Expression of type I interferon-associated genes at point of antiretroviral therapy interruption predicts time to HIV virological rebound

Authors: Panagiota Zacharapoulou[1*], Emanuele Marchi[1*], Ane Ogbe[1], Nicola Robinson[1], Helen Brown[1], Mathew Jones[1], Lucia Parolini[1], Matthew Pace[1], Nicholas Grayson[1,2], Pontiano Kaleebu[3], Helen Rees[4], Sarah Fidler[5,6], Philip Goulder[2], Paul Klenerman[1,7], John Frater[1,7†]

[1] Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] Department of Paediatrics, University of Oxford, Oxford, UK

[3] Medical Research Council/Uganda Virus Research Institute, Entebbe, Uganda

[4] Wits Reproductive Health and HIV Institute of the University of the Witwatersrand in Johannesburg, South Africa

[5] Division of Medicine, Wright Fleming Institute, Imperial College, London, UK

[6] Imperial College NIHR Biomedical Research Centre, London, UK

[7] National Institute of Health Research Biomedical Research Centre, Oxford, UK

[†]Correspondence to: John Frater; Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK. John.frater@ndm.ox.ac.uk

*These authors contributed equally.

Word Count: Abstract: 150 words; Text (inc references): 4565 words.

One Sentence Summary: Expression of genes associated with type I interferons predicts time to rebound of HIV viraemia after stopping antiretroviral therapy.

Abstract: Although certain individuals with HIV infection can stop antiretroviral therapy (ART) without evidence of viral load rebound, the mechanisms under-pinning 'post-treatment control' remain unclear. Twelve individuals who had received 12 months of ART from primary HIV infection and then undertook a TI were sampled at the time of stopping therapy. Using RNA-Seq we explored gene expression in CD4 T cells to look for evidence of a mechanism that might underpin virological rebound and lead to discovery of an associated biomarker. Using independent analysis tools, genes associated with the type I interferon response were strongly associated with a delayed time to viral rebound following TI. These are the first data we are aware of that link transcriptomic signatures associated with innate immunity with control following TI. While these results need to be confirmed in larger trials, they could help define a strategy for new therapies and identify new biomarkers for remission.

Introduction

Unless suppressed by antiretroviral therapy (ART), HIV infection leads to persistent viremia, progressive loss of CD4+ T cells and eventually AIDS. However ART is not a cure for HIV infection, requires lifelong adherence, and has been associated with side-effects, drug resistance and stigma(Peluso et al., 2019). Inherent to testing novel HIV cure strategies(Dash et al., 2019; Davenport et al., 2019; Mendoza et al., 2018) is analytical treatment interruption (TI) in which ART is stopped and any delay or prevention in viral rebound is examined. However there is still much controversy around TIs, and how to best implement and interpret them(Julg et al., 2019). Identifying biomarkers which predict outcomes following TI would provide enormous value to both understanding mechanisms of virological remission and identifying new drug candidates.

A rare proportion of ART-treated individuals are able to stop therapy and maintain undetectable viremia for months and, in some cases, years(Etemad et al., 2019; Martin et al., 2017; Sáez-Cirión et al., 2013). Multiple studies have been attempted to elucidate the biological mechanisms underlying these outcomes, and numerous clinical factors and molecular biomarkers which correlate with time to rebound have been proposed (Conway et al., 2019; Hurst et al., 2015; Krebs & Ananworanich, 2016; Martin et al., 2017; Sharaf et al., 2018; Stöhr et al., 2013; Williams et al., 2014). However, the host factors that affect the responsiveness of T cells have not been thoroughly characterised (Hyrcza et al., 2007), especially at a transcriptome level, possibly due to small cohort sizes and the availability of samples at key timepoints. Additionally, many gene expression studies have employed microarray technology(Lee et al., 2019; L.-L. Zhang et al., 2017) which may be associated with limitations compared with more recent RNA-Seq approaches(Zhao et al., 2014).

SPARTAC (Short Pulse Antiretroviral Treatment at HIV-1 Seroconversion)(Fidler et al., 2013) was one of the largest randomized clinical trials to study different durations of ART administered to participants with Primary HIV Infection (PHI) and the impact on disease progression. Previous analyses revealed that although the majority of participants experienced HIV viral rebound shortly after stopping ART (i.e <4 weeks post TI), some maintained undetectable viraemia to the end of the study (>500 days after TI). We studied participants who received 12 months of ART started in primary HIV infection (PHI), and analysed CD4+ T cell mRNA at the time of TI. Our aim was to identify genes or gene-sets expressed by CD4+ T cells that might associate with longer periods of remission after TI. Our results could be integrated into personalised diagnostic algorithms, to inform therapeutic strategies towards HIV cure and to better understand the mechanism of HIV latency and remission.

Results

Clustering of participants based on clinical response.

We identified 18 SPARTAC participants who had received 12 months of ART commenced during PHI, and who had viable samples of PBMCs taken at the point of TI. Clustering analysis of RNA expression revealed distinct groups based on ethnicity and gender (See Table 1 and Supplementary Figure S1), which might have contributed to the clinical outcomes. Therefore, to avoid potential confounding effects, we excluded six individuals resulting in a study group comprising twelve women, all from South Africa. 'Days to viral load rebound after TI' was used to define clinical phenotypes and structure the comparative analysis (Table 1).

4

Table 1: Participant Demographics

		Original Cohort		**Final Cohort**
Sex		Female (n=14)	Male (n=4)	**Female (n=12; 100%)**
Rebound	Early (<100 days to rebound)	5 (35.70%)	2 (50%)	**4 (33.3%)**
	Intermediate (>100 & <500days to rebound)	4 (28.50%)	2 (50%)	**3 (25%)**
	Late (> 500 days to rebound)	5 (35.70%)	0	**5 (41.6%)**
Country	Europe	1 (7.10%)	4 (100%)	**0**
	Africa	13 (92.80%)	0%	**12 (100%)**

Based on observations in SPARTAC and other cohorts(Colby et al., 2018, 2020; Etemad et al., 2019; Stöhr et al., 2013) we defined 'early' rebounders following TI (<100 days) as representing normal progressors, 'intermediates' (100-500 days) as representing a phenotype reflecting transient remission, and 'late' rebounders (>500 days) to be more characteristic of elite or 'sustained' controllers. As these cut-offs (i.e 100 and 500 days) have not been previously ratified - and there might be some overlap between them - we tested their validity by dividing the dataset based on these phenotypes in two different ways. First, we grouped the 'intermediates' and 'lates' together as Post-Treatment Controllers (PTC) (i.e all rebound >100 days) and compared these with the Progressors (<100 days). Next, we grouped the 'early' and 'intermediate' groups (i.e all rebound <500 days) into a group of non-sustained controllers (non-SC) and compared these with the sustained controllers (SC; >500 days). (Table S1). For this dataset, grouping the "intermediates" with either "early" or "late" rebounders helps identify genes associated with the more extreme phenotypes (i.e early or late rebounders) and improves the sample sizes for comparison, but at the potential risk of increasing variability and noise.

Differential Gene Expression and Gene Set Enrichment Analysis shows a strong association of type I interferon pathways with sustained control of viremia

Differential Gene Expression (DGE) analysis was performed to identify genes that were differentially expressed when comparing clinical groups. Genes with a reported adjusted p-value (padj) <0.05 and log(2) fold change >1 were considered as significantly differentially expressed. DGE was performed to compare the three clinical phenotypes: Late vs Early, Late vs Intermediate and Intermediate vs Early (Supplementary Figure S2c,d,e), and 1, 49 and 4 genes, respectively, were found to be differentially expressed in the three groups. It is likely that due to increased diversity of the signal in the extreme groups (i.e Early and Late), the small patient numbers and the stringency of the parameters, DGE was only able to return a few significant genes. When combining the patients into the larger groups, DGE identified four genes that were differentially expressed between non-SC versus SC (i.e rebound < vs > 500 days), and five genes between Progressors and PTCs (i.e rebound <vs > 100 days) (Supplementary Figure S2a,b). However, in view of these limitations we turned to alternative approaches.

We identified functionally linked 'gene sets' which can offer a more comprehensive insight than single genes into the mechanisms underpinning a phenotype. Gene Set Enrichment Analysis (GSEA) was performed on all genes ranked by DESeq2 Wald statistics (with the Wald test being applied to each gene), for each group of phenotypes. For the non-SC vs SC (i.e. rebound < vs > 500 days) analysis the gene set representing interferon signaling pathways was clearly enriched in the SC phenotype (Figure 1a). The PTC phenotype (rebound > 100 days) was associated with 51 gene sets and the most statistically significant enriched biological processes were related to immunoregulation, platelet activation and interferon alpha and beta signaling (Figure 1b). DGE and GSEA were also performed on the samples comparing the

6

distinct phenotypes (Early, Intermediate and Late), and also showed interferon alpha and beta signaling to be enriched in the late rebounders (Supplementary Figure S3), however, as noted above, due to small sample sizes we were not able to take these analyses further.

Figure 1: Pathway enrichment according to clinical phenotype using DESeq2 with GSEA and Rank Prod with ReactomePA
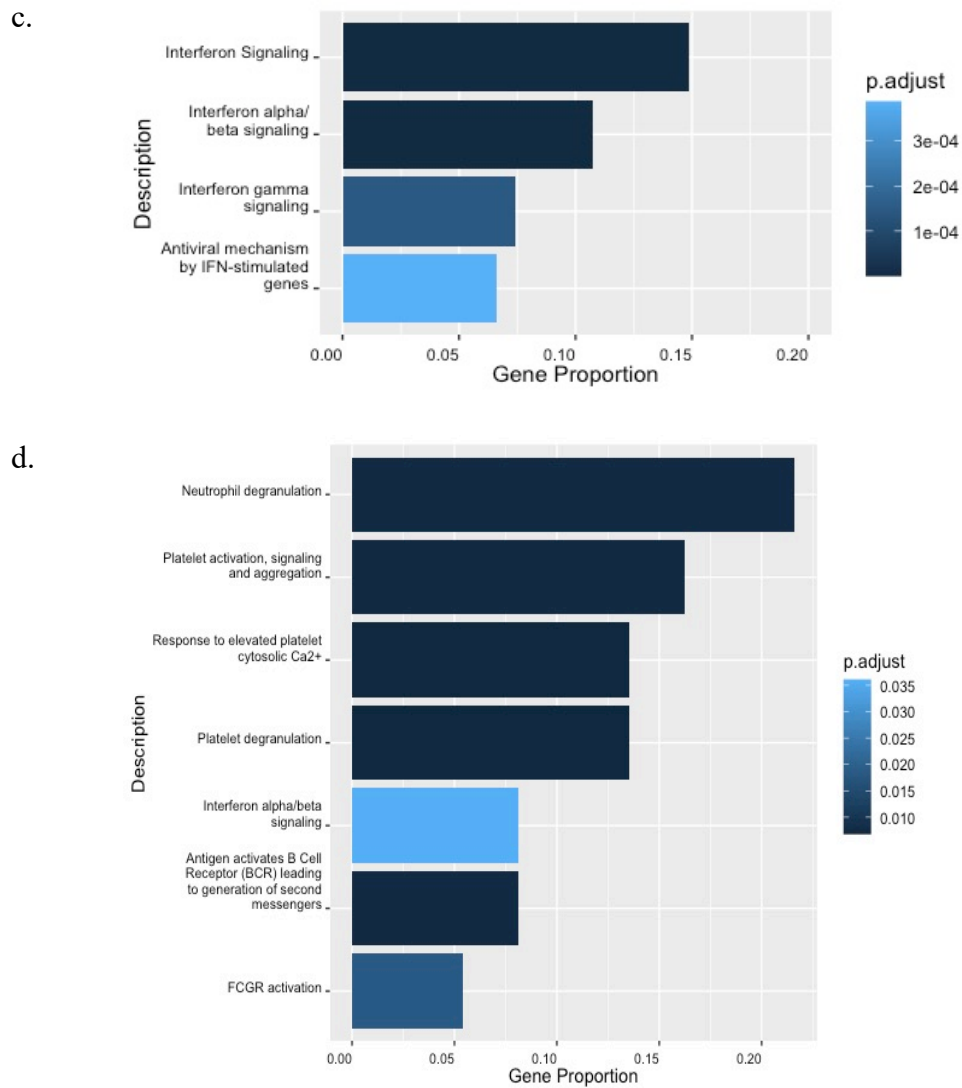
a.

b.

c.

d.
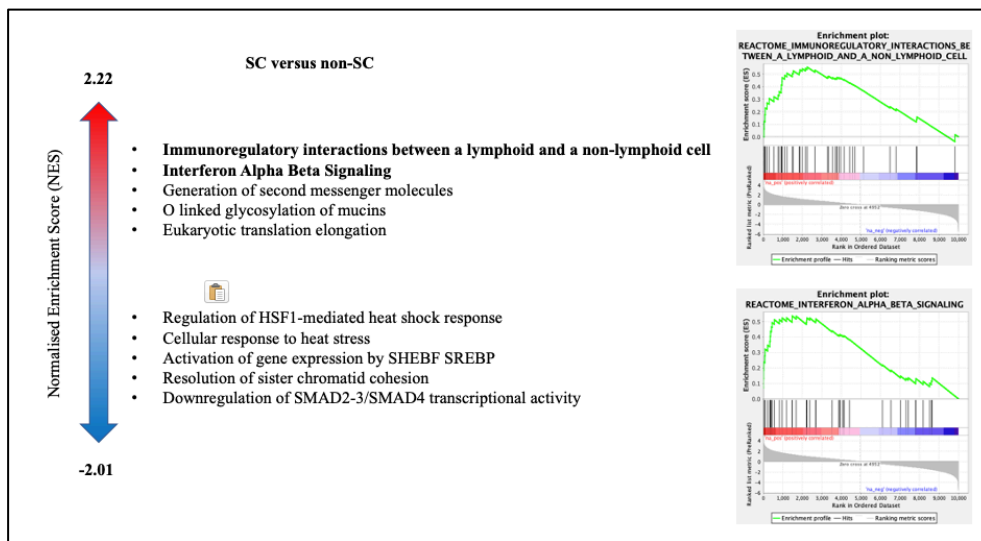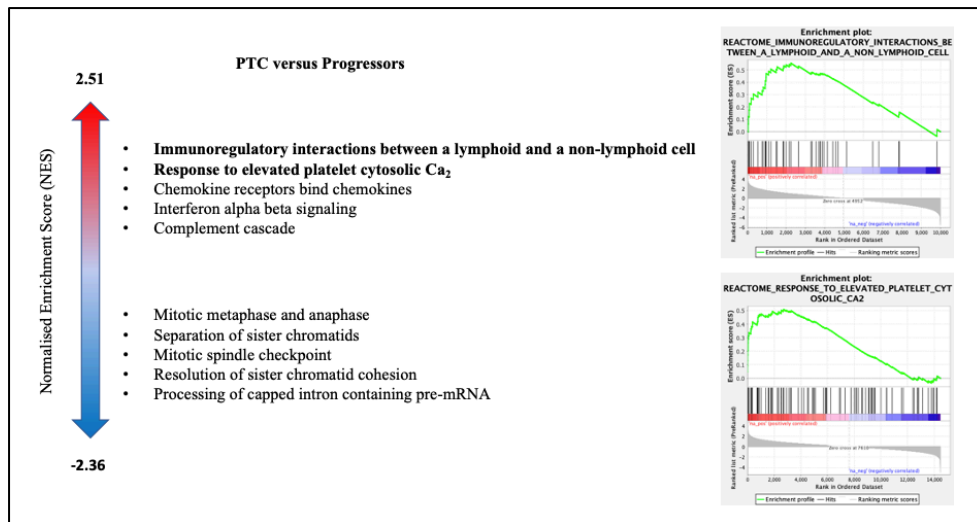
Figure 2: Pathway enrichment according to clinical phenotype using DESeq2 with GSEA and Rank Prod with ReactomePA Top five up- and down-regulated pathways and GSEA pre-ranked enrichment plots for pathways (FDR<0.25) for SC versus non-SC phenotypes (a) and PTC versus Progressors (b). The range of Normalised Enrichment Scores (NES) shown on the left. Bar-plots presen plots presenting ReactomePA pathway enrichment for SC versus nonSC (c) and PTC versus Progressors (d).

8

To validate the results presented above, a non-parametric approach, RankProd, was employed as an alternative method for DGE followed by pathway enrichment with ReactomePA. This analysis also showed enrichment of type I Interferon pathways for the SC phenotype and for Platelet and Fc-gamma receptor (FCGR) activation pathways for PTCs compared to Progressors (Figure 1c-d), consistent with the GSEA.

Weighted Gene Co-expression Network Analysis (WGCNA) identifies modules associated with rebound phenotypes and time to rebound

For an alternative interrogation of the transcriptomics data, Weighted Gene Co-expression Network Analysis (WGCNA) was employed to identify gene co-expression modules that relate to the clinical features of interest and inform on pathway enrichment. Aside from the different phenotype groups (SC versus non-SC and PTC versus Progressors), Time to Rebound was also used as a clinical trait of interest, as it offered the opportunity to explore the genetic correlations without having to make an a priori decision on sample grouping. WGCNA constructs a weighted network that represents the interaction patterns among genes, by emphasizing the strong gene-gene correlations at the expense of the weak ones in order to reduce noise. Here, the co-expression network was built using the expression data of a total of 4457 out of 28395 genes, that were retained after filtering for low variability. All pairwise correlations were calculated by raising the correlation values to the power of $\beta=18$, for which the scale-free topology fit index was 0.8 (Figure 2a). WGCNA then clustered all genes with similar co-expression patterns into modules, conventionally denoted by colour names (Figure S4). The modules were then associated with three external sample traits, namely SC (rebound >500 days), PTC (rebound >100 days) and 'Days to Rebound' as a continuous variable.

Four modules were positively correlated with the traits of interest. 'PaleTurquoise' (labelled as Module 1 in Figure 2b) was correlated with both Time to Rebound (cor=0.64, p=0.02) and the SC phenotype (cor=0.62, p=0.03). 'Salmon' (Module 2; cor=0.58, p<0.05) and 'Yellow' (Module 3; cor=0.63, p=0.03) correlated with the SC phenotype. 'Grey60' (Module 4) was found to be associated with the PTC phenotype (cor=0.65, p=0.02) (Figure 2b). A significantly high correlation of Module Membership (MM) and Gene Significance (GS) is reported for these modules showing that the genes that are significantly associated with the respective traits are also important elements of each module (Figure 2c-g).

Functional Enrichment of WGCNA modules associated with SC phenotype and Days to Rebound indicates Interferon Type I pathway involvement

The modules obtained from WGCNA were then uploaded on STRING for downstream enrichment analysis (Figure 3a-b). To be consistent with the GSEA analysis, the Reactome database was used to report the enrichment. The Module 1 genes (PaleTurquoise) showed an enrichment (p<0.05) in the Interferon Type I pathway (Figure 3c). Based on the GS and MM scores, 17 genes in Module 1 were identified as hub genes. Genes in modules 2 and 3 that correlated with the SC phenotype were shown to have non-significant enrichment values (0.06 and 0.3, respectively), indicating that these identified random proteins are not well connected. Module 4 (Grey60) that was associated with the PTC phenotype showed an enrichment (p<0.01) in platelet activation and degranulation (Figure 3d).
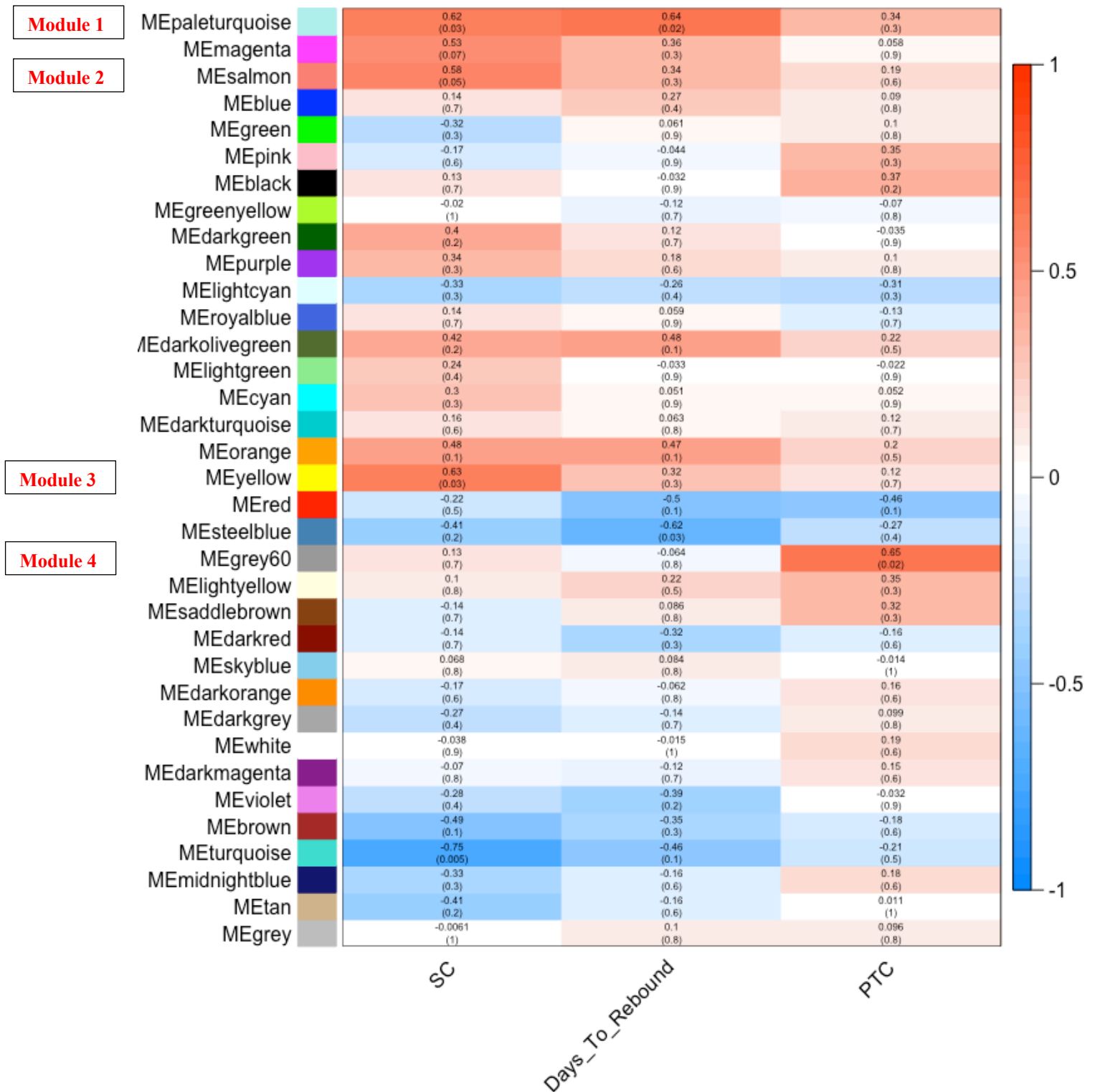
10

Figure 3: Network topology analysis and module identification using WGCNA

a.
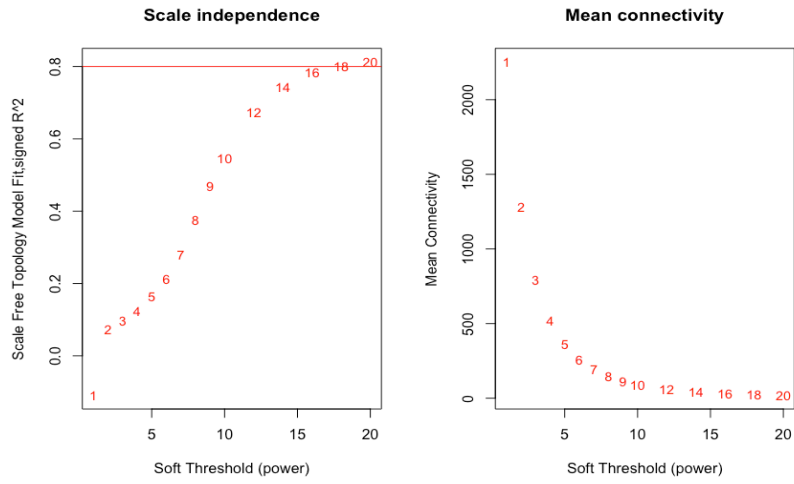
b.

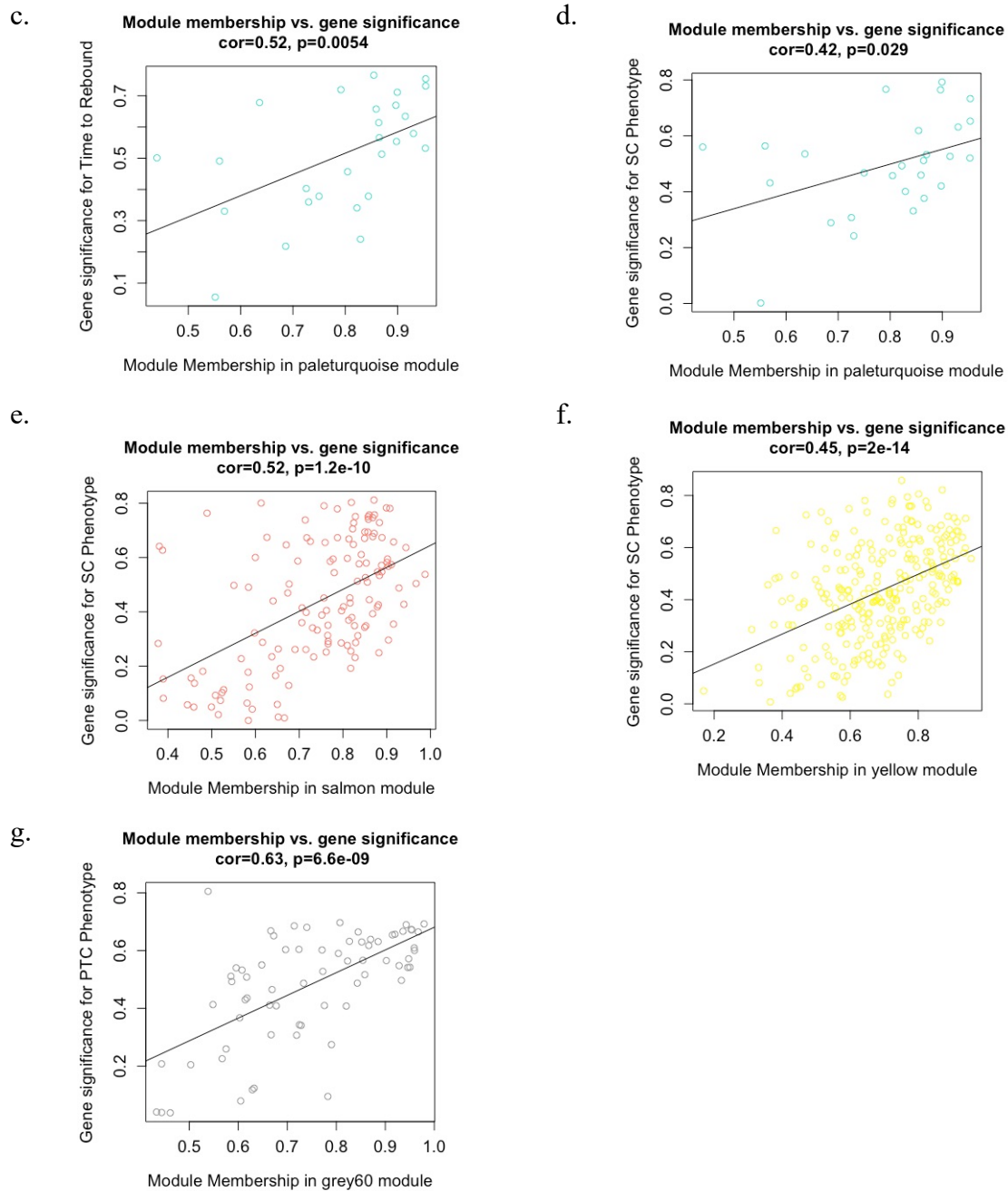## Module-trait relationships

c.

d.

e.

f.

g.

Figure 4: Network topology analysis and module identification using WGCNA a. Network topology analysis for different soft thresholding powers. b. Heatmap plot of module-trait correlation and its statistical significance in parenthesis. The table is colour-coded based on direction and intensity of correlation, which is shown in red boxes. c-g: Scatterplots of gene significance (GS) for trait of interest versus module membership (MM): paleturquoise (days to rebound) (c); paleturquoise (SC phenotype) (d); salmon (SC phenotype) (e); yellow (SC phenotype) (f); grey60 (PTC phenotype) (g).

13

Figure 5: Gene Network and Pathway Enrichment Visualisation of selected WGCNA modules.

a.

b.

c.

d.
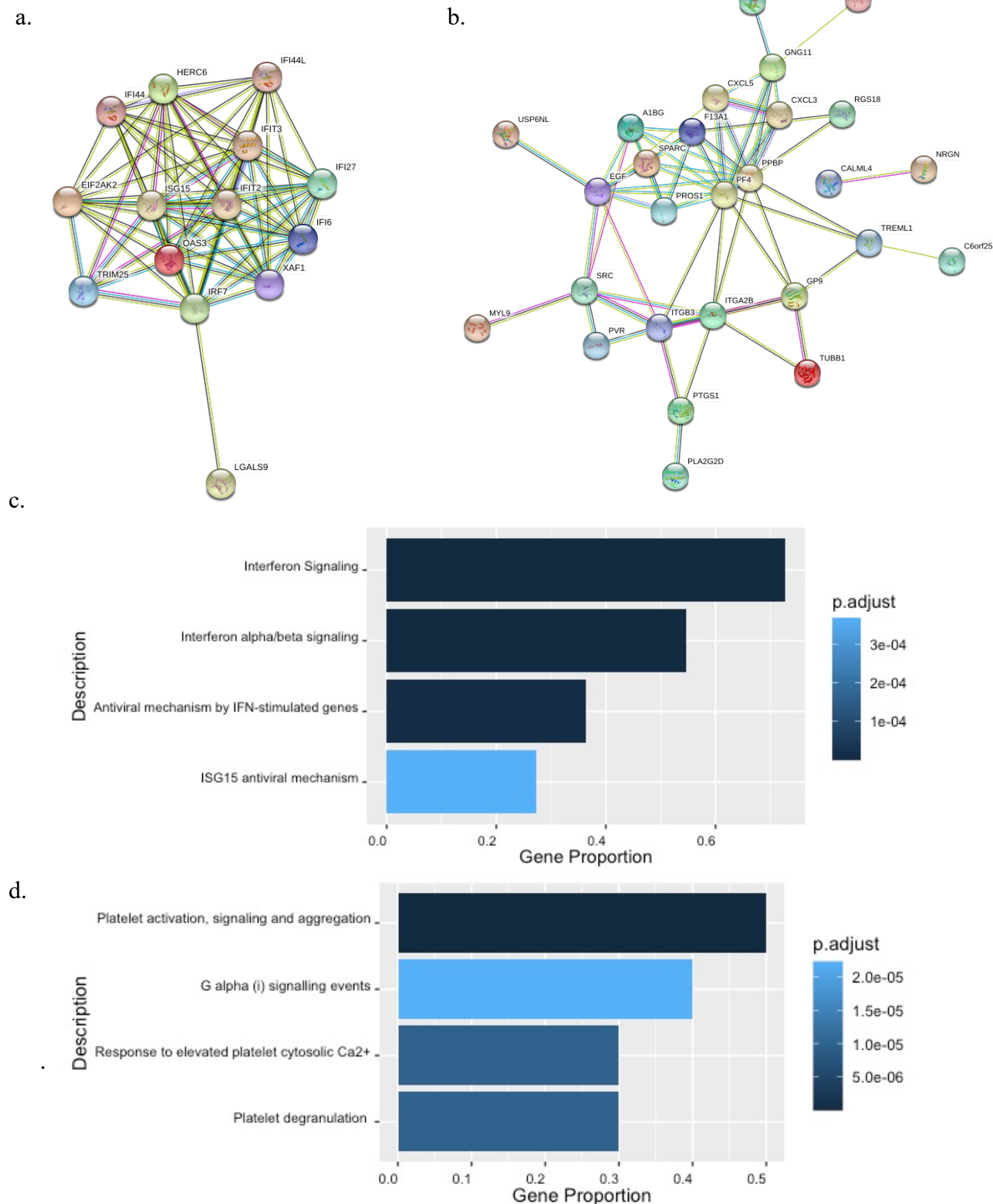
Figure 6: Gene Network and Pathway Enrichment Visualisation of selected WGCNA modules. Visualisation of the Protein-Protein Interaction network of genes within the PaleTurquoise (a) and the Grey60 (b) modules, generated by STRING. c-d. Top pathways enrichment bar plots for PaleTurquoise (c) and Grey60 (d) hub genes.

Risk score calculation based on the expression of two genes can potentially predict time to rebound

As a next step, we looked to see if any of the individual genes identified in the WGCNA results were more closely linked to progression, potentially leading to a signature that might predict longer post-TI viral suppression. Univariable analysis showed that the expression of 5 out of 17 hub genes contained in Module 1, namely ISG15, TRIM25, IFIT3, IFI6 and XAF1, was associated with a protective Hazard Ratio (HR) for early rebound <1 (Table 2).

Table 2. Univariable Cox Regression for Individual Protective Genes.

Gene	HR	95%CI	p-value
ISG15	0.11	0.02-0.66	0.01
IFIT3	0.088	0.00-0.89	0.04
TRIM25	0.021	0.00-0.75	0.03
IFI6	0.074	0.00-0.84	0.02
XAF1	0.19	0.04-0.90	0.03

HR= Hazard Ratio; 95% CI = 95% Confidence Intervals

Interestingly, all five are interferon stimulated genes. A multivariable Cox Regression analysis with LASSO was performed using the five significant genes and only TRIM25 and ISG15 were retained (Figure 4a). A Risk Score (RS) was calculated for all participants using the expression of these two genes and their LASSO coefficients (Table S2) were classified as either low or high relative to the mean RS. Kaplan Meier analysis showed that a lower RS with higher expression of both ISG15 and TRIM25 was found to be significantly associated with longer suppression post-TI, compared to participants with a high RS (P<0.001) (Figure 4b).

Figure 7: The expression of ISG15 and TRIM25 predict risk of earlier viral rebound.

a.


```
(Intercept)   .
ISG15        -1.533708
TRIM25       -2.044780
IFIT3         .
IFI6          .
XAF1          .
```
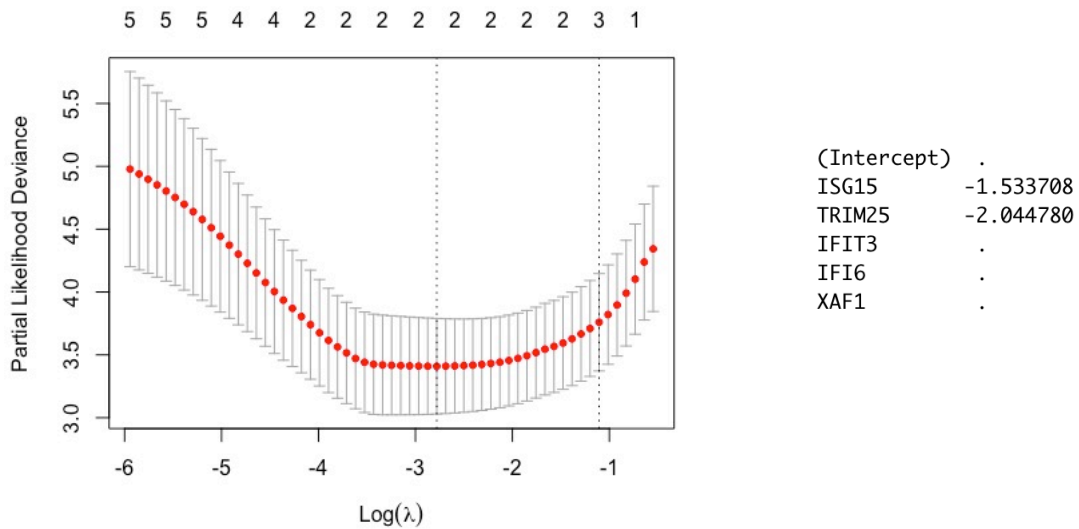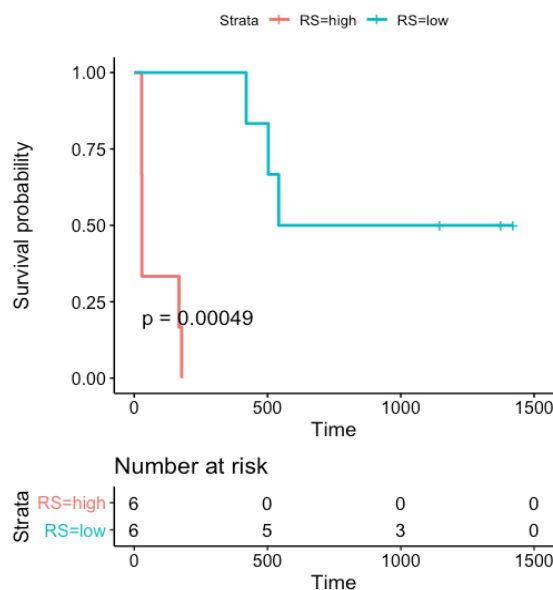
b.

Figure 8: The expression of ISG15 and TRIM25 predict risk of earlier viral rebound. a. Estimation of λ and coefficient penalisation for gene selection with LASSO. Two of the five genes are selected as correlating with a lower risk of early rebound. Negative coefficient indicates the effect is protective. b. Kaplan-Meier curve of gene expression-based Risk Score (RS), predicting the likelihood of early and late post-TI rebound. Blue and red lines represent low and high-risk score respectively and '+' marks represent the censored samples.

16

Discussion

The majority of HIV positive individuals tend to rebound quickly after stopping ART, although it is evident that for a few some degree of virological remission is established. Understanding the mechanisms of post-treatment control through the identification of biomarkers that would help stratify patients into viral rebound risk groups would be invaluable for clinical decision making. Molecular studies constitute a powerful tool for identifying signatures that determine different response phenotypes. High-throughput RNA sequencing and subsequent bioinformatics analysis render an excellent opportunity to characterise the host immune profile of ART responders and provide a mechanism for identifying biomarkers that predict control post-ART cessation.

Our aim was to gain a deeper understanding of how host gene expression affected the duration of post-TI viral control using samples collected during the SPARTAC trial. This cohort offered a unique opportunity to study samples from participants with PHI that have been followed-up for an average of 4 years post-TI(Fidler et al., 2013). The samples used in this study were collected at the day of TI. To minimise the noise caused by averaging the expression of different cell types, only CD4+ T cells were selected and sequenced. This was because based CD4+ T cells comprise the majority of the HIV reservoir, and therefore any variability in transcriptional activity might impact the likelihood of proviral expression and rebound viraemia, and also related to the significance of the CD4+ T-cell response in the control of HIV(Frater et al., 2014). In addition, we restricted our analyses to women from Africa to exclude potential confounding factors from sex or ethnicity.

The results of DGE/GSEA revealed a distinct transcriptomic difference in those who controlled for over 500 days (i.e. between the SC phenotype and the non-SC phenotype), which was

associated with the Type I Interferon Pathway (IFN-I). IFN-I pathways comprise a family of cytokines playing a dual role in the regulation of the innate immune response in infection(McNab et al., 2015). While continuous exposure to IFN-I inflicts chronic immune activation and subsequent exhaustion of T-cells and progression to AIDS, studies on SIV models illustrate the benefit of IFN-I administration at the early stage of infection in inhibiting viral replication(Deeks et al., 2017; Sandler et al., 2014; Van der Sluis et al., 2020). Similarly, HIV develops strategies to evade IFN-I and to inhibit the functionality of the proteins regulated by IFN-I (Fenton-May et al., 2013), suggestive of a fundamental role of this pathway in viraemic control. Additionally, the presence of less fit, adapted viruses - well described for T cell immune escape(Roberts et al., 2015; Zimbwa et al., 2007) - could also contribute to the difference in IFN-I response observed in this study(Adland et al., 2020; Cohn et al., 2018; Iyer et al., 2017). By defining post-treatment control ('PTC') participants as those that rebounded more than 100 days after TI(Etemad et al., 2019), we identified an enrichment of platelet activation pathways as well as IFN-I. The strong statistical association with platelet activation is intriguing and possibly consistent with reports that release of chemokines by activated platelets can function as an extra barrier at the early stages of HIV infection(Solomon Tsegaye et al., 2013). However, further investigation is required to determine the role of platelets in the defense against HIV.

WGCNA was then performed to determine how gene modules associate with specific clinical traits. For an unbiased analysis, 'time to rebound' was included with the SC and PTC phenotypes, to examine whether WGCNA findings corresponded with the DGE/GSEA analysis. The analysis identified two dominant modules of which one was associated with the 'Time to Rebound' and the SC trait, and one with the PTC phenotype. The gene ontology and pathway analysis showed a clear module enrichment in IFN-I associated with time to rebound,

supporting the previous results for the phenotype traits. The majority of genes reported as hub genes were interferon stimulated genes (ISG), again consistent with the argument that the response to type I interferons is impacting rebound. A univariable Cox analysis identified five ISGs associated with remission in the participants (Table 2). Based on a multivariable Cox regression with LASSO screening of the five genes, a risk score based on the expression of ISG15 and TRIM25 only was developed to predict the likelihood of viral rebound. Higher expression of both genes was strongly protective for post-TI rebound. ISG15 has been reported to contribute to the inhibition of HIV virion release from infected cells(Gargan et al., 2018; Perng & Lenschow, 2018) and TRIM25 is an E3 ligase that positively regulates ISG15 conjugation to pathogen proteins, in a process with reported antiviral effects called ISGylation(Martín-Vicente et al., 2017). Although more work is underway to characterise the role of this pathway, this preliminary independent identification of these two genes would be consistent with a role in maintaining virological remission.

The small sample size on which this analysis was performed is a key limitation. In addition, our analysis was limited to bulk CD4+ T cells at a single timepoint, with no enrichment for those cells that were HIV-specific or contained proviral DNA. At this stage, the latter is not possible due to technological limitations. The other factor to consider is whether our choice of clinical phenotypes accurately discriminated between 'post-treatment controllers' and 'elite controllers'. The latter have been well characterised(Martin et al., 2017), and more generally associated with effective HLA Class I-restricted T cell responses(Kiepiela et al., 2007), whereas it is still unclear as to which mechanisms are driving PTC. Larger studies will be needed to tease out these differences. However, that three independent analyses conferred the same statistically significant findings should be taken into consideration. Furthermore, this is a transcriptomic snapshot from the time of TI, and information about the expression of ISG is

not available for other timepoints, before and after TI. This finding of a strong type I interferon signal associating with delayed rebound in this small study needs to be confirmed in larger clinical trials incorporating TI. If these data are reproducible, they could help with the identification of a valuable biomarkers of remission and pathways to drug discovery for the HIV cure field.

Materials and Methods

Participant characteristics and trial design

We analysed 18 SPARTAC participants who had received 48 weeks of suppressive ART commenced shortly after seroconversion. Participants were selected based on availability of viable peripheral blood mononuclear cells (PBMCs) and stratified by time to rebound. Participants were sampled at the time of TI, 48 weeks after commencing ART during Primary HIV Infection (PHI). The demographics of the participants are presented in Table 1, and Supplementary Table S1. Based on time to rebound, participants were classified as those who rebounded Early (<100 days after TI), Intermediate (100-500 days after TI) and Late (>500 days after TI). Initial analyses revealed gene expression varied significantly dependent on country of origin (African vs Non-African) (Supplementary Figure S1), and so - to avoid confounders - participants were analysed separately according to origin. Additionally, there were no male participants in the Late category, and so analyses including this parameter were restricted to female participants.

RNA Isolation and Sequencing

CD4+ T cells were isolated using negative selection (Stem Cell Technologies CD4 enrichment kit) according to the manufacturer's recommendations, and samples were processed in one batch. Total RNA was extracted using the Qiagen RNA Kit with Qiashredder columns and was

sent for library preparation and sequencing. Material was quantified using RiboGreen (Invitrogen) on the FLUOstar OPTIMA plate reader (BMG Labtech) and the size profile and integrity analysed on the 2200 TapeStation (Agilent, RNA ScreenTape). Input material was normalised to 100ng prior to library preparation. Polyadenylated transcript enrichment and strand specific library preparation was completed using NEBNext Ultra II mRNA kit (NEB) following manufacturer's instructions. Libraries were amplified on a Tetrad (Bio-Rad) using in-house unique dual indexing primers(Lamble et al., 2013). Individual libraries were normalised using Qubit, and the size profile was analysed on the 2200 TapeStation. Individual libraries were normalised and pooled together accordingly. The pooled library was diluted to ~10 nM for storage. The 10 nM library was denatured and further diluted prior to sequencing. Paired end sequencing was performed using an Illumina Novaseq6000 platform at 150 paired end mode (Illumina, San Diego, CA).

Differential Gene Expression Screening and Gene Set Enrichment Analysis

DESeq2(Love et al., 2014) was used to compute the differential gene expression (DGE) between phenotypes using a featurecounts table. Only genes with a DGE of adjusted p-value of <0.05 were considered statistically significant. Gene Set Enrichment Analysis (GSEA)(Mootha et al., 2003; Subramanian et al., 2005) was used to detect differences in biologically relevant pathways in the dataset. The datasets were pre-ranked by the DESeq2 Wald statistic value. Permutation was set at 1000. Gene sets in the Reactome database(Jassal et al., 2020) were used as reference. Results satisfying a False Discovery Rate (FDR) cut off of <25% were considered statistically significant. Due to the small number of replicates, RankProd(Del Carratore et al., 2017) – an alternative nonparametric statistical method to identify differentially expressed genes - was also used. This method ranks the genes that consistently rank high as up- or downregulated in a number of experiments. The test was run

on data transformed with the DESeq2 variance stabilising transformation. The percentage of false prediction (pfp) cut off was set to 0.25. The statistically significant upregulated genes for each phenotype, ranked by an increasing RP score, were further analysed with the ClusterProfiler(Yu et al., 2012) and ReactomePA(Yu & He, 2016) R packages, for pathway enrichment.

Co-Expression Network Construction

Weighted Gene Co-expression Network Analysis (WGCNA)(Langfelder & Horvath, 2008; B. Zhang & Horvath, 2005) is an R package used for gene expression profiling and was applied to the identification of genes associated with time to rebound. After pre-processing for low variance filtering and outlier removal, an appropriate soft-threshold power was selected to promote and penalise the strong and weak gene connections, respectively. Following this, a signed network was created using the one-step approach, according to the package guidance. Genes were organised in modules, based on a common expression pattern. The expression of the genes in each module was summarised with an eigengene value, and a colour-label attributed to each to assist identification.

Identification and Annotation of Important Modules and Hub genes

Module eigengenes were correlated with the clinical traits, in order to identify the most biologically relevant modules. The modules that were selected for downstream analysis were the ones that had a significant correlation with the trait. The most connected genes (hub genes), which are of functional significance, were defined by their Module Membership (MM>0.8) and their Gene-trait Significance (GS>0.2). All selected modules were uploaded on STRING(Szklarczyk et al., 2019) for pathway enrichment, gene ontology annotation and PPI visualisation.

Identification of predictor genes for time to rebound

Survival analysis was performed for the identified hub genes, using the survival(Terry, 2020) package on R, to assess the prognostic value of the hub genes. The first viral rebound after TI was used as the event of interest. A univariable Cox Proportional Hazard Regression was performed on each gene. Genes with a statistically significant correlation to time to rebound were selected for a multivariable Cox regression with LASSO after selecting an optimal regularization parameter λ, to penalise the variable coefficients. A Risk Score (RS) to predict prognosis was calculated according to the formula below, where β is the multivariate Cox coefficient and exp is the expression value for all significant genes.

$$RS = \sum_{n=1}^{gene} \beta_{gene} \times exp_{gene}$$

All participants that did not report post-TI rebound during the follow-up period of SPARTAC were classified as censored. All genes with a statistically significant association with a longer remission were then used to calculate a prediction score for time to rebound, by multiplying the gene coefficient with the gene expression.

Supplementary Materials

Table S1. Detailed demographics

Figure S1. Dendrogram showing clustering of participants from different countries based on gene expression.

Figure S2. Volcano plots showing the differentially expressed genes according to patient clinical phenotype.

Figure S3. GSEA analyses according to clinical phenotype

Figure S4. Dendrogram of genes clustering into modules.

References

Adland, E., Millar, J., Bengu, N., Muenchhoff, M., Fillis, R., Sprenger, K., Ntlantsana, V., Roider, J., Vieira, V., Govender, K., Adamson, J., Nxele, N., Ochsenbauer, C., Kappes, J., Mori, L., van Lobenstein, J., Graza, Y., Chinniah, K., Kapongo, C., … Goulder, P. (2020). Sex-specific innate immune selection of HIV-1 in utero is associated with increased female susceptibility to infection. *Nature Communications*, *11*(1), 1767. https://doi.org/10.1038/s41467-020-15632-y

Cohn, L. B., da Silva, I. T., Valieris, R., Huang, A. S., Lorenzi, J. C. C., Cohen, Y. Z., Pai, J. A., Butler, A. L., Caskey, M., Jankovic, M., & Nussenzweig, M. C. (2018). Clonal CD4+ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. *Nature Medicine*, *24*(5), 604–609. https://doi.org/10.1038/s41591-018-0017-7

Colby, D. J., Sarnecki, M., Barouch, D. H., Tipsuk, S., Stieh, D. J., Kroon, E., Schuetz, A., Intasan, J., Sacdalan, C., Pinyakorn, S., Grandin, P., Song, H., Tovanabutra, S., Shubin, Z., Kim, D., Paquin-Proulx, D., Eller, M. A., Thomas, R., de Souza, M., … Ananworanich, J. (2020). Safety and immunogenicity of Ad26 and MVA vaccines in acutely treated HIV and effect on viral rebound after antiretroviral therapy interruption. *Nature Medicine*, *26*(4), 498–501. https://doi.org/10.1038/s41591-020-0774-y

Colby, D. J., Trautmann, L., Pinyakorn, S., Leyre, L., Pagliuzza, A., Kroon, E., Rolland, M., Takata, H., Buranapraditkun, S., Intasan, J., Chomchey, N., Muir, R., Haddad, E. K., Tovanabutra, S., Ubolyam, S., Bolton, D. L., Fullmer, B. A., Gorelick, R. J., Fox, L., … group, T. R. study. (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nature Medicine*, *24*(7), 923–926. https://doi.org/10.1038/s41591-018-0026-6

Conway, J. M., Perelson, A. S., & Li, J. Z. (2019). Predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers. *PLoS Computational Biology*, *15*(7), e1007229. https://doi.org/10.1371/journal.pcbi.1007229

Dash, P. K., Kaminski, R., Bella, R., Su, H., Mathews, S., Ahooyi, T. M., Chen, C., Mancuso, P., Sariyer, R., Ferrante, P., Donadoni, M., Robinson, J. A., Sillman, B., Lin, Z., Hilaire, J. R., Banoub, M., Elango, M., Gautam, N., Mosley, R. L., … Gendelman, H. E. (2019). Sequential LASER ART and CRISPR Treatments Eliminate HIV-1 in a Subset of Infected Humanized Mice. *Nature Communications*, *10*(1), 2753. https://doi.org/10.1038/s41467-019-10366-y

Davenport, M. P., Khoury, D. S., Cromer, D., Lewin, S. R., Kelleher, A. D., & Kent, S. J. (2019). Functional cure of HIV: the scale of the challenge. *Nature Reviews Immunology*, *19*(1), 45–54. https://doi.org/10.1038/s41577-018-0085-4

Deeks, S. G., Odorizzi, P. M., & Sekaly, R.-P. (2017). The interferon paradox: can inhibiting an antiviral mechanism advance an HIV cure? *The Journal of Clinical Investigation*, *127*(1), 103–105. https://doi.org/10.1172/JCI91916

Del Carratore, F., Jankevics, A., Eisinga, R., Heskes, T., Hong, F., & Breitling, R. (2017). RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics (Oxford, England)*, *33*(17), 2774–2775. https://doi.org/10.1093/bioinformatics/btx292

Etemad, B., Esmaeilzadeh, E., & Li, J. Z. (2019). Learning From the Exceptions: HIV Remission in Post-treatment Controllers . In *Frontiers in Immunology* (Vol. 10, p. 1749). https://www.frontiersin.org/article/10.3389/fimmu.2019.01749

Fenton-May, A. E., Dibben, O., Emmerich, T., Ding, H., Pfafferott, K., Aasa-Chapman, M. M., Pellegrino, P., Williams, I., Cohen, M. S., Gao, F., Shaw, G. M., Hahn, B. H., Ochsenbauer, C., Kappes, J. C., & Borrow, P. (2013). Relative resistance of HIV-1

founder viruses to control by interferon-alpha. *Retrovirology*, *10*(1), 146. https://doi.org/10.1186/1742-4690-10-146

Fidler, S., Porter, K., Ewings, F., Frater, J., Ramjee, G., Cooper, D., Rees, H., Fisher, M., Schechter, M., Kaleebu, P., Tambussi, G., Kinloch, S., Miro, J. M., Kelleher, A., McClure, M., Kaye, S., Gabriel, M., Phillips, R., Weber, J., & Babiker, A. (2013). Short-course antiretroviral therapy in primary HIV infection. *The New England Journal of Medicine*, *368*(3), 207–217. https://doi.org/10.1056/NEJMoa1110039

Frater, J., Ewings, F., Hurst, J., Brown, H., Robinson, N., Fidler, S., Babiker, A., Weber, J., Porter, K., & Phillips, R. E. (2014). HIV-1-specific CD4+ responses in primary HIV-1 infection predict disease progression. *AIDS*, *28*(5). https://journals.lww.com/aidsonline/Fulltext/2014/03130/HIV_1_specific_CD4__respon ses_in_primary_HIV_1.7.aspx

Gargan, S., Ahmed, S., Mahony, R., Bannan, C., Napoletano, S., O'Farrelly, C., Borrow, P., Bergin, C., & Stevenson, N. J. (2018). HIV-1 Promotes the Degradation of Components of the Type I IFN JAK/STAT Pathway and Blocks Anti-viral ISG Induction. *EBioMedicine*, *30*, 203–216. https://doi.org/10.1016/j.ebiom.2018.03.006

Hurst, J., Hoffmann, M., Pace, M., Williams, J. P., Thornhill, J., Hamlyn, E., Meyerowitz, J., Willberg, C., Koelsch, K. K., Robinson, N., Brown, H., Fisher, M., Kinloch, S., Cooper, D. A., Schechter, M., Tambussi, G., Fidler, S., Babiker, A., Weber, J., … Frater, J. (2015). Immunological biomarkers predict HIV-1 viral rebound after treatment interruption. *Nature Communications*, *6*, 8495. https://doi.org/10.1038/ncomms9495

Hyrcza, M. D., Kovacs, C., Loutfy, M., Halpenny, R., Heisler, L., Yang, S., Wilkins, O., Ostrowski, M., & Der, S. D. (2007). Distinct Transcriptional Profiles in Ex Vivo CD4+ and CD8+ T Cells Are Established Early in Human Immunodeficiency Virus Type I Infection and Are Characterized by a Chronic Interferon Response as Well as Extensive Transcriptional Changes in CD8+ T Cells. *Journal of Virology*, *81*(7), 3477–3486. https://doi.org/10.1128/JVI.01552-06

Iyer, S. S., Bibollet-Ruche, F., Sherrill-Mix, S., Learn, G. H., Plenderleith, L., Smith, A. G., Barbian, H. J., Russell, R. M., Gondim, M. V. P., Bahari, C. Y., Shaw, C. M., Li, Y., Decker, T., Haynes, B. F., Shaw, G. M., Sharp, P. M., Borrow, P., & Hahn, B. H. (2017). Resistance to type I interferons is a major determinant of HIV-1 transmission fitness. *Proceedings of the National Academy of Sciences*, *114*(4), E590 LP-E599. https://doi.org/10.1073/pnas.1620144114

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., … D'Eustachio, P. (2020). The reactome pathway knowledgebase. *Nucleic Acids Research*, *48*(D1), D498–D503. https://doi.org/10.1093/nar/gkz1031

Julg, B., Dee, L., Ananworanich, J., Barouch, D. H., Bar, K., Caskey, M., Colby, D. J., Dawson, L., Dong, K. L., Dubé, K., Eron, J., Frater, J., Gandhi, R. T., Geleziunas, R., Goulder, P., Hanna, G. J., Jefferys, R., Johnston, R., Kuritzkes, D., … Walker, B. D. (2019). Recommendations for analytical antiretroviral treatment interruptions in HIV research trials-report of a consensus meeting. *The Lancet. HIV*, *6*(4), e259–e268. https://doi.org/10.1016/S2352-3018(19)30052-9

Kiepiela, P., Ngumbela, K., Thobakgale, C., Ramduth, D., Honeyborne, I., Moodley, E., Reddy, S., de Pierres, C., Mncube, Z., Mkhwanazi, N., Bishop, K., van der Stok, M., Nair, K., Khan, N., Crawford, H., Payne, R., Leslie, A., Prado, J., Prendergast, A., … Goulder, P. (2007). CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Medicine*, *13*(1), 46–53. https://doi.org/10.1038/nm1520

Krebs, S. J., & Ananworanich, J. (2016). Immune activation during acute HIV infection and the impact of early antiretroviral therapy. *Current Opinion in HIV and AIDS*, *11*(2). https://journals.lww.com/co-hivandaids/Fulltext/2016/03000/Immune_activation_during_acute_HIV_infection_and.7.aspx

Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B., & Piazza, P. (2013). Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnology*, *13*(1), 104. https://doi.org/10.1186/1472-6750-13-104

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Lee, S. Y., Park, Y. K., Yoon, C.-H., Kim, K., & Kim, K.-C. (2019). Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1. *BMC Medical Genomics*, *12*(1), 3. https://doi.org/10.1186/s12920-018-0443-x

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Martín-Vicente, M., Medrano, L. M., Resino, S., García-Sastre, A., & Martínez, I. (2017). TRIM25 in the Regulation of the Antiviral Innate Immunity. *Frontiers in Immunology*, *8*, 1187. https://doi.org/10.3389/fimmu.2017.01187

Martin, G. E., Gossez, M., Williams, J. P., Stöhr, W., Meyerowitz, J., Leitman, E. M., Goulder, P., Porter, K., Fidler, S., Frater, J., & Investigators, the S. T. (2017). Post-treatment control or treated controllers? Viral remission in treated and untreated primary HIV infection. *AIDS (London, England)*, *31*(4), 477–484. https://doi.org/10.1097/QAD.0000000000001382

McNab, F., Mayer-Barber, K., Sher, A., Wack, A., & O'Garra, A. (2015). Type I interferons in infectious disease. *Nature Reviews Immunology*, *15*(2), 87–103. https://doi.org/10.1038/nri3787

Mendoza, P., Gruell, H., Nogueira, L., Pai, J. A., Butler, A. L., Millard, K., Lehmann, C., Suárez, I., Oliveira, T. Y., Lorenzi, J. C. C., Cohen, Y. Z., Wyen, C., Kümmerle, T., Karagounis, T., Lu, C.-L., Handl, L., Unson-O'Brien, C., Patel, R., Ruping, C., … Nussenzweig, M. C. (2018). Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature*, *561*(7724), 479–484. https://doi.org/10.1038/s41586-018-0531-2

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., … Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3), 267–273. https://doi.org/10.1038/ng1180

Peluso, M. J., Deeks, S. G., & McCune, J. M. (2019). HIV "cure": A shot in the arm? *EBioMedicine*, *42*, 3–5. https://doi.org/10.1016/j.ebiom.2019.04.011

Perng, Y.-C., & Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nature Reviews Microbiology*, *16*(7), 423–439. https://doi.org/10.1038/s41579-018-0020-5

Roberts, H. E., Hurst, J., Robinson, N., Brown, H., Flanagan, P., Vass, L., Fidler, S., Weber, J., Babiker, A., Phillips, R. E., McLean, A. R., Frater, J., & investigators, S. trial. (2015). Structured Observations Reveal Slow HIV-1 CTL Escape. *PLOS Genetics*, *11*(2), e1004914. https://doi.org/10.1371/journal.pgen.1004914

Sáez-Cirión, A., Bacchus, C., Hocqueloux, L., Avettand-Fenoel, V., Girault, I., Lecuroux, C.,

Potard, V., Versmisse, P., Melard, A., Prazuck, T., Descours, B., Guergnon, J., Viard, J.-P., Boufassa, F., Lambotte, O., Goujard, C., Meyer, L., Costagliola, D., Venet, A., … Group, A. V. S. (2013). Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathogens*, *9*(3), e1003211–e1003211. https://doi.org/10.1371/journal.ppat.1003211

Sandler, N. G., Bosinger, S. E., Estes, J. D., Zhu, R. T. R., Tharp, G. K., Boritz, E., Levin, D., Wijeyesinghe, S., Makamdop, K. N., del Prete, G. Q., Hill, B. J., Timmer, J. K., Reiss, E., Yarden, G., Darko, S., Contijoch, E., Todd, J. P., Silvestri, G., Nason, M., … Douek, D. C. (2014). Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature*, *511*(7511), 601–605. https://doi.org/10.1038/nature13554

Sharaf, R., Lee, G. Q., Sun, X., Etemad, B., Aboukhater, L. M., Hu, Z., Brumme, Z. L., Aga, E., Bosch, R. J., Wen, Y., Namazi, G., Gao, C., Acosta, E. P., Gandhi, R. T., Jacobson, J. M., Skiest, D., Margolis, D. M., Mitsuyasu, R., Volberding, P., … Li, J. Z. (2018). HIV-1 proviral landscapes distinguish posttreatment controllers from noncontrollers. *The Journal of Clinical Investigation*, *128*(9), 4074–4085. https://doi.org/10.1172/JCI120549

Solomon Tsegaye, T., Gnirß, K., Rahe-Meyer, N., Kiene, M., Krämer-Kühl, A., Behrens, G., Münch, J., & Pöhlmann, S. (2013). Platelet activation suppresses HIV-1 infection of T cells. *Retrovirology*, *10*(1), 48. https://doi.org/10.1186/1742-4690-10-48

Stöhr, W., Fidler, S., McClure, M., Weber, J., Cooper, D., Ramjee, G., Kaleebu, P., Tambussi, G., Schechter, M., Babiker, A., Phillips, R. E., Porter, K., & Frater, J. (2013). Duration of HIV-1 viral suppression on cessation of antiretroviral therapy in primary infection correlates with time on therapy. *PloS One*, *8*(10), e78287–e78287. https://doi.org/10.1371/journal.pone.0078287

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545 LP – 15550. https://doi.org/10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Terry, T. (2020). A package for Survival Analysis in R. *CRAN*. https://cran.r-project.org/package=survival

Van der Sluis, R. M., Zerbato, J. M., Rhodes, J. W., Pascoe, R. D., Solomon, A., Kumar, N. A., Dantanarayana, A. I., Tennakoon, S., Dufloo, J., McMahon, J., Chang, J. J., Evans, V. A., Hertzog, P. J., Jakobsen, M. R., Harman, A. N., Lewin, S. R., & Cameron, P. U. (2020). Diverse effects of interferon alpha on the establishment and reversal of HIV latency. *PLoS Pathogens*, *16*(2), e1008151–e1008151. https://doi.org/10.1371/journal.ppat.1008151

Williams, J. P., Hurst, J., Stöhr, W., Robinson, N., Brown, H., Fisher, M., Kinloch, S., Cooper, D., Schechter, M., Tambussi, G., Fidler, S., Carrington, M., Babiker, A., Weber, J., Koelsch, K. K., Kelleher, A. D., Phillips, R. E., & Frater, J. (2014). HIV-1 DNA predicts disease progression and post-treatment virological control. *ELife*, *3*, e03821. https://doi.org/10.7554/eLife.03821

Yu, G., & He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway

analysis and visualization. *Mol. BioSyst.*, *12*(2), 477–479. https://doi.org/10.1039/C5MB00663E

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article17. https://doi.org/10.2202/1544-6115.1128

Zhang, L.-L., Zhang, Z.-N., Wu, X., Jiang, Y.-J., Fu, Y.-J., & Shang, H. (2017). Transcriptomic meta-analysis identifies gene expression characteristics in various samples of HIV-infected patients with nonprogressive disease. *Journal of Translational Medicine*, *15*(1), 191. https://doi.org/10.1186/s12967-017-1294-5

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLOS ONE*, *9*(1), e78644. https://doi.org/10.1371/journal.pone.0078644

Zimbwa, P., Milicic, A., Frater, J., Scriba, T. J., Willis, A., Goulder, P. J. R., Pillay, T., Gunthard, H., Weber, J. N., Zhang, H.-T., & Phillips, R. E. (2007). Precise Identification of a Human Immunodeficiency Virus Type I Antigen Processing Mutant. *Journal of Virology*, *81*(4), 2031 LP – 2038. https://doi.org/10.1128/JVI.00968-06

Acknowledgements

Supplementary Materials

STUDY ID	Sex	Category	Days to Rebound	Country	Control Category	
SIU001	Male	Early	24	Italy	normal progressors (Progressors)	'non-SC'
SJR002	Female	Early	28	South Africa		
SJZ003	Female	Early	28	South Africa		
SUN004	Female	Early	28	UK		
SUU005	Male	Early	28	UK		
SJA006	Female	Early	29	South Africa		
SJS007	Female	Early	29	South Africa		
SJV008	Female	Intermediate	168	Uganda	Post Treatment Controllers (PTC)	
SJK009	Female	Intermediate	168	South Africa		
SJJ010	Female	Intermediate	178	South Africa		
SUA011	Male	Intermediate	288	UK		
SUX012	Male	Intermediate	301	UK		
SJH013	Female	Intermediate	420	South Africa		
SJL014	Female	Late	503	South Africa		Sustained Controllers (SC)
SJC015	Female	Late	542	South Africa		
SJE016	Female	Late	1145*	South Africa		
SJE017	Female	Late	1376*	South Africa		
SJV018	Female	Late	1421*	South Africa		

*censored at this timepoint.

Table S 1: Detailed Demographics

Participant	Rebound	Days to Rebound	Incidence	Risk Score (RS)
SJR002	Early	28	1	-29.88
SJZ003	Early	28	1	-31.63
SJA006	Early	29	1	-29.96
SJS007	Early	29	1	-31.43
SJE016	Late	1145	0	-31.99
SJE017	Late	1376	0	-32.63
SJV018	Late	1421	0	-34.68
SJL014	Late	503	1	-31.74
SJC015	Late	542	1	-32.99
SJK009	Med	168	1	-30.37
SJJ010	Med	178	1	-30.81
SJH013	Med	420	1	-32.03
Mean Risk Score				-31.68

Table S 2: Individual Risk Score per participant, based on the expression of ISG15 and TRIM25
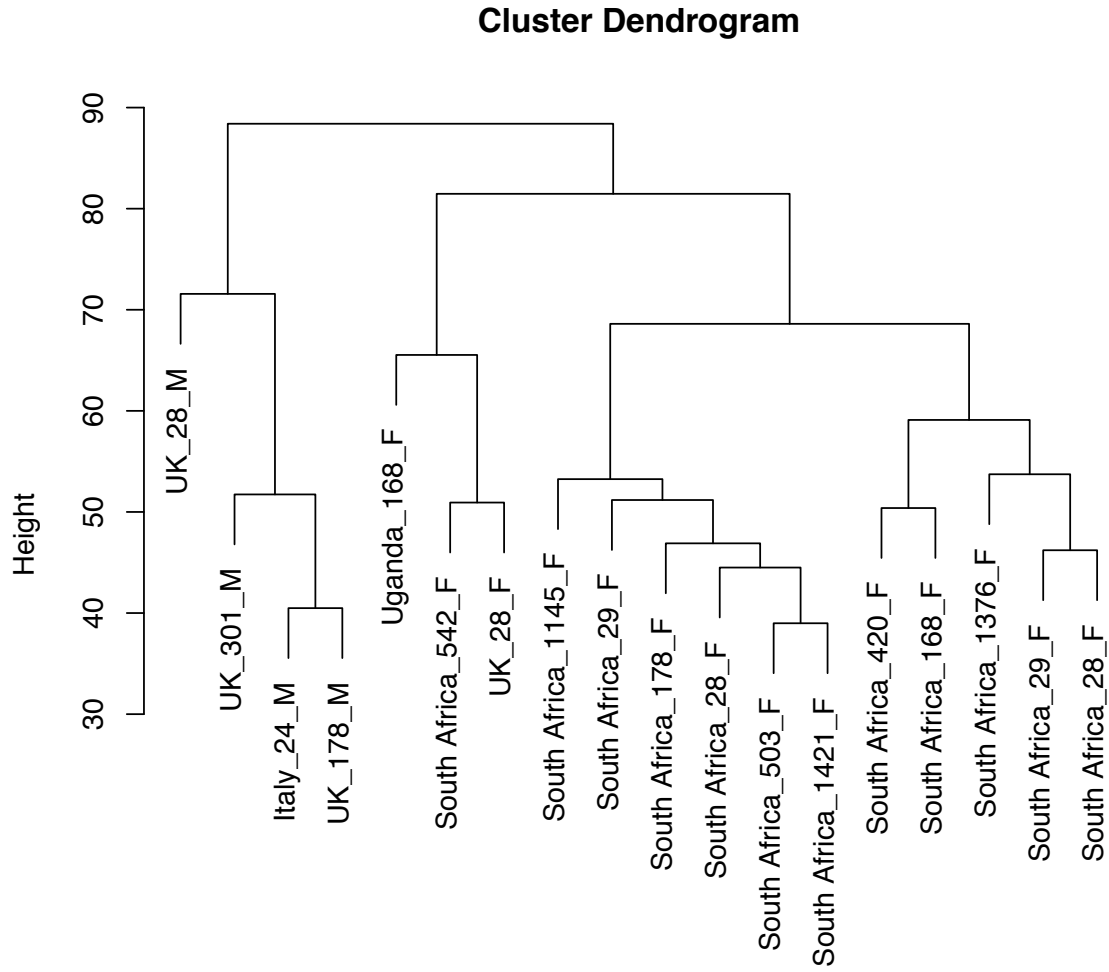
Cluster Dendrogram

Figure S 1: Dendrogram showing clustering of participants from different countries based on gene expression.
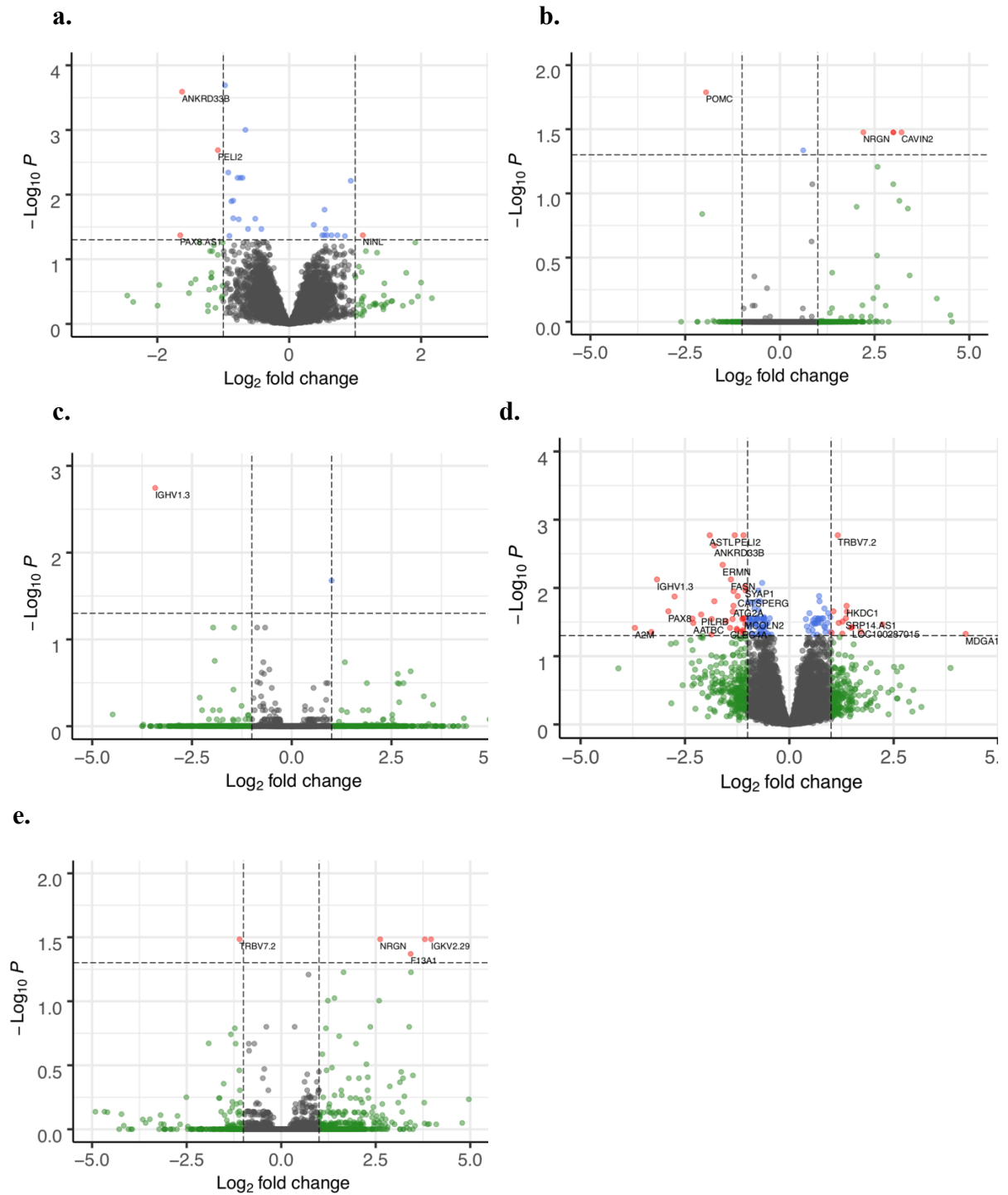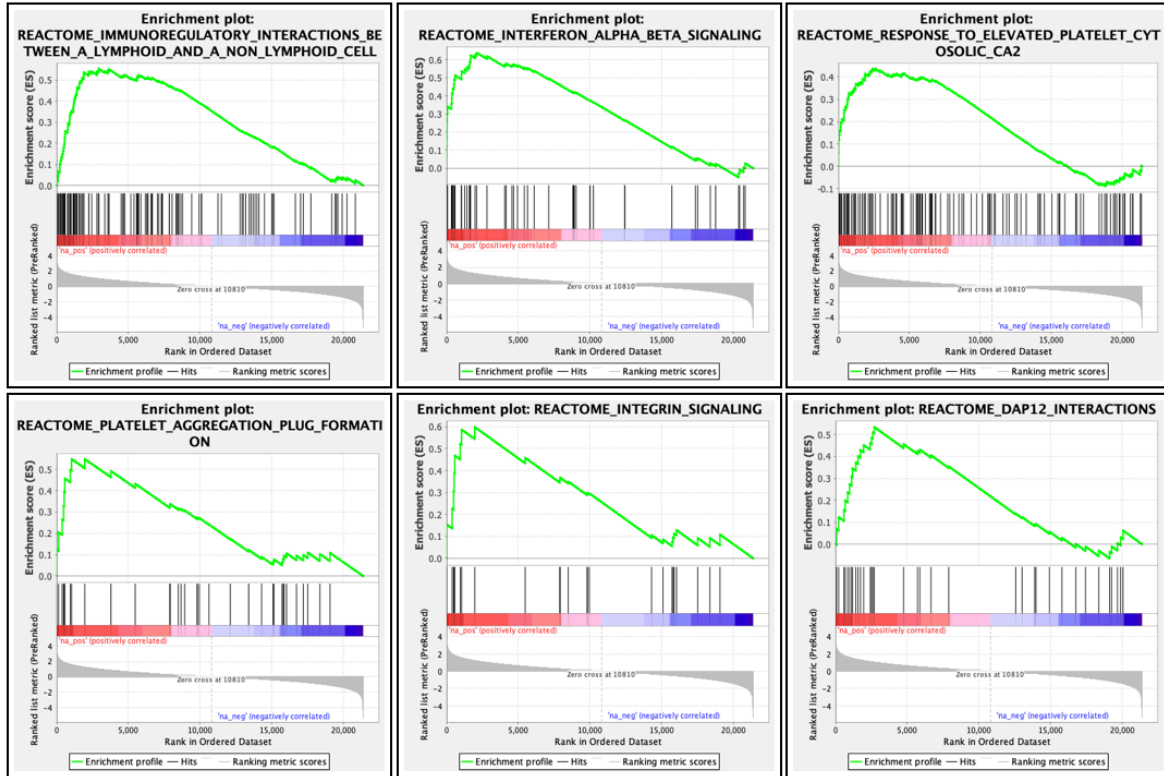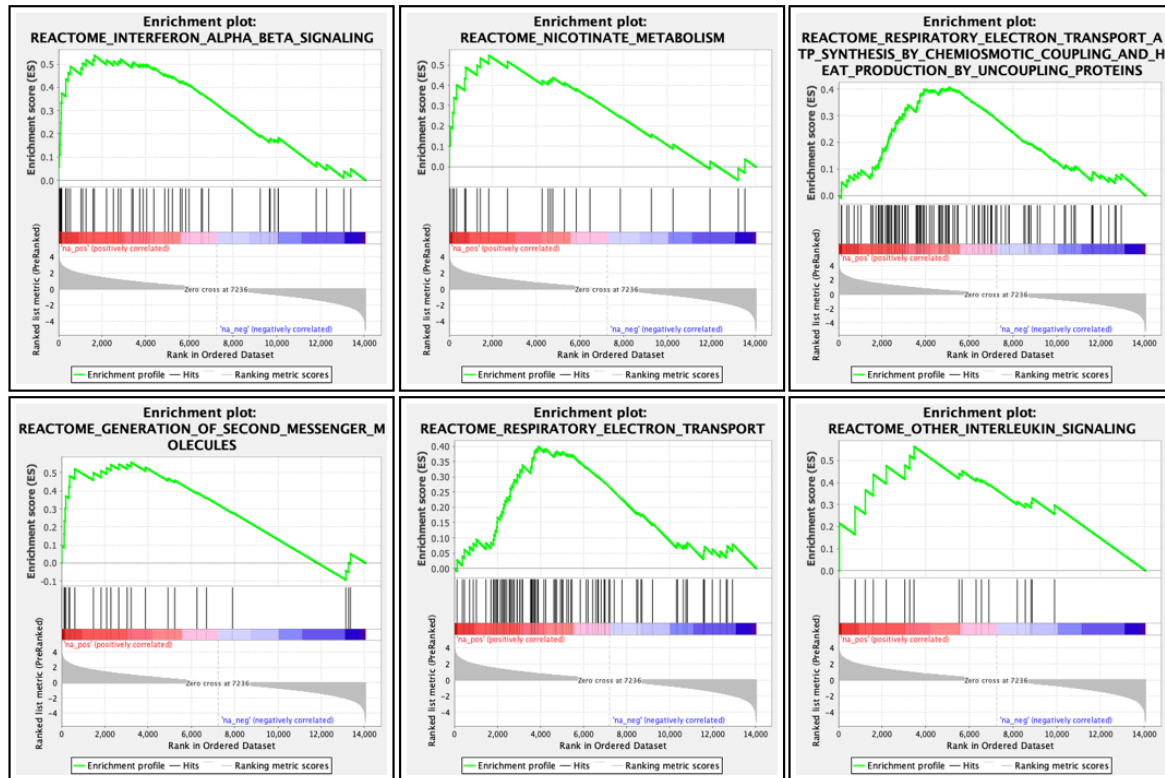
Figure S 2: Volcano plots showing the differentially expressed genes according to patient clinical phenotype. SC versus nonSC (a), PTC versus Progressors (b), Late versus Early (c), Late versus Intermediate (d) and Intermediate versus Early (e). Red is used for genes significantly differentially expressed.
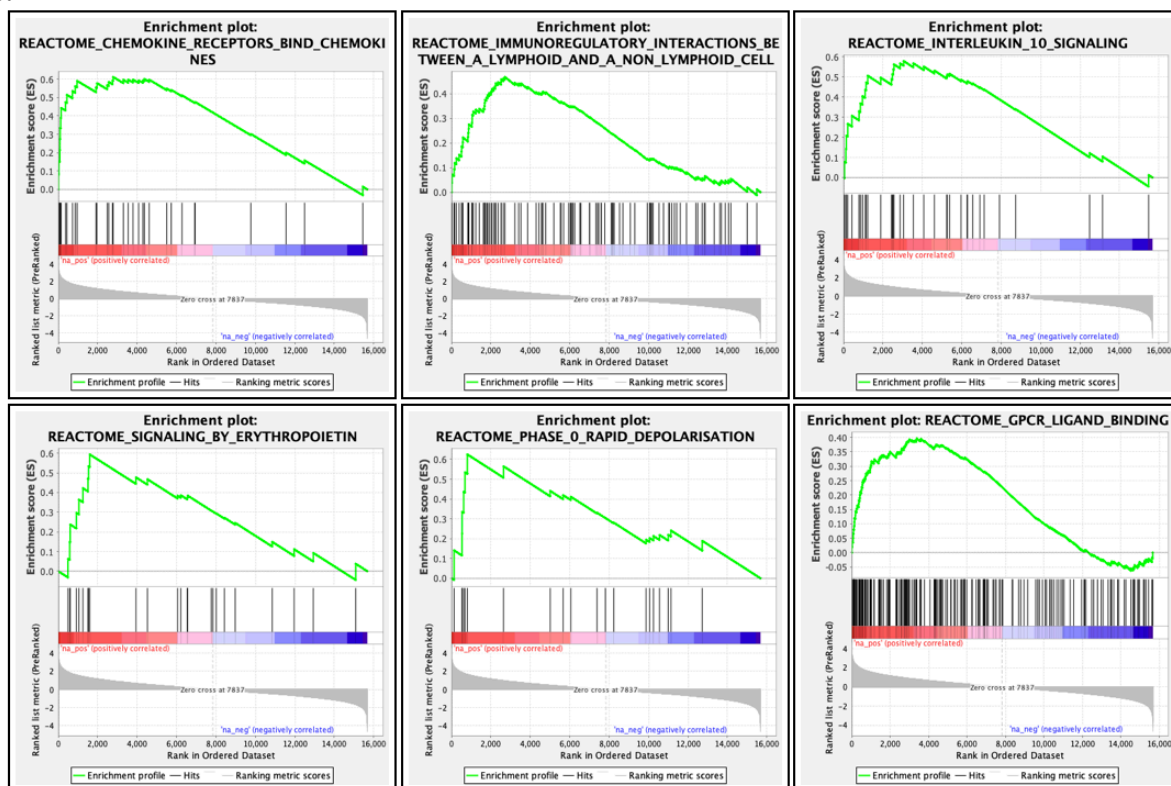
a.

Figure S 3: GSEA analyses according to clinical phenotype. top 6 pathways for Late versus Early (a), Late versus Intermediate (b) and Intermediate versus Early(c).
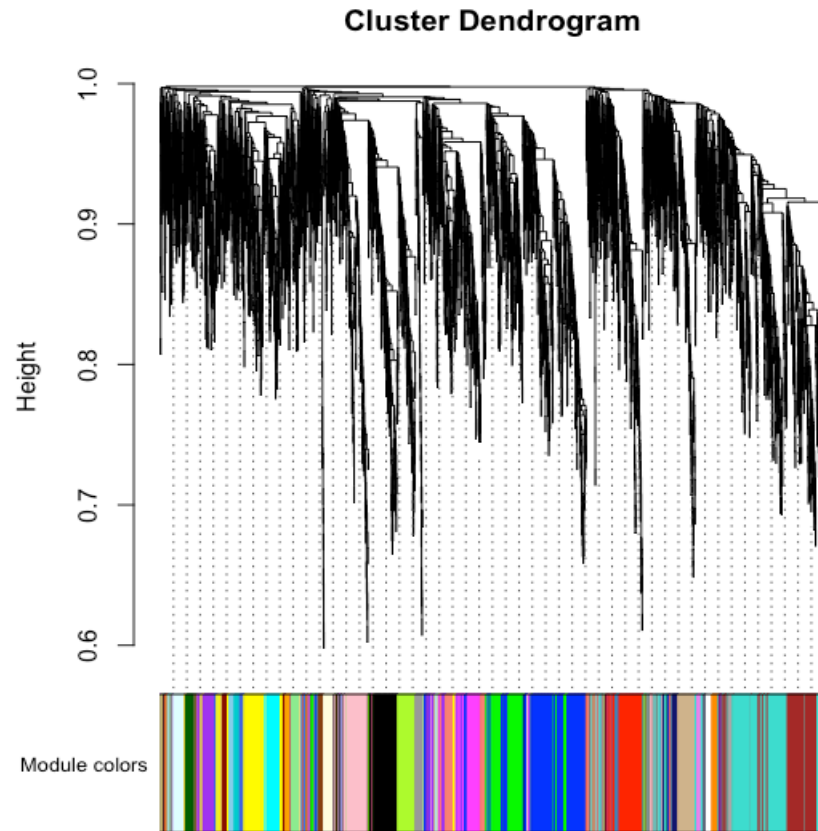
Figure S 4: GSEA analyses according to clinical phenotype. top 6 pathways for Late versus Early (a), Late versus Intermediate (b) and Intermediate versus Early(c).