Oral tenofovir disoproxil fumarate/emtricitabine for HIV pre-exposure prophylaxis increases expression of type I/III interferon-stimulated factors in the gastrointestinal tract but not in the blood

Author list

Sean M. Hughes[1], Claire N. Levy[1], Fernanda Calienes[2], Joanne D. Stekler[3,15,17], Urvashi Pandey[1], Lucia Vojtech[1], Alicia R. Berard[4,5], Kenzie Birse[4,5], Laura Noel-Romas[4,5], Brian Richardson[6], Jackelyn Golden[6], Michael Cartwright[6], Ann C. Collier[15], Claire E. Stevens[15], Marcel E. Curlin[7,8,9], Timothy H. Holtz[7,8], Nelly Mugo[10], Elizabeth Irungu[10], Elly Katabira[11], Timothy Muwonge[11], Javier R. Lama[12], Jared M. Baeten[3,15,17], Adam Burgener[4,5,13], Jairam R. Lingappa[14,15,17], M. Juliana McElrath[2,15,16,17], Romel Mackelprang[17], Ian McGowan[18,19], Ross D. Cranston[20], Mark J. Cameron[6] and Florian Hladik[1,2,15]

[1] Department of Obstetrics and Gynecology, University of Washington, Seattle, Washington, United States of America.
[2] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America.
[3] Department of Epidemiology, University of Washington, Seattle, Washington, United States of America.
[4] National HIV and Retrovirology Labs, JC Wilt Infectious Disease Research Centre, Public Health Agency of Canada, Winnipeg, Manitoba, Canada.
[5] Department of Medical Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada.
[6] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, United States of America.
[7] Thailand Ministry of Public Health-U.S. Centers for Disease Control and Prevention Collaboration, Nonthaburi, Thailand.
[8] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America.
[9] Department of Medicine, Division of Infectious Diseases, Oregon Health and Sciences University, Portland, Oregon, United States of America
[10] Partners in Health Research and Development, Kenya Medical Research Institute, Thika, Kenya
[11] Infectious Disease Institute, Makerere University, Kampala, Uganda
[12] Asociación Civil Impacta Salud y Educación, Lima, Peru.
[13] Unit of Infectious Diseases, Department of Medicine Solna, Centre for Molecular Medicine, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden.
[14] Department of Pediatrics, University of Washington, Seattle, Washington, United States of America.
[15] Department of Medicine, University of Washington School of Medicine, Seattle, Washington, United States of America.
[16] Department of Laboratory Medicine, University of Washington, Seattle, Washington, United States of America.
[17] Department of Global Health, University of Washington, Seattle, Washington, United States of America.
[18] Magee-Womens Research Institute and Foundation, Pittsburgh, Pennsylvania, United States of America.
[19] Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America.
[20] Department of Infectious Diseases and Dermatology, University of Barcelona, Barcelona, Spain

Abstract

Tenofovir disoproxil fumarate and emtricitabine are used for HIV treatment and pre-exposure prophylaxis. Previously, we found that topical rectal application of tenofovir 1% gel caused many gene expression changes. Here, we measured RNA and protein expression in several clinical trials of oral administration in HIV-uninfected individuals (using microarrays, RNAseq, droplet digital PCR, mass spectrometry, and microscopy). We found tens to hundreds of differentially expressed genes in the gastrointestinal tract, but none in the blood or female reproductive tract. In rectal samples from one trial, most of the 13 upregulated genes were related to type I/III interferon signaling. Similar changes were seen at the protein level in the same trial and in the duodenum and rectum in another trial. We conclude that tenofovir disoproxil fumarate and emtricitabine have little effect on gene expression in the blood or female reproductive tract but increase type I/III interferon signaling in the gut. This effect may enhance their anti-viral efficacy when used as pre-exposure prophylaxis, in particular to prevent rectal HIV transmission. However, it may also contribute to chronic immune activation and HIV reservoir maintenance in chronically treated people living with HIV.

Introduction

Tenofovir disoproxil fumarate (TDF) and emtricitabine (FTC) are nucleotide/nucleoside reverse transcriptase inhibitors (NRTIs) used for HIV treatment and oral pre-exposure prophylaxis (PrEP). Treatment with TDF/FTC and other combination antiretroviral therapy drugs has changed the lives of HIV-infected individuals, allowing lifespans approaching those of HIV-uninfected individuals. In addition, when used as PrEP, TDF/FTC is extremely effective at preventing HIV infection among individuals at high risk of infection[1,2].

Our studies in MTN-007, a phase 1 randomized, placebo-controlled safety study of topical rectal application of tenofovir 1% gel, revealed changes to rectal gene expression after once-daily application for 7 days[3,4]. The genes affected by tenofovir in MTN-007 suggested perturbations in mucosal immune homeostasis, mitochondrial function, and regulation of epithelial cell differentiation and survival. In MTN-007, tenofovir was formulated as a 1% reduced glycerin gel and applied topically to the rectum. In clinical practice, TDF (the prodrug of tenofovir), and especially co-formulated TDF/FTC, are widely used in oral form for both treatment and prevention. We therefore sought determine if oral administration also causes gene expression changes. Because lifelong treatment is required, even small drug effects on gene expression could have significant consequences, possibly even related to HIV infection and persistence.

To determine whether TDF or TDF/FTC cause gene expression changes when delivered orally, we performed transcriptome-level gene expression studies with samples from several trials of oral PrEP in HIV-uninfected individuals. We looked for gene expression changes in the blood as well as in gastrointestinal and female reproductive tract biopsies. The trials included TDF used alone as well as in combination with FTC, with samples taken after about two months of treatment. In some of these studies, we performed additional experiments using RNA sequencing, droplet digital PCR (ddPCR), mass spectrometry, and microscopy.

We found differentially expressed genes in the upper and lower gastrointestinal tract, but not in the blood or the female reproductive tract. In both the duodenum and rectum, we found increased expression of genes related to type I/III interferon signaling. Similar changes were found at the protein level by mass spectrometry and microscopy. Given the widespread use of TDF/FTC, as well as other NRTI drugs, the relatively limited gene expression changes were reassuring. However, we were surprised by the consistency of increased expression of type I/III interferon genes in the gastrointestinal tract. The long-term health importance of this change in gene expression is unclear: it could cause greater antiviral efficacy, in particular when used to prevent rectal HIV transmission, or it could contribute to chronic immune activation in people living with HIV.

Results

We measured the effect of TDF/FTC or TDF alone on blood, gastrointestinal biopsies, and female reproductive tract biopsies from several trials of oral PrEP in HIV-uninfected individuals. We measured gene expression by microarray, RNAseq, and ddPCR and protein expression by mass spectrometry and microscopy. Samples were obtained from the Genital Mucosal Substudy (GMS)[5] of the Partners PrEP Study[6] (paired blood and female reproductive tract biopsies during and after treatment, GMS A, and unpaired placebo vs. treatment blood, GMS B), the Microbicide Trials Network trial 017[7] (MTN-017; paired rectal biopsies before and during treatment); and ACTU-3500 (paired rectal and duodenal biopsies and blood before and during treatment). Sample sizes are given in Table 1.

Differentially expressed genes

Gene expression was measured with microarrays. The expression of each probe was compared between no treatment and treatment with oral TDF or TDF/FTC, generally paired within individuals as described in the methods. Differential expression was defined by an FDR-adjusted p-value less than 0.05, with up-regulation meaning higher expression during treatment. As shown in Figure 1 and Table 1, differentially expressed genes were found in two studies: MTN-017 rectal samples (13 genes up, n = 36 paired samples) and ACTU-3500 duodenal samples (116 genes up and 135 genes down, n = 8 paired samples). No differentially expressed genes were found in any of the other study arms to our chosen FDR-adjusted p-value threshold of 0.05. Gene lists for every study arm and sample type are in the supplemental data.

Study	Drug	Sample	N	Genes up	Genes down
MTN-017	TDF/FTC	Rectum	36 pairs	13	0
ACTU-3500	TDF/FTC	Duodenum	8 pairs	116	135
ACTU-3500	TDF/FTC	Rectum	8 pairs	0	0
ACTU-3500	TDF/FTC	Whole blood	8 pairs	0	0
ACTU-3500	TDF/FTC	PBMC	8 pairs	0	0
GMS A	TDF/FTC	Vagina	3 pairs	0	0
GMS A	TDF	Vagina	12 pairs	0	0
GMS A	TDF	Ectocervix	9 pairs	0	0
GMS A	TDF	PBMC	10 pairs	0	0
GMS B	TDF	PBMC	36 drug 20 placebo	0	0
GMS B	TDF/FTC	PBMC	26 drug 20 placebo	0	0

Table 1. Differentially expressed genes as defined by an FDR-adjusted p-value less than 0.05, with up indicating higher expression during drug treatment.

All 13 genes differentially expressed in the rectum in MTN-017 were expressed more highly during treatment with TDF/FTC than prior to treatment. As shown in Table 2, seven of these 13 genes related to type I interferon signaling as defined by membership in the gene ontology biological process “type I interferon signaling pathway” (GO:0060337). Of the six genes not included in GO:0060337, several have been identified in the literature as nonetheless induced by type I interferon: DDX60[42,43], SAMD9[42], IFI27L1[44], and HERC6[45]. Thus only two of the thirteen genes (the pseudogene MROH3P and CCDC77) have no reported roles related to type I interferon. Gene ontology overrepresentation analysis of these thirteen genes revealed biological processes related to type I interferon and response to virus as the overrepresented processes with the

smallest adjusted p-values (complete overrepresentation analysis results are provided as supplemental data files).

Entrez ID	Gene	Gene name	Link to type I/III interferon	Fold change (log2)	FDR
3429	IFI27	interferon alpha inducible protein 27	Gene set	0.76	3.54E-7
2537	IFI6	interferon alpha inducible protein 6	Gene set	0.80	2.27E-4
3434	IFIT1	interferon induced protein with tetratricopeptide repeats 1	Gene set	0.75	2.49E-3
9636	ISG15	ISG15 ubiquitin-like modifier	Gene set	0.98	2.49E-3
91543	RSAD2	radical S-adenosyl methionine domain containing 2	Gene set	0.53	3.14E-3
4599	MX1	MX dynamin like GTPase 1	Gene set	0.92	3.60E-3
4938	OAS1	2'-5'-oligoadenylate synthetase 1	Gene set	0.32	0.016
122509	IFI27L1	interferon alpha inducible protein 27 like 1	Literature[44]	0.25	2.27E-4
55601	DDX60	DExD/H-box helicase 60	Literature[42,43]	0.39	2.49E-3
54809	SAMD9	sterile alpha motif domain containing 9	Literature[42]	0.35	0.012
55008	HERC6	HECT and RLD domain containing E3 ubiquitin protein ligase family member 6	Literature[42]	0.31	0.049
647215	MROH3P	maestro heat like repeat family member 3, pseudogene	None reported	0.22	0.021
84318	CCDC77	coiled-coil domain containing 77	None reported	0.12	0.048

Table 2. Genes differentially expressed (higher during daily oral TDF/FTC use) in rectal biopsies in MTN-017. “Gene set” indicates membership in GO:0060337. “Literature” indicates that a link to type I/III interferon has been reported in the indicated articles.

While the dominant role of genes differentially expressed in rectal tissue during daily oral TDF/FTC was related to type I/III interferon signaling, the differentially expressed genes in the duodenum had different roles. For the genes that were downregulated in the duodenum during treatment, the top biological processes related to cellular metabolism, suggesting that oral TDF/FTC caused reduced expression of genes involved in metabolism, especially of lipids and carbohydrates. The upregulated genes were related to a variety of biological processes including RNA splicing and phospholipid transport (complete overrepresentation analysis results are provided as supplemental data files).

Correlation of fold changes across study arms

While no genes met the FDR < 0.05 threshold for differential expression in the rectum in ACTU-3500, the fold changes of the 13 differentially expressed genes from the rectum in MTN-017 were strongly correlated with the fold changes of these same genes in the rectum in ACTU-3500 and two, ISG15 and DDX60, had unadjusted p-values less than 0.05 (Figure 2A, left). The Spearman correlation coefficient of the fold changes between the two studies for those 13 genes was 0.94, as compared to 0.07 for all other genes detected in the rectum in both studies.

Similarly, there was a strong correlation between the fold changes of the 13 differentially expressed genes in the rectum from MTN-017 with the fold changes of the same genes in the duodenum in ACTU-3500, as shown

in Figure 2A (right). Only one of these genes had an FDR-adjusted p-value of less than 0.05 in the duodenum, but all 13 had positive fold changes and unadjusted p-values less than 0.05. The Spearman correlation coefficient for these 13 genes was 0.81, as compared to 0.06 for genes that were not differentially expressed in either study. The correlation of the differentially expressed genes from the duodenum with the same genes in the rectum in MTN-017 was 0.13.

Gene set testing of gene sets of differentially expressed genes

To further compare expression patterns across study arms, we created three gene sets from the differentially expressed genes: the 13 genes that were higher during treatment in the rectum, the 116 that were higher in the duodenum, and the 135 that were lower in the duodenum. We then used gene set testing to test whether these gene sets were enriched in other study arms. As shown in Figure 2B, the set of genes upregulated in the rectum was positively enriched (higher expression during treatment) in the rectum and duodenum in the ACTU-3500 study as well as the ectocervix (Figure 2B, left) in the GMS A study. This means that the genes that were upregulated by TDF/FTC in the rectum in MTN-017 tended to also be expressed more highly during TDF/FTC in the rectum and duodenum in ACTU-3500. The same gene set was up in one PBMC (Figure 2B, right) arm in the GMS A study, down in whole blood, and neither in three other PBMC arms. This result suggests that changes in gene expression seen in the rectum in MTN-017 also occurred in other mucosal tissues in other studies, but not in the blood.

The 116 genes that were upregulated in the duodenum were also enriched in the same direction in both rectal study arms and in one PBMC study arm. They were enriched in the opposite direction in the vagina and ectocervix. The 135 genes that were downregulated in the duodenum were only enriched in two study arms, both of which were PBMC. This result suggests that gene expression changes seen in the duodenum were less generalizable to other other tissues and study arms than the changes in the rectum. Complete gene set testing results are provided as supplementary data files.

Taken together, the results shown in Figures 1 and 2 suggest that the oral use of TDF or TDF/FTC affects the expression of relatively few genes. In particular, we did not find any evidence of differential gene expression in the blood, despite assessing both PBMC and whole blood in several distinct cohorts. We did detect differentially expressed genes in the gastrointestinal tract, specifically the rectum and the duodenum. The 13 genes that were differentially expressed in the rectum in MTN-017 also increased in expression in the duodenum and rectum in ACTU-3500, with large correlation coefficients for their fold changes. The genes that were differentially expressed in the duodenum, by contrast, did not have highly correlated fold changes in other study arms. Gene set testing supported this finding. This suggests that those 13 genes, most of which are interferon-related, may represent an underlying biological pathway that is affected by TDF/FTC in the gastrointestinal tract.

Gene set testing of Hallmark gene sets

To assess higher level biological effects, we tested the fifty Hallmark gene sets[46] against each study arm and specimen type. Each gene set comprises genes that are involved in a biological state or process. The gene sets that had adjusted p-values below 0.05 in at least two study arms for tissue and blood are shown in Figure 3A-B. Complete gene set testing results are provided as supplementary data files.

For tissue (Figure 3A), thirteen gene sets had adjusted p-values below 0.05 in at least two study arms. Of these gene sets, four were related to immunity (allograft rejection, interferon-α and -γ responses, and TNF-α signaling via NF-κB), and four were related to cell proliferation (E2F targets, G2M checkpoint, and two MYC targets gene sets). In all but one case, the immune related gene sets were enriched in the positive direction,

meaning that the genes in the sets tended to be expressed at higher levels during product use. All four of the immune-related gene sets were enriched in the ectocervix, and several were enriched in the gastrointestinal tract or vagina, in particular the interferon-α response gene set. All four proliferation-related sets were enriched in the vagina and duodenum during TDF/FTC use, being higher in the duodenum and lower in the vagina. Two proliferation-related gene sets (Myc targets v1 and v2) went up in the rectal biopsies from ACTU-3500. The same two gene sets went down in the rectal biopsies from MTN-017. Overall, TDF/FTC seemed to induce interferon-α responses in the gastrointestinal tract and to suppress inflammatory responses in the blood.

Only two gene sets had adjusted p-values below 0.05 in at least two study arms for blood samples (Figure 3B). Both of the gene sets were immune-related (complement, interferon-γ response, and TNF-α signaling via NF-κB). In all cases, these gene sets were enriched in the negative direction, meaning that the genes in the sets tended to be expressed at lower levels during product use. This result suggests that TDF/FTC has a somewhat dampening effect on interferon-γ and tumor necrosis factor-α-related gene expression in the blood.

ddPCR validation of microarray data

We used ddPCR to measure the expression of three genes—ISG15, IFI6, and MX1—chosen to represent the type I/III interferon-related genes that increased in expression in the duodenum and rectum during TDF/FTC treatment. Out of the 13 genes that were differentially expressed in MTN-017, these three genes had the highest fold changes in the rectum in MTN-017 and ACTU-3500 and were among the top five in the duodenum in ACTU-3500. The expression of each gene of these three genes was normalized to the expression of ubiquitin C (UBC), which was chosen as reference due to the stability of its expression across tissues and treatments in the microarray data.

The fold changes for ddPCR and microarray are shown in Figure 4. In all cases, the changes were in the same direction by ddPCR as they were by microarray, and the sizes of the changes were similar between the two methodologies, though the magnitudes tended to be larger by ddPCR than by microarray. The participant-level fold changes calculated from ddPCR measurements correlated well with the microarray data. The Pearson correlations were 0.93 for MX1, 0.92 for ISG15, and 0.90 for IFI6. Taking each gene separately and stratifying by study, sample type and gene, Pearson correlations ranged from 0.54 to 0.99, with a mean of 0.86 and median of 0.92. These data suggest that the microarray data provide reasonable estimates of gene expression levels and fold changes and add confidence to our conclusions that genes related to type I/III interferon are induced in the gut, but not the blood, by oral TDF/FTC.

RNA sequencing of MTN-017

In addition to microarrays, the RNA from MTN-017 was analyzed by RNA sequencing, to potentially take advantage of its wider dynamic range and ability to pick up novel transcripts and/or variants. As shown in Figure 5AB, two genes were differentially expressed in the RNAseq data (at an FDR-adjusted p-value threshold of 0.05), both higher during TDF/FTC treatment than at baseline: IFIT1, which was also differentially expressed by microarray, and SLC6A20 (solute carrier family 6 member 20), which was not. When we looked at correlations of fold changes between the microarray and the RNAseq data from MTN-017, we saw an overall Spearman correlation coefficient of 0.34. The correlation coefficient was 0.84 when only looking at the genes differentially expressed by microarray (Figure 5B). In other words, the genes that were differentially expressed by microarray tended to have similar fold changes by RNAseq.

By virtue of having both microarray and RNAseq data on the same samples, we were able to look at genes that had similar fold changes by both methodologies. We looked at genes that had log2-fold changes below -0.25

or above 0.25 by both assays. Only twelve genes fell into the downregulated group and none had adjusted p-values below 0.05 by either assay but, strikingly, eight were metallothioneins (MT1A, 1E, 1F, 1G, 1H, 1M, 1X, and 2A), which bind to heavy metal ions. There were 28 genes with fold changes above 0.25 by both microarray and RNAseq, including 11 of the 13 genes that were differentially expressed (as defined by adjusted p-value) by microarray. Many of the additional genes were also related to type I/III interferon signaling: IRF7, IRF9, OAS2, OAS3, IFITM1, and IFI44L, for example. Overrepresentation analysis of these 28 genes (with or without the 11 differentially expressed genes from the microarray) again yielded many gene ontology biological processes related to type I/III interferon signaling.

Mass spectrometry-based proteomics of MTN-017

Rectal biopsies from MTN-017 were run by mass spectrometry for protein identification. The biopsies were run in two batches, with the samples from American participants in one batch and samples from Thai participants in the other. In both cases, no proteins were detected as differentially expressed after adjustment for multiple comparisons (Figure 5C). A complete list of protein fold changes is provided as supplementary data.

When we compared the fold changes of all proteins or the proteins that had unadjusted p-values below 0.05 between transcript and protein (Figure 5D), we found little correlation. However, we were particularly interested in assessing the fold changes of the protein forms of the 13 differentially expressed genes we identified by microarray in the same study. None of these proteins were present in the American participant batch. However, five (ISG15, MX1, OAS1, DDX60 and SAMD9) were present in the samples from the Thai participants. Four of the five proteins had positive fold changes (all except SAMD9) and there was a strong correlation between the fold changes detected by microarray and mass spectrometry (Spearman correlation coefficient of 0.8; Figure 5D). These data suggest that the upregulation of certain type I/III interferon genes by TDF/FTC seen at the RNA level may extend to the protein level.

Microscopy

Because we found increased expression of some type I IFN-related genes in the gut during TDF/FTC use, we evaluated the expression of ISG15 and MX1 by immunofluorescence microscopy in duodenal and rectal biopsies from ACTU-3500. Slides were evaluable from 8 pairs of duodenal biopsies and 6 pairs of rectal biopsies (before and during treatment). ISG15 was expressed by few if any stromal cells. In contrast, we found ISG15 expression in essentially all epithelial cells (Figure 6A) and its expression could be divided into dim and bright cells (Figure 6A). The intensity of ISG15 staining did not change within either category of cell with TDF/FTC use (Figure 6B). However, the percentage of cells falling into the bright category (Figure 6C) increased in both the rectum (increase of 0.43 percentage points, Bonferroni-adjusted p-value = 0.009) and the duodenum (0.43 percentage points, Bonferroni-adjusted p-value = 0.25).

Because MX1 was expressed in the epithelium and the stroma, we assessed its expression in both locations (Figure 6D). Within the stroma, preliminary test staining of biopsies from two individuals showed that MX1 was associated mostly with DC-SIGN$^+$ dendritic cells (Figure 5D), whereas few CD68$^+$ macrophages or CD3$^+$ T lymphocytes expressed MX1 (not shown). We therefore co-stained for DC-SIGN in all biopsies and stratified stromal cell analysis by DC-SIGN$^+$ and DC-SIGN$^-$ cells. We did not detect a change in the intensity of MX1 expression in any of the groups (Figure 5E; all Bonferroni-adjusted p-values > 0.45). The differences were small and increased in most cases. In summary, ISG15 and MX1 protein levels were mostly stable by immunofluorescence microscopy, but TDF/FTC use increased the proportion of cells expressing high levels of ISG15.

Signature of TDF/FTC effect on the gastrointestinal mucosa

In total, we had data on gene expression after TDF/FTC treatment for two months from two different anatomical sites in the gastrointestinal tract (rectum and duodenum), two different studies (MTN-017 and ACTU-3500), and several different assays (microarray, RNAseq, ddPCR, and mass spectrometry-based proteomics). As shown above (Figures 2, 4 and 5), the differentially expressed genes from the rectum in MTN-017 by microarray had similar fold changes when assessed by different assays, in different studies, and at different sites. These data are summarized in Figure 7 and show that in general the 13 genes were consistently upregulated during TDF/FTC treatment regardless of study, assay, or site in the gastrointestinal tract. In particular, IFI6 and IFI27 (which with the also differentially expressed IFI27L1 form three of the four FAM14 family members[47]) as well as ISG15, MX1, DDX60, RSAD2, and IFIT1 showed very consistent upregulation. On the other hand, CCDC77 and MROH3P—the only two genes in the set that have not been shown to be related to type I/III interferon signaling—do not show convincing evidence of being affected by treatment, suggesting that these two genes may have been false positives. The overall consistency of the results, in particular for the genes related to type I/III interferon, strengthens the evidence for TDF/FTC-induced changes in the gastrointestinal mucosa.

Discussion

In the studies described here, we found that oral treatment with TDF/FTC or TDF alone had limited effects on host gene expression, with no differentially expressed genes in the blood or female reproductive tract and tens to hundreds in the gastrointestinal tract. Notably, though, genes related to type I/III interferon signaling were common among the TDF/FTC-induced genes in the gut, with good agreement between the microarray hybridization and RNA sequencing platforms. ddPCR testing of ISG15, IFI6, and MX1 expression levels confirmed the microarray/RNAseq findings. Protein-level data by mass spectrometry-based proteomics, as well as focused immunohistology of gut sections for MX1 and ISG15, were largely in congruence.

The limited gene expression changes we found here indicate that TDF/FTC have few off-target effects on host gene expression. TDF/FTC's limited effects on host gene expression are reassuring in light of its widespread use among HIV-infected individuals for treatment and HIV-uninfected individuals for prevention.

The main gene expression change caused by TDF/FTC was increased expression of interferon-stimulated genes in the gastrointestinal tract. We found 13 differentially expressed genes in the rectal samples from MTN-017, eleven of which were type I/III interferon-related. As shown in Figure 7, the differentially expressed genes from the rectum from the MTN-017 microarrays changed in similar ways in other studies, anatomical sites, and assays. The fold changes of these genes in MTN-017 strongly correlated with their fold changes in the rectum and duodenum in ACTU-3500, a smaller, independent cohort, though in that study most of these genes had FDR-adjusted p-values > 0.05. In addition, similar changes were seen to these genes as measured by ddPCR and RNA sequencing, as well as to protein forms as measured by proteomics and immunohistology. These eleven type I/III interferon-related genes may represent a signature of TDF/FTC-induced changes in the gastrointestinal tract. A targeted assay measuring expression of some of these genes could provide a tool for future studies to monitor the effect of TDF/FTC use at other sites and time points, including in treated HIV-infected individuals, a group not studied here.

Type I and type III interferons stimulate very similar interferon-stimulated genes, making it difficult to differentiate which type of interferon is being affected on the basis of changes to interferon-stimulated genes alone[48]. We have therefore referred to these genes as type I/III interferon-related throughout. Type III interferon signaling plays an important role in the epithelium and its receptors are predominantly expressed on epithelial cells[49]. The fact that we saw changes in interferon-stimulated genes only in mucosal samples may indicate that the changes seen here are type III interferon signaling, but determining with confidence whether TDF/FTC affects primarily type I or type III interferons will require further research.

Moreover, we cannot yet explain how TDF/FTC induce the type I/III interferon gene expression changes. Our data did not provide information about whether the gene expression changes were initiated by IFN-α, IFN-β (type I interferons) or IFN-λ (type III interferon) secretion, as these cytokines were not detectable by microarray, RNAseq, or mass spectrometry. Nevertheless, our data indicate that the finely tuned balance between type I/III activating signals that induce antiviral states and promote immune responses and suppressive signals that limit toxicity are perturbed by TDF/FTC[50]. Intracellular upregulation of the three most highly and consistently induced factors, ISG15, MX1 and IFI6, increases the antiviral state of a cell[51]. However, ISG15, for example, can also be secreted, and adopt cytokine-like functions that may affect neighboring and even distant immune cells[52].

Thus, stimulation of interferon pathways by TDF/FTC could have beneficial or detrimental effects. On the one hand, increased type I/III interferon signaling could enhance innate immune response readiness, in particular anti-viral preventative or treatment efficacy. On the other hand, it may play a role in the persistent immune

activation seen in HIV-infected patients. The causes of chronic immune activation despite fully suppressive antiretroviral therapy (ART) remain unclear; it is possible that long-term ART may itself contribute through stimulation of type I interferon pathways in the gastrointestinal tract. The effects we observed after two months of drug use were relatively mild (less than 2-fold in the rectum, less than 4- to 6-fold in the duodenum). However, their biological impact may accumulate over long periods of treatment, especially in HIV-infected individuals that develop comorbidities already associated with chronic inflammation as they age, such as cardiovascular disease[53].

By changing type I/III interferon signaling, TDF/FTC may affect the latent HIV reservoir as well. Blocking of interferon-stimulated gene signaling reduced the frequency of HIV reservoir cells and caused delayed HIV rebound after ART discontinuation in two studies in humanized mice[54–56]. In contrast, interferon-α treatment of ART-treated macaques failed to reduce the size of the HIV reservoir[57]. Thus, type I/III interferon pathway stimulation by TDF/FTC could have a stabilizing effect on the HIV reservoir.

It is notable that the type I/III interferon-pathway stimulation we observed with TDF/FTC occurred in the gastrointestinal tract and was not detected in the blood. The gastrointestinal tract is where HIV spreads most devastatingly during initial infection and where a large latent reservoir is established[58]. On ART the vast majority of latently infected cells capable of producing virus resides in the gut[59]. Thus, the gene expression changes we observed occur at precisely the site where they may be most likely to affect the HIV reservoir. The explanation for gene expression changes being restricted to the gastrointestinal tract may be the high concentrations achieved there with oral dosing. However, the detection of some gene expression changes (at the gene set level and perhaps by ddPCR) in the ectocervix (where drug concentrations are presumably lower) argues against this. Optimization of formulation to avoid excessive drug concentrations in the gut may reduce off-target effects. Tenofovir alafenamide, for example, yields significantly lower drug levels in the gut but retains equivalent efficacy compared to TDF, the formulation studied here[60].

Oral PrEP more effectively prevents rectal than vaginal HIV transmission[61]. The gene expression changes caused by TDF/FTC occurred primarily in the gastrointestinal mucosa and not in the female genital tract. If the induction of type I/III interferon genes by TDF/FTC causes increased antiviral efficacy, this could be an explanation for oral PrEP's greater efficacy in the rectum than in the vagina. Other explanations that have been proposed include pharmacokinetic differences in tenofovir levels between vaginal and rectal tissues[62] and perturbations of tenofovir metabolism by a dysbiotic vaginal microbiome[63].

In addition to changes to immune-related gene expression in the gastrointestinal tract, the other notable gene expression changes we observed related to cell proliferation. Specifically, we saw some evidence of increased expression of gene sets related to cell proliferation in the duodenum, and reduced expression of these gene sets in the vagina. Proliferation-related gene sets were conflicting in the rectum (two sets up in one study, one set down in the other). It is difficult to speculate about the clinical relevance of changes to cell proliferation pathways. Factors associated with cell cycle regulation are crucial to balance cell proliferation with cell death, and for cells to respond to DNA damage. TDF/FTC's effect on cell cycle processes in the duodenum may not be surprising given the increased drug concentrations likely achieved in the upper gastrointestinal tract. Long-term, high concentration oral administration of TDF to mice has been reported to cause a low incidence of duodenal tumors (Canadian product monograph for VIREAD) and liver adenomas (US prescribing information for VIREAD). It is unclear whether our findings with human duodenal biopsies relate in any way to these outcomes in rodents and no such findings have been reported during human use.

Our study has a few limitations. One is the use of within-person comparisons between on- and off-drug. As participants were aware of when they were or were not taking an intervention, behavioral changes or other factors than the drug itself could explain the gene expression changes we observed. The only study where placebo was compared to treatment is the GMS B study, in which the only available samples were PBMC and no differentially expressed genes were observed. Another limitation is that the ddPCR and RNAseq confirmations were performed on the same RNA samples as the microarrays. We were also limited in our ability to differentiate effects of FTC from those of TDF, given that only a limited set of participants received TDF alone and none received FTC alone. Importantly, we also do not know if the effects of TDF and FTC extend to other drugs of the NRTI class. This is an important question, which we will attempt to answer in a planned clinical trial comparing NRTI-based ART regimens to an NRTI-sparing one. The NRTI-sparing arm will reveal whether switching away from NRTI-based therapy may decrease chronic immune activation and affect the dynamics of the latent reservoir.

In conclusion, our results are reassuring in terms of safety of oral TDF or TDF/FTC for therapy and prevention and thought-provoking in terms of the potential ramifications of increased type I/III interferon-related gene expression in the gastrointestinal tract. Increased expression of the 13 differentially expressed genes from MTN-017—or of ISG15, IFI6, and MX1 by themselves—may represent a signature of the effect of TDF/FTC on the gastrointestinal tract. Follow-up studies are required to clarify whether this newly uncovered effect has any bearing on chronic immune activation or HIV reservoir dynamics.

Methods

Studies

Samples were used from four studies, which are described in Table 3. Two sets of samples were used from the Genital Mucosal Substudy (GMS)[5] of the Partners PrEP Study[6]: paired samples during and after treatment (GMS A) and unpaired placebo vs. treatment samples (GMS B). The samples all came from the same parent study, but were processed separately. The Microbicide Trials Network trial 017 (MTN-017) included oral TDF/FTC as well as topical tenofovir; only the oral TDF/FTC samples are included in this analysis. Rectal biopsies taken after two months of oral TDF/FTC use were compared to baseline samples. ACTU-3500 followed nine men initiating oral PrEP with TDF/FTC in Seattle, with a baseline visit and a visit after two months of PrEP use. Eight men completed both visits; the ninth moved out of state before the second visit. Ethics reviews are published in the primary manuscripts for the GMS and MTN-017 studies (listed in Table 3). ACTU-3500 was reviewed through the University of Washington institutional review board, number 49167. Sample sizes varied within each trial depending on drug (TDF/FTC or TDF) and sample type. Complete sample size information is listed in Table 1.

Study	Drug	Gender	Control	Treatment	Samples	References
GMS A	TDF or TDF/FTC	Women	2 months after treatment cessation	24-36 months of treatment	Ectocervical Vaginal PBMC	NCT02621242[5,6]
GMS B	TDF or TDF/FTC	Women	Placebo	24-36 months of treatment	PBMC	NCT02621242[5,6]
MTN-017	TDF/FTC	Men, transgender women	Pre-treatment	2 months of treatment	Rectal	NCT01687218[7]
ACTU-3500	TDF/FTC	Men	Pre-treatment	2 months of treatment	Rectal Duodenal PBMC Whole blood	NCT02621242

Table 3. Characteristics of studies from which samples were obtained. GMS Genital Mucosal Substudy of the Partners PrEP study, MTN-017 Microbicide Trials Network study 017, ACTU-3500 AIDS Clinical Trial Unit Study 3500, TDF tenofovir disoproxil fumarate, FTC emtricitabine, PBMC peripheral blood mononuclear cells.

Adherence

Tenofovir levels were measured in the serum for the GMS A and GMS B studies as previously described[8]. Samples from the treatment arm without detectable tenofovir were removed, as was one post-treatment sample, where drug was unexpectedly detected.

Tenofovir levels were measured in serum in the MTN-017 study. Adherence was high in the oral arm of this study, with 94% of participants taking the daily pill at least 80% of the time[9]. Participant report of pill use was largely concordant with serum tenofovir levels, with only 4.4% of serum samples having no detectable tenofovir when participants reported product use. All samples from this study were included.

Adherence was determined for the ACTU-3500 study by participant self-report. All participants reported daily use of TDF/FTC throughout the study period.

Sample processing and storage

Vaginal, cervical, and rectal biopsies were obtained in GMS A and B and MTN-017 as described in the primary manuscripts[5–7]. Rectal biopsies were obtained in ACTU-3500 by anoscopy using Radial Jaw 4 biopsy forceps

(Boston Scientific). Duodenal biopsies were obtained by esophagogastroduodenoscopy under light anesthesia also using Radial Jaw 4 biopsy forceps. Biopsies were placed into RNALater Stabilization Solution (ThermoFisher Scientific, Waltham, MA, USA) and held at 4°C for 24 h and then frozen at -80°C. PBMC were isolated from whole blood by density gradient centrifugation and then cryopreserved and stored in the vapor phase of a liquid nitrogen freezer. Whole blood was drawn into PAXgene tubes (PreAnalytiX, Hombrechtikon, Switzerland), which were frozen at -20°C for 24 h and then stored at -80°C.

RNA extraction and quality control

Biopsies were thawed at room temperature. They were transferred with forceps into 600 µL of Buffer RLT (Qiagen, Hilden, Germany) and homogenized using a Bio-Gen PRO200 homogenizer (PRO Scientific, Oxford, CT, USA) followed by passing 10 times through a needle and syringe. RNA was extracted using the RNeasy fibrous tissue mini kit (Qiagen) automated on a QIAcube (Qiagen). PBMC were thawed, washed by centrifugation, counted, and RNA was extracted using the RNeasy mini kit (Qiagen) on a QIAcube. PAXgene tubes were thawed and held at room temperature for 3 h with occasional mixing, in order to completely lyse red blood cells. RNA was extracted from PAXgene samples using the PAXgene Blood RNA kit (PreAnalytiX) according to the manufacturer's instructions. RNA was stored at -80°C until use.

Sample quality control

Cell viability was measured prior to RNA extraction for PBMC samples (Guava Viacount, EMD Millipore, Burlington, MA, USA). For the GMS A study, viabilities were available from 49/70 PBMC samples. Of these, the mean viability was 85%, and 95% of samples had viabilities of 71% or greater. For the ACTU-3500 study, the mean cell viability was 94%, and 95% of samples had a viability of 88% or greater. For the GMS B study, the mean cell viability was 84%, and 95% of samples had a viability of 62% or greater.

The quality of all RNA was determined using the RNA integrity number as calculated from the TapeStation R6K assay (Agilent, Santa Clara, CA, USA) and the concentration was determined by NanoDrop (ThermoFisher). Samples were deemed of sufficient quality for use with an RNA integrity number of six or greater. For the GMS A study, the mean RIN was 8.3 and 95% of samples had a RIN of 7 or greater. For the MTN-017 study, the mean RIN was 8.3 and 95% of samples had a RIN of 7.4 or greater. For the ACTU-3500 study, the mean RIN was 8.5 and 95% of samples had a RIN of 7.4 or greater. For the GMS B study, the mean RIN was 9.1 and 95% of samples had a RIN of 8.1 or greater.

Microarray labeling and hybridization

For the GMS A and GMS B studies, samples were prepared for microarray using 50 ng of total RNA with the Ovation PicoSL WTA System V2 kit (NuGEN, San Carlos, CA, USA) and labeled with the Encore BiotinIL kit (NuGEN). The Illumina TotalPrep RNA Amplification kit (ThermoFisher) was used to prepare samples for microarray from the ACTU-3500 study (275 ng total RNA as input) and the MTN-017 study (500 ng input), with input sizes chosen based on available RNA.

750 ng of the labeled cDNA (from the NuGEN kits) or cRNA (from the Thermo kit) was hybridized to HumanHT-12 v4 Expression BeadChips (Illumina, San Diego, CA, USA) and scanned by the Fred Hutch Genomics Core facility. Images were converted to expression data using GenomeStudio (Illumina).

Microarray analysis

Microarray analysis was done using R. All microarrays were pre-processed within study and sample type using variance stabilizing transformation[10] and robust spline normalization from the lumi package[11]. Probes that were rarely expressed in a given study arm were removed.

Differential gene expression was assessed using the limma package[12] and gene set testing was done using the camera function from the same package[13]. In general, paired models were fit, with modeling done separately for each sample type and study (due to the separate preprocessing of different sample types). For the ACTU-3500 and MTN-017 study, within-participant comparisons were done comparing baseline samples to those obtained at the end of two months of treatment. Similar within-participant comparisons were done for the GMS A study, with the difference that samples during treatment were compared to samples taken two months after the end of treatment. For the GMS B study, treatment samples were compared to samples taken at the same time point from placebo recipients. Probe-level p-values were adjusted for multiple comparisons to control the false discovery rate using the method of Benjamini and Hochberg[14]. Individual probes with an adjusted p-value of less than 0.05 were defined as differentially expressed. The same threshold was used for gene sets, where p-values were adjusted using the false discovery rate for the number of gene sets times the number of study arms tested (e.g. 50 Hallmark gene sets * 11 study arms). Because there are multiple probes for some genes, probes were collapsed into genes by taking the probe with the lowest adjusted p-value for each gene[15].

Thirteen cervical samples from the GMS A study clustered separately on principal components analysis plots from the rest of the cervical samples. Differential gene expression analysis comparing this cluster to the rest of the cervical samples revealed tens of thousands of genes to be differentially expressed and suggested that these samples may have included endocervical tissue, rather than only ectocervical as intended, possibly due to cervical ectopy. Keratin genes and gene ontology processes related to keratinocyte and epidermal development were higher in the main group of cervical samples, while the small cluster had higher expression of processes related to cilia movement and development, consistent with the ciliated epithelial cells of the endocervix. Because gene expression differed so dramatically in the thirteen cervical samples in question, we removed them from the analysis.

Reverse transcription and ddPCR of selected genes

We used all the paired samples with sufficient RNA available to repeat measurements of transcript levels for MX dynamin like GTPase 1 (MX1), ISG15 ubiquitin-like modifier (ISG15), and interferon-α inducible protein 6 (IFI6) by ddPCR assay. As the reference gene, we used Ubiquitin C (UBC), selected based on a comparison of commonly used reference genes in the microarray data. Among those genes, UBC was expressed in all sample types and the average fold change (across different sample types and studies) was close to 0 and the standard deviation was small. The samples used for ddPCR were the same as shown in Table 1, except that the GMS B samples were not used (because they were unpaired), the number of sample pairs was reduced by one each for the vaginal and ectocervical samples from the TDF arm of the GMS A study because of insufficient RNA remaining from those samples. The vaginal samples from the TDF/FTC arm of the GMS A study were not tested by ddPCR because the sample size was so low (only three pairs of samples).

Reverse transcription was performed using 100 ng of RNA per sample in a 20 µL reaction mixture using qScript cDNA Synthesis Kit (QuantaBio, Beverly, MA, USA) according to the manufacturer's instructions. The incubation conditions were 22°C for 5 minutes, 42°C for 30 min, and then 85°C for 5 min. After reverse transcription, the samples were diluted to 100 µL with water and 5 µL (cDNA equivalent of 5 ng RNA) was used per ddPCR well.

The primers and probes used for ddPCR are shown in Table 4 and were purchased from Integrated DNA Technologies (Skokie, IL, USA). Assays were run in duplex (IFI6 on the FAM channel with MX1 on the HEX channel in one set of wells and ISG15 on the FAM channel with UBC on the HEX channel in a second set of

wells). Each sample was run in duplicate for each assay. Sample pairs (i.e. on- and off-treatment) were always run on the same plates. ddPCR was performed using ddPCR Supermix for Probes (no dUTP), with droplets generated on a QX200 Automated Droplet Generator and droplets read on a QX200 Droplet Reader according to the manufacturer's instructions (Bio-Rad, Hercules, CA, USA).

Target	Sequence Name	Component	Sequence
IFI6	Hs.PT.58.4407609	Probe	/56-FAM/CCA AGG TCT /ZEN/AGT GAC GGA GCC C/3IABkFQ/
		Primer 1	GTA GCA CAA GAA AAG CGA TAC C
		Primer 2	CTG CTG TGC CCA TCT ATC AG
MX1	Hs.PT.58.38362411	Probe	/5HEX/CTT GGA ATG /ZEN/GTG GCT GGA TGG C/3IABkFQ/
		Primer 1	CAT TCA GTA ATA GAG GGT GGG A
		Primer 2	TGA AAT CTG GAG TGA AGA ACG C
ISG15	Hs.PT.58.39185901.g	Probe	/56-FAM/CAC CTG GAA /ZEN/TTC GTT GCC CGC /3IABkFQ/
		Primer 1	GCC TTC AGC TCT GAC ACC
		Primer 2	CGA ACT CAT CTT TGC CAG TAC A
UBC	Hs.PT.39a.22214853	Probe	/5HEX/TCG ATG GTG /ZEN/TCA CTG GGC TCA AC/3IABkFQ/
		Primer 1	CCT TAT CTT GGA TCT TTG CCT TG
		Primer 2	GAT TTG GGT CGC AGT TCT TG

Table 4. Primers and probes used for ddPCR.

The ddPCR data was analyzed using QuantaSoft version 1.7.4.0917 (Bio-Rad). The same fluorescence thresholds were applied to all samples across all plates. Wells with fewer than 10,000 droplets were removed. Concentrations of IFI6, MX1 and ISG15 were divided by the concentration of UBC from the corresponding sample to yield copies of each gene per copy of UBC. This value was log2-transformed to convert it to a normal distribution and place it on a comparable scale to the microarray data. Replicate wells were then averaged. Fold changes were calculated by subtracting the expression level from the off-treatment sample from the on-treatment sample.

RNA sequencing in MTN-017

Total RNA prepared above was normalized to 300 ng input for library preparation with the TruSeq Stranded Total RNA with Ribo-Zero Globin kit (Illumina). The resulting libraries were assessed on the Agilent Fragment Analyzer with the HS NGS assay (Agilent) and quantified using the KAPA Library Quantification Kit (Roche) on a ViiA 7 Real Time PCR platform (Thermo Fisher). High depth sequencing (50 million reads per sample) was performed with a HiSeq 2500 (Illumina) on two High Output v4 flow cells as a 50 base pair, paired-end run. Raw demultiplexed fastq paired end read files were trimmed of adapters and filtered using the program skewer[16] to remove any reads with an average phred quality score of less than 30 or a length of less than 36 bp. Trimmed reads were aligned using the HISAT2[17] aligner to the Homo sapiens NCBI reference genome assembly version GRCh38 and sorted using SAMtools[18]. Aligned reads were counted and assigned to gene meta-features using the program featureCounts[19] as part of the Subread package. Counts data were analyzed analogously to the microarray data, using the voom function from limma and then fitting models for each transcript. Because the samples were processed in two batches, batch number was included in the model in addition to participant ID and treatment.

Proteomics in MTN-017

Frozen rectal biopsies from MTN-017 were processed as described previously[20]. For protein extraction, tissues were washed 3 times with 10 mM Tris (pH 7.6), placed in 5 mL of a lysis solution consisting of 7 M Urea, 2 M Thiourea, 40 mM Tris, and 10 mM DTT, and homogenized with a gentleMACS Octo Dissociator (RNA02-01M setting, Miltenyi Biotec, Bergisch Gladbach, Germany). Precipitates were removed by centrifugation at 9000 g for 20 minutes at 4ºC, transfer of supernatant to a new tube, and a second round of centrifugation at 15,000 g for 20 minutes. Supernatants were stored at -80ºC. Trypsin digestion was performed as described previously[21]. Briefly, for each sample, 600 µL of tissue lysate was denatured in urea exchange buffer (8 M Urea in 1:10 0.5 M HEPES:water solution, GE HealthCare, Uppsala, Sweden) and filtered through a 10 kDa membrane. Filtered lysates were alkylated with 50 mM iodoacetamide for 20 minutes, and then washed with 50 mM HEPES buffer. Nucleic acids were removed by treatment with benzonase (150 units/µL in HEPES with $MgCl_2$, Novagen, Darmstadt, Germany) for 30 minutes, and then lysates were washed with HEPES buffer. Trypsin digestion (2 µg trypsin per 100 µg protein, Promega, WI, USA) was performed overnight at 37ºC. Eluted peptides were dried using a speed vacuum and then stored at -80ºC. Reverse-phase liquid chromatography using a step-wise gradient was used to remove salts and detergents. Peptide quantification was performed with the LavaPep Fluorescent Protein and Peptide Quantification Kit (Gel Company, San Francisco, CA, USA).

Mass spectrometry was performed using a nano flow liquid chromatography system (Easy nLC, Thermo Fisher) connected inline to a Velos Orbitrap mass spectrometer as described previously[22,23]. One µg of peptide was run for each sample. Feature detection, normalization, and quantification were performed using Progenesis LC-Mass Spectrometry software (Nonlinear Dynamics, Newcastle upon Tyne, UK) with default settings. Peptides were found using Mascot v.2.4.0 (Matrix Science, Boston, MA, USA) to search against the SwissProt database[24] restricting taxonomy to Human. Search results were imported into Scaffold (Proteome Software, Portland, OR, USA) for peptide identification, requiring ≤0.1 FDR for protein identification, ≤0.01 FDR for peptide identification, and at least 2 unique peptides identified per protein. Samples were run in two batches, with US participants in one batch and Thai participants in the other. Data from the two cohorts were combined using Combat[25].

Immunofluorescence microscopy

Pairs of rectal and duodenal biopsies from eight subjects were stained for ISG15 or for MX1 and DC-SIGN protein for immunofluorescence microscopy. Each pair consisted of one pre- and one on-treatment (~60 days) sample from the ACTU-3500 study. Two rectal biopsies were of poor quality, so they and their pairs were excluded from analysis, reducing the sample size for the rectal biopsies to six pairs. The biopsies were collected into RNAlater Stabilization Solution (ThermoFisher), held at 4 ºC overnight, and then stored at -80 ºC. Prior to use, biopsies were thawed, fixed in 10% neutral buffered formalin for 3 days, and stored in 70% ethanol until paraffin embedding. Four micron thick sections were cut, attached to positively-charged slides and baked at 60 ºC for 1 h, with each slide holding one pre- and one on-treatment tissue section from the same participant. The histopathologist and data analyst were blinded to treatment status. Staining was performed using the procedure described previously[26]. Primary antibodies were anti-MX1 antibody (Atlas Antibodies Cat#HPA030917, RRID: AB_10602010), anti-DC-SIGN (Santa Cruz (DC28): sc-65740), and anti-ISG15 (Atlas Antibodies Cat#HPA004627, RRID: AB_1079152).

Slides were scanned on an Aperio FL (Version 2; Leica Biosystems). Exposure times were 125 ms for ISG15 and 64 ms for DAPI. For the triple stain, exposure times were 20 ms for MX1, 125 ms for DC-SIGN and 80 ms DAPI. Images were analyzed with HALO v2.2 image analysis software (Indica Labs, Albuquerque, NM, USA) with the CytoNuclear FL v1.4 algorithm. Images were annotated manually to select stroma or epithelium, which were

analyzed separately. Individual cells were identified by the software via DAPI-stained nuclei in conjunction with cell-defining parameters including nuclear contrast threshold, minimum nuclear intensity, nuclear segmentation aggressiveness, nuclear size, minimum nuclear roundness and maximum nuclear radius. These parameters were set to be optimal for each pair (i.e. settings were the same for the on- and off-treatment pairs for each person and sample type). The ISG15 signal was so bright that the software identified surrounding cells as positive for ISG15, despite manual inspection clearly showing that only one central cell was positive. To correct for this, hierarchical clustering was used on the spatial positions of the cells identified as positive. A distance threshold was empirically determined to count adjacent cells as a single positive cell, while still identifying nearby but distinct positive cells as distinct. The output of this analysis was verified by comparison to manual counting. Moreover, there was a strong correlation between the percentage of cells that were bright for ISG15 before and after adjustment for falsely positive surrounding cells ($r = 0.98$ for duodenum and 0.95 for rectum), the numbers were simply lower (and more reflective of manual inspection) after adjustment.

Data analysis

Data were initially processed using instrument-specific software as described in the above sections. Following export from instrument-specific software, data were analyzed using R version 3.5.2[27] and the following packages from CRAN or Bioconductor[28]: AnnotationDbi[29], Biobase[28], broom[30], conflicted[31], edgeR[32], ggrepel[33], here[34], limma[12], lumi[11], msigdbr[35], org.Hs.eg.db[36], pander[37], patchwork[38], plater[39], RColorBrewer[40], and tidyverse[41]. R was run through RStudio version 1.1.463.

Acknowledgments

We wish to express gratitude to all study volunteers for their participation. We are grateful to Max Abou and Lauren Girard for proteomic wet lab support as well as Stuart McCorrister and Garrett Westmacott for mass spectrometry technical support. We acknowledge the Fred Hutch Experimental Histopathology (Sunni Farley, Savanh Chanthaphavong, Li-Ya Huang) and Genomics (Cassie Sather, Crissa Bennett) core facilities for their assistance.

This work was funded by NIH R01AI116292 (to FH), NIH R01AI111738 (to JRL), Bill and Melinda Gates Foundation #47674 (to JRL), NIH R01AI134293 (to RM), NIH AI027757 (to JMB). MJC and IM were supported by the Microbicide Trials Network (UM1AI068633, Sharon Hillier, PI). The ddPCR portion of this work was supported by a grant from the James B. Pendleton Charitable Trust. The NCI 5 P30 CA015704-44 Cancer Center Support Grant supported the Fred Hutch Experimental Histopathology core facility.

Citations

1. Okwundu, C. I., Uthman, O. A. & Okoromah, C. A. Antiretroviral pre-exposure prophylaxis (PrEP) for preventing HIV in high-risk individuals. *Cochrane Database Syst. Rev.* **7**, CD007189 (2012).
2. US Preventive Services Task Force. Preexposure Prophylaxis for the Prevention of HIV Infection: US Preventive Services Task Force Recommendation Statement. *JAMA* **321**, 2203–2213 (2019).
3. McGowan, I. *et al.* A phase 1 randomized, double blind, placebo controlled rectal safety and acceptability study of tenofovir 1% gel (MTN-007). *PLoS One* **8**, e60147 (2013).
4. Hladik, F. *et al.* Mucosal effects of tenofovir 1% gel. *Elife* **4**, (2015).
5. Lund, J. M. *et al.* HIV-1-Neutralizing IgA Detected in Genital Secretions of Highly HIV-1-Exposed Seronegative Women on Oral Preexposure Prophylaxis. *J. Virol.* **90**, 9855–9861 (2016).
6. Baeten, J. M. *et al.* Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N. Engl. J. Med.* **367**, 399–410 (2012).
7. Cranston, R. D. *et al.* MTN-017: A Rectal Phase 2 Extended Safety and Acceptability Study of Tenofovir Reduced-Glycerin 1% Gel. *Clin. Infect. Dis.* **64**, 614–620 (2017).
8. Baeten, J. M. *et al.* Single-agent tenofovir versus combination emtricitabine plus tenofovir for pre-exposure prophylaxis for HIV-1 acquisition: an update of data from a randomised, double-blind, phase 3 trial. *Lancet. Infect. Dis.* **14**, 1055–1064 (2014).
9. Carballo-Diéguez, A. *et al.* High levels of adherence to a rectal microbicide gel and to oral Pre-Exposure Prophylaxis (PrEP) achieved in MTN-017 among men who have sex with men (MSM) and transgender women. *PLoS One* **12**, e0181607 (2017).
10. Lin, S. M., Du, P., Huber, W. & Kibbe, W. a. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* **36**, e11 (2008).
11. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–8 (2008).
12. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
13. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).
14. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1994).
15. Ramasamy, A., Mondry, A., Holmes, C. C. & Altman, D. G. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* **5**, e184 (2008).
16. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
17. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–60 (2015).
18. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
19. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–30 (2014).
20. Burgener, A. *et al.* A systems biology examination of the human female genital tract shows compartmentalization of immune factor expression. *J. Virol.* **87**, 5141–50 (2013).
21. Birse, K. M. *et al.* Unbiased proteomics analysis demonstrates significant variability in mucosal immune factor expression depending on the site and method of collection. *PLoS One* **8**, e79505 (2013).
22. Birse, K. *et al.* Molecular Signatures of Immune Activation and Epithelial Barrier Remodeling Are Enhanced during the Luteal Phase of the Menstrual Cycle: Implications for HIV Susceptibility. *J. Virol.* **89**, 8793–8805 (2015).
23. Birse, K. D. *et al.* Genital Injury Signatures and Microbiome Alterations Associated With Depot

Medroxyprogesterone Acetate Usage and Intravaginal Drying Practices. *J. Infect. Dis.* **215**, 590–598 (2017).
24. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
25. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–27 (2007).
26. Paulson, K. G. *et al.* Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. *Nat. Commun.* **9**, 3868 (2018).
27. R Core Team. R: A Language and Environment for Statistical Computing. (2018).
28. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
29. Pagès, H., Carlson, M., Falcon, S. & Li, N. AnnotationDbi: Annotation Database Interface. (2018).
30. Robinson, D. broom: Convert Statistical Analysis Objects into Tidy Data Frames. (2016).
31. Wickham, H. conflicted: An Alternative Conflict Resolution Strategy. (2018).
32. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
33. Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. (2018).
34. Müller, K. here: A Simpler Way to Find Your Files. (2017).
35. Dolgalev, I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. (2018).
36. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. (2018).
37. Daróczi, G. & Tsegelskyi, R. pander: An R 'Pandoc' Writer. (2018).
38. Pedersen, T. L. patchwork: The Composer of ggplots. (2017).
39. Hughes, S. M. plater: Read, Tidy, and Display Data from Microtiter Plates. *J. Open Source Softw.* **1**, (2016).
40. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. (2014).
41. Wickham, H. tidyverse: Easily Install and Load 'Tidyverse' Packages. (2016).
42. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–5 (2011).
43. Miyashita, M., Oshiumi, H., Matsumoto, M. & Seya, T. DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol. Cell. Biol.* **31**, 3802–19 (2011).
44. Parker, N. & Porter, A. C. G. Identification of a novel gene family that includes the interferon-inducible human genes 6-16 and ISG12. *BMC Genomics* **5**, 8 (2004).
45. Kane, M. *et al.* MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* (2013). doi:10.1038/nature12653
46. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
47. Cheriyath, V., Leaman, D. W. & Borden, E. C. Emerging roles of FAM14 family members (G1P3/ISG 6-16 and ISG12/IFI27) in innate immunity and cancer. *J. Interferon Cytokine Res.* **31**, 173–81 (2011).
48. Stanifer, M. L., Pervolaraki, K. & Boulant, S. Differential regulation of type I and type III interferon signaling. *Int. J. Mol. Sci.* **20**, 1–22 (2019).
49. Pervolaraki, K. *et al. Differential induction of interferon stimulated genes between type I and type III interferons is independent of interferon receptor abundance. PLoS pathogens* **14**, (2018).
50. Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nat. Rev. Immunol.* **14**, 36–49 (2014).
51. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* **32**, 513–45 (2014).
52. Perng, Y.-C. & Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* **16**, 423–439

(2018).

53. Longenecker, C. T., Sullivan, C. & Baker, J. V. Immune activation and cardiovascular disease in chronic HIV infection. *Curr. Opin. HIV AIDS* **11**, 216–25 (2016).
54. Zhen, A. *et al.* Targeting type I interferon-mediated activation restores immune function in chronic HIV infection. *J. Clin. Invest.* **127**, 260–268 (2017).
55. Deeks, S. G., Odorizzi, P. M. & Sekaly, R.-P. The interferon paradox: can inhibiting an antiviral mechanism advance an HIV cure? *J. Clin. Invest.* **127**, 103–105 (2017).
56. Cheng, L. *et al.* Blocking type I interferon signaling enhances T cell recovery and reduces HIV-1 reservoirs. *J. Clin. Invest.* **127**, 269–279 (2017).
57. Palesch, D. *et al.* Short-Term Pegylated Interferon α2a Treatment Does Not Significantly Reduce the Viral Reservoir of Simian Immunodeficiency Virus-Infected, Antiretroviral Therapy-Treated Rhesus Macaques. *J. Virol.* **92**, (2018).
58. Cohen, M. S., Shaw, G. M., McMichael, A. J. & Haynes, B. F. Acute HIV-1 Infection. *N. Engl. J. Med.* **364**, 1943–54 (2011).
59. Estes, J. D. *et al.* Defining total-body AIDS-virus burden with implications for curative strategies. *Nat. Med.* **23**, 1271–1276 (2017).
60. Fletcher, C. V. *et al.* Lymphoid tissue pharmacokinetics of tenofovir-alafenamide vs -disoproxil fumarate. *Conf. Retroviruses Opportunistic Infect. Seattle.* 103 (2019).
61. Buchbinder, S. P. Maximizing the Benefits of HIV Preexposure Prophylaxis. *Top. Antivir. Med.* **25**, 138–142 (2018).
62. Cottrell, M. L. *et al.* A Translational Pharmacology Approach to Predicting Outcomes of Preexposure Prophylaxis Against HIV in Men and Women Using Tenofovir Disoproxil Fumarate With or Without Emtricitabine. *J. Infect. Dis.* **214**, 55–64 (2016).
63. Klatt, N. R. *et al.* Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938–945 (2017).

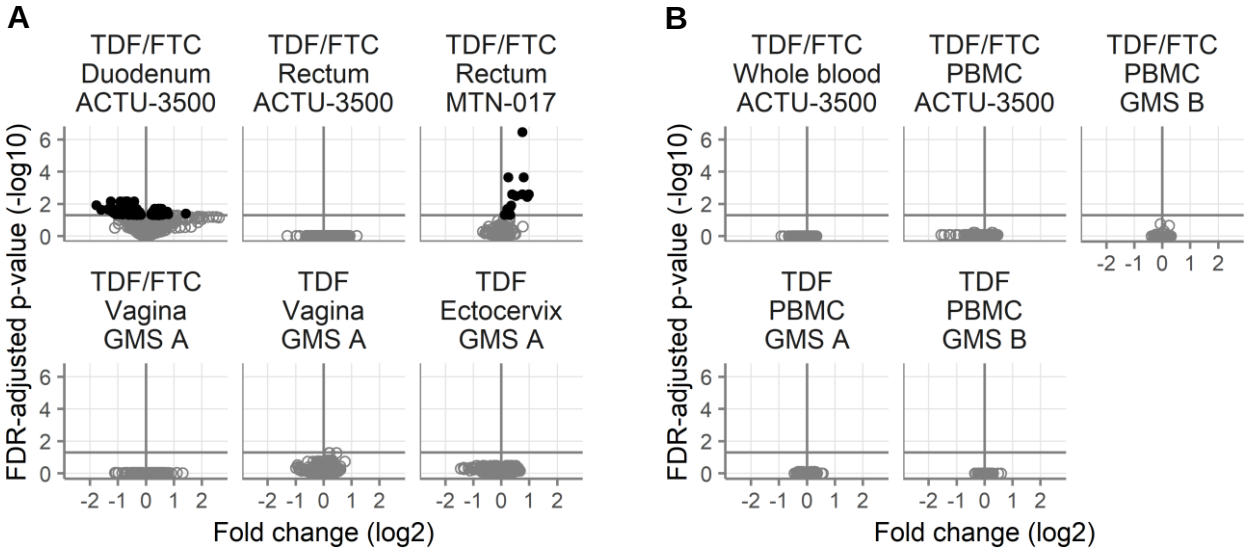

Figure 1. Gene expression changes during oral TDF or TDF/FTC product use. Changes in gene expression in tissue (A) and blood (B) samples. Each point represents the expression of a single gene. Positive fold changes indicate higher expression during product use and negative fold changes indicate lower expression during product use. Filled black symbols indicate genes with FDR-adjusted p-values less than 0.05, while open gray symbols indicate genes with FDR-adjusted p-values greater than 0.05.

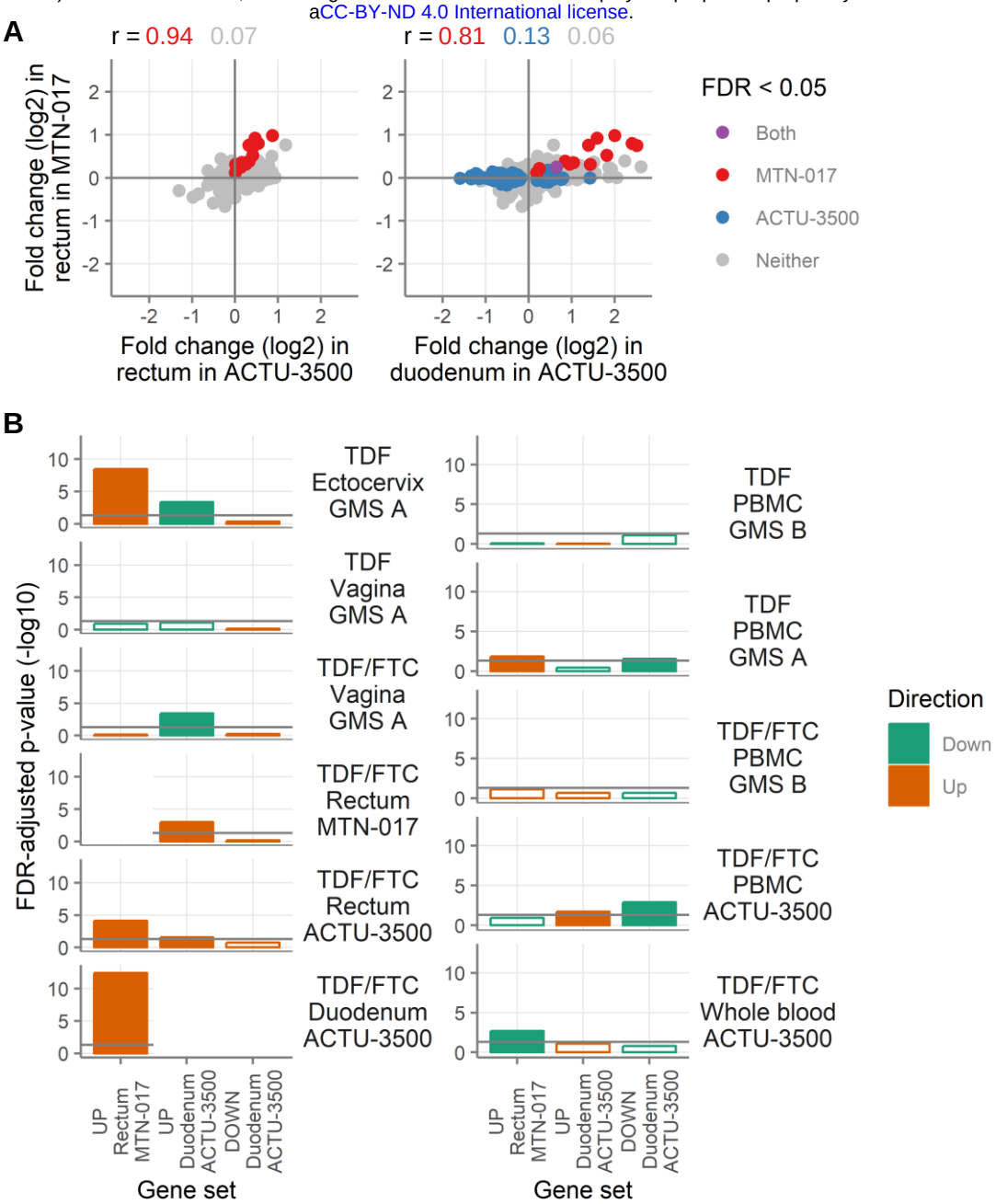

Figure 2. Consistent changes to differentially expressed genes. (A) Fold changes of all genes detected in the rectal samples from MTN-017 compared to the rectal samples from ACTU-3500 (left) and the duodenal samples from ACTU-3500 (right). Colors indicate genes with FDR-adjusted p-values less than 0.05 in MTN-017 (red), ACTU-3500 (blue), both (purple), or neither (gray). Spearman correlation coefficients for the genes falling into each subset are shown. (B) Gene sets composed of differentially expressed genes from the rectum in MTN-017 or the duodenum in ACTU-3500 were tested for enrichment against other study arms in the mucosa (left) and blood (right). Bars indicate the result of a gene set test for the gene set shown on the x-axis tested against the study labeled at right. Filled bars indicate an FDR-adjusted p-value less than 0.05 and open symbols the opposite, with bar length showing the -log10 of the FDR-adjusted p-value. Colors indicate the direction of change, with “up” meaning more expression during product use and “down” meaning the opposite. The horizontal grey lines show an FDR-adjusted p-value of 0.05. Gene sets were not tested against the study arm they came from; this is indicated by a lack of bar and of gridlines.

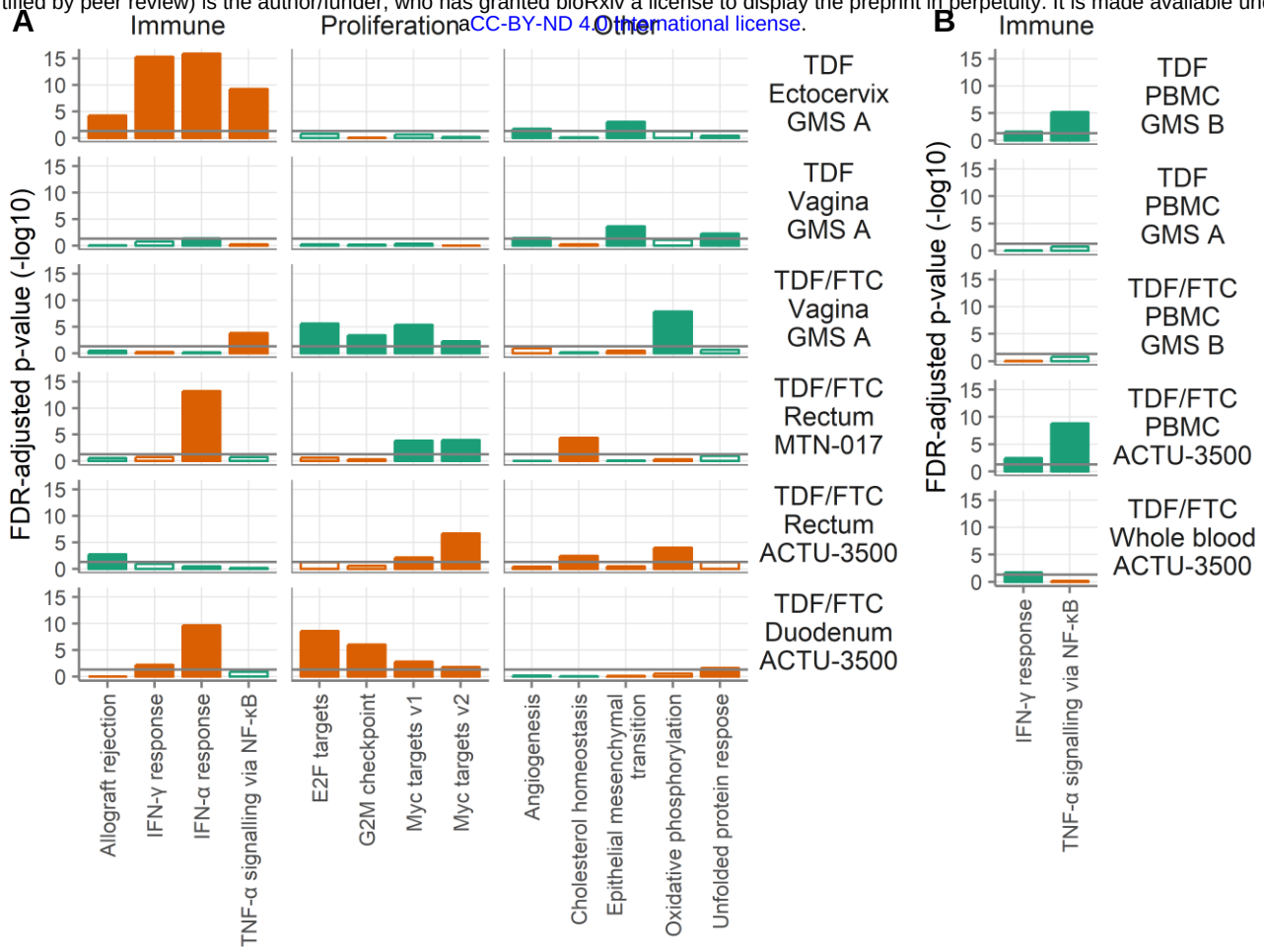

Figure 3. Gene set testing of Hallmark gene sets. The Hallmark gene sets were tested for enrichment in the mucosa (A) and the blood (B). All gene sets that had an FDR-adjusted p-value less than 0.05 in at least two of the mucosa (A) or blood (B) data sets. Bars indicate the result of a gene set test for the gene set shown on the x-axis tested against the study shown at right. Filled bars indicate an FDR-adjusted p-value less than 0.05 and open bars the opposite, with bar length proportional to the -log10 of the FDR-adjusted p-value. Colors indicate the direction of change, with orange meaning more expression during product use and green meaning less. The horizontal grey lines show an FDR-adjusted p-value of 0.05. The gene sets are grouped into categories as labeled at the top.

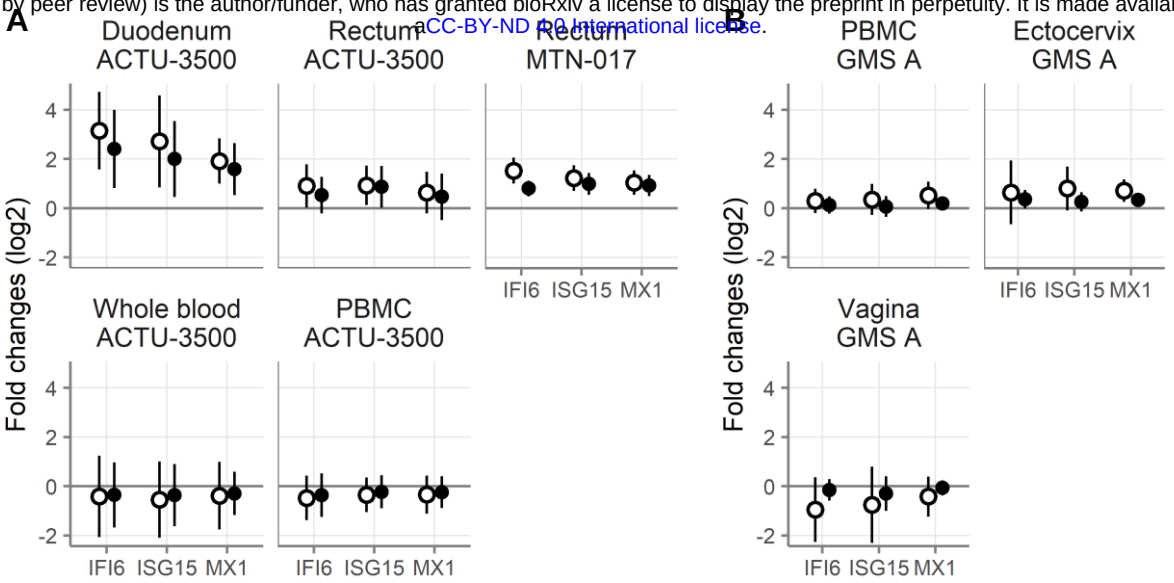

Figure 4. Comparison of gene expression changes measured by ddPCR or microarray. Fold changes in gene expression of three genes (IFI6, ISG15, or MX1) as detected by ddPCR (open symbols) and microarray (filled symbols) after treatment with TDF/FTC (A) or TDF alone (B). Symbols show the mean across all participants and vertical lines show the 95% confidence intervals of the mean. IFI6 stands for interferon alpha inducible protein 6, ISG15 for ISG15 ubiquitin-like modifier, and MX1 for MX dynamin like GTPase 1. A positive fold change means higher expression during treatment and a negative fold change means higher expression off of treatment.

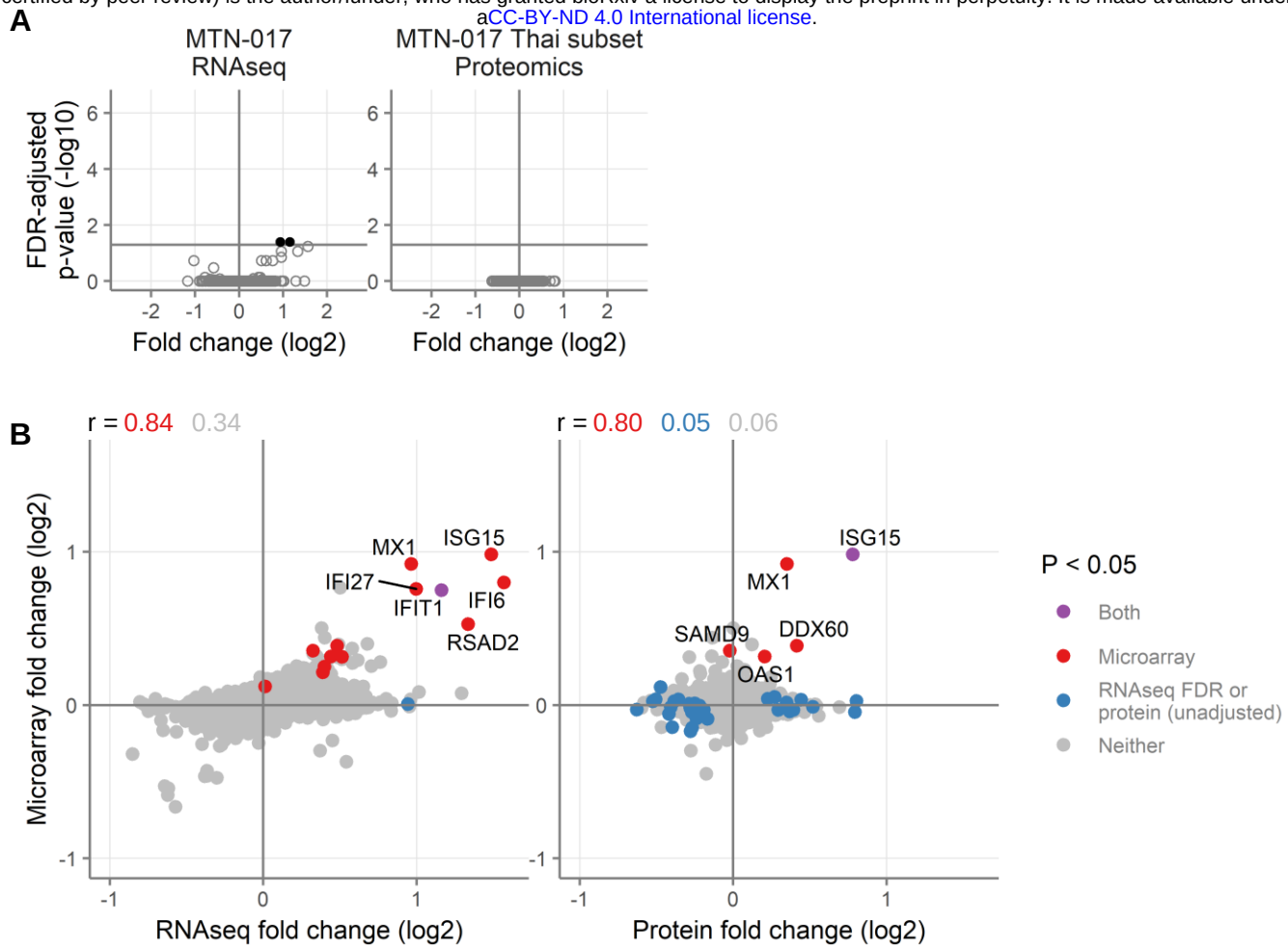

Figure 5. RNA sequencing and protein expression changes during oral TDF/FTC product use. (A) TDF/FTC-induced changes in RNAseq transcript (left) or protein (right) expression. Each point represents the expression of a single gene/protein (averaged across subjects as described in the Methods). Positive fold changes indicate higher expression during product use and negative fold changes indicate lower expression during product use. Open gray symbols indicate genes with FDR-adjusted p-values greater than 0.05. (B) Correlation of fold changes of genes as detected by microarray (y-axis) with genes/proteins detected by RNAseq (left) or proteomics (right) from the rectal samples from MTN-017. Colors indicate genes with FDR-adjusted p-values less than 0.05 in microarray transcripts (red), RNAseq genes (blue, by FDR, left) or proteins (blue, unadjusted p-value, right), both (purple, FDR for microarray and RNAseq, unadjusted for protein), or neither (gray). Spearman correlation coefficients for the genes falling into each subset are shown. Selected genes are labeled.

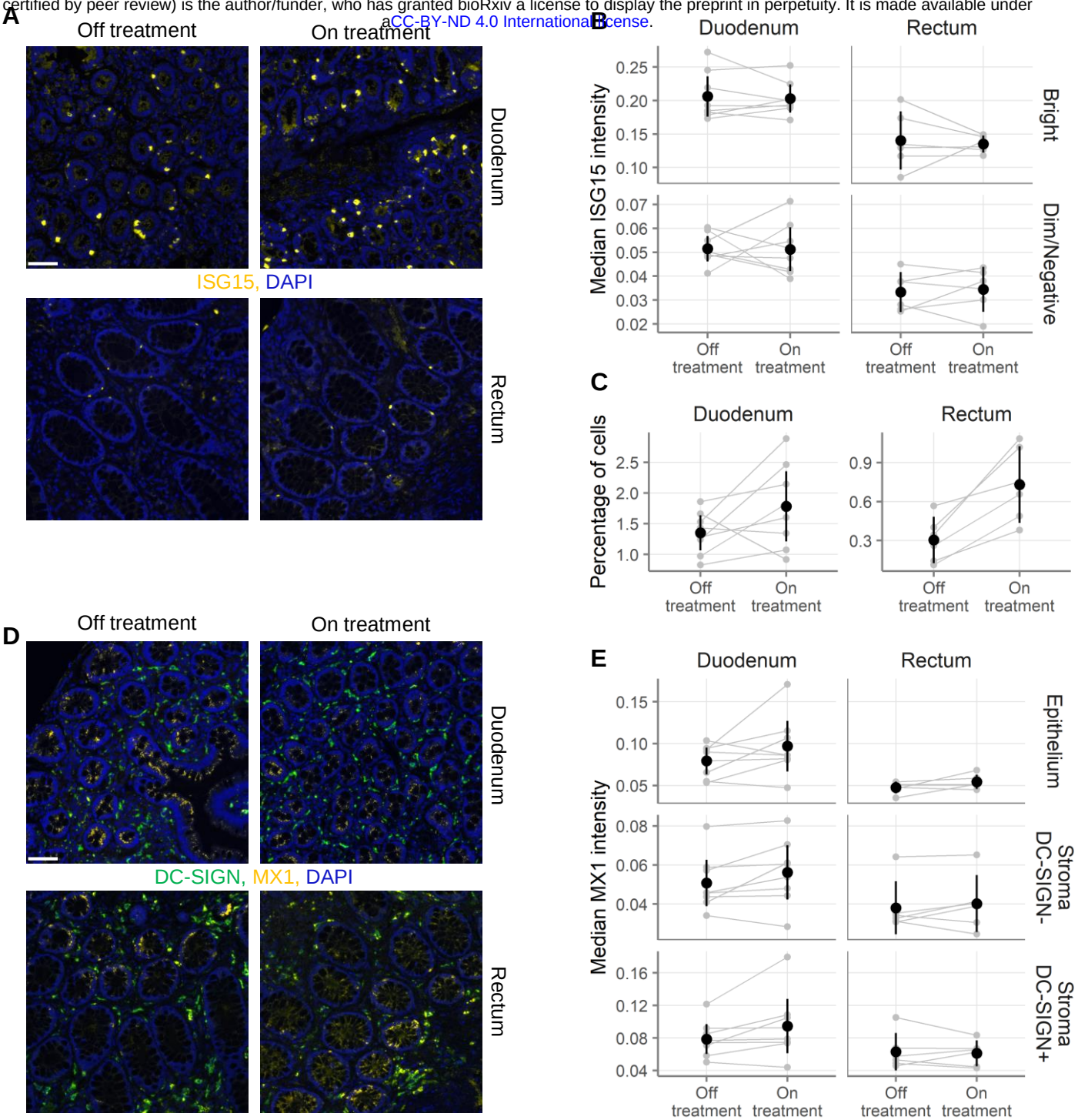

Figure 6. Immunofluorescence microscopy staining for ISG15 and MX1. (A) 20X magnification images of duodenal (top) and rectal (bottom) biopsies, stained for ISG15 (yellow) and DAPI (blue). Biopsies from pre-treatment (left) and at the end of two months of treatment (right) are shown. Scale bar is 50 µm. Duodenal biopsies came from one donor and rectal biopsies came from a second. (B) ISG15 intensity was measured in paired duodenal (n = 8 donors) and rectal (n = 6) biopsies from ACTU-3500. Cells were divided into bright and dim/negative groups. The intensity of the median cell is shown for each biopsy. (C) The percentage of all cells that fell into the bright ISG15 group is shown for each biopsy. (D) 20X magnification images of duodenal (top) and rectal (bottom) biopsies, stained for MX1 (yellow), DC-SIGN (green), and DAPI (blue). Biopsies from prior to treatment initiation (left) and at the end of two months of treatment (right) are shown. Scale bar is 50 µm. Duodenal biopsies came from one donor and rectal biopsies came from a second. (E) MX1 and DC-SIGN were measured in paired duodenal (n = 8 donors) and rectal (n = 6) biopsies from ACTU-3500. Cells were divided into epithelial, stromal without DC-SIGN and stromal with DC-SIGN. The MX1 intensity of the median cell within each group is shown for each biopsy. Gray points indicate measurements from a single biopsy and cell group, with gray lines connecting the matching observation from the same donor. Black symbols and vertical lines show the mean and 95% confidence interval of the mean.

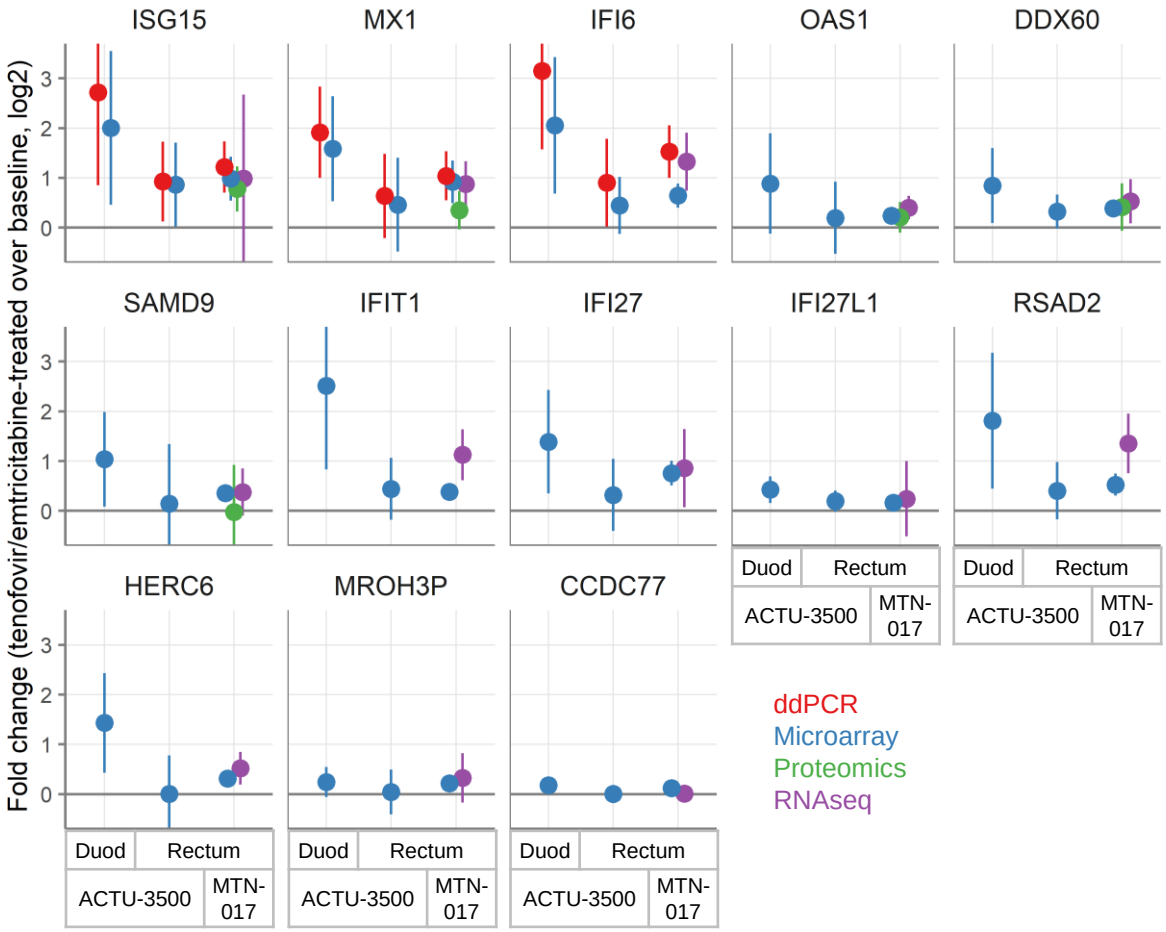

Figure 7. Comparison of expression changes in the gastrointestinal tract across assay of the thirteen differentially expressed genes from MTN-017 microarray data. Fold changes in gene expression of the thirteen genes that were differentially expressed in the rectal microarray data of MTN-017 after treatment with TDF/FTC. Symbols show the mean across all participants and vertical lines show the 95% confidence intervals of the mean. The colors indicate the assay by which the fold changes were detected (red = ddPCR, blue = microarray, green = proteomics, purple = RNAseq). There were 8 pairs of samples for all tissues in ACTU-3500 and 36 pairs for MTN-017. A positive fold change means higher expression during treatment, where a negative fold change means higher expression off of treatment. 95% confidence interval bars are truncated at the top of the plots for ddPCR for ISG15 and IFI6 and for microarray for IFIT1.