFQ values. Heat map built using hierarchical clustering revealed six distinct expression patterns of the proteins from the TN to the TEMRA subsets. Green shades correspond to a decreased expression level and red shades to an increased expression level. B. To visualize the known physical and functional interactions between the differentially expressed proteins, interaction network analysis of the significantly different genes was performed using STRING (version 10). Active interaction sources Experiments, Databases, Neighborhood and minimum required interaction score: 0.400 were used. The identified network was uploaded into Cytoscape (version 3.2.1). STRING analysis was done to visualize the known physical and functional interactions between the differentially expressed proteins. Node color corresponds to cluster in which these proteins were located while node size corresponds to the degree of edges.

As a second validation, we compared our data set with recently published data by Böttcher et al. In this study, instead of using the expression of CD45RA, CCR7, CD28 and CD27 to define and isolate functional CD8+ T cell subsets in humans, the expression...
of CD62L and CX3CR1, was employed. As CX3CR1 is mainly expressed on T\textsubscript{EM} and T\textsubscript{EM RA} cells \textsuperscript{28}, CX3CR1\textsuperscript{−} cells would be closest to the T\textsubscript{CM}, and T\textsubscript{EM1} and T\textsubscript{EM2} subsets, whereas cells expressing CX3CR1 would approximate the other T\textsubscript{EM} and the T\textsubscript{EM RA} subsets. Because these data were analysed in a manner identical to our approach, we could test how these data would compare and whether the combined data set would support our initial conclusions.

Unsupervised hierarchical clustering of the combined data showed that the proteome of the CD62L\textsuperscript{−}CX3CR1\textsuperscript{−} populations was placed in between the T\textsubscript{N} and T\textsubscript{CM} and the T\textsubscript{EM1} populations, suggesting that this concerns a different and possibly even earlier stage in CD8\textsuperscript{+} T cell differentiation than was indicated by the T\textsubscript{CM} subset proteome (Figure 5A). CD62L\textsuperscript{−}CX3CR1\textsuperscript{−} cells clustered together with the T\textsubscript{CM}, T\textsubscript{EM1} and the T\textsubscript{EM2} populations, whereas CD62L\textsuperscript{−}CX3CR1\textsuperscript{+} populations clustered together with the T\textsubscript{EM4} and the T\textsubscript{EM RA} subsets (Figure 5A).

Also when the combined data was abstracted into a principal component analysis, we found that the CD62L\textsuperscript{−}CX3CR1\textsuperscript{−} population was indeed entirely distinct from the other CD8\textsuperscript{+} T cell subsets defined in the current study (Figure 5B). Otherwise, the T\textsubscript{N} populations from either study clustered adequately together, which was also the case for the T\textsubscript{EM1} and CD62L\textsuperscript{−}CX3CR1\textsuperscript{−} T\textsubscript{EM} subsets (Figure 5B). Finally, the CD62L\textsuperscript{−}CX3CR1\textsuperscript{+} populations also clustered strongly with the T\textsubscript{EM RA} subset.

When comparing the differentially expressed proteins identified in the current study with those found in the study by Böttcher et al., the transitions in the proteome were also largely gradual. However, also in that study, non-linear proteomic transitions were seen, particularly concerning the CD62L\textsuperscript{−}CX3CR1\textsuperscript{−} population and the T\textsubscript{EM2}, T\textsubscript{EM3} and the T\textsubscript{EM4} subsets (Figure 5B).
Figure 5. Validation through integration with CD62L/CX3CR1-defined subset proteomes. A. Integration of the proteomes using the database of Böttcher et al. Heatmap and hierarchical clustering (with k-means) of 549 proteins that were identified in both datasets and significantly different in either of the datasets using the built-in analysis-of-variance function in PERSEUS (FDR of 5% and S0 of 0.4). Heat map colors are based on the combined Z-scored (log2) LFQ values. Validation of the proteome using the database of Böttcher et al. Heat map built using unsupervised hierarchical clustering. Green shades correspond to a decreased expression level and red shades to an increased expression level.

B. Principle component analysis was performed to further assess how the CD45RA/CCR7/CD28/CD27-defined subset proteomes in the current study correspond to the CD62L/CX3CR1-defined subset proteomes from Böttcher et al. Symbol and colour corresponds to the respective subsets.
Together, this combined analysis supports the notion that changes in the proteomes of CD8+ T cell populations belonging to distinct stages of differentiation occur in a stepwise, gradual manner.

Distinct metabolic regulation of the different CD8+ T cell subsets

As already mentioned above, metabolic changes have been identified as critical processes in the development of specific effector and memory functions of CD8+ T cells. For this reason, we explored how the proteins identified by MS, and known to be involved in regulation of metabolism, are expressed by different subsets. Unsupervised hierarchical clustering revealed three distinct clusters of proteins that were either expressed in a gradually increasing manner from the T_N to the T_EM RA state (metabolism (M) cluster 1), proteins that showed a sudden drop in expression in three T_EM (2-4) subsets (cluster M2), or proteins that were expressed in a gradually decreasing manner from the T_N to the T_EM RA subsets (cluster M3) (Figure 6).

Cluster M1 contains proteins such as ACLY (ATP citrate lyase) and ACAA2 (acetyl-CoA acyltransferase 2) that are involved in the cytosolic generation of acetyl-CoA, the biosynthesis of fatty acids and in fatty acid oxidation (FAO). It also holds various enzymes that are acting in lysosomes to degrade various types of molecules, like HEXB (β-hexosaminidase), GNS (glucosamine (N-acetyl)-6-sulfatase), NAGA (N-Acetylgalactosaminidase-α) and GLB1 (β-galactosidase) (Figure 6). Remarkably, cluster M2 specifically contains proteins involved in the electron transport chain across the mitochondrial inner membrane like ATP5EP2/ATP5E (ATP synthase, H+ transporting, mitochondrial F1 complex-ε), ATP5D (ATP synthase, H+ transporting, mitochondrial F1 complex-δ) ENO1 (enolase 1), NDUFA7 and NDUFB10 (NADH:ubiquinone reductase). Cluster M3 comprises many molecules already mentioned in the protein kinase C cluster. Proteins like ALDOC (aldolase C), FBP1 (Fructose-1,6-bisphosphatase 1) and LDHB (lactate dehydrogenase B) are involved in glycolysis and gluconeogenesis. Furthermore, this cluster also holds various molecules involved in nucleotide biosynthesis like AMPD3 (adenosine monophosphate deaminase 3), ADK (adenosine kinase), CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase), PAICS (phosphoribosylaminomimidazole carboxylase) and PFAS (phosphoribosylformylglycinamidase synthase).

DNA and RNA synthesis would indeed be essential to the vigorous proliferation of cells seen during the acute phase of the T cell response, a functional trait that is specifically displayed by T_N cells during primary infection and T_CM and T_EM cells during antigenic recall. T_EM RA cells have higher activation thresholds and also do not proliferate as vigorously once activated. These data also suggest that antigen-experienced cells rely less on glycolysis and instead move towards fatty acids as an energy source. Recently, Pearce et al. reported that murine CD8+ memory T cells indeed rely on FAO and oxidative phosphorylation as their energy source. They also showed that the substrates for FAO in memory cells appear to derive from cell-intrinsic lipolysis occurring via an, as of yet, poorly understood mechanism.

In conclusion, these data show that CD8+ T cell subsets alter their metabolic programs in a largely gradual fashion. Interestingly, a selection of T_EM subsets also displays a non-
Figure 6. The CD8+ T cell metabolome. Heat map and built using unsupervised hierarchical clustering (with k-means) of revealed three distinct expression patterns of proteins involved in metabolism (metabolome based on RECON1 database supplemented with the genes involved in the main metabolism pathways (TCA cycle, Glycolysis, Lipid digestion, mobilization and transport, Pentose phosphate, Electron transport and Fatty acid biosynthesis) revealed three distinct expression patterns.). Heat map colors are based on the Z-scored (log2) LFQ values. Green shades correspond to a decreased expression level and red shades to an increased expression level.
linear drop in the way they appear to utilize mitochondrial respiratory burst capacity. Finally, we regard it highly likely that these metabolic changes are essential to the distinct functional profiles of CD8+ T cell subsets found in humans.

**Distinct expression patterns of adhesion-related proteins per human CD8+ T cell subset**

Recently, Basu et al. demonstrated how mechanical force exerted locally at the immunological synapse significantly potentiates the cytotoxic capacity of the T cells through a coordinated tension-augmented pore forming mechanism by perforin. These findings corroborated an essential role for the cytoskeleton and cell adhesion molecules in CD8+ T cell killing of target cells, other than just the formation of an immunological synapse through which serine proteases are delivered into target cells. For this reason we here used the protein database to identify differences between CD8+ T cell subsets in the expression of proteins involved in cell-cell adhesion, also called the ‘adhesome’. As a reference we used the meta-adhesome database generated by Horton et al., which included a list of 2412 proteins identified in six different proteome studies on integrin adhesion complexes (IAC) formed by different human and murine cell types. In addition, we added several proteins that were previously described to have specific roles in CD8+ T cell adhesion but which were not identified in the meta-analysis from Horton et al., probably because their analyses did not contain immune cells.

The CD8+ T cell adhesome comprised 117 proteins that were expressed at significantly different levels between different CD8+ T cell subsets. The adhesome clustered the subsets in the same manner as the complete proteome (not shown). The 117 proteins could be divided into fourteen categories, four of which were quite sizeable (actin regulators, adaptors, metabolism and RNA/DNA regulation) and ten much smaller categories (Figure 7). When regarding proteins either as adhesome-intrinsic (structural) versus adhesome-associated (regulatory), we find that among the structural proteins not only the previously discussed ITGB1 and ICAM3, but also LGALS1 and LGALS3 (Lectin, Galactoside-Binding, Soluble, 1 and 3, respectively) were present. Interestingly, LGALS1-mediated signalling was recently also shown to induce a regulatory T cell phenotype in a murine infection model whereas LGALS3 seems to do the opposite. Both are being expressed in an increasing manner from the T_N to the T_EM RA state, although LGALS3 is expressed earlier in this spectrum than LGALS1 (Figure 7).

The CD8+ T cell adhesome comprises a large group of proteins closely involved in regulating actin. Alpha actinins 1 and 4 are known to bind actin strands to the cellular membrane in non-muscle cells, and have also been implicated in the formation of the immunological synapse. ACTN1 (alpha actinin 1) was expressed in a decreasing manner from the T_N to the T_EM RA subset, whereas ACTN4 (alpha actinin 4) was expressed in an increasing manner along this subset spectrum, indicating that specific changes are made to the cytoskeleton during differentiation. Together with the before-mentioned TLN1, but also with VCL (vinculin), alpha-actinins interact with intracellular domains of integrins like ITGB, thus not only tethering the cytoskeleton to the membrane but also to sites of cell-cell adhesion. In turn, actin itself interacts with myosins to mediate cellular...
A large part of the proteins in the CD8⁺ T cell adhesome are involved in regulating the many aspects of cell-cell adhesion rather than being part of the adhesion apparatus themselves. One major category concerns RNA/DNA regulation, comprising many proteins that are located in cluster 6 where the expression levels dip in T_{EM}², T_{EM}³ and T_{EM}⁴. Curiously, these are almost all involved in mRNA transport, mRNA turnover and pre-mRNA processing (Figure 7). Another large category concerns proteins involved in cellular metabolism, most of which have been discussed already in the previous paragraph (Figure 7). The adaptor category holds proteins that mediate intracellular signalling, which mainly increase from the T_N to the T_{EM}⁴ state. In this category, AHNAK (Desmoyokin) was shown to be important to be specifically important to CD8⁺ T cell-mediated cytotoxicity by mediating intracellular calcium fluxes (Figure 7) ⁵⁴. Finally, GTPases and guanine exchange factors are essential components acting in the signalling cascades regulating actin dynamics. For most of these proteins the specific roles in CD8⁺ T cells remain to be elucidated as of yet.

Together, these data show how the adhesome is made up of various structural and associated proteins of which the expression is regulated in a seemingly ordered manner over the different CD8⁺ T cell subsets detected in the human circulation.

**Concluding Remarks**

We here show that label-free MS approaches can be used to depict changes that occur in circulating CD8⁺ T cells as a consequence of activation and differentiation processes after encountering viruses. It is important to note that the cells that we have analysed are quiescent and therefore it is not surprising that the vast majority of the proteome is shared between all subsets, including the naïve cells. Our analyses reveal that most of the changes within these seven circulating CD8⁺ subsets seem to occur in a gradual fashion. This is in particular true when the major subsets are being considered: T_N, T_{CM}.
Figure 7. The CD8+ T cell adhesome. Interaction network analysis Protein-protein interaction network of 117 proteins involved in CD8+ T cell migration and adhesion (based on meta-adhesome supplemented with T cell specific adhesion molecules) using . STRING (version 10). Active interaction sources Textmining, Experiments, Databases, Co-expression, Neighborhood, Gene Fusion, Co-occurrence and minimum required interaction score 0.150 were used. The identified network was uploaded into Cytoscape (version 3.2.1). Node color corresponds to cluster in which these proteins were located.

\( \text{T}_{\text{EM}1} \) and \( \text{T}_{\text{EM}RA} \) (effector-type) populations. This gradual change would in our opinion be in support of a progressive differentiation model where persistent or repetitive stimulation induce a further acquisition of effector functions while at the same time limiting clonogenic potential. This paradigm has been derived from experimental studies in inbred mice using only a limited set of viruses. It is noteworthy that the unbiased proteomic approach in humans, being exposed for variable times to a plethora of viruses, supports this hypothesis. On the other hand, it is clear that when focusing on minor subsets, alterations in the proteome occur that are non-linear. These particular subsets and the protein they express certainly warrant further experimentation. However, because of their low frequency it is a challenge to obtain enough material for in depth analysis. Still, as the majority of their protein content follows the pattern of the large subsets we feel that also from these minor pools proteomes have been reliably generated. Although resting, it is unclear if these low frequent populations contain transitional cells, i.e. cells migrating between tissues, or whether they use (like naïve, central memory and effector-type cells)
the circulation, and not specific tissues, as a major site of residency. Although we did not find enrichment for tissue residence markers (CD69, CD103) on these cells it cannot be excluded that these cells are passers-by. The technical innovation in acquiring reliable proteomes from subsets of immune cells opens the opportunity to further scrutinize the regulation of differentiation. A question that has not been answered in our study is the relation between transcriptomes and proteomes, which is a highly relevant issue when considering immediate versus induced functions of previously primed T cells. Further, it will be highly relevant to take this type of analyses beyond the blood to determine whether T cells in tissues fall within same gradual type of differentiation pattern, or rather if (long term) tissue residency impacts the expressed proteome in a specific manner.

MATERIALS AND METHODS

Subjects

Buffycoats from 8 different healthy asymptomatic adults were obtained from the local bloodbank. PBMCs were separated using lymphoprep.

Immunofluorescence staining, sorting

CD8 T cells were enriched using the MagniSort™ Human CD8 T cell Enrichment Kit (negative selection, eBioscience) according to manufacturing conditions and stored frozen until the day of sorting. The enriched T cells were subsequently thawed and stained in RPMI 5% FCS with the following antibodies: CD3 BV510, CCR7 BV421, CD45RA FITC, CD27 APC-eFluor 780, CD8 APC, CD28 PE. From the CD3+CD8+lymphocyte gate the following subsets were sorted on a 6-way influx sorter (BD). The gating strategy as well as a representative example of the immunofluorescent staining can be found in supplemental figure 1. Some subsets could not be obtained from all donors owing to a natural variation in which these subsets are present in healthy adults 5: T N isolates were obtained from 5 different donors; T CM isolates were obtained from 6 different donors; T EM1 isolates were obtained from 4 different donors, 2 of which were pooled together to obtain sufficient cells shortly before protein isolation; T EM2 isolates were obtained from 6 different donors, 4 of which were pooled together to obtain sufficient cells shortly before protein isolation; T EM3 isolates were obtained from 6 different donors, 4 and 2 of which, respectively, were pooled together to obtain sufficient cells shortly before protein isolation; T EM4 isolates were obtained from 6 different donors; 5 of which were pooled together to obtain sufficient cells shortly before protein isolation; T EM RA isolates were obtained from 5 different donors, one of which was analysed in duplicate (Supplemental Table I). After sorting cells were washed 3 times with PBS after which cell pellets were frozen in liquid nitrogen and stored at -80 °C until sample processing.

Immunofluorescence staining, analysis

PBMCs were washed in phosphate-buffered saline containing 0.01% (wt/vol) NaN3 and 0.5% (wt/vol) bovine serum albumin. Two million PBMCs were incubated with a combination
of the following antibodies: CD3 BV510, CCR7 BV421 (Biolegend, San Diego, CA, USA), CD45RA FITC (BD Biosciences, San Jose, CA, USA), CD27 APC-eFluor 780, CD8 APC, CD28 PE or CD28 FITC supplemented with either CD2 PE, CD26 PE, CD150 PE, CD74 PE, CD11a FITC, KLRG1 FITC, GPR56 PE, CD58 PE. The FOX-p3 staining kit (eBioscience) was used for intracellular stainings with each of the following antibodies: PDK1 PE, FOXP1 PE (BD Biosciences), granzyme B PE, granzyme A PE, perforin PE (BD Pharmingen, San Diego, CA, USA), Eomesodermin PerCP-eFluor 710, granzyme K PerCP-eFluor 710 (eBioscience Inc, San Diego, CA, USA), granulysin PE, T-Bet PE and CD74 PE (BioLegend). Measurements were done on a LSR Fortessa flow cytometer (BD) and analysis was performed with FlowJo software (FlowJo, Ashland, OR, USA)).

Sample preparation for mass spectrometry analysis
Cell pellets (1-2 ×10^6) were lysed in 30 μl 4% SDS, 100 mM DTT, 100 mM Tris.HCl pH 7.5 supplemented with HALT protease and phosphatase inhibitor cocktail (Thermo Scientific) and processed into tryptic peptides using the Filter Aided Sample Preparation method 56. Peptides were desalted and concentrated using Empore-C18 StageTips 57 and eluted with 0.5% (v/v) acetic acid, 80 % (v/v) acetonitrile. Sample volume was reduced by SpeedVac and supplemented with 2% acetonitrile, 0.1% TFA to a final volume of 12 μl. For each sample, 3 technical replicates were analysed by injecting 3 μl of the sample.

Mass spectrometry data acquisition
Tryptic peptides were separated by nanoscale C18 reverse chromatography coupled on line to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) via a nanoelectrospray ion source (Nanospray Flex Ion Source, Thermo Scientific). Peptides were loaded on a 20 cm 75–360 μm inner-out diameter fused silica emitter (New Objective) packed in-house with ReproSil-Pur C18-AQ, 1.9 μm resin (Dr Maisch GmbH). The column was installed on a Dionex Ultimate3000 RSLC nanoSystem (Thermo Scientific) using a MicroTee union formatted for 360 μm outer diameter columns (IDEX) and a liquid junction. The spray voltage was set to 2.15 kV. Buffer A was composed of 0.5 % acetic acid and buffer B of 0.5 % acetic acid, 80% acetonitrile. Peptides were loaded for 17 min at 300 nl/min at 5% buffer B, equilibrated for 5 minutes at 5% buffer B (17-22 min) and eluted by increasing buffer B from 5-15% (22-87 min) and 15-38% (87-147 min), followed by a 10 minute wash to 90 % and a 5 min regeneration to 5%. Survey scans of peptide precursors from 400 to 1500 m/z were performed at 120K resolution (at 200 m/z) with a 1.5 × 105 ion count target. Tandem mass spectrometry was performed by isolation with the quadrupole with isolation window 1.6, HCD fragmentation with normalized collision energy of 30, and rapid scan mass spectrometry analysis in the ion trap. The MS2 ion count target was set to 104 and the max injection time was 35 ms. Only those precursors with charge state 2–7 were sampled for MS2. The dynamic exclusion duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 3 s cycles. All data were acquired with Xcalibur software.
Mass spectrometry data analysis

The RAW mass spectrometry files were processed with the MaxQuant computational platform, 1.5.2.8. Proteins and peptides were identified using the Andromeda search engine by querying the human Uniprot database (downloaded February 2015). Standard settings with the additional options match between runs, Label Free Quantification (LFQ), and unique peptides for quantification were selected. The generated ‘proteingroups. txt’ table was filtered for potential contaminants, reverse hits and ‘only identified by site’ using Perseus 1.5.1.6. The LFQ values were transformed in log2 scale, the three technical replicates per experimental condition grouped and averaged based on the median, and proteins were filtered for at least two valid values in one of the experimental groups. Missing values were imputed by normal distribution (width=0.3, shift = 1.8), assuming these proteins were close to the detection limit. Quantitative significance (Pearson correlation coefficient, Principle Component Analysis and ANOVA) was performed by Perseus software.

To identify the proteins with the most prominent differences expression profiles within the different CD8+ T cell subsets, we used the built-in analysis-of-variance functions in PERSEUS using an FDR of 5% and S0 of 0.4 (the S0 parameter sets a threshold for minimum fold change). This identified a total of 286 differentially expressed proteins. Enrichment analysis of the significantly different genes was performed using the Cytoscape (version 3.2.1) plug-in BiNGO (version 3.0.3) with an FDR threshold of 0.05. Interaction network analysis of the significantly different genes was performed using STRING (version 10) using the parameters Active prediction method: Experiments, Databases, Neighbourhood and Score: 0.400. The identified network was uploaded into Cytoscape (version 3.2.1). For metabolic annotations we used the human metabolic network reconstruction RECON 1 supplemented with the genes involved in the main metabolism pathways (TCA cycle, Glycolysis, Lipid digestion, mobilization and transport, Pentose phosphate, Electron transport and Fatty acid biosynthesis). For adhesome annotations, we used the meta-adhesome database generated by Horton et al., which included a list of 2412 proteins identified in six different proteome studies on integrin adhesion complexes (IAC) formed by different human and murine cell types. Adhesome categories are literature based. Adhesome proteins included in the current analysis that were not identified by Horton et al. comprise: MYO1F, MYO1G, VAV3, COTL1, CAPG, PSTIP1, CX3CR1, ITGAL, KLRG1, GPR56, ITGAM, FGR, BIN2, OSBPL3, CHN2 and ICAM3. For comparison with Böttcher et al. we downloaded Supplementary Data 3. We combined Z-scored log2 LFQ values of 549 proteins that were identified in both datasets and significantly different in either of the datasets using the built-in analysis-of-variance functions in PERSEUS (FDR of 5% and S0 of 0.4) as described above.

ACKNOWLEDGMENTS

We would like to thank Berend Hooibrink for his help with the cell sorter to isolate the CD8+ T cell subsets.
FUNDING

This research funded in part by the Dutch Kidney Foundation (14OKG05, Kolff Physician Researcher Grant to M. van Aalderen).
REFERENCES


22. Hogg N, Patzak I, Willenbrock F. The insider's guide to leukocyte integrin


SUPPLEMENTARY TABLES AND FIGURES

Supplemental Table I. High resolution MS measurements identified a total of 6114 proteins.  
File too large to include in this dissertation. Please see http://www.ebi.ac.uk/pride. Project  
Name: Label-free analysis of CD8+ T cell subset proteomes supports a progressive  
File can be requested via m.c.vanaalderen@amc.uva.nl

Supplemental Table II. Identification of 4816 proteins in at least 2 samples in one of the subsets.  
File too large to include in this dissertation. Please see http://www.ebi.ac.uk/pride. Project  
Name: Label-free analysis of CD8+ T cell subset proteomes supports a progressive  
File can be requested via m.c.vanaalderen@amc.uva.nl

Supplemental Table III. Absolute copy number values of the quantified proteins.  
File too large to include in this dissertation. Will be published online (to follow). Until that  
moment, file can be requested via m.c.vanaalderen@amc.uva.nl.

Supplemental Table IV. Enrichments of differentially-expressed proteins for specific biological processes.  
File too large to include in this dissertation. Will be published online (to follow). Until that  
moment, file can be requested via m.c.vanaalderen@amc.uva.nl.

Supplemental Table V. Enrichments of proteins for specific biological processes per cluster.  
File too large to include in this dissertation. Will be published online (to follow). Until that  
moment, file can be requested via m.c.vanaalderen@amc.uva.nl.
Supplemental Figure 1. Gating strategy used for cell-sorting. Dot plots showing the gating strategy used to isolate the seven different CD45RA/CCR7/CD28/CD27-defined subsets used for the analyses by cell-sorting.

Supplemental Figure 2. Protein expression range and abundance per CD8+ T cell subset. A. Median protein copy number (estimated by the proteomic ruler methodology (juiste referentie) and log transformed) Proteins expression range per CD8+ T cell subset. B. Protein rank by abundance plotted versus cumulative protein abundance Cumulative protein abundance per CD8+ T cell subset . Proteins show a wide range of expression spanning several orders of magnitude and similar abundance distribution for all CD8+ T cell subsets.
Supplemental Figure 3. Protein expression levels in cluster 1. Selection of protein expression levels (log2 LFQ intensity) from cluster 1 as determined by mass spectrometry.

Supplemental Figure 4. Protein expression levels in cluster 2. Selection of protein expression levels (log2 LFQ intensity) from cluster 2 as determined by mass spectrometry.
Supplemental Figure 5. Protein expression levels in cluster 3. Selection of protein expression levels (log2 LFQ intensity) from cluster 3 as determined by mass spectrometry.
Supplemental Figure 6. Protein expression levels in cluster 4. Selection of protein expression levels (log2 LFQ intensity) from cluster 4 as determined by mass spectrometry.

Supplemental Figure 7. Protein expression levels in cluster 5. Selection of protein expression levels (log2 LFQ intensity) from cluster 5 as determined by mass spectrometry.
Supplemental Figure 8. Protein expression levels in cluster 6. Selection of protein expression levels (log2 LFQ intensity) from cluster 6 as determined by mass spectrometry.
Supplemental Figure 9. Impaired detection of some molecules by MS. A. Protein expression levels (log2 LFQ intensity) from different subunits of the CD3 and CD8 molecule, the CD45 protein (which is expressed by CD8+ T cells in the different isoforms R0 and RA that were not specifically detected by MS), CD27, CD28 and CCR7 (the latter two, which were partially or not picked up by MS). B. Diagram showing the amino acid sequence of PTPRC (CD45), which can be differentially spliced to form different isotypes (shown in red) as well as the peptides picked up by MS in the current study (shown in blue). C. Peptide intensity values of CCR7-derived peptides Detection of peptides that form CCR7 by MS in the current study per CD8+ T cell subset.