



## Data in Brief

# Genome-wide RNA profiling of long-lasting stem cell-like memory CD8 T cells induced by Yellow Fever vaccination in humans



Silvia A. Fuertes Marraco<sup>a,b,1</sup>, Charlotte Sonesson<sup>c,1,2</sup>, Mauro Delorenzi<sup>a,b,c</sup>, Daniel E. Speiser<sup>a,b,\*</sup>

<sup>a</sup> Ludwig Cancer Center, University of Lausanne, Switzerland

<sup>b</sup> Department of Oncology, Lausanne University Hospital (CHUV), Switzerland

<sup>c</sup> Bioinformatics Core Facility, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

## ARTICLE INFO

## Article history:

Received 11 June 2015

Accepted 19 June 2015

Available online 2 July 2015

## ABSTRACT

The live-attenuated Yellow Fever (YF) vaccine YF-17D induces a broad and polyfunctional CD8 T cell response in humans. Recently, we identified a population of stem cell-like memory CD8 T cells induced by YF-17D that persists at stable frequency for at least 25 years after vaccination. The YF-17D is thus a model system of human CD8 T cell biology that furthermore allows to track and study long-lasting and antigen-specific human memory CD8 T cells. Here, we describe in detail the sample characteristics and preparation of a microarray dataset acquired for genome-wide gene expression profiling of long-lasting YF-specific stem cell-like memory CD8 T cells, compared to the reference CD8 T cell differentiation subsets from total CD8 T cells. We also describe the quality controls, annotations and exploratory analyses of the dataset. The microarray data is available from the Gene Expression Omnibus (GEO) public repository with accession number GSE65804.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications

Organism/cell line/tissue	Homo sapiens
Sex	Male or female (Table 1)
Array type	Agilent Whole Human Genome Microarray 8x60K (v2; one color, Cy3) – Agilent Technologies
Data format	Raw and processed (normalized expression levels)
Experimental factors	RNA samples originate from CD8 T cell populations (Table 2) isolated from cryopreserved peripheral blood mononuclear cells (PBMC). RNA was extracted and amplified by Miltenyi Biotec SuperAmp™ technology.
Experimental features	CD8 T cell populations were purified by flow cytometry from 8 different healthy donors vaccinated with Yellow Fever 17D (donors d1 to d8, Table 1); the populations (“cell types”, Table 2) include: A2/NS4b tetramer positive CCR7 + CD45RA + CD8 T cells (“A2_NS4b Naïve-like”), “Total Naïve” (CCR7 + CD45RA +), “Total Tscm” (Stem-cell like memory T: CCR7 + CD45RA + CD58 + CD95 +), “Total CM” (Central Memory: CCR7 + CD45RA-) and “Total Effector” (CCR7-).

(continued)

Specifications	
Consent	Samples were obtained from participants under written informed consent following study protocol approval by the Human Research Ethics Committee of the Canton de Vaud (Switzerland).
Sample source location	Lausanne, Switzerland

## 1. Direct link to deposited data

<http://www.dtd.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65804>

## 2. Experimental design, materials and methods

## 2.1. Experimental design

The microarray experiment detailed here was performed in the framework of our recent study on the characterization of CD8 + stem cell-like memory T (CD8 Tscm) cells identified in Yellow Fever YF-17D vaccinees that persist for decades following vaccination [1]. The design of the microarray analysis was aimed to characterize the genome-wide RNA expression profiles of this population of long-lasting YF-specific CD8 Tscm. The cells of interest were isolated from 8 different vaccinees, spanning up to 17 years of vaccination history, male or female, as detailed in Table 1. YF-specific CD8 Tscm were analyzed and isolated from peripheral blood of the eight vaccinees based on the

\* Corresponding author at: Clinical Tumor Biology & Immunotherapy Group, Ludwig Cancer Center, University of Lausanne, Biopôle 3, Chemin des Boveresses 155, CH-1066 Lausanne, Switzerland.

E-mail address: [doc@dspeiser.ch](mailto:doc@dspeiser.ch) (D.E. Speiser).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Current address: University of Zurich, Switzerland and SIB Swiss Institute of Bioinformatics, Switzerland.

**Table 1**  
Donors (YF-17D vaccinees) from whom the 5 different T cell populations were isolated.

Donor	Gender	Age	Years since YF-17D vaccination	Isolation (sort) day
d1	Female	25.4	2.4	I
d2	Female	25.9	4.1	II
d3	Female	26.8	7.2	I
d4	Male	33.6	7.9	I
d5	Female	30.7	8.7	II
d6	Female	23.6	12.1	I
d7	Male	41.4	12.7	II
d8	Male	26.6	17.8	II

**Table 2**  
T cell populations (“cell types”) isolated from each donor for microarray analysis. All populations were sorted from CD8 T-enriched samples by gating within live (Vivid aqua-) CD8+ CD16- cells after doublet exclusion.

Sample title in GEO database	T cell population	Markers used for flow cytometry sorting
A2_NS4b_ naive/like	YF-17D-specific CD8 T cells, naïve-like (Tscm)	A2/NS4b + CCR7 + CD45RA +
total_naive	Naïve in total CD8 T	CCR7 + CD45RA +
total_Tscm	Tscm in total CD8 T	CCR7 + CD45RA + CD95 + CD58 +
total_cm	CM in total CD8 T	CCR7 + CD45RA-
total_effector	Effectors (EM + EMRA) in total CD8 T	CCR7-

detection of T cells bearing a TCR specific for the HLA-A\*02-restricted epitope NS4b<sup>214–222</sup> (LLWNGPMAV) of YF-17D [1,2], within the conventional Naïve gate: for the microarray, they are referred to as “A2/NS4b Naïve-like”. As reference populations, various differentiation subsets were also isolated from the total pool of circulating CD8 T cells, including: total Naïve, total CM, and total effectors (CCR7- cells, i.e. a mixture of effector memory (EM) and effector memory CD45RA+ (EMRA)). We also isolated the more recently described total population of CD8 Tscm [3,4]. The nomenclature and markers used to isolate these five populations are listed in Table 2.

## 2.2. Sample collection, preparation & cryopreservation

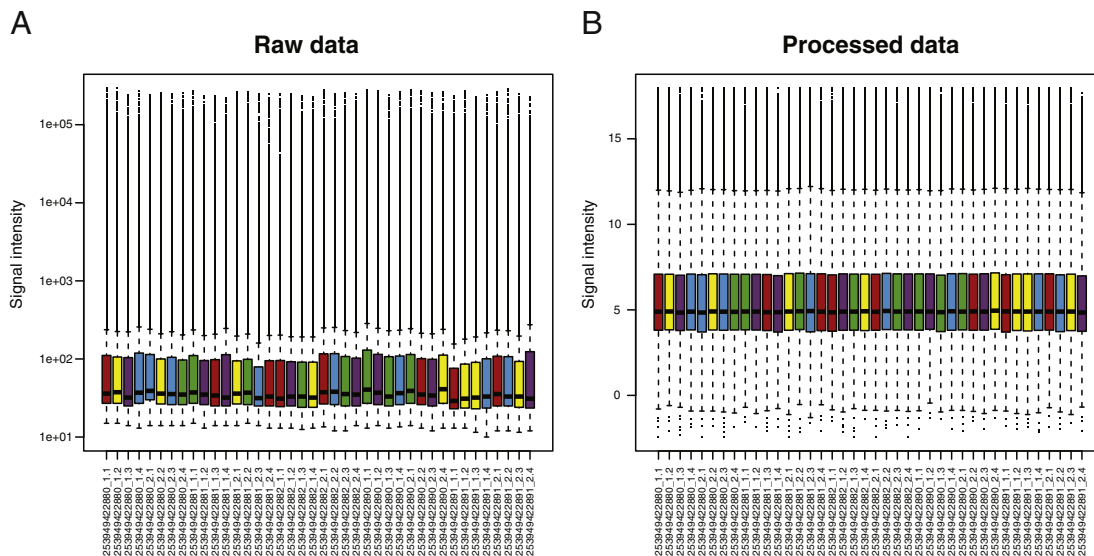
Healthy volunteers vaccinated with YF-17D donated circulating leukocytes according to the standards of the Blood Transfusion Center in Epalinges, Switzerland (Service Vaudois de Transfusion Sanguine). Donations were performed either via leukapheresis (all donors except d2) or a standard complete blood donation where the leukocyte-rich fraction (buffy coat) was recovered for the study (d2 only). Peripheral blood mononuclear cells (PBMC) were separated by density gradient fractionation using Lymphoprep™. All samples were immediately cryopreserved in RPMI 1640 supplemented with 40% FCS and 10% DMSO.

## 2.3. Isolation of CD8 T cell populations

Freshly thawed PBMC were first enriched for CD8 T cells using the human CD8 T cell enrichment kit from StemCell™ Technologies (negative selection, per manufacturer's instructions). For flow cytometry-based isolation of CD8 T cell populations, stainings were performed using PBS with 5 mM EDTA, 0.2% BSA and 0.2% Na-azide (FACS buffer) and all steps were carried out at 4 °C. Enriched CD8 T cells were first stained with a PE-conjugated HLA-A\*02/NS4b<sup>214–222</sup> tetramer (TCmetrix Sàrl.) for 40 min, followed by surface staining for 30 min using CD8-Alexa700 (clone HIT8a, Biolegend), CD45RA-ECD (clone 2H4LDH11LDB9, Beckman Coulter), CCR7-BrilliantViolet421 (clone G043H7, Biolegend), CD16-KromeOrange (clone 3G8, Beckman Coulter), CD58-FITC (clone 1C3, BD Biosciences), CD95-PECy7 (clone DX2, Biolegend) and live/dead cell-marker Vivid-Aqua (Molecular Probes®, Invitrogen) for 30 min. The flow cytometry gating and sorting strategy is described in [1]. From each population of CD8 T cells (Table 2) and each donor (Table 1), 1000 cells were isolated by flow cytometry-based purification using a BD FACSAria I flow cytometer. Purifications were performed on two separate days, with samples from donors d1, d3, d4 and d6 processed in Sort I, and samples from donors d2, d5, d7 and d8 processed in Sort II.

## 2.4. RNA extraction, amplification (SuperAmp miltenyi biotech) and hybridization onto microarray platform for data collection

RNA extraction and amplification were performed using the SuperAmp™ technology from Miltenyi Biotec. Per sample, 1000

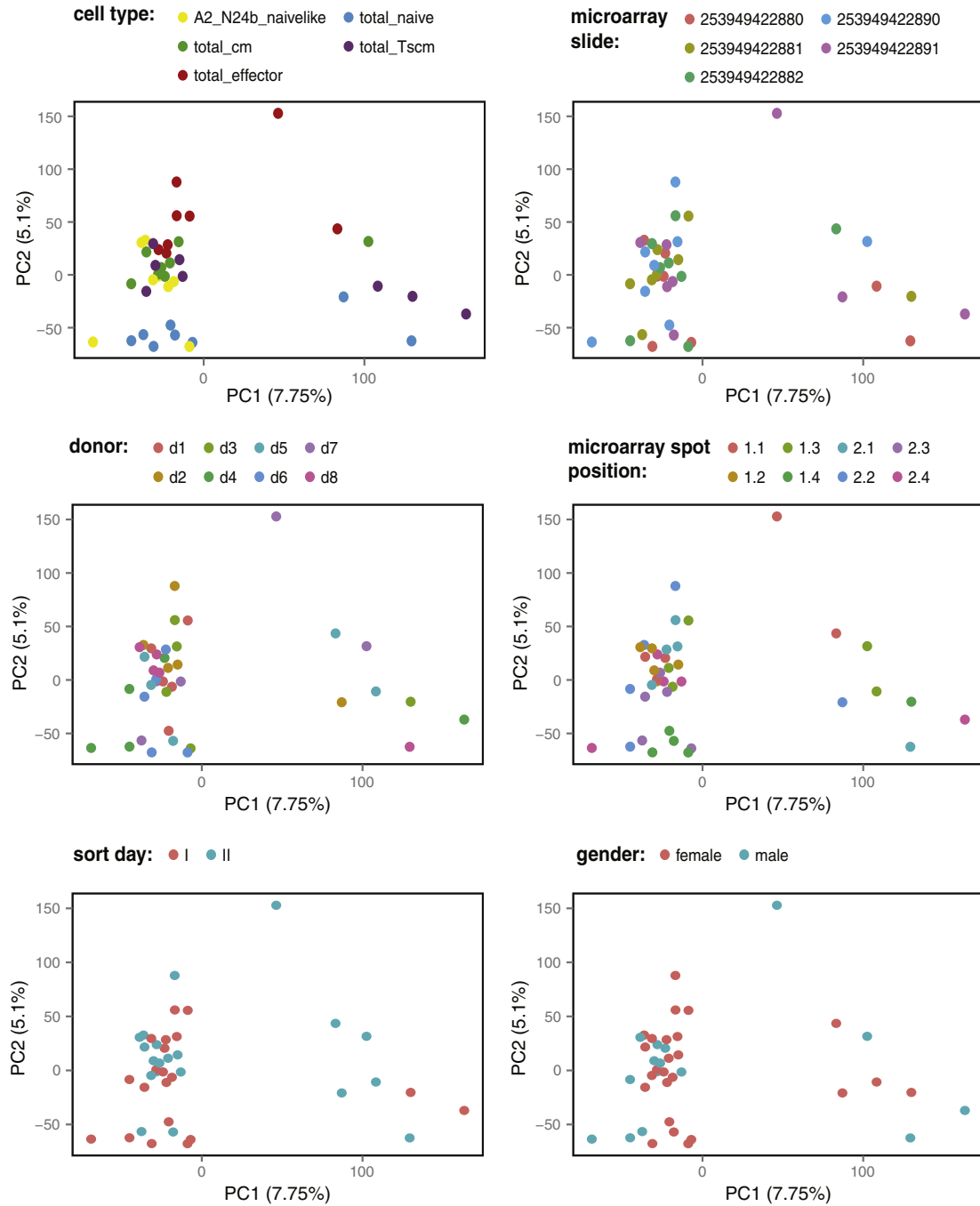


**Fig. 1.** Signal intensity distribution (raw and processed data). A. Signal intensity distribution of raw data from all microarray probes, vertically aligned per sample. B. Signal intensity distribution after the processing steps described in the text (background subtraction, quantile normalization, filtering, summarization and exclusion of control probes), vertically aligned per sample. The samples are colored by the “cell type” annotation.

cells were purified directly into 13.5  $\mu$ l of SuperAmp™ lysis buffer (Miltenyi Biotec). RNA lysates were incubated for 10 min at 45 °C and immediately frozen at –20 °C. Thereafter, frozen RNA lysates were shipped to and processed and analyzed by the Genomic Services of Miltenyi Biotec. The SuperAmp™ technology performs in-column immunomagnetic-based mRNA extraction using paramagnetic oligo(dT) MicroBeads, with bead-bound 1st strand cDNA synthesis and 3'-end tailing for subsequent elution and single-primer global PCR amplification and Klenow labeling. The cDNA samples were controlled for integrity by analysis on Agilent 2100 Bioanalyzer. Samples were hybridized onto Agilent

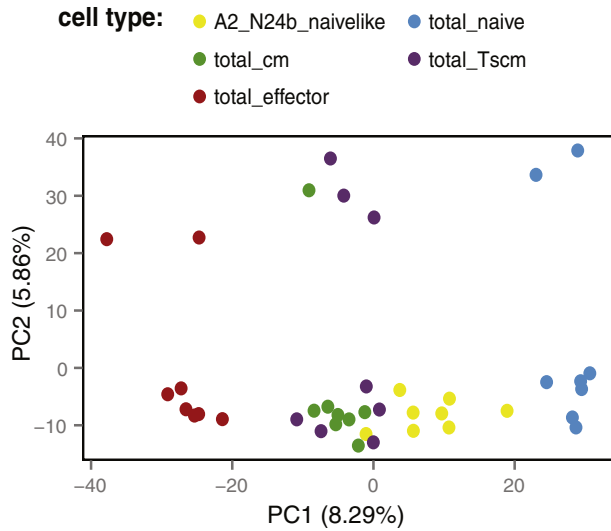
Whole Human Genome Microarray 8x60K (V2, one color: Cy3) and fluorescence signals were detected using Agilent's Microarray Scanner System (Agilent Technologies). The image files were processed with the Agilent Feature Extraction Software to yield probe intensities. The 40 samples (8 donors  $\times$  5 cell populations) were randomly distributed on a total of 5 microarray slides with 8 spots per slide. In the files uploaded into the GEO repository (GSE65804), the name of the raw data file contains the 12-digit code of the microarray slide (e.g. 253949422880) as well as the spot position of the sample within that slide (1.1–1.4 and 2.1–2.4). Supplementary Material 2 provides a sample annotation table,

PCA with “all genes”



**Fig. 2.** Principal component analysis based on all genes: no detectable influence of observed technical or experimental parameters. Two-dimensional representations of PC1 versus PC2 for all samples colored as indicated according to “Day of cell sorting”, “microarray slide”, “microarray spot”, “donor” (ordered by increasing years since vaccination), “gender” of the donor and “cell type”.

## PCA with “top 10% most variable genes”



**Fig. 3.** PCA based on the 10% most variable probes: differentiation gradient in PC1 according to cell type. Two-dimensional representation of PC1 versus PC2 for all samples colored according to “cell type”. Additional colorings, showing no association between the first two principal components and the other observed technical and experimental parameters, are available in Supplementary Material 1.

relating the raw data file names to the name and characteristics of the donor (d1 to d8, Table 1) and the “cell type” (Table 2).

### 3. Microarray data: quality controls and exploratory analysis

#### 3.1. Data pre-processing

The raw data from the Agilent Feature Extraction software was processed and analyzed in R, using mainly functions from the limma package. The code that was used to generate the figures in this manuscript, starting from the raw data deposited in GEO, is available in Supplementary Material 1. Supplementary Materials 2 and 3 contain the sample annotation tables that are used in the computational analyses, with ‘sample info’ and ‘targets’ specified, respectively.

**Table 3**

Probes (and corresponding genes) with largest weight on the second principal component in PCA based on the 1% probes with highest variance. This component discriminates between male and female donors, and is dominated by probes measuring the expression of genes located on the sex chromosomes. Probes with positive weight are highly expressed in the female donors, and probes with negative weight are highly expressed in the male donors.

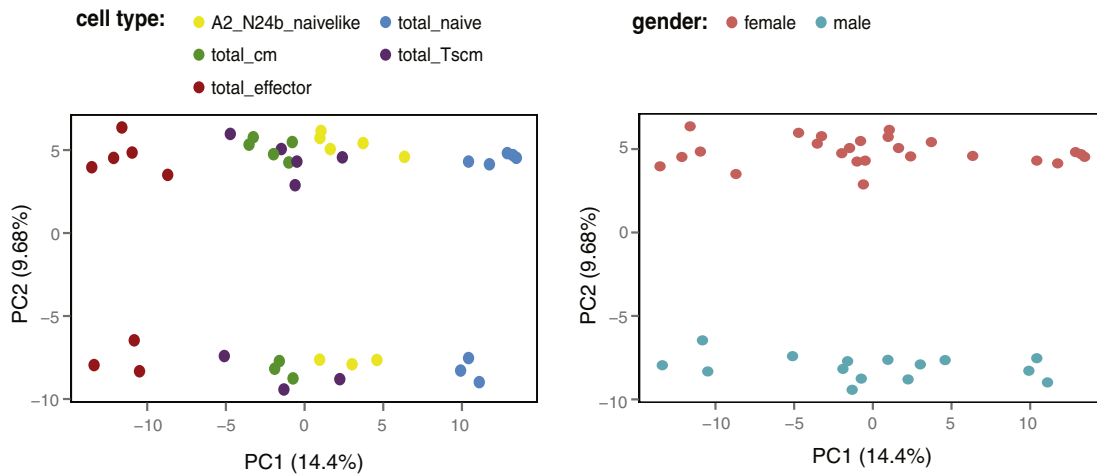
PC2, positive weight	PC2, negative weight
A_33_P3405911 (TSIX, chrX)	A_23_P364792 (TXLNG2P, chrY)
A_19_P00316565 (TSIX, chrX)	A_21_P0006594 (TTY15, chrY)
A_19_P00327297 (XLOC_008015, chrX)	A_23_P324384 (RPS4Y2, chrY)
A_19_P00320438 (TSIX, chrX)	A_21_P0006606 (XLOC_008323, chrY)
A_19_P00326132 (TSIX, chrX)	A_23_P259314 (RPS4Y1, chrY)
A_19_P00321129 (TSIX, chrX)	A_33_P3725324 (USP9Y, chrY)
A_19_P00331623 (XIST, chrX)	A_23_P137238 (KDM5D, chrY)
A_19_P00321917 (TSIX, chrX)	A_23_P160004 (UTY, chrY)
A_19_P00806762 (TSIX, chrX)	A_33_P3261353 (BCORP1, chrY)
A_21_P0006456 (XLOC_008185, chrX)	A_21_P0006651 (XLOC_008386, chr10)

The raw data (Fig. 1 A) was background-corrected using the “normexp” algorithm implemented in limma, and thereafter normalized between the arrays using quantile normalization. All samples were considered to be of good quality. In addition to the expression values, the Agilent Feature Extraction software provides (among other things) flags indicating whether or not a given feature is detected above the background level. We used this information to filter out all probes that were not detected well above the background (defined as the estimated background level + 2.6 times the estimated standard deviation of the background) for any of the samples. This reduced the number of features from the original 62,976 to 50,477. Finally, we averaged the expression of replicated probes and filtered out control probes. This further reduced the number of analyzed features to 41,923. Mapping from probe names to gene identifiers was done using the information provided by the Agilent Feature Extraction software. Fig. 1 B displays the distribution of intensity values after the pre-processing.

#### 3.2. Exploratory analysis

We used Principal Component Analysis (PCA) to perform exploratory analysis in order to summarize and visualize the data, as well as to detect potential artifacts and effects of experimental and technical parameters. The panels of Fig. 2 show the same two-dimensional PCA sample representation (PC1 versus PC2) based on all 41,923 probes, colored

## PCA with “top 1% most variable genes”



**Fig. 4.** PCA based on the 1% most variable probes: differentiation gradient in PC1 according to cell type and gender differences in PC2. Two-dimensional representations of PC1 versus PC2 for all samples colored according to “cell type” (showing a differentiation gradient in PC1) and to “gender” (showing gender differences in PC2). Additional colorings, showing no association between the first two principal components and the other observed technical and experimental parameters, are available in Supplementary Material 1.

according to various sample attributes. No association with any of the recorded technical parameters can be seen, including: “the day of cell sorting” (sort I vs. sort II), the position of the samples depending on the “microarray slide” or “microarray spot”, the “donor” or the “gender” of the donor. The second principal component shows some discrimination between the cell populations (“cell type”, top-left panel of Fig. 2).

Since we expected the differences between the different T-cell populations to affect a relatively small set of genes, we further performed exploratory analysis after restricting the data set to the 10% probes with the highest variance across the 40 samples. Again, no clear impact of any of the recorded technical parameters was seen (Supplementary Material 1), but the first principal component clearly ordered the five cell populations along a gradient of differentiation (Fig. 3) – this observation is discussed in our recently published study [1]. After reducing the variable set further to keep only the top 1% probes with highest variance across the samples, the cell population gradient still dominated the first principal component. Considering all the other technical and experimental parameters enumerated above, the analysis showed that the second principal component distributed samples according to gender (Fig. 4). The highest weights in the second principal component were assigned to probes corresponding to genes located on the sex chromosomes (Table 3), several of which have been previously shown to be strongly associated with gender [5–7].

### Acknowledgments

We thank all blood donors for their participation, the personnel of the Blood Transfusion Center in Epalinges (Sophie Waldvogel, Christine Thibaud for coordinating the Leukaphereses and Jocelyne Conne for the laboratory processing), Danny Labes (Flow Cytometry Facility of the Ludwig Cancer Center) for cell sorting, and our laboratory colleagues Laurène Cagnon, Philippe O. Gannon, Mathilde Allard, Samia Abed Maillard, Nicole Montandon, and Nathalie Rufer for excellent collaboration.

**Funding:** Ludwig Cancer Research Center, the Cancer Vaccine Collaborative, the Cancer Research Institute (all N.Y., U.S.A), the Swiss Cancer League (02836–08–2011), the Swiss National Science Foundation (310030\_135553, 320030\_152856 and CRSII3\_141879), and

a grant of the Swiss State Secretariat for Education, Research and Innovation to the SIB for Service and infrastructure resources.

**Author Contributions:** SAFM and DES conceived and designed the experiments. SAFM performed the experiments. SAFM, CS, MD and DES analyzed the data (bioinformatic analyses: CS and MD). SAFM, CS, MD and DES wrote the paper. All authors revised and accepted the final version of the manuscript.

**Competing interests:** the authors declare that this study was conducted in the absence of any potential conflict of interest.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gdata.2015.06.024>.

### References

- [1] S.A. Fuertes Marraco, C. Soneson, L. Cagnon, P.O. Gannon, M. Allard, S.A. Maillard, et al., Long-lasting stem cell-like memory CD8+ T cells with a naïve-like profile upon yellow fever vaccination. *Sci. Transl. Med.* 7 (2015) 282ra48, <http://dx.doi.org/10.1126/scitranslmed.aaa3700>.
- [2] R.S. Akondy, N.D. Monson, J.D. Miller, S. Edupuganti, D. Teuwen, H. Wu, et al., The yellow fever virus vaccine induces a broad and polyfunctional human memory CD8+ T cell response. *J. Immunol.* 183 (2009) 7919–7930, <http://dx.doi.org/10.4049/jimmunol.0803903>.
- [3] L. Gattinoni, E. Lugli, Y. Ji, Z. Pos, C.M. Paulos, M.F. Quigley, et al., A human memory T cell subset with stem cell-like properties. *Nat. Med.* 17 (2011) 1290–1297, <http://dx.doi.org/10.1038/nm.2446>.
- [4] E. Lugli, L. Gattinoni, A. Roberto, D. Mavilio, D.A. Price, N.P. Restifo, et al., Identification, isolation and in vitro expansion of human and nonhuman primate T stem cell memory cells. *Nat. Protoc.* 8 (2012) 33–42, <http://dx.doi.org/10.1038/nprot.2012.143>.
- [5] F. Staedtler, N. Hartmann, M. Letzkus, S. Bongiovanni, A. Scherer, P. Marc, et al., Robust and tissue-independent gender-specific transcript biomarkers. *Biomarkers* 18 (2013) 436–445, <http://dx.doi.org/10.3109/1354750X.2013.811538>.
- [6] B. Payer, J.T. Lee, X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* 42 (2008) 733–772, <http://dx.doi.org/10.1146/annurev.genet.42.110807.091711>.
- [7] Y. Tang, A. Lu, R. Ran, B.J. Aronow, E.K. Schorry, R.J. Hopkin, et al., Human blood genomics: distinct profiles for gender, age and neurofibromatosis type 1. *Brain Res. Mol. Brain Res.* 132 (2004) 155–167, <http://dx.doi.org/10.1016/j.molbrainres.2003.10.014>.