1 **SETDB2 is differentially expressed following infection with the coronavirus**
2 **family.**

3 Shahan Mamoor¹
4 ¹shahanmamoor@gmail.com
5 East Islip, NY USA

6 Epigenetic mechanisms are important means by which the host defends itself against
7 pathogenic insults. We mined published and public microarray datasets¹⁻⁶ to discover
8 conserved epigenetic machinery mobilized by the coronavirus family and identified
9 significant differential expression of the H3K9me3 histone lysine methyltransferase
10 SETDB2⁷ by multiple human coronaviruses, including the Severe Acute Respiratory
11 Syndrome virus SARS-CoV-1. These data suggest induction of SETDB2 - and perhaps
12 the H3K9me3 mark - may be important for the coronavirus life cycle or for immune
13 defense against coronaviruses.

14
15
16
17
18
19
20
21
22
23
24
25
26 Keywords: SARS-CoV-1, SARS-CoV-2, MERS-CoV, HCoV-229E, coronavirus, host-
27 pathogen interactions, systems biology of viral infection, targeted therapeutics in
28 coronavirus infection.

Nearly 20,000,000 people have been infected with the novel human coronavirus SARS-CoV-2 in less than one year^{8,9}. Deciphering the host transcriptional response is important to understanding host-pathogen interactions. We mined published and public microarray datasets¹⁻⁶ to identify genes most differentially expressed in the host following infection with a series of human coronaviruses: the Middle East Respiratory Syndrome (MERS) coronavirus MERS-CoV, the Severe Acute Respiratory Syndrome (SARS) coronavirus 1 SARS-CoV-1, and the human coronavirus 229E (HCoV-229E). This integrated systems-level transcriptome analysis revealed conserved transcriptional modulation of the histone lysine methyltransferase SETDB2 by the coronavirus family.

Methods

We used datasets GSE100509¹, GSE56677², GSE59185³, GSE68820⁴, GSE22581⁵ and GSE89167⁶ for this analysis of gene expression changes in mammalian cells and tissues following coronavirus infection. GSE100509 was generated using Agilent-026652 Whole Human Genome Microarray 4x44K v2 technology with $n=5$ each for primary human microvascular endothelial cells infected with MERS-CoV at 0 hours, 12 hours, 24 hours, 36 hours and 48 hours. GSE56677 was generated using Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 technology with $n=3$ each for CALU3 2B4 cells infected with MERS-CoV London (Eng1): at 0 hours, 3 hours, 7 hours, 12 hours, 18 hours and 24 hours post-infection. GSE59185 was generated using Agilent-028005 SurePrint G3 Mouse GE 8x60K Microarray technology with $n=3$ lung tissue from mock-infected BALB/c mice and

1 $n=3$ lung tissue from SARS-CoV-1-infected BALB/c mice, both at 2 days post-infection.
2 GSE68820 was generated using Agilent-014868 Whole Mouse Genome Microarray
3 4x44K G4122F technology with $n=5$ for lung tissue from mock-infected
4 C57BL6/NJ mice and $n=5$ for lung tissue from SARS-CoV-1-infected C57BL6/NJ mice,
5 both at 2 days post-infection. GSE22581 was generated using Affymetrix Canine
6 Genome 2.0 Array technology with $n=3$ each for ferret lung on day 0, day 1 and day 2
7 post-infection through the intranasal route with SARS-CoV-1. GSE89167 was
8 generated using 039494 SurePrint G3 Human GE v2 8x60K Microarray 039381
9 technology with $n=2$ HuH-7 cells and $n=2$ HCoV-229E infected HuH-7 cells. The
10 Benjamini and Hochberg method of p -value adjustment was used for ranking of
11 differential expression but raw p -values were used for assessment of statistical
12 significance of global differential expression. Log-transformation of data was auto-
13 detected, and the NCBI generated category of platform annotation was used.

14
15 A statistical test was performed to evaluate the significance of difference in
16 SETDB2 mRNA expression levels in CALU3 2B4 cells with MERS-CoV infection at 3
17 hours, 7 hours, 12 hours, 18 hours and 24 hours post-infection as compared to CALU3
18 2B4 cells at baseline (0 hours) using a one-way ANOVA with Dunnett's multiple
19 comparisons test. A statistical test was performed to evaluate the significance of
20 difference in SETDB2 mRNA expression levels in the lungs of BALB/c mice at 2 days
21 post-infection with SARS-CoV-1 using an unpaired, two-tailed t-test with Welch's
22 correction. Only p -values less than 0.05 were considered statistically significant. We
23 used PRISM for all statistical analyses (Version 8.4.0)(455).

Results

We mined six independent microarray datasets¹⁻⁶ to understand at the systems-level and in an unbiased fashion the transcriptional response to coronavirus infection.

SETDB2 is differentially expressed in primary human cells and in a human cell line following infection with MERS-CoV.

We identified the histone lysine methyltransferase SETDB2 as among the genes most differentially expressed following infection of primary human microvascular endothelial cells with MERS-CoV¹. When sorting each of the transcripts measured by microarray based on significance of change in expression between uninfected and MERS-CoV infected cells, SETDB2 ranked 1240 out of 34127 total transcripts (Table 1). Differential expression of SETDB2 in MERS-CoV-infected primary human microvascular endothelial cells was statistically significant (Table 1; $p=1.08E-20$).

We queried a second microarray dataset² to determine if differential expression of SETDB2 could be observed following infection of a human cell line (CALU3 2B4) with a separate isolate of MERS-CoV, MERS-CoV London (Eng1). Again, SETDB2 was among the genes most differentially expressed following infection of human cells with a MERS-CoV (Table 2). When sorting each of the transcripts measured by microarray based on significance of change in expression between uninfected and MERS-CoV London infected CALU3 cells, SETDB2 ranked 168 out of 28653 total transcripts (Table 2). Differential expression of SETDB2 in MERS-CoV-London-infected CALU3 cells was statistically significant (Table 2; $p=5.89E-12$).

1 SETDB2 is differentially expressed in the lungs of mice infected with SARS-CoV-1.

2 We next queried microarray data³⁻⁴ generated from the lungs of mice infected
3 with SARS-CoV-1 to determine if differential expression of SETDB2 could be observed
4 following infection with a different coronavirus type and whether this could be observed
5 *in vivo*. SETDB2 was among the most differentially expressed in the lungs following
6 infection of BALB/c mice with SARS-CoV-1. When sorting each of the transcripts
7 expressed in the lung measured by microarray based on significance of change in
8 expression between mock-infected and SARS-CoV-1 infected mice, SETDB2 ranked 80
9 out of 62976 total transcripts (Table 3). Differential expression of SETDB2 in the lungs
10 of SARS-CoV-1-infected BALB/c mice was statistically significant (Table 3; $p=1.27E-08$).
11

12 We also queried microarray data⁴ generated from the lungs of a different strain of
13 mice infected with SARS-CoV-1. Again, SETDB2 was among the most differentially
14 expressed in the lungs following infection of C57BL6/NJ mice with SARS-CoV-1. When
15 sorting each of the transcripts measured by microarray expressed in the lung based on
16 significance of change in expression between mock-infected and SARS-CoV-1 infected
17 C57BL6/NJ mice, SETDB2 ranked 2852 out of 29649 total transcripts (Table 4).
18 Differential expression of SETDB2 in the lungs of SARS-CoV-1-infected C57BL6/NJ
19 mice was statistically significant (Table 4; $p=2.27E-07$).
20

21 SETDB2 is differentially expressed in the lungs of ferrets infected with SARS-CoV-1.

22 We next queried another microarray dataset⁵ to determine if *in vivo* differential
23 expression of SETDB2 could be observed in a separate small animal model of SARS-
24
25
26
27
28

CoV-1 infection. SETDB2 was differentially expressed in the ferret lung following intranasal infection with SARS-CoV-1. When sorting each of the transcripts expressed in the ferret lung measured by microarray based on change in expression between mock-infected and infected ferrets, SETDB2 ranked 2179 out of 43035 total transcripts (Table 5). Differential expression of SETDB2 in the ferret lung was statistically significant (Table 5; $p=1.09E-04$). SETDB2 was not differentially expressed in the blood of ferrets following intranasal infection with SARS-CoV-1 (Table 6).

SETDB2 is differentially expressed in a human cell line following infection with the human coronavirus HCoV-229E.

Finally, we queried a sixth microarray dataset⁶ to determine if SETDB2 differential expression could be observed following infection with a third human coronavirus, the human coronavirus HCoV-229E. We found conserved differential expression of SETDB2 in the Huh-7 cell line following infection with HCoV-229E in vitro. When sorting each of the transcripts measured by microarray based on change in expression between uninfected and HCoV-229E-infected HuH-7 cells, SETDB2 ranked 564 out of 62976 total transcripts (Table 7). Differential expression of SETDB2 in HCoV-229E-infected Huh-7 cells was statistically significant (Table 7; $p=0.0073182$).

Rapid transcriptional induction of SETDB2 in human cells following infection with MERS-CoV London.

We assessed the kinetics of SETDB2 transcription in CALU3 cells following infection with MERS-CoV London. At 3 hours post-infection, statistically significant

induction of SETDB2 could be observed (Figure 1; $p < 0.0001$), with a mean fold increase of 1.152 ± 0.004 in SETDB2 mRNA levels (Table 1). By 7 hours post-infection, SETDB2 levels dropped to an extent but remained significantly elevated as compared to baseline (0 hours) (Figure 1; $p < 0.0001$), with a mean fold increase of 1.092 ± 0.005 in SETDB2 mRNA levels (Table 1). From 12 to 24 hours, SETDB2 levels continued to progressively increase, with significantly increased SETDB2 mRNA levels at 12 hours (Figure 1; $p < 0.0001$), 18 hours (Figure 1; $p < 0.0001$), and 24 hours (Figure 1; $p < 0.0001$); we calculated mean fold induction of SETDB2 of 1.125 ± 0.004 at 12 hours, 1.158 ± 0.010 at 18 hours, and 1.179 ± 0.004 at 24 hours (Table 1).

SETDB2 mRNA levels are significantly increased in the mouse lung following infection with SARS-CoV-1.

We obtained exact mRNA levels of SETDB2 in the mouse lung to determine whether SETDB2 was differentially expressed because mRNA levels were increasing or decreasing during infection. At 2 days post-infection, SETDB2 levels in the lungs of BALB/c mice were increased as compared to mock-infected lungs (Figure 2); this difference in SETDB2 mRNA quantity was significantly different (Figure 2; $p = 0.0001$).

Thus, we found here conserved differential expression of the histone lysine methyltransferase SETDB2 in human cells following infection with two types of MERS-CoV, in the lungs of two separate strains of mice following infection with SARS-CoV-1, and in a human cell line following infection with HCoV-229E. SETDB2 was transcribed within 3 hours of infection with MERS-CoV London *in vitro* and induction of SETDB2 could be observed in the mouse lung at 2 days post-infection.

Discussion

SETDB2 is a histone lysine methyltransferase that catalyzes tri-methylation of lysine 9 on histone H3, the H3K9me3 mark. The SET domain, originally found in members of the Polycomb and Trithorax family members and named after Su(var)3-9, Enhancer of Zeste and trithorax, is the catalytic domain which performs the methyltransferase catalytic function in SET domain-containing enzymes^{10,11}. SETDB2 also contains a methyl-CpG-binding domain, as well as a pre-SET domain amino-terminal to the SET domain⁷. In zebrafish, depletion of SETDB2 using morpholino oligonucleotides results in significant global reduction of the H3K9me3 mark⁷.

SETDB2 is described as the only histone lysine methyltransferase whose expression is induced by infection with influenza virus¹². SETDB2 expression could be induced by Type I interferons, which resulted in SETDB2 expression and occupancy at chemokine Cxcl1 gene, resulting in repression of CXCL1 expression; in addition to Cxcl1, SETDB2 occupancy and gene repression was observed at other NF-kB target genes¹². In the absence of SETDB2, the Cxcl1 promoter was marked by significantly reduced quantities of H3K9me3¹². Mice with a hypomorphic allele displayed higher neutrophil infiltration in the context of sterile lung inflammation; these mice were also less susceptible to bacterial superinfection¹². Together the data suggested that the SETDB2 functioned in concert with NF-kB and type I interferon signaling to “fine-tune” the host immune response during viral infection.

We found here that the H3K9me3 histone lysine methyltransferase SETDB2 was among the most differentially expressed genes following infection with multiple human

1 coronaviruses, including MERS-CoV, SARS-CoV-1, and HCoV-229E. As utilization of
2 SETDB2 appears to be a conserved feature of the human coronaviruses, future studies
3 can determine whether deposition of the H3K9me3 mark on chromatin is important for
4 SARS-CoV-2 infection and whether modulation of SETDB2 methyltransferase function
5 can influence outcomes in SARS-CoV-2 (COVID-19) infection.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. GSE100509. Baric R, Sims A, Heller N, Waters KM, Einfeld AJ, Kawaoka Y. PNNL. 902 Battelle Blvd. Richland, USA.
2. Selinger, C., Tisoncik-Go, J., Menachery, V.D., Agnihothram, S., Law, G.L., Chang, J., Kelly, S.M., Sova, P., Baric, R.S. and Katze, M.G., 2014. Cytokine systems approach demonstrates differences in innate and pro-inflammatory host responses between genetically distinct MERS-CoV isolates. *BMC genomics*, 15(1), p.1161.
3. Regla-Nava, J.A., Nieto-Torres, J.L., Jimenez-Guardeño, J.M., Fernandez-Delgado, R., Fett, C., Castaño-Rodríguez, C., Perlman, S., Enjuanes, L. and DeDiego, M.L., 2015. Severe acute respiratory syndrome coronaviruses with mutations in the E protein are attenuated and promising vaccine candidates. *Journal of virology*, 89(7), pp.3870-3887.
4. Totura, A.L., Whitmore, A., Agnihothram, S., Schäfer, A., Katze, M.G., Heise, M.T. and Baric, R.S., 2015. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *MBio*, 6(3), pp.e00638-15.
5. Danesh, A., Cameron, C.M., León, A.J., Ran, L., Xu, L., Fang, Y., Kelvin, A.A., Rowe, T., Chen, H., Guan, Y. and Jonsson, C.B., 2011. Early gene expression events in ferrets in response to SARS coronavirus infection versus direct interferon-alpha2b stimulation. *Virology*, 409(1), pp.102-112.
6. Poppe, M., Wittig, S., Jurida, L., Bartkuhn, M., Wilhelm, J., Müller, H., Beuerlein, K., Karl, N., Bhujju, S., Ziebuhr, J. and Schmitz, M.L., 2017. The NF-κB-dependent and-independent transcriptome and chromatin landscapes of human coronavirus 229E-infected cells. *PLoS pathogens*, 13(3), p.e1006286.
7. Xu, P.F., Zhu, K.Y., Jin, Y., Chen, Y., Sun, X.J., Deng, M., Chen, S.J., Chen, Z. and Liu, T.X., 2010. Setdb2 restricts dorsal organizer territory and regulates left-right asymmetry through suppressing fgf8 activity. *Proceedings of the National Academy of Sciences*, 107(6), pp.2521-2526.
8. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. and Garry, R.F., 2020. The proximal origin of SARS-CoV-2. *Nature medicine*, 26(4), pp.450-452.
9. World Health Organization. Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.

- 1 10. Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G., 1994.
2 The protein encoded by the Drosophila position-effect variegation suppressor gene
3 Su (var) 3-9 combines domains of antagonistic regulators of homeotic gene
4 complexes. *The EMBO journal*, 13(16), pp.3822-3831.
- 5 11. Dillon, S.C., Zhang, X., Trievel, R.C. and Cheng, X., 2005. The SET-domain protein
6 superfamily: protein lysine methyltransferases. *Genome biology*, 6(8), p.227.
- 7 12. Schliehe, C., Flynn, E.K., Vilagos, B., Richson, U., Swaminathan, S., Bosnjak, B.,
8 Bauer, L., Kandasamy, R.K., Griesshammer, I.M., Kosack, L. and Schmitz, F., 2015.
9 The methyltransferase Setdb2 mediates virus-induced susceptibility to bacterial
10 superinfection. *Nature immunology*, 16(1), pp.67-74.

Rank	ID	p-value	F	Gene	Gene name
1240	A_23_P341532	1.08E-20	361.3031	SETDB2	SET domain, bifurcated 2

Table 1: Differential expression of SETDB2 in primary human microvascular endothelial cells following infection with MERS-CoV.

The rank of differential expression, probe ID, p-value of global differential expression, F statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	F	FC	Gene	Gene name
168	A_23_P341532	5.89E-12	134.6287	<u>3hr</u> : 1.152 ± 0.004 <u>7hr</u> : 1.092 ± 0.005 <u>12hr</u> : 1.125 ± 0.004 <u>18hr</u> : 1.158 ± 0.010 <u>24hr</u> : 1.179 ± 0.004	SETDB2	SET domain, bifurcated 2

Table 2: Differential expression of SETDB2 in the CALU3 2B4 human cell line following infection with MERS-CoV London.

The rank of differential expression, probe ID, p-value of global differential expression, F statistic, fold change of SETDB2 expression as compared to baseline (0 hours), gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name
80	28625	1.27E-08	2.65E+01	10.63449	Setdb2	SET domain, bifurcated 2

Table 3: Differential expression of SETDB2 in the lungs of BALB/c mice following infection with SARS-CoV-1.

The rank of differential expression, probe ID, p-value of global differential expression, F statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name
2852	A_51_P192116	2.27E-07	11.87352	7.041532	Setdb2	SET domain, bifurcated 2

Table 4: Differential expression of SETDB2 in the lungs of C57BL6/NJ mice following infection with SARS-CoV-1.

The rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared (mock-infected and infected mouse lung), gene and gene name are listed in this chart.

Rank	ID	p-value	F	Gene	Gene name
2179	CfaAffx.7429.1.S1_at	1.09E-04	3.93E+01	SETDB2	SET domain, bifurcated 2

Table 5: Differential expression of SETDB2 in the lungs of ferrets following intranasal infection with SARS-CoV-1.

The rank of differential expression, probe ID, p-value of global differential expression, F statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name
6011	CfaAffx.7429.1.S1_at	1.04E-01	-1.883271	-4.2819	SETDB2	SET domain, bifurcated 2

Table 6: SETDB2 is not differentially expressed in the blood of ferrets infected with SARS-CoV-1.

The rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared (mock-infected and infected ferret blood), gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name
564	16903	0.0073182	3.8705838	-2.463	SETDB2	SET domain, bifurcated 2

Table 7: SETDB2 is differentially expressed in the Huh-7 human cell line following infection with HCoV-229E.

The rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared (uninfected and infected Huh-7 cells), gene and gene name are listed in this chart.

Figure 1: Rapid transcriptional induction of SETDB2 in human cells following infection with MERS-CoV London.

SETDB2 mRNA expression levels are graphically represented from 3 hours to 24 hours with mean mRNA level marked and the result of a statistical test evaluating significance of difference in SETDB2 expression as compared to baseline (0 hours), a p-value, listed above each time point.

Figure 2: Transcriptional induction of SETDB2 in the lungs of mice following infection with SARS-CoV-1.

SETDB2 mRNA expression levels are graphically represented in the lungs of BALB/c mice after mock infection (left) and in the lungs 2 days post-infection with SARS-CoV-1.