An ID card for T cells

Ning Jiang, Shuqi Zhang & Keyue Ma

Simultaneous measurement of paired TCR sequences and gene expression in single T cells improves the characterization of T cells.

The T-cell repertoire in the immune system of a healthy adult includes tens of millions of unique cells[1]. Understanding the identity of each cell remains an outstanding problem in immunology. Traditionally, a T cell has been defined either by the particular T-cell antigen receptor (TCR) it expresses or by its functional phenotype, such as naive, memory, helper, regulatory and effector T cell, each of which is identified by a certain combination of three to five phenotypic markers. However, no methods have been available to combine both types of analysis. In this issue, Han *et al.*[2] present an approach to simultaneously determine the paired TCRα and TCRβ sequences and the expression of selected phenotypic genes in single cells (**Fig. 1**). The resulting 'ID card' for every cell allows tracking of T-cell lineage and functional state, which should aid the development of T cell–based diagnostics and therapeutics.

The diversity of the T-cell repertoire is generated through recombination of the V(D)J genes encoding the TCR α- and β-chains[3]. Sequencing of these genes separately in a large number of T cells has provided useful information, but until recently it has not been possible to determine pairs of TCR α- and β-chains from single cells. Now, approaches based on emulsion PCR allow paired sequencing of TCR α/β chains on a large scale[4]. However, additional work is still needed to increase amplification efficiency, especially for the detection of rare clones. At the single-cell level, the TCR sequencing studies performed so far were at a much smaller scale and achieved only about 50% efficiency in determining paired TCR sequence data.

Technologies for characterizing T-cell phenotypes have also improved, enabling increased throughput and simultaneous measurement of larger numbers of markers. Most notably, mass cytometry can measure at least 45 (and potentially up to 100) surface and intracellular markers per cell with high throughput[5]. Microfluidics-based qPCR systems can determine expression levels of up to 96 genes in 96 single cells[6]. An imaging-based fluorescent

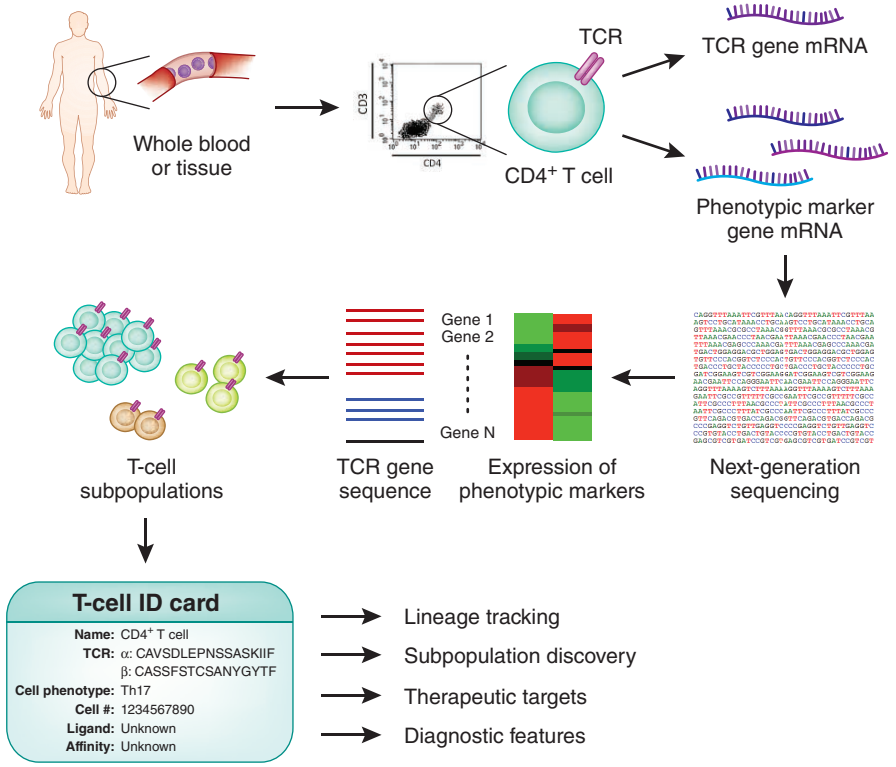

Figure 1 The method of Han *et al.*[2] determines paired TCR α- and β-sequences and phenotypic-gene expression in single cells. T cells are sorted by FACS into single cells. Gene-specific primers and nested PCR are used to amplify the genes encoding the TCR α- and β-chains and a set of phenotypic markers. TCR gene sequences and the presence or absence of phenotypic markers are revealed by next-generation sequencing. Combining these data with the results of single-cell functional assays, such as assays of antigen specificity and affinity, yields an ID card bearing all the information for each T cell. This correlated information can be used to track T-cell lineage, discover subpopulations, and screen for therapeutic targets and diagnostic features.

Ning Jiang is in the Department of Biomedical Engineering, Cockrell School of Engineering, University of Texas at Austin, Austin, Texas, USA, and the Institute for Cellular and Molecular Biology, College of Natural Sciences, University of Texas at Austin, Austin, Texas, USA; Shuqi Zhang is in the McKetta Department of Chemical Engineering, Cockrell School of Engineering, University of Texas at Austin, Austin, Texas, USA; and Keyue Ma is at the Institute for Cellular and Molecular Biology, College of Natural Sciences, University of Texas at Austin, Austin, Texas, USA.
e-mail: jiang@austin.utexas.edu

hybridization technology allows simultaneous measurement of up to 800 genes in 12 cells per run[6]. However, none of these approaches provides TCR sequence information.

Efforts to understand the role of T cells in health and disease will greatly benefit from both TCR sequence and phenotypic information for each T-cell clone. When a T cell is activated by an antigen, it divides and proliferates. The daughter cells all share the same TCR with their parent. Thus, the TCR sequence profile is important for identifying antigen specificities and the proliferation history of a TCR lineage. Phenotypic information is crucial for defining the effects of T cells on disease initiation and development, which has diagnostic and therapeutic value.

Current approaches for obtaining both TCR sequence and phenotype information are inadequate because of limits on the numbers of markers that can be monitored. For example, single-cell TCR sequencing coupled with fluorescence-activated cell sorting (FACS) is limited to ~13 markers owing to fluorescence spectral overlapping. As shown in B cells, a larger number of markers can be analyzed by using microfluidic chip–based qPCR in combination with Sanger sequencing to determine immune receptor sequences[7]. However, this method requires the use of specialized devices. In addition, Sanger sequencing is limited in its ability to resolve the second productive light (in B-cell) or α (in T-cell) chains.

The method developed by Han *et al.*[2] requires only common laboratory equipment and access to a next-generation sequencer, streamlining the correlated measurement of TCR sequences and phenotypic markers in thousands of cells. Using a nested PCR approach, the authors obtained the expression level for a set of 17 phenotypic markers along with TCR α- and β-chain sequences in over 80% of the cells in a population, the highest value reported so far[2]. Although their method is not designed to quantify transcript abundance, the presence or absence of a given transcript matched nicely with the protein secretion measured using FACS for a few cytokines.

Having validated their method in peripheral blood mononuclear cells and a T-cell line, the authors went on to study tumor-infiltrating lymphocytes. $CD4^+$ T cells were isolated from the colon tissue of a patient with colorectal cancer. Significant clonal expansion of T cells bearing the same TCR was observed in cells isolated from tumor tissue compared to those obtained from adjacent normal tissue. Interestingly, even within populations with identical TCRs, substantial heterogeneity in the expression of the 17 marker genes was observed, underscoring the importance of single-cell analysis[2].

In colorectal carcinoma, there is considerable interest in the therapeutic and prognostic value of T helper 17 (Th17) cells because these cells have an inflammatory anti-microbial role that promotes cancer growth[8]. However, little is known about the lineage of tumor-infiltrating Th17 cells. By comparing the TCR sequences shared between different subpopulations of these cells, Han *et al.*[2] found that $FOXP3^+RORC^+IL17^+$ T cells and $FOXP3^-RORC^+IL17^+$ T cells share a common ancestor. This suggests that the landscape of tumor-infiltrating lymphocytes is very complicated when both T-cell lineage and phenotype are considered, and that determining the origin of Th17 cells might reveal avenues for new therapies. Future research should use a larger patient cohort at different stages of tumor progression.

T-cell biology holds many questions to be addressed in the future. What antigens do T cells bind to? What are the affinities of polyclonal TCRs, and how does TCR affinity affect T-cell development and differentiation? What is the regulatory network that controls T-cell development and differentiation? T cells are implicated in many diseases such as cancers, infections and autoimmune diseases. Possessing correlated information on the TCR, its ligands and the phenotype of the T cell will likely be important in the diagnosis, treatment and prevention of a wide range of conditions. As high-throughput technologies for the analysis of antigen binding[9], single-cell transcriptomes[10] and other cellular attributes become available, the T-cell ID card will only become more informative.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Arstila, T.P. *et al. Science* **286**, 958–961 (1999).
2. Han, A., Glanville, J., Hansmann, L. & Davis, M.M. *Nat. Biotechnol.* **32**, 684–692 (2014).
3. Davis, M.M. & Bjorkman, P.J. *Nature* **334**, 395–402 (1988).
4. Turchaninova, M.A. *et al. Eur. J. Immunol.* **43**, 2507–2515 (2013).
5. Bendall, S.C. & Nolan, G.P. *Nat. Biotechnol.* **30**, 639–647 (2012).
6. Yosef, N. *et al. Nature* **496**, 461–468 (2013).
7. Weinstein, J.A., Zeng, X., Chien, Y.H. & Quake, S.R. *PLoS ONE* **8**, e67624 (2013).
8. Ladoire, S., Martin, F. & Ghiringhelli, F. *Cancer Immunol. Immunother.* **60**, 909–918 (2011).
9. Birnbaum, M.E. *et al. Cell* **157**, 1073–1087 (2014).
10. Shapiro, E., Biezuner, T. & Linnarsson, S. *Nat. Rev. Genet.* **14**, 618–630 (2013).

A genealogy of the citrus family

Riccardo Velasco & Concetta Licciardello

Clarification of the genetic relationships among species opens new possibilities for enhancing citrus diversity and disease resistance.

The family relationships among citrus fruits—which include sweet and sour oranges, mandarins, clementines, tangerines, grapefruits, pummelos, kumquats, lemons and limes—are famously obscure. Thousands of years of citrus cultivation and interbreeding have yielded some 25 species and at least 250 commercial varieties. Following on the recently published draft genome of sweet orange[1], Wu *et al.*[2] report in this issue the genome sequences of clementines, mandarins, pummelos, sweet oranges and sour oranges. By combining these data with some elegant sleuth work, they also make important progress in elucidating the phylogenetic history of citrus domestication. Given the dependence of the $9 billion/year global citrus industry on large-scale monoculture, the findings provide much-needed opportunities for *de novo* breeding of citrus varieties with increased resistance to pathogens and environmental extremes. Breeding guided by genomic information will also provide a greater diversity of fruit for consumers in terms of new combinations of taste, aroma, size, color, shape and ease of peeling.

Wu *et al.*[2] began prudently by choosing to assemble the 301-Mb genome of a clementine with a haploid genome to sidestep the complications of assembling outbred diploid genomes[3]. Whereas the 367-Mb sweet-orange genome, assembled using Illumina paired-end tags, was only 76% complete[1], the haploid genome of the clementine assembled using Sanger sequencing is 96% complete. The clementine genome has an L50 value more than twice that of the sweet-orange genome (119 versus 49.9 Kb)

Riccardo Velasco is at Fondazione Edmund Mach, di San Michele a/Adige, Trento, Italy. Concetta Licciardello is at the Consiglio di Ricerca e Sperimentazione in Agricoltura, CRA-ACM, Acireale, Italy.
e-mail: riccardo.velasco@fmach.it