

**DNA methylation from a Type I restriction modification system influences gene
expression and virulence in *Streptococcus pyogenes***

Taylor M. Nye¹, Kristin M. Jacob^{2#}, Elena K. Holley², Juan M. Nevarez², Suzanne
Dawid², Lyle A. Simmons¹, and Michael E. Watson, Jr.^{2*}

¹Department of Molecular, Cellular, and Developmental Biology, University of Michigan,
Ann Arbor, MI, USA

²Division of Pediatric Infectious Diseases, Department of Pediatrics and Communicable
Diseases, University of Michigan, Ann Arbor, MI, USA

#Current Address: Michigan State University, Department of Microbiology and Molecular
Genetics, East Lansing, MI, USA

***Corresponding Author:**

Email: mewats@med.umich.edu (MEW)

Abstract

DNA methylation is pervasive across all domains of life. In bacteria, the presence of N6-methyladenosine (m6A) has been detected among diverse species, yet the contribution of m6A to the regulation of gene expression is unclear in many organisms. Here, we investigated the impact of DNA methylation on gene expression and virulence within the human pathogen *Streptococcus pyogenes*, or Group A Streptococcus. Single Molecule Real-Time sequencing and subsequent methylation analysis identified 412 putative m6A sites throughout the 1.8 Mb genome. Deletion of the Restriction, Specificity, and Methylation gene subunits (Δ RSM strain) of a putative Type I restriction modification system lost all detectable m6A at the recognition sites and failed to prevent transformation with foreign-methylated DNA. RNA-sequencing identified 20 genes out of 1,895 predicted coding regions with significantly different gene expression. All of the differentially expressed genes were down regulated in the Δ RSM strain relative to the parent strain. Importantly, we found that the presence of m6A DNA modifications affected expression of Mga, a master transcriptional regulator for multiple virulence genes, surface adhesins, and immune-evasion factors in *S. pyogenes*. Using a murine subcutaneous infection model, mice infected with the Δ RSM strain exhibited an enhanced host immune response with larger skin lesions and increased levels of pro-inflammatory cytokines compared to mice infected with the parent or complemented mutant strains, suggesting alterations in m6A methylation influence virulence. Further, we found that the Δ RSM strain showed poor survival within human neutrophils and reduced adherence to human epithelial cells. These results demonstrate that, in addition to restriction of foreign DNA, gram-positive bacteria also use restriction

modification systems to regulate the expression of gene networks important for virulence.

Word Count: 271 (300 words max)

Author summary

DNA methylation is common among many bacterial species, yet the contribution of DNA methylation to the regulation of gene expression is unclear outside of a limited number of gram-negative species. We characterized sites of DNA methylation throughout the genome of the gram-positive pathogen *Streptococcus pyogenes* or Group A *Streptococcus*. We determined that the gene products of a functional restriction modification system are responsible for genome-wide m6A. The mutant strain lacking DNA methylation showed altered gene expression compared to the parent strain, with several genes important for causing human disease down regulated. Furthermore, we showed that the mutant strain lacking DNA methylation exhibited altered virulence properties compared to the parent strain using various models of pathogenesis. The mutant strain was attenuated for both survival within human neutrophils and adherence to human epithelial cells, and was unable to suppress the host immune response in a murine subcutaneous infection model. Together, these results show that bacterial m6A contributes to differential gene expression and influences the ability of Group A *Streptococcus* to cause disease. DNA methylation is a conserved feature among bacteria and may represent a potential target for intervention in effort to interfere with the ability of bacteria to cause human disease.

Word Count: 200 (150–200 words, non-technical summary)

Introduction

DNA methylation has been shown to regulate diverse pathways across all domains of life [1]. In eukaryotes, cytosine methylation regulates developmental gene expression and aberrant DNA methylation patterns have been implicated in many disease states, including cancer [2, 3]. Although studied in a limited number of prokaryotic organisms, DNA methylation has been implicated in a myriad of cellular processes, including protection from the invasion of foreign DNA, cell cycle regulation, DNA mismatch repair, and the regulation of gene expression [4]. It was recently shown that within the genomes of over 200 prokaryotes surveyed that greater than 90% contained N6-methyladenosine (m6A), N4-methylcytosine (m4C), or 5-methylcytosine modifications (m5C) [5]. These results demonstrate that DNA methylation among prokaryotes is more pervasive than originally anticipated. What remains uncertain is if DNA methylation imparts any regulatory controls influencing virulence properties or other phenotypes amongst the array of diverse prokaryotic species.

DNA methylation in bacteria has been well characterized in the context of restriction modification (RM) systems [4, 5]. RM systems are a mechanism of bacterial host defense to prevent the invasion of foreign DNA. RM systems are generally comprised of a site-specific restriction endonuclease (REase), methyltransferase (MTase), and, in some cases, a specificity subunit that together form a protein complex that cleaves foreign DNA after it enters the cell. Methylation of the host DNA at the

same recognition site serves to safeguard the host chromosome from cleavage. In addition to RM systems, DNA can also be methylated by orphan MTases. Orphan MTases methylate DNA in site-specific sequences and lack an active cognate endonuclease [5, 6]. In bacteria, the two most well studied orphan MTases are *Escherichia coli* DNA adenosine methyltransferase (Dam) and *Caulobacter crescentus* cell cycle regulated methyltransferase (CcrM) [5, 6]. Site-specific DNA methylation by Dam and CcrM has been shown to regulate DNA mismatch repair, cell cycle progression, origin sequestration, and gene expression, demonstrating that DNA methylation imparts critical regulatory functions [6].

Despite the importance of RM systems and orphan MTases, the lack of genome-wide detection tools has hindered the identification of DNA base modifications and characterization of the physiological consequences resulting from MTase inactivation in bacteria. The use of methylation-sensitive restriction endonucleases to identify sites of DNA base modifications is limited by the sequence specificity of the recognition site, potentially missing many base modifications that could occur outside of a particular sequence context ([5] and references therein). While bisulfite sequencing allows for genome-wide detection of m5C in sequence specific-contexts, no such genome-wide detection tool has been available for the detection of m6A or m4C until the recent advent of Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing platform [7-11]. SMRT sequencing relies on differences in DNA polymerase kinetics to detect base modifications in the template strand in a sequence-context specific manner without *a priori* knowledge of the modification.

Our group previously used the PacBio SMRT sequencing platform to complete whole genome sequencing and reference genome assembly of two strains of the bacterial human pathogen *Streptococcus pyogenes*, or Group A Streptococcus (GAS) [12, 13]. *S. pyogenes* causes a wide variety of human infections, ranging from the relatively common streptococcal pharyngitis and cellulitis to the relatively uncommon, but severe, streptococcal toxic shock syndrome and necrotizing fasciitis, which have high morbidity and mortality rates [14-16]. *S. pyogenes* is a model bacterial pathogen, not only for the infections it produces, but also for the great diversity of toxins and virulence factors expressed by the organism and the highly complex nature of regulatory mechanisms employed to control virulence factor expression [14, 16-18]. Indeed, *S. pyogenes* utilizes over 30 recognized transcriptional regulatory proteins and 13 two-component regulatory systems to coordinate virulence factor expression in response to varying environmental signals (e.g., carbohydrate availability, temperature, pH, oxygen tension, salt concentrations, osmolality, etc.), growth phase, intracellular metabolite concentrations, and signaling pheromones involved in quorum sensing [17, 18]. DNA methylation has not been previously investigated as a significant mechanism influencing virulence factor expression within *S. pyogenes*, and DNA methylation may represent an unrecognized target for therapeutic intervention to help prevent or treat severe streptococcal disease.

In this study, we show that in *S. pyogenes* strain MEW123, a representative derivative of a serotype M28 clinical pharyngitis isolate, the active Type I RM system SpyMEW123I is responsible for the bipartite m6A motif identified throughout the genome. We show that deletion of the RM system and subsequent loss of m6A from *S.*

pyogenes results in the down regulation of a distinct set of operons involved in streptococcal virulence. Importantly, our study shows that methylation by a Type I RM system correlates with differential expression of Mga, a major transcriptional regulator of multiple virulence factors, surface adhesins, and immune evasion factors in *S. pyogenes*. The results presented here demonstrate that RM systems can integrate their methylation signal to influence the expression of gene networks important for bacterial virulence.

Results

SMRT sequencing and methylation analysis identifies m6A modifications in a bipartite recognition sequence in the *S. pyogenes* genome. Previously we completed whole genome assembly using PacBio SMRT sequencing with *S. pyogenes* strain MEW123, a representative serotype M28 isolate used by our group to investigate streptococcal mucosal colonization [12] (for strain list refer to Table 1). To begin our investigation, we performed methylation analysis of the SMRT sequencing data. We identified m6A DNA base modifications in the MEW123 genome at the consensus sequence 5' GCANNNNNTTYG and its corresponding partner motif 5' CRAANNNNNNTGC, consistent with m6A modification motifs previously reported by Blow *et al.* (Table 2) [5]. Within the MEW123 genome, 412 occurrences of each m6A site within the bipartite recognition motif were identified; the majority occurred in predicted coding (92%) and intergenic (6%) regions of the MEW123 genome. The bipartite recognition motif is characteristic of Type I RM systems, which are typically comprised of three separate subunits, including a restriction endonuclease, a specificity

subunit, and a methyltransferase subunit, that act together as a single protein complex and typically act at large distances from the methylation site. The RM system annotation pipeline used in Blow *et al.* identified the putative Type I restriction modification system, annotated as SpyMEW123I, consisting of a three-gene cluster with separate restriction endonuclease (*hsdR*), specificity (*hsdS*), and methyltransferase (*hsdM*) genes, as a predicted match for modification of the identified m6A motif in *S. pyogenes* [5, 19] (Fig 1A and Fig 1B). This three-gene cluster exhibits high amino acid sequence homology to the Type I RM system identified in *S. pyogenes* SF370 at Spy_1904 (*hsdR*), Spy_1905 (*hsdS*), and Spy_1906 (*hsdM*), with 99%, 87%, and 99% identity, respectively [20]. This Type I RM system is present in virtually all sequenced *S. pyogenes* strains to date, with rare exception reported in some *emm1* strains from Japan with spontaneous deletion of a two-component regulatory system and the adjacent Type I RM system [21]. Notably, we did not detect the 5mC modifications at C^mCNCGG reported by Euler *et al.* in our PacBio SMRT sequencing results, which is not surprising given the MTase, M.SpyI, is absent from the *S. pyogenes* M28 serotype [22]. The REase and MTase activities of SpyMEW123I are annotated as R.SpyMEW123I and M.SpyMEW123I, respectively.

Table 1. Strains used in this study.

Strain/Plasmids	Description	Source
Strains		
<u><i>Escherichia coli</i></u>		
DH5α	Standard cloning vector, <i>recA1 endA1 hsdR17</i>	Invitrogen
<u><i>Streptococcus pyogenes</i></u>		
MEW123	Streptomycin-resistant clone of pediatric throat isolate, serotype M28	[12]

MEW380	MEW123 transformed by plasmid pIL09 to insertionally-inactivate <i>scpA</i> gene by spectinomycin resistance marker (Ω_{scpA})	This study
MEW409	MEW123 transformed with plasmid pIL03 to insertionally-inactivate M protein, <i>emm</i> gene, by spectinomycin resistance marker (Ω_{emm28})	This study
MEW480	MEW123 with in-frame deletion of <i>mga</i> (Δmga) by allelic exchange after transformation with plasmid pIL01	This study
MEW489	MEW123 transformed with plasmid pKJ19 to insertionally-inactivate restriction endonuclease with spectinomycin resistance marker (Ω_{RE})	This study
MEW513	MEW123 with restriction-modification system in-frame deletion (ΔRSM) by allelic exchange after transformation with pKJ24	This study
MEW552	MEW513 with RSM operon complemented <i>in trans</i> on plasmid pEH01 ($\Delta RSM/pRSM$)	This study
HSC5	Serotype M14 reference strain	[23]
HSC5 Ω_{emm}	HSC5 mutant with the M protein, <i>emm14</i> gene, disrupted by spectinomycin resistance marker (Ω_{emm14})	[24]
Plasmids		
pJRS233	Low-copy <i>E. coli</i> to <i>S. pyogenes</i> temperature-sensitive vector for allelic replacement (erythromycin-resistant, Erm^R)	[25]
pGCP213	High-copy <i>E. coli</i> to <i>S. pyogenes</i> temperature-sensitive vector for allelic replacement (Erm^R)	[26]
pJoy3	<i>E. coli</i> to <i>S. pyogenes</i> shuttle vector (chloramphenicol-resistant, $Chlor^R$)	[27]
pSpc18	integration vector containing <i>aad9</i> (spectinomycin resistance gene from <i>Enterococcus faecalis</i>) (Spc^R)	[28]
pIL01	pJRS233 with in-frame deletion of <i>mga</i> (Erm^R)	This study
pIL03	pSpc18 with <i>emm28</i> gene fragment disruption (Spc^R)	This study
pIL09	pSpc18 with <i>scpA</i> gene fragment disruption (Spc^R)	This study
pKJ19	pSpc18 with restriction endonuclease gene fragment disruption (Spc^R)	This study
pKJ24	pGCP213 with in-frame deletion of restriction-modification gene cluster (Erm^R)	This study
pEH01	pJoy3 with intact RSM operon cloned for complementation ($Chlor^R$)	This study

170

171 **Table 2. Motif analysis of modified bases from PacBio SMRT sequencing in wild**
172 **type *S. pyogenes* strain MEW123.**

Motif*	Type	Motifs in genome	Mean modQV	Mean coverage
--------	------	------------------	------------	---------------

GC <u>A</u> NNNNNNNTTY G	m6A	412	332.64	245.02
CRA <u>A</u> NNNNNNNTG C	m6A	412	282.18	237.32
<u>I</u> HTWGAAGA	unknown	410	44.74	240.57
<u>A</u> NDYVGCAD	m6A	3502	86.17	241.02
<u>I</u> NRDDDG	unknown	34390	44.88	235.68
<u>I</u> NNNDNNH	unknown	746710	47.81	235.15
<u>I</u> HRCNTWNH	unknown	3107	43.39	237.84
<u>A</u> GNNAVNW	m6A	32122	78.13	239.29
<u>I</u> NNNCRV	unknown	77468	45.11	235.03
VA <u>A</u> HNBAYVW	m6A	27467	76.34	242.45
<u>I</u> HNNDVNG	unknown	96451	43.43	237.89

*modified base is bolded and underlined

M.SpyMEW123I is responsible for m6A modifications in the *S. pyogenes* genome.

To determine if the SpyMEW123I RM system was responsible for the observed m6A modifications in strain MEW123, an in-frame deletion mutation was constructed using a plasmid vector designed for allelic replacement (pGCP213) as previously described [26] (Table 1 and Fig 1A). Approximately 95% of the three-gene sequence encoding the *hsdR*, *hsdS*, and *hsdM* genes was deleted producing strain MEW513 (referred to as

181 Δ RSM); the in-frame deletion was confirmed by PCR amplification and Sanger DNA
182 sequencing (Table 1). Growth of the MEW123 parent strain, referred to as wild-type
183 (WT) and the Δ RSM mutant were not significantly different in rate or final growth density
184 when measured in either the nutrient rich Todd-Hewitt medium with 0.2% yeast extract
185 (THY broth) or the low-carbohydrate C-medium (S1 Fig). To confirm a reduction in m6A
186 base modifications and to determine the sequence context lacking m6A base
187 modifications in the Δ RSM strain, genomic DNA was isolated and sequenced via
188 PacBio SMRT sequencing. Modification analysis showed loss of detectable m6A base
189 modifications at 5' GCANNNNNTTYG and 5' CRAANNNNNNTGC sites, demonstrating
190 that streptococci with a SpyMEW123I deletion no longer have m6A DNA base
191 modifications at the consensus sequence identified in the WT strain (Fig 1B, Table 3). A
192 number of additional methylation events were identified in MEW513; however, these
193 occurred at far lower frequencies compared to the modifications at the consensus
194 sequences in the parent strain and the quality of the read scores (Mod QV) were low
195 compared to the RSM-dependent modifications. Based on these low quality read
196 scores, we feel it is unlikely that these additional modifications reflect compensatory
197 methylation events. Furthermore, SMRT sequencing of the MEW513 genome did not
198 identify any unforeseen mutations outside of the in-frame deletion within *hsdRSM* that
199 we anticipated.

200 To further confirm that the MTase component of the RSM gene cluster,
201 M.SpyMEW123I, was indeed responsible for producing m6A DNA modifications,
202 genomic DNA was harvested from the WT and the Δ RSM strain and spotted onto a
203 nitrocellulose membrane for immunodetection using an α -m6A antibody. We found that

the α -m6A signal was substantially reduced in genomic DNA blots from the Δ RSM strain compared to the WT parent, suggesting a significant and near complete reduction in m6A base modifications in the Δ RSM strain (Fig 1C). Complementation *in trans* of the Δ RSM mutant with a plasmid encoded copy of the three gene cluster (*hsdRSM*) produced strain MEW552 (referred to as Δ RSM/pRSM) and successfully restored detection of the α -m6A signal to levels comparable to the WT strain (Fig 1C). These results demonstrate that the MTase activity of SpyMEW123I is responsible for base modifications at 5' GCANNNNNTTYG and 5' CRAANNNNNNTGC sites *in vivo*.

Table 3. Motif analysis of modified bases from PacBio SMRT sequencing in Δ RSM strain MEW513.

Motif*	Type	Motifs in genome	Mean modQV	Mean coverage
<u>I</u> YTWGARGR	unknown	701	46.56	263.00
D <u>A</u> GKBANYW	m6A	5826	88.94	256.78
<u>A</u> NNYRGYA	m6A	8371	85.54	259.83
G <u>A</u> HBBAACA	m6A	499	125.47	268.90
<u>I</u> NNNDNNH	unknown	746710	49.02	251.20
<u>I</u> NRDDDDG	unknown	34390	45.12	251.97
<u>I</u> HRGCNTH	unknown	5464	43.69	253.61
<u>I</u> NNNCRV	unknown	77468	45.81	251.94
D <u>I</u> NRVCBNHHH	unknown	29343	44.13	251.93
<u>A</u> HSBAMYW	m6A	9198	75.01	265.11
<u>I</u> HNNDVNG	unknown	96451	44.13	256.11

*modified base is bolded and underlined

The SpyMEW123I RM system influences *S. pyogenes* transformation efficiency demonstrating functional restriction of foreign DNA acceptance. Deletion of the three-gene cluster, *hsdRSM*, containing the predicted endonuclease, specificity, and methylation gene subunits abolished m6A base modifications in the Δ RSM mutant strain. In Type I RM systems, DNA cleavage is dependent on the MTase and specificity subunits, in addition to the REase subunits which are often independently regulated by a separate promoter [29]. Fully unmethylated recognition motifs induce REase activity that results in DNA cleavage typically between two fully unmethylated motifs at sites distant from the recognition sequence; this distance may range from 40 base pairs to several kilobases away from the RM site. Type I MTases can function to add m6A *de novo* on fully unmethylated DNA or act as maintenance MTases at hemi-methylated recognition sites [29-31]. Additional mechanisms also protect DNA from restriction, including proteolysis of the REase subunits or protection by DNA binding proteins that can protect unmethylated sites from cleavage in the host chromosome [32]. To establish the functionality of the REase component of SpyMEW123I, a transformation efficiency assay was performed using pJoy3 plasmid DNA methylated in an *E. coli* host (Table 1). This 6.3 kb plasmid contains eight predicted Dam MTase RM sites (5' GATC) and is delivered in its native double-stranded circular form via electroporation into electrocompetent *S. pyogenes* where the plasmid is maintained and replicates extrachromosomally [27]. In addition to testing the effect of deleting the entire *hsdRSM* gene cluster in the Δ RSM mutant strain, we constructed an additional strain derivative of MEW123 with a spectinomycin-resistance cassette disrupting the *hsdR* REase gene subunit alone producing strain MEW489 (referred to as Ω RE, Table1). If the

239 SpyMEW123I RM system has true restriction enzyme activity to foreign-modified DNA,
240 then we would expect that inactivating the *hsdR* gene subunit, either individually or
241 within the entire RSM gene cluster, would enhance the transformation efficiency of the
242 plasmid. Indeed, we found that the rates of transformation with foreign-methylated
243 plasmid DNA increased significantly for both the Δ RSM mutant and the Ω RE mutant
244 strains compared to the WT parent strain, providing evidence that the restriction
245 endonuclease component of SpyMEW123I is active and functional (Fig 2A). We were
246 unable to compare our complementation strain Δ RSM/pRSM for transformation
247 efficiency as this strain already carries the pJoy3 plasmid encoding the *hsdRSM* gene
248 cluster. As a control, we undertook transformation of a MEW123 mutant in the gene
249 encoding the C5a peptidase, *scpA* (strain 489 or Ω *scpA*), as mutants in this gene would
250 not be expected to show enhanced transformation efficiency; as expected, the
251 transformation efficiency of Ω *scpA* was not significantly different than the WT (Fig 2A).
252 Interestingly, inactivation of the endonuclease subunit *hsdR* alone in the Ω RE mutant
253 strain conferred significantly greater transformation efficiency than that observed in the
254 Δ RSM mutant (Fig 2A). In many Type I RM systems the restriction subunit is generally
255 under control of a separate promoter than the specificity and methylation subunits in the
256 RSM gene cluster [29]. We found that the α -m6A signal generated by dot blot of
257 genomic DNA from the Ω RE strain was intermediate in intensity between the WT and
258 Δ RSM strains (Fig 1C). This result suggests that the methyltransferase subunit was still
259 functional in the Ω RE strain, but that there may have been some degree of polar effect
260 from the spectinomycin-resistance cassette used to inactivate *hsdR* that was reducing
261 transcription of the *hsdS* and *hsdM* gene products compared to WT levels. We

speculate that the residual functional activities of the specificity and methyltransferase subunits in the Ω RE mutant strain, even though less than WT levels, may have conferred additional stability to the incoming foreign-methylated plasmid DNA, possibly offering protection from other minor endonucleases, thereby enhancing overall transformation efficiency.

Variation in m6A base modification occurrence at RM recognition sites identified

by SMRT sequencing in the *S. pyogenes* genome. As discussed above, the MTase activity of Type I RM systems may function to maintain the state of hemi-methylated or fully methylated DNA, whereas REase cleavage only occurs on fully unmethylated DNA. Thus, RM sites can exist in the genome in a hemi-methylated state while still conferring protection from digestion [29]. Having shown that SpyMEW123I functions as an active RM system, we sought to establish the fraction of reads that were called as methylated at each recognition site. Our analysis of sequencing data from WT *S. pyogenes* MEW123 found substantial variation in the fraction of sequencing reads modified at RM sites (Fig 2B). The fraction of reads called as modified at a given RM site did not appear to be dependent on orientation or genome position. Of the m6A modifications called at RM sites, 4.9% of sites were called as m6A modified in less than 50% of sequencing reads, 23.7% of sites were called as modified in between 50-75% of sequencing reads, and finally 71.4% of sites were called as modified at greater than 75% of aligned reads. Previous studies have also reported heterogeneity in the frequency of SMRT sequencing reads with base modifications; it has been hypothesized that these differences are due to timing in DNA replication and subsequent methylation [33-35]. Whether there is a temporal component accounting for the heterogeneity in m6A DNA

285 modifications, and whether this impacts other functions of m6A modifications, such as in
286 influencing gene transcription in *S. pyogenes*, is unknown. Given the heterogeneity
287 observed in the fraction of reads called as m6A methylated, we hypothesized that m6A
288 modifications produced by the SpyMEW123I RM system might have additional functions
289 outside of host protection from foreign DNA prompting the experiments below.

290 **RNA-sequencing shows that deletion of the SpyMEW123I RM system results in**
291 **the down regulation of several transcripts involved in streptococcal immune**
292 **evasion and adherence.** In addition to functioning in RM systems, m6A base
293 modifications from orphan MTases have been shown to function in cell cycle regulation,
294 DNA mismatch repair, and the regulation of transcription [4]. In the pathogenic
295 *Escherichia coli* serotype O104:H4 strain C227-11 associated with hemolytic uremic
296 syndrome, deletion of the ϕ Stx104 RM system results in the differential expression of
297 over 38% of the genes, including genes involved in motility, cell projection, and cation
298 transport [33]. Mismatch repair is not coupled to methylation in *S. pyogenes* or most
299 other gram-positive bacteria [36]. Therefore, we asked if m6A originating from the
300 SpyMEW123I RM system in *S. pyogenes* might have additional functions outside of
301 host defense from foreign DNA. We isolated RNA from streptococcal cells during mid-
302 exponential growth phase in C media broth culture from WT and Δ RSM strains followed
303 by RNA-sequencing. The results of the differential expression analysis showed that 20
304 genes were differentially expressed in the Δ RSM strain compared to WT (adjusted p.
305 val < 0.05, log₂ fold change >1, data set available at NCBI repository). Interestingly, all
306 20 genes were down regulated in Δ RSM relative to WT suggesting a common
307 regulatory mechanism (Fig 3A and 3B, Table 4). The three genes (*hsdRSM*) of the

SpyMEW123I RM gene cluster showed the greatest log₂ fold change in expression of -10.8, -10.7, and -11.7, respectively, which was expected because these genes were deleted in the ΔRSM strain.

Table 4. Differentially expressed genes in ΔRSM strain compared to WT.

<u>Cluster</u>	<u>MEW123 Locus</u>	<u>Fold-Change (Log2)</u>	<u>Ave Expr</u>	<u>adj.P.Val</u>	<u>Prob</u>	<u>Product Annotation</u>
1	AWM59_00940	-3.88	1.98	0.00	1.00	hypothetical protein
1	AWM59_00945	-3.16	1.30	0.00	0.98	hypothetical protein
2	AWM59_02865	-1.14	6.80	0.01	0.89	cell surface protein
2	AWM59_02875	-1.12	3.04	0.05	0.45	conjugal transfer protein
2	AWM59_02880	-1.29	4.30	0.04	0.46	cell division protein (FtsK)
2	AWM59_02900	-1.14	2.87	0.01	0.87	hypothetical protein
3	AWM59_07395	-1.23	6.09	0.03	0.55	hypothetical protein
3	AWM59_07430	-4.12	5.14	0.00	1.00	hypothetical protein
4	AWM59_07895	-10.73	4.10	0.00	1.00	Endonuclease <i>hsdR</i>
4	AWM59_07900	-11.69	3.07	0.00	1.00	Specificity <i>hsdS</i>
4	AWM59_07905	-10.79	3.98	0.00	1.00	Methyltransferase <i>hsdM</i>
5	AWM59_08190	-5.60	5.26	0.00	0.99	Collagen-like surface protein A (ScIA)
5	AWM59_08310	-2.38	6.45	0.00	0.99	LPXTG anchor domain surface protein
5	AWM59_08315	-4.32	5.53	0.00	1.00	peptidase C5 (ScpA)
5	AWM59_08325	-8.65	5.31	0.00	1.00	M28 protein (M28)
5	AWM59_08330	-8.01	6.54	0.00	1.00	emm-like protein
5	AWM59_08335	-1.24	6.39	0.04	0.45	Trans-Acting Positive Regulator (Mga)
6	AWM59_08385	-4.57	6.93	0.00	1.00	fibronectin-binding protein (SfbX)
6	AWM59_08390	-5.11	7.31	0.00	1.00	Serum Opacity Factor (SOF)
6	AWM59_08395	-6.47	3.57	0.00	1.00	hypothetical protein

314 The majority of the differentially expressed genes are located in approximately 6
315 separate operons or gene clusters as indicated in Table 4. Interestingly, several of
316 these gene groups are transcriptionally regulated, at least in large part, by activity of the
317 multiple gene regulator protein, Mga [37-39]. During mid-exponential growth phase,
318 Mga acts as a transcriptional activator to regulate a core set of virulence factors at the
319 *mga* locus [37]. The *mga* locus consists of several components: a) the M protein (*emm*
320 gene) a major surface protein involved in resistance to phagocytosis and intracellular
321 killing by neutrophils and used to distinguish *S. pyogenes* isolates, b) a fibronectin-
322 binding protein that binds host complement regulator factors, c) an *emm*-like protein that
323 binds IgG and fibrinogen, d) the C5a peptidase (*ScpA*) which cleaves C5a chemotxin,
324 e) the *enn* protein that binds IgA, and f) the *mga* gene itself. All genes at the *mga* locus
325 displayed log₂ fold changes ranging from -1.2 to -8.7 in the Δ RSM strain relative to WT
326 (Fig 3A). To confirm this differential expression, we again isolated total RNA from
327 strains during mid-exponential growth phase in C media broth culture and performed
328 quantitative RT-PCR for detection of transcripts *mga*, *emm28*, and *scpA* (Fig 3B).
329 Consistent with the RNA-seq results, the qRT-PCR results showed that these genes
330 analyzed were significantly down regulated in the Δ RSM strain, with approximately 5-
331 fold to over 300-fold decreased expression in the Δ RSM strain relative to WT (Fig 3B).
332 Complementation *in trans* in the Δ RSM/pRSM strain restored transcript expression
333 patterns similar to WT values. Deletion of the *mga* gene produced qRT-PCR results in a
334 similar trend to the Δ RSM strain for the examined transcripts, with significantly
335 decreased detection of *emm28* and *scpA* transcripts; *mga* transcript was not detected in
336 the Δ *mga* strain (Fig 3B). Examination of these transcripts in the *hsdR* insertional

337 inactivation mutant Ω RE showed transcript detection of *mga* and *emm28* comparable to
338 WT levels, with detection of *scpA* transcript approximately four to five-fold of WT levels.
339 This transcript pattern was very different than those of the Δ RSM and Δ *mga* mutant
340 strains and more similar to the WT pattern. Even though the spectinomycin resistance
341 cassette insertion into *hsdR* may have produced some polar effect with slightly
342 decreased methyltransferase activity as noted on the anti-m6A dot blot (Fig 1C), it
343 seems sufficient residual m6A base modifications persisted to not significantly disrupt
344 gene expression (Fig 3B). Taken together, these results from RNA sequencing and
345 qRT-PCR provide evidence that m6A base modifications correlate with patterns of
346 differential gene expression in *S. pyogenes*, including those of several recognized
347 virulence factors and major regulators of virulence gene expression.

348 **Disruption of m6A DNA modifications enhances the host inflammatory response**
349 **to streptococcal infection in a murine subcutaneous ulcer model.** Given that the
350 genes in the Mga regulon were significantly down regulated in the Δ RSM strain relative
351 to WT, we were interested in determining the impact of disrupting m6A DNA
352 modifications on *S. pyogenes* virulence using a murine subcutaneous infection model
353 [40, 41]. C57BL/6J mice were inoculated at the shaved flank with 1×10^7 CFUs of either
354 MEW123 (WT) or the Δ RSM mutant strain and resulting skin ulcers were photographed
355 daily for sizing the skin ulcer area. As shown in Fig 4A, there was no significant
356 difference in skin lesion size at day two post-infection in comparison of the mice infected
357 with either the WT or the Δ RSM strains. However, by three to four days post-infection,
358 and for the remainder of the experiment, the skin lesions of mice infected with the
359 Δ RSM strain were significantly larger than those of mice infected by the WT strain (Fig

4A). No strain caused a lethal infection among any of the mice with the 1×10^7 CFU inoculum. Representative images of skin lesions for mice infected with the WT, the Δ RSM strain, and the complemented Δ RSM/pRSM strain over time are shown in Fig 4B, with skin lesions of mice infected with the Δ RSM strain notably larger on average at 4 and 6 days compared to those of mice infected with the WT or complemented strain. Complementation of the Δ RSM mutation *in trans* by strain MEW552 (Δ RSM/pRSM) produced murine skin lesions smaller than the Δ RSM mutant but not significantly different than the WT strain throughout the duration of the experiment (Fig 4C).

Skin lesion sizes reached a mean peak size at four to six days post infection. To determine if the difference in skin lesion size correlated with the concentration of viable streptococci at the site of infection, the skin lesions of mice were dissected and homogenized at day four post-infection to obtain viable CFU counts. Upon dissection, we made the observation that skin lesions from mice infected with the Δ RSM strain were grossly more purulent than lesions of mice infected by the WT and complemented Δ RSM/pRSM strains. The skin lesions contained on average CFU counts of approximately 1×10^6 to 1×10^7 CFUs; while there was a slight trend to higher CFU counts on day four post-infection for the Δ RSM streptococci compared to the WT and complemented strain CFUs, there were no statistically significant differences in CFU counts between these groups (S2 Fig). We noted that skin lesions of mice infected with the WT and complemented strain Δ RSM/pRSM strains seemed to heal more quickly than those of mice infected with the Δ RSM strain (Fig 4C).

With the subcutaneous ulcer model, skin lesion size tends to correlate closely with the degree of the host immune response, with particular regards to the neutrophil

influx [40, 41]. To compare the inflammatory response in skin lesions of mice infected with the WT and the Δ RSM strain, we performed skin biopsies for cytokine analysis and histologic examination at six-days post-infection; this time point was chosen as it was the time point with the greatest difference in skin lesion size between the experimental groups. Measurements of interleukin-1 beta (IL-1 β), interleukin-6 (IL-6), interleukin-17A (IL-17A), and tumor necrosis factor alpha (TNF α), were obtained as evidence of pro-inflammatory cytokine activity. Cytokine concentrations for all four cytokines measured were significantly greater from mice infected with WT streptococci than mice mock-infected with sterile phosphate-buffered saline (PBS) (Fig 5A). Cytokine concentrations from mice infected with the Δ RSM strain were significantly greater than mock-infected or mice infected with the WT strain (Fig 5A). Furthermore, histologic analysis of skin lesions shows predominantly increased neutrophil influx, but also a modest increase in the number of macrophages in the subcutaneous tissue of mice infected with the Δ RSM strain compared with WT (Fig 5B). Infiltration of T lymphocytes was not appreciably different between skin lesions of mice infected with WT or the Δ RSM strain (Fig 5B). Cytokines IL-6 and IL-17A, in particular, are important for coordinating neutrophil trafficking to areas of infection [42-44]. Our results in mice infected with the Δ RSM strain showing enhanced pro-inflammatory cytokine detection, increased neutrophil infiltration, and larger skin lesions, suggests an effect of altered gene transcription patterns in the Δ RSM strain and a more robust host inflammatory response compared to mice infected with the WT parent strain. Given the known association of several of the streptococcal gene transcripts down regulated in the Δ RSM strain, including *mga*, *emm28*, and *scpA*, with immune evasion properties, we hypothesized that m6A DNA modifications and

proper regulation of gene expression are important contributors to immune evasion strategies and/or disruption of host immune responses by *S. pyogenes*.

To determine if the loss of specific virulence factors recapitulates the phenotype of the Δ RSM strain in the murine subcutaneous ulcer model, we infected mice with derivatives of strain MEW123 with in-frame deletions of *mga* (strain MEW480, Δ Mga), and spectinomycin-resistance cassette disruption mutations of *emm28* (strain Ω emm28) and *scpA* (strain 380, Ω scpA). As shown in Fig 4C, infection of mice by the Δ Mga strain produced skin lesions significantly larger than the WT strain and comparable to the Δ RSM strain in size throughout the experiment. Infection by the Ω emm28 strain was not statistically different than the WT strain at day 2 and day 4 post-infection; however, by day 6 and day 8 post-infection, the Ω emm28 strain produced lesions that were statistically significantly larger than the WT (Fig 4C). Infection of mice by the Ω scpA strain produced the widest range of murine skin lesion sizes, with some mice having very large lesions following infection (Fig 4C); however, at no time point was the average size of the lesions produced by the Ω scpA strain statistically different than WT. Overall, these results suggest that the presence of m6A DNA base modifications produced by M.SpyMEW123 activity correlate with differential transcriptional expression of several *S. pyogenes* virulence factors, especially those within the Mga operon, and that these seem to influence host-pathogen interactions at the site of infection.

Disruption of m6A DNA modifications inhibits streptococcal survival within human neutrophils. A major function of the *S. pyogenes* M protein is to promote streptococcal survival, resisting killing by human leukocytes by interfering with

bactericidal activity within neutrophils following phagocytosis [45, 46]. Staali *et al.* found that *S. pyogenes* strains with or without M protein underwent phagocytosis by neutrophils to similar levels, but only strains expressing M protein survived intracellularly whereas strains lacking M protein expression were rapidly killed [45]. Given our findings that elimination of m6A DNA modifications was associated with decreased transcript expression for *mga* and *emm28*, we wished to compare survival within human neutrophils. Purified human neutrophils were incubated with WT or Δ RSM *S. pyogenes* strains using a neutrophil bactericidal assay similar to a previous report [45]. Briefly, streptococci and neutrophils were mixed together allowing the neutrophils to internalize *S. pyogenes* strains followed by elimination of extracellular bacteria with penicillin and gentamicin. It was previously determined that there was no significant difference in susceptibility to penicillin and gentamicin at the high concentrations used in these experiments between the WT or Δ RSM strains (S1 Fig). Streptococcus surviving within neutrophils were liberated by treatment with the detergent saponin and plated for viable CFUs. As shown in Figure 6, we utilized serotype M14 HSC5 and a derivative strain with disruption in the M14 *emm* gene (Ω *emm14*) as positive and negative controls, respectively. As expected, the Ω *emm14* mutant was significantly attenuated for intracellular survival within neutrophils compared to the M14 parent strain (Fig 6). Similarly, we compared survival of the MEW123 parent strain (M28) and its cognate strain with disruption of the M28 *emm* gene (Ω *emm28*) or the Δ RSM mutant. We found that both the Ω *emm28* and the Δ RSM mutant were significantly attenuated for intracellular survival compared to the M28 parent strain, further confirming the role of M protein in promoting intracellular neutrophil survival by the serotype M28 MEW123

strain, in addition to demonstrating correlation of m6A DNA base modifications with differential expression of M protein (Fig 6). These results provide further support for m6A DNA base modifications in *S. pyogenes* as important for promoting streptococcal virulence, possibly by influencing virulence factor expression.

Disruption of m6A DNA modifications inhibits adherence to human vaginal epithelial cells *in vitro*, but does not appear to impair carriage *in vivo* in a murine vaginal colonization model. From the RNA-Seq results we found that the Δ RSM strain had significantly decreased transcript expression of several recognized and known adhesin proteins, including M28, M-like protein, collagen-binding protein, and fibronectin-binding proteins, as well as several hypothetical surface proteins [38, 47-49]. As a group, serotype M28 *S. pyogenes* are overrepresented in cases of human infection within the female urogenital tract, including vulvovaginitis and puerperal sepsis (a.k.a. “childbed fever”) [50-53]. Serotype M28 *S. pyogenes* have a particular predilection for cervical and vaginal epithelium due to surface proteins, including protein R28 among others, which may explain the overrepresentation of this serotype with infections in this niche [15, 54]. Therefore, we asked if m6A DNA modifications influenced adherence of the serotype M28 MEW123 strain to human vaginal epithelial cells.

As shown in Figure 7A, disruption of m6A DNA modifications in the Δ RSM strain was indeed associated with significantly decreased adherence to human vaginal epithelial cells *in vitro* compared to the WT parent strain. The attenuation in vaginal epithelial cell adherence by the Δ RSM strain was comparable to a strain lacking expression of the M protein (Δ *emm28*), suggesting that decreased expression of M protein, among other adhesins, by the Δ RSM strain was at least partly responsible for

decreased adherence (Fig 7A). To determine if impaired adherence to human vaginal cells *in vitro* translated to impaired vaginal mucosal colonization *in vivo*, we utilized a murine vaginal model and compared streptococcal carriage burdens over time [40, 55]. In contrast to the results of the *in vitro* adherence assay, using the murine vaginal carriage model we found no significant difference in vaginal streptococcal burdens in comparison of mice inoculated with either the WT or the Δ RSM strains over the course of the 28-day experiment (Fig 7B). Given that human cells are the natural hosts of *S. pyogenes*, this may be an example of the human-restricted nature of *S. pyogenes* in which a murine model cannot adequately replicate the natural human environment in which this pathogen evolved to survive. Nevertheless, our overall results showed several key differences in virulence phenotypes correlating with alterations in gene transcription associated with streptococcal m6A DNA methylation.

Discussion

In this report, we provide evidence that m6A DNA base modifications influence gene transcription patterns and overall virulence properties in a major gram-positive bacterial pathogen of humans, *S. pyogenes*. The *S. pyogenes* RM system, SpyMEW123I, is a Type I RM system and is responsible for the majority of m6A base modifications distributed throughout the *S. pyogenes* genome. The target consensus sequences identified by our study, 5' GCANNNTTYG and its corresponding partner motif 5' CRAANNNTGC, were consistent with m6A motifs identified in *S. pyogenes* previously reported by Blow *et al* [5]. We found approximately 412 occurrences of each m6A site with the majority found within coding regions. Interestingly, we found that not

498 all m6A sequence motifs were consistently modified to the same extent; only about 70%
499 of consensus sites were modified in at least 75% of sequencing reads, suggesting that
500 m6A modifications may be intermittently present with additional functions beyond simple
501 protection from restriction, including influencing gene expression patterns based on
502 timing of hemi- or full-methylation status. It is not known at this time whether all of the
503 m6A sites, or only the sites within the intergenic regions, would participate in influencing
504 transcriptional expression, but methylation events modifying access of transcriptional
505 regulators to intergenic promoter regions would be a potential mechanism.

506 With the introduction of SMRT sequencing, groups have now identified m6A DNA
507 modifications within a diversity of prokaryotes, including *E. coli*, *Campylobacter jejuni*,
508 *Salmonella enterica* serovar Typhimurium, *Vibrio breoganii*, *Geobacter metallireducens*,
509 *Chromhalobacter salexigens*, *Bacillus cereus*, and *Borrelia burgdorferi* [33-35, 56, 57].
510 Additional evidence of 5-methylcytosine (m5C) DNA modifications influencing
511 transcriptional expression of multiple genes with an impact on several phenotypic traits
512 has recently been described in *Helicobacter pylori*, further expanding the recognized
513 influence of prokaryotic methylation modifications [58]. Some of the DNA modifications
514 described have been linked to orphan MTases without an associated endonuclease,
515 such as DNA Adenine Methyltransferase (Dam) of *S. enterica*, *E. coli*, and *Haemophilus*
516 *influenzae* [56, 59, 60]. Uncoupling DNA methylation from restriction endonuclease
517 protection is conceptually easier to envision with an orphan MTase, freeing the orphan
518 MTase to have roles in DNA mismatch repair and influencing gene expression of
519 potential virulence factors [6]. Indeed, Dam-dependent DNA modifications in *S. enterica*
520 have been linked to alterations of gene expression and virulence [56]. However, two

521 examples have recently been reported in *C. jejuni* and *B. burgdorferi* of intact RM
522 systems also influencing gene expression patterns [34, 35]. Both of the RM systems in
523 these organisms are representatives of Type IIG RM systems, which differ significantly
524 from the Type I RM system described here for *S. pyogenes* in that they consist of a
525 single polypeptide with both REase and MTase activity [34, 35, 61]. The effects on gene
526 expression conferred by these systems in *C. jejuni* and *B. burgdorferi* were noted by
527 Casselli *et al.* to be more modest in terms of numbers of genes influenced by m6A base
528 modifications when compared to the larger number of transcriptional changes found
529 from the standalone activity of Dam MTase in Salmonella [34, 35, 56]. It would seem
530 that with an intact RM system the conditions involved in determining gene expression is
531 more stringent and regulates a fewer number of genes than orphan MTases. DNA
532 methylation from Type I RM systems has also been well-established in phase variation
533 in a number of Gram-positive pathogens, including *Streptococcus pneumoniae*,
534 *Streptococcus suis*, *Listeria monocytogenes*, and *Mycoplasma pulmonis*, which can
535 have downstream effects on gene expression [62-65]. In phase variation, switching of
536 specificity subunits of Type I RM systems results in cells with different sites of
537 methylation within the population, which can create heterogeneity in gene expression.
538 The role of methylation in phase variation differs from our findings here as we show that
539 loss of methylation at a single site (i.e. not switching of specificity subunits to create
540 methylation at diverse sites) results in the down regulation of a very defined subset of
541 genes.

542 The M.SpyMEW123I MTase activity we describe here modifies 412 sites in the
543 MEW123 genome, whereas Dam-modified recognition sites approximate 19,000 per

chromosome [56]. Perhaps the context of the m6A recognition motif in a particular intergenic promoter region, combined with specific transcription factors sensitive to the presence or absence of m6A modifications, determines the specificity of which genes an intact RM system will influence. Our results reported here demonstrate that the *S. pyogenes* Type I RM system is functional as a protective mechanism with restricting uptake of foreign DNA (Fig 2A). Similar results were found by Okada et al. in a series of *emm1* *S. pyogenes* isolates from Japan with spontaneous deletions in their Type I RM systems; isolates lacking the Type I RM system had significantly increased rates of transformation with foreign plasmid DNA [21]. While their study did not specifically address virulence properties of isolates lacking the RM system, the authors speculated that enhanced rates of DNA uptake and transformation exhibited by strains lacking REase activity may be beneficial by allowing uptake of potentially advantageous genes from the environment contributing to overall fitness.

Inactivation of the SpyMEW123I RM system was associated with significant dysregulation of gene transcript expression in broth culture, with 20 genes from at least six separate gene clusters/operons significantly down regulated (Table 4). Notable among the down regulated genes were the trans-acting regulator Mga, the M-like protein, M28 protein, C5a peptidase (ScpA), a cell surface protein, a collagen-like surface protein (SclA), the Serum Opacity Factor (SOF), and a fibronectin-binding protein (SfbX). Most of these genes are regulated by the Mga transcriptional regulator in serotypes that have been investigated. Mga is a ubiquitous stand-alone regulator primarily active during exponential growth phase and is responsible for influencing expression of over 10% of the *S. pyogenes* genome, primarily genes involved in

567 metabolism, but also many virulence factors including adhesins and surface proteins
568 involved in immune evasion [37, 38]. Mga binds to upstream promoter regions to
569 activate high-level transcription of genes in the Mga core regulon [66]. The majority of
570 Mga-regulated promoters, including most of the genes in the core Mga regulon, contain
571 a single Mga binding site centered around position -54 and overlapping the -35 region of
572 the gene promoter, likely interacting with the α -subunit of RNA polymerase [67]. In
573 theory, m6A base modifications at or around this site could potentially influence Mga
574 and RNA polymerase binding to the promoter region, perhaps by stabilizing or localizing
575 Mga to the proper site, promoting activation of gene transcription. Consistent with this
576 hypothesis, examination of the genome sequences upstream of the Mga open reading
577 frame for *S. pyogenes* strains MEW123, MEW427, and SF370, all reveal the existence
578 of the m6A consensus motifs approximately 800 bp upstream of the *mga* start codon
579 [12, 13, 20]. It is unclear if, or how, this m6A motif site located upstream of the predicted
580 Mga promoter region activates Mga expression. The mechanism of m6A-dependent
581 regulation of the *mga* locus is the subject of active investigation by our group.

582 Regulation of virulence factor expression in response to different environmental
583 cues and stresses is critical to the success of *S. pyogenes* survival and pathogenesis.
584 Over 30 recognized transcriptional regulatory proteins and 13 two-component regulatory
585 systems must function to coordinate virulence factor expression properly [17, 18]. We
586 found that loss of m6A DNA modifications in our Δ RSM mutant correlated with
587 significant changes in virulence properties of *S. pyogenes*. In a murine model of
588 subcutaneous ulcer formation, we noted that mice infected with the Δ RSM mutant
589 displayed enhanced inflammatory responses compared to mice infected with the WT

590 strain, with comparatively larger skin lesions, increased detection of pro-inflammatory
591 cytokine levels, and enhanced neutrophil infiltrates on histologic examination (Fig 4 and
592 5). Disruption of m6A DNA modifications and an associated dysregulation of gene
593 transcript expression may result in failed activation of multiple important adhesins and
594 streptococcal proteins involved in evading host immunity (Fig 3 and Table 4). For
595 example, neutrophilic infiltration in response to bacterial infections is enhanced by
596 activity of host chemotaxins, chiefly complement protein C5a. A major virulence
597 determinant of *S. pyogenes* aiding immune evasion is to degrade complement C5a
598 through activity of ScpA, a surface-expressed, serine-protease specifically degrading
599 host C5a and interfering with neutrophil recruitment [68]. We found that the Δ RSM
600 mutant exhibited significantly decreased transcript expression for ScpA which may
601 partly explain a more exaggerated neutrophil response to infection with the Δ RSM
602 mutant strain, resulting in more inflammation and larger skin lesions (Fig 4C and 5).
603 Previous investigation into the contribution of ScpA to host immune responses was
604 performed using a murine air sac model of subcutaneous infection performed by Ji et al.
605 [69]; air sacs infected with *S. pyogenes* lacking ScpA expression exhibited a
606 significantly enhanced host inflammatory response compared to the WT parent, with a
607 neutrophil predominance analogous to our results. Another report found similar to
608 slightly larger skin lesions in mice infected subcutaneously with *S. pyogenes* lacking
609 ScpA compared to WT [70]. The effect of *S. pyogenes* virulence factors in murine
610 models is not always similar to activity in the human environment; it is known that ScpA
611 does cleave murine C5a, but at slower rates compared to human C5a, and these
612 differences may impact our ability to detect phenotypes in these non-human systems

[71]. Similar to our own results with the Ω_{scpA} strain infections, the results reported by Li et al. were not statistically significant suggesting that the individual contribution of ScpA in this murine model may be modest, but when the expression of multiple virulence factors is disrupted the effects may be more apparent. Indeed, our experiments in the skin lesion model with the ΔRSM and the Δmga strains showed significant differences in lesion size and inflammatory response compared to the WT and complemented mutant strains. Both of these mutant strains would be expected to have similar patterns of differential gene expression and as result they phenocopy each other in this model. Decreased expression of several adhesins and other factors may have contributed to enhanced spread of the infection together with an exaggerated host inflammatory response resulting in larger areas of inflammation and larger skin lesion formation.

Decreased M protein expression, among other adhesins, also explains the decreased *in vitro* adherence of the ΔRSM mutant to human vaginal epithelial cells. Interestingly, the decreased adherence to human vaginal epithelial cells *in vitro* did not correlate with disrupted carriage in the murine vaginal mucosa colonization model. This suggests that there are additional adhesins not influenced by m6A DNA modifications that are important for promoting and maintaining carriage *in vivo*. One example would be the R28 adhesin of serotype M28 *S. pyogenes* strains, which is a major streptococcal adhesin to human cervical epithelial cells [54]. Our RNA-sequencing experiments did not find significant differences in the transcription of the MEW123 R28 gene (AWM59_02815) between WT and the ΔRSM mutant (full data set available in NCBI repository). With only 20 genes significantly downregulated in the ΔRSM mutant

clearly not all major *S. pyogenes* adhesins and virulence factors are impacted by m6A DNA modifications. Our data show that only a few gene operons, or regulons as in the case of Mga, are differentially expressed in the absence of m6A base modifications in *S. pyogenes* and that down regulation of these genes impacts virulence.

In this study, we have demonstrated that the SpyMEW123I RM system and m6A DNA modifications in *S. pyogenes* significantly influence DNA restriction activity, in addition to correlating with differential gene transcription and virulence properties of this important human pathogen. Disruption of the SpyMEW123I Type I RM in *S. pyogenes* altered the transcriptional profile of the mutant strain resulting in attenuated virulence and impaired evasion of the host immune response in both *in vitro* and *in vivo* models. Similar to our results, disruption of Type IIG RM systems in *C. jejuni* and *B. burgdorferi* also interfered with genetic regulation of virulence factors of those pathogens [34, 35]. Together, these findings demonstrate that intact RM systems in these bacterial pathogens, and likely many other prokaryotes, can exert multiple functions, including restriction-mediated protection from foreign DNA in addition to influencing gene expression. Understanding how m6A DNA modifications influence virulence properties in these organisms could potentially identify targets for therapeutic intervention, potentially changing patterns of virulence factor expression resulting in strain attenuation helping to prevent human disease. Further investigation is necessary to fully comprehend the many functions of DNA methylation and the complex nature of bacterial physiology and pathogenesis.

Materials and methods

Ethics statement. Experimental protocols involving the use of mice were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Michigan Medical School (Ann Arbor, MI, USA). The University of Michigan IACUC complies with the policies and standards as outlined in the Animal Welfare Act and the “Guide for the Care and Use of Laboratory Animals,” [72]. The protocol numbers approved by the University of Michigan IACUC are as follows: Skin and Soft Tissue Infection Model of *Streptococcus pyogenes* Virulence (PRO00007495), and Murine Vaginal Colonization Model for *Streptococcus pyogenes* (PRO00007218). For consistency, all experiments utilized female C57BL/6J mice at approximately 6 weeks of age at the time of use. Mice were purchased from The Jackson Laboratories (catalog #000664) (Bar Harbor, ME, USA), and maintained in a University of Michigan animal facility with biohazard containment properties. Following arrival, mice were allowed to acclimatize in the facility for one week prior to beginning experiments. When manipulated, mice were briefly sedated by inhalation of isoflurane via drop jar dosing. Animals were inspected at least once daily for evidence of suffering, manifested by significantly diminished or no activity, decreased appetite, poor grooming, increased respiratory rate, or weight loss greater than 15% of body weight; if evidence of suffering was identified, then the mouse was euthanized. Euthanasia was primarily through carbon dioxide asphyxiation with a subsequent secondary method of euthanasia, including induction of bilateral pneumothorax, decapitation, and/or removal of a vital organ.

Bacterial strains, media, and growth conditions. The principal strain used in this study was *S. pyogenes* MEW123, a streptomycin-resistant (*rpsL*_{K56T}), serotype M28

pharyngeal isolate [55]. Other strains used are listed in Table 1. Growth rates and yields of MEW123 and associated mutant strains were measured using a Synergy HTX plate reader (BioTek, Winooski, VT, USA) in 96 well plates (Greiner Bio-One, Monroe, NC, USA). Briefly, 4µl of overnight culture grown in THY broth was inoculated into 200µl of the described fresh media, with identical strains and conditions measured in at least triplicate. Growth was at 37°C, room air, in static conditions for 12 hours and OD_{620nm} was measured every 3 seconds. Unless otherwise noted, all *S. pyogenes* strains had equivalent growth rates and yields under all *in vitro* conditions tested (S1 Fig). Routine culture of *S. pyogenes* was performed in Todd-Hewitt medium (Becton Dickinson, Franklin Lakes, NJ, USA) supplemented with 0.2% yeast extract (Difco Laboratories, Detroit, MI, USA) (THY media). Where required, Bacto agar (Difco) was added to a final concentration of 1.4% (w/v) to produce solid media. Gene expression experiments used C-Medium, a lower-glucose, higher-protein media that more closely resembles *in vivo* conditions [73]. Incubation was performed at 37°C under anaerobic conditions (GasPack™, Becton Dickinson) for solid media, or in sealed tubes without agitation for broth media. Aerobic culture was conducted as described [74]. For inoculation of mice, *S. pyogenes* was harvested from culture in THY broth at early logarithmic-phase (OD₆₀₀ 0.2), washed once in PBS, briefly sonicated on ice to break up long streptococcal chains, and resuspended in PBS to 10⁸ CFU/ml. Molecular cloning used *Escherichia coli* strain DH5a (Invitrogen, Grand Island, NY, USA) cultured in LB broth. When appropriate, antibiotics were added at the following concentrations: erythromycin, 500 µg/mL for *E. coli* and 1 µg/mL for *S. pyogenes*; chloramphenicol, 20 µg/mL for *E. coli* and 3 µg/mL for *S. pyogenes*; spectinomycin, 100 µg/mL for both *E. coli*

and *S. pyogenes*; and streptomycin, 1000 µg/mL for *S. pyogenes*. In some experiments, growth was monitored in THY broth supplemented with either penicillin, gentamicin, or erythromycin at concentrations ranging from 0.05 µg/mL to 100 µg/mL. All antibiotics were obtained from Sigma Chemical Co., St. Louis, MO, USA.

Gene cloning and mutant construction. *Streptococcus pyogenes* MEW123 was used as a source strain for DNA, Genbank CP014139.1 [12]. Bacterial strains and plasmid vectors are listed in Table 1. The primers used for PCR amplification and cloning are listed in Table 5. For cloning and routine DNA Sanger sequencing, the Phusion High-Fidelity DNA Polymerase (New England Biolabs, Inc., Ipswich, MA, USA) was used. For routine endpoint PCR amplification standard Taq DNA Polymerase was used (New England Biolabs, Inc.). Polymerase chain reaction products were digested with indicated restriction enzymes and ligated to pJRS233 or pGCP213 for in-frame deletions, pSPC18 for insertional mutations, or pJoy3 as a plasmid vector for transformation efficiency assays. In-frame deletion mutants and insertional mutants were constructed essentially as described [25, 26], [28], and [27], respectively.

a) Construction of an in-frame deletion of *SpyMEW123I* gene cluster.

The Δ RSM in-frame deletion allele was cloned by splice overlap extension (SOE) PCR [75]. Corresponding GenBank accession numbers for the MEW123 restriction endonuclease gene *hsdR*, specificity subunit *hsdS*, and the methyltransferase subunit *hsdM*, are AWM59_07895, AWM59_07900, and AWM59_07905, respectively. The upstream region of the gene cluster was PCR amplified using primers MEW123 Del-RSM F1 and MEW123 Del-RSM R2, producing a 1.02 kb amplicon. The downstream region of the gene cluster was

PCR amplified using primers MEW123 Del-RSM F3 and MEW123 Del-RSM R4, producing a 1.02 kb amplicon. These two amplicons contain complementary ends that anneal together and essentially will produce an in-frame deletion of the three-gene restriction endonuclease, specificity subunit, and DNA methyltransferase open reading frames. The two amplicons were mixed together as template and further amplified using primers MEW123 Del-RSM F1 and MEW123 Del-RSM R4, the resulting amplicon was approximately 2.04 kb and contained a unique EcoRI site at the 5' end and a unique HindIII site at the 3' end. The resulting amplicon was digested with EcoRI and HindIII, and inserted within same restriction sites of the *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacements, pGCP213 [26], producing plasmid pKJ24. The pKJ24 plasmid was confirmed by Sanger DNA sequencing using primers MEW M13 F and MEW M13 R, which bind just outside and flank the multiple cloning site region within pGCP213. Electrocompetent cells of MEW123 were prepared and transformation was performed essentially as previously described [76]. The pKJ24 plasmid carrying the RSM in-frame deletion was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Erythromycin-resistant transformants were handled according to the temperature-sensitive selection protocol as previously described [26]. Final clones of *S. pyogenes* that had successfully replaced the full-length genomic RSM gene cluster with the in-frame deletion allele were screened by endpoint PCR and confirmed by Sanger DNA sequencing. The resulting strain containing the in-frame deletion allele (Δ RSM) was identified as MEW513.

b) Construction of Δ RSM strain complemented *in trans* with plasmid-encoded RSM operon. GenBank accession numbers for the MEW123 restriction endonuclease gene *hsdR*, specificity subunit gene *hsdS*, and the methyltransferase subunit gene *hsdM*, are AWM59_07895, AWM59_07900, and AWM59_07905, respectively. The operon was cloned by PCR using primers pJoy3_123_RSM_F and pJoy3_123_RSM_R, producing an amplicon of approximately 6 kb. This fragment was inserted into plasmid pJoy3 linearized by digestion with EcoRI and SphI using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs, Inc.), producing plasmid pEH01. This plasmid was transformed into electrocompetent *S. pyogenes* MEW513 through electroporation with conditions as described above. Chloramphenicol-resistant clones were selected and screened by endpoint PCR, with restoration of m6A methylation activity confirmed by dot blot. The resulting strain containing the plasmid encoded RSM operon for complementation (Δ RSM/pRSM) was identified as strain MEW552.

c) Construction of spectinomycin-cassette disruption mutant of restriction endonuclease gene, *hsdR*. The GenBank accession number for the restriction-endonuclease subunit gene, *hsdR*, is AWM59_07895. A fragment of the endonuclease open reading frame was cloned by PCR using primers 123_7895_F and 123_7895_R, producing an amplicon of approximately 950 bp. This fragment was inserted into plasmid pSpc18 linearized by digestion with HindIII and BamHI using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs, Inc.), producing plasmid pKJ19. This plasmid was transformed into

electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the restriction endonuclease gene *hsdR* (Ω RE) was identified as strain MEW489.

d) Construction of in-frame deletion of *mga*. The GenBank accession number for the MEW123 Mga protein, gene *mga*, is AWM59_08335. An in-frame deletion allele of *mga* was cloned by splice-overlap extension (SOE) PCR [75]. The upstream region of the *mga* gene was cloned using primers M28 Mga 5' Sall and M28 Mga 5' SOE R, producing an amplicon of approximately 420 bp. The downstream region of the *mga* gene was cloned using primers M28 Mga 3' BamHI and M28 Mga 3' SOE F, producing an amplicon of approximately 410 bp. The two amplicons are mixed together as template and amplified using the outside primers M28 MGA 5' Sall and M28 Mga 3' BamHI, producing an amplicon of approximately 830 bp. This amplicon was subsequently digested with BamHI and Sall and ligated into the *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacements, plasmid pJRS233 [25], cut similarly with BamHI and Sall. The resulting plasmid was named pIL01, with confirmation by Sanger DNA sequencing and PCR verification. Electrocompetent cells of MEW123 were prepared and transformation with plasmid pIL01 was performed essentially as previously described [76]. Erythromycin-resistant transformants were handled according to the temperature-sensitive selection protocol as

previously described [26]. Final clones of *S. pyogenes* that had successfully replaced the full-length genomic *mga* allele with the in-frame deletion allele were screened by endpoint PCR and confirmed by Sanger DNA sequencing. The resulting strain containing the in-frame deletion allele (Δmga) was identified as MEW480.

e) Construction of spectinomycin-cassette disruption mutant of strain MEW123 *scpA* gene (ScpA protein). The GenBank accession number for the MEW123 *scpA* gene is AWM59_08315. A fragment of the *scpA* open reading frame was cloned by PCR using primers M28 ScpA Sall F and M28 ScpA SacI R, producing an amplicon of approximately 1.1 kb. The amplicon was digested with Sall and SacI and ligated into plasmid pSpc18 linearized with Sall and SacI, producing plasmid pIL09. This plasmid was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the *scpA* gene ($\Omega scpA$) was identified as strain MEW380.

f) Construction of spectinomycin-cassette disruption mutant of strain MEW123 *emm28* gene (M28 protein). The GenBank accession number for the M28 protein, gene *emm28*, is AWM59_08325. A fragment of the *emm28* open reading frame was cloned by PCR using primers M28 Emm HindIII F and M28 Emm BamHI R, producing an amplicon of approximately 1.1 kb. This amplicon incorporated unique sites for HindIII and BamHI, and the amplicon was

accordingly restriction digested and ligated into plasmid pSpc18 opened with HindIII and BamHI, producing plasmid pIL03. This plasmid was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the *emm28* gene (Ω_{emm28}) was identified as strain MEW409.

Table 5. Primers used in this study.

Primer	Sequence 5'-3' (restriction sites underlined)
MEW123 Del-RSM F1	ATATGAATTCGGTTTTTTGGTAAAAAACTTTTTTAGCA
MEW123 Del-RSM R2	TTTTTTGGTCTTTTTTAATCCCCATTTCGACATGATA
MEW123 Del-RSM F3	TGGGGATTAAAAAAGACCAAAAAACACCACAGTAGA
MEW123 Del-RSM R4	ATATAAGCTTTTAATTTAACAAATATTTCTAAAGAAAAATGGATTGG
pJoy3 123 RSM F	ACCATTATTGTGAGGAAGTGCCTTACCGATCCCTTAAAAG
pJoy3 123 RSM R	ACCGATAGCACCCGCGCATGGATGAGATGATTCTATTTTGATTTATAG
123_7895 F	TAAACGACGGCCAGTGCCAATCTCTTAGAAACAGGTGAAAG
123_7895 R	ATTCGAGCTCGGTACCCGGGAGATATCATTTTGCGCATAG
M28 Emm HindIII F	CCCAAGCTTATAAACAGTATTCGCTTAGAAAATTTAAAAACAGG
M28 Emm BamHI R	CGCGGATCCGTTAGCTGCTTCGCCTGTTGACGGTAACG
M28 ScpA Sall F	CCCGTTCGACCTCAATGCACAATCAGACATTAAAGC
M28 ScpA SacI R	CCCAGAGCTCTCAATATCGCCACGTTCAATAAGG
M28 Mga 5' Sall	CCCGTTCGACTGACAATAATGTCACAGAT
M28 Mga 5' SOE R	TTGTTGGCTAGTAAACAATTTACTTACATGC
M28 Mga 3' SOE F	GTTTACTAGCCAACAAGCAACATCATCATAGGATTTTCAGACG
M28 Mga 3' BamHI	CGCGGATCCCGCTCTTCGAATACTTTGTT
Emm28 RT-PCR F	CAGACTTAGCAGAAGCAAATAGC
Emm28 RT-PCR R	CAGCTTGTTTAGCCAATTGCTC
Mga RT-PCR F	CTTATCTACCCTCAAACGCCTC
Mga RT-PCR R	CGAATTTGCCTCTCATCTCCTG
ScpA RT-PCR F	CACTGATTTTGATGTGATTGTAGACAA
ScpA RT-PCR R	ATGCAAGTGTCAAACGACGATCT
recA RT-PCR F	ATTGATTGATTCTGGTGCGG
recA RT-PCR R	ATTTACGCATGGCCTGACTC
MEW M13 F	CAGGGTTTTCCCAGTCACGAC
MEW M13 R	GAGCGGATAACAATTCACACAGG

829

830 **Transformation efficiency assay.** Transformation efficiency was assessed by
831 electroporation of electrocompetent *S. pyogenes* strains with 0.5 µg plasmid pJoy3
832 conferring chloramphenicol resistance isolated from *E. coli* DH5α. Electroporation was
833 performed using a Gene Pulser II system (BioRad, Hercules, CA, USA) under the
834 following settings; Volts at 1.75 kV, capacitance at 400Ω, and resistance at 25 µf.
835 Transformants were plated onto THY agar supplemented with chloramphenicol. In
836 addition, a separate aliquot of the sample was plated onto THY agar with no antibiotics
837 to determine the total viable cell count. Transformation efficiency was determined as the
838 number of chloramphenicol resistant cells per total viable cell count.

839 **SMRT sequencing.** Genomic DNA was purified from *S. pyogenes* strains MEW123
840 (WT) and MEW513 (ΔRSM) using the Wizard Genomic DNA Purification Kit (Promega,
841 Madison, WI). Genomic DNA preparation, library preparation, and sequencing of
842 MEW123 was performed as previously described [12]. For MEW513, one Single
843 Molecule Real-Time (SMRT) cell was used to sequence the library prepared with 5 kb
844 mean insert size on the Pacific Biosciences RSII sequencer by the University of
845 Michigan Sequencing Core (<https://brcf.medicine.umich.edu/cores/dna-sequencing>).
846 Modification and motif analysis was performed using
847 RS_Modification_and_Motif_Analysis.1 version 2.3.0 using the published MEW123
848 reference genome with an average reference coverage of 501 and 539 for MEW123
849 and MEW513, respectively. Data generated in this analysis have been deposited in
850 NCBI's Gene Expression Omnibus [77] and are accessible through GEO Series

851 accession number GSE130428

852 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130428>).

853 **Dot blot assay for m6A modification.** Genomic DNA was isolated from *S. pyogenes*
854 strains MEW123 (WT) and MEW513 (Δ RSM), as described above. DNA was treated
855 with RNase during purification to remove any contaminating mRNA or rRNA potentially
856 containing m6A base modifications. DNA was denatured by heating at 98°C for 10 min
857 and then placed immediately on ice for 5 minutes. Denatured DNA or unmodified
858 oligonucleotides as a negative control was then spotted at 500 ng per spot onto
859 nitrocellulose membranes and allowed to air dry. Membranes were then placed onto
860 Whatman paper soaked with PBS containing 0.5% Tween 20 (PBST), and DNA was
861 cross-linked to the membranes using a Bio-Rad GS Genelinker using two 125 mJ
862 delivery cycles. Membranes were blocked in 5% milk protein in PBS for 1 h at room
863 temperature and then incubated with a dilution of anti-m6A primary rabbit antibody (2
864 μ g/ml) (EMD Millipore ABE572 Anti-N6-methyladenosine (m6A) Antibody) in 5% milk
865 PBS overnight at 4°C. Primary antibody was removed and the membrane was washed
866 three times with PBST for 5 minutes each wash. The membrane was then incubated
867 with a 1:5,000 dilution of horseradish peroxidase-conjugated anti-rabbit secondary
868 antibody in 5% milk PBS at room temperature for 1 hour. The secondary antibody was
869 removed, and the membrane washed with PBST three times for 5 minutes each wash.
870 Chemiluminescent substrate (Pierce SuperSignal West Femto HRP Substrate,
871 ThermoFisher Scientific, Waltham, MA) was applied and the membrane was visualized.

872 **RNA-sequencing.** Streptococcus from fresh overnight growth on THY agar plates was
873 inoculated into 40 ml of C-media broth and grown statically to mid-log phase OD_{600nm} of

874 0.6. RNA was then purified using the RiboPure RNA Purification Kit (Life Technologies),
875 for bacteria according to the manufacturer's recommendations. The University of
876 Michigan Sequencing Core performed ribosomal rRNA depletion using the Ribo-Zero
877 Magnetic Kit, bacteria and subsequent library preparation. Fifty-base single end reads
878 were sequenced on the Illumina HiSeq 4000. Sequence alignment was performed using
879 the Burrows-Wheeler Aligner (BWA) version 0.7.8-r455 to the MEW123 reference
880 genome [12]. Subsequent differential expression analysis was performed using the
881 limma package in R [78]. Differentially expressed genes were called as those that had a
882 Benjamini-Hochberg adjusted p-value less than 0.05 and a log₂ fold change greater
883 than 1. Log₂ CPM values were computed using edgeR and were subsequently used to
884 construct the heatmap using the *aheatmap* function as part of the NMF package in R
885 [79, 80]. Data generated in this analysis have been deposited in NCBI's Gene
886 Expression Omnibus [77] and are accessible through GEO Series accession number
887 GSE130427 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130427>).

888 **Real-time PCR for comparison of transcript expression.** Based on results of the
889 most significantly differentially expressed genes between WT (MEW123) and Δ RSM
890 (MEW513), three genes were selected for independent reverse-transcription cDNA
891 preparation and real-time PCR amplification for relative comparison of transcript
892 expression; *mga*, *emm28*, and *scpA*. RNA was isolated as described above from strains
893 grown in C-media broth to mid-log phase OD_{600nm} of 0.6. Synthesis of cDNA was
894 performed using the iScript™ cDNA Synthesis Kit (BioRad). Real time amplification of
895 select genes was performed using an iCycler Thermocycler (BioRad) and iQ SYBR
896 Green Supermix (BioRad). Sequences for RT-PCR primers are as shown in Table 5.

897 Relative transcript levels were determined using the *recA* transcript as reference by the
898 $2^{(-\Delta\Delta Ct)}$ method [81]. All RNA was stored at -80°C. All cDNA was stored at -20°C or
899 utilized directly for comparative RT-PCR analysis. For each experiment, three biological
900 replicates were analyzed in duplicate. Statistical significance was examined using the
901 paired t-test in Prism 6 (GraphPad).

902 **Murine subcutaneous infection model.** Inflammatory infection of murine
903 subcutaneous tissue was conducted as described in detail [41]. On the day of infection,
904 mice sedated by inhalation of isoflurane received a subcutaneous injection of 100 μ l
905 PBS containing 1×10^7 *S. pyogenes* into the shaved flank. Following infection, the
906 resulting ulcers were photographed over several days and the areas of the irregular
907 lesions were calculated using ImageJ software as described in detail elsewhere [82,
908 83]. Skin biopsies were obtained from euthanized mice and homogenized in 1 ml ice
909 cold PBS using a FastPrep-24 homogenizer (MP Biomedicals, LLC., Santa Ana, CA);
910 tissue was homogenized in 2 ml conical screw top vials with 3.2 mm stainless steel
911 beads (Fisher Scientific, Pittsburg, PA) with two FastPrep cycles of speed 6.0 for 45
912 sec, with a 5 min ice incubation between pulses to prevent overheating.

913 **Murine tissue cytokine analysis.** Samples of mouse skin and subcutaneous tissue
914 homogenates were harvested at six-days post-infection. Cytokine protein
915 concentrations were determined by a multiplex murine ELISA assay (EMD Millipore,
916 Billerica, MA, USA) according to the manufacturer's protocol.

917 **Murine tissue histologic analysis.** Murine skin biopsies were obtained at six-days
918 post-infection and were fixed in 4% formalin and dehydrated up to 70% ethanol prior to
919 paraffin embedding through the University of Michigan Pathology Core for Animal

920 Research (PCAR). H&E staining and immunohistochemistry services were performed
921 by the PCAR using commercially available anti-CD3 (T lymphocytes), and anti-F4/80
922 (macrophages) antibodies (Abcam, Cambridge, MA, USA). Digital images were
923 obtained with an EC3 digital imaging system (Leica Microsystems, Buffalo Grove, IL,
924 USA) using Leica Acquisition Software (Leica Microsystems). Adjustments to contrast in
925 digital images were applied equally to all experimental and control images.

926 **Human vaginal epithelial cell *in vitro* assays.** Adherence of *S. pyogenes* strains was
927 assessed to an established human vaginal epithelial cell line, VK2/E6E7, using methods
928 similar to those previously described [84-86]. The human vaginal epithelial cell line
929 VK2/E6E7 was purchased from the American Type Culture Collection (ATCC,
930 Manassas, Virginia), and cells were grown and maintained in media and conditions as
931 recommended by ATCC. Human cells were grown to confluence in 12-well tissue
932 culture-treated plates and washed with sterile PBS prior to inoculation with bacteria. *S.*
933 *pyogenes* strains were grown in THY broth to early stationary phase (OD_{600nm} 0.6),
934 washed twice in sterile PBS, and adjusted to give an inoculum of $\sim 5 \times 10^6$ CFU in 1 ml
935 per well, for a multiplicity of infection (MOI) of ~ 5 . Bacteria and human cells were
936 incubated at 37°C in 5% carbon dioxide for 60 min, after which time the supernatants
937 were removed and cells were washed four times with 2 mL sterile PBS to remove non-
938 adherent organisms. To recover *S. pyogenes* from the epithelial cells, each well was
939 treated with 0.2 mL 0.25% Trypsin-EDTA (Invitrogen) and incubated at 37°C for 5 min,
940 and then lysed by addition of 0.8 ml sterile water at pH 11. Lysis in water at pH 11 was
941 shown to result in a more complete eukaryotic cellular breakdown with maximal
942 recovery of bacteria from the surface in addition to intracellular reservoirs [87]. This

943 method recovers all cell-associated streptococci, predominantly extracellular adherent
944 cells with a relatively smaller amount of intracellular cells. The cell suspension was
945 serially diluted in PBS and plated onto THY agar for determination of viable CFU count.
946 The total cell-associated CFU percentage was calculated as (total CFU recovered from
947 the well/CFU of the original input inoculum) x 100%.

948 **Murine vaginal colonization model.** Experiments were performed as previously
949 described [55]. To synchronize estral cycles, sedated mice were estrogen
950 supplemented by intra-peritoneal injection with 0.5 mg β -estradiol 17-valerate (Sigma)
951 dissolved in 0.1 mL sterile sesame oil (Sigma) 2 days prior to streptococcal inoculation
952 and again on the day of inoculation (considered day #0). On day #0, sedated mice were
953 inoculated with $\sim 1 \times 10^6$ colony forming units (CFUs) instilled into the vaginal vault using
954 a P20 micropipetter (Gilson, Inc., Middleton, WI) in a total volume of 20 μ L PBS. At
955 successive intervals over a 1-month period post-inoculation, the vaginal vaults of
956 sedated mice were gently washed with 50 μ L sterile PBS and serial dilutions in sterile
957 PBS were plated onto THY agar plates supplemented with 1000 μ g/mL streptomycin to
958 determine viable CFUs. This concentration of streptomycin suppressed growth of
959 normal mouse vaginal flora but had no effect on the plating efficiency of the
960 streptomycin-resistant *S. pyogenes* strains. For colonization experiments, between 5 to
961 20 mice were tested per *S. pyogenes* strain, as indicated in the relevant figure legends.

962 **Human neutrophil bactericidal activity assay.** Human neutrophils were purchased
963 from a commercial supplier (Astarte Biologics, Bothell, WA, USA) and prepared
964 according to supplier recommendations to a concentration of 5×10^6 cells/ml in room
965 temperature Hank's Balanced Salt Solution (HBSS). Neutrophil bactericidal assay was

performed similar to that reported by Staali *et al.* [45]. Briefly, streptococcal strains were grown in fresh C-media to mid-log phase (OD_{600nm} of 0.6) and were washed twice in HBSS with calcium and magnesium, but without Phenol Red (Sigma, St. Louis, MO, USA). Streptococci were counted using a hemocytometer and adjusted to a concentration of 5×10^7 CFU/mL in room temperature HBSS. Neutrophils and streptococci were mixed in a 1:10 ratio of neutrophils to bacteria, and were incubated together for 10 minutes at 37°C. Next, extracellular streptococci were eliminated by addition of gentamicin (100 μ g/mL) and penicillin (5 μ g/mL) in HBSS for 20 minutes at 37°C. Next, cells were diluted in 1 mL of HBSS, centrifuged at 400g x 5 min, and washed with 1 mL fresh HBSS. The wash was repeated a second time in HBSS and the final cell pellet was resuspended in 50 μ L of 2% saponin in distilled water at pH 11 and allowed to remain at room temperature for 20 minutes to lyse neutrophils and release viable intracellular streptococci. The cells were diluted in distilled pH 11 water and aliquots plated onto fresh THY agar media for CFU counts. Three biological replicate experiments for each strain were performed.

Statistical analyses. Comparison of nonparametric data sets was performed using the Mann-Whitney U-test to determine significant differences. Differences between groups for recovery of CFU in vaginal washes were tested using a repeated measures analysis of variance. Differences in relative transcript levels were tested for significance with a two-tailed paired t-test. Differences in VK cell adherence and in neutrophil bactericidal survival assays were compared using a non-paired t-test. For all tests, the null hypothesis was rejected for $P < 0.05$. Computation utilized the resources available in GraphPad Prism™ (GraphPad Software, Inc., San Diego, CA).

References

1. Siegfried Z, Simon I. DNA methylation and gene expression. Wiley Interdiscip Rev Syst Biol Med. 2010;2(3):362-71. doi: 10.1002/wsbm.64. PubMed PMID: 20836034.
2. Gray SG, Eriksson T, Ekstrom TJ. Methylation, gene expression and the chromatin connection in cancer (review). Int J Mol Med. 1999;4(4):333-50. Epub 1999/09/24. PubMed PMID: 10493973.
3. Virani S, Colacino JA, Kim JH, Rozek LS. Cancer epigenetics: a brief review. ILAR J. 2012;53(3-4):359-69. Epub 2013/06/08. doi: 10.1093/ilar.53.3-4.359. PubMed PMID: 23744972; PubMed Central PMCID: PMC4021643.
4. Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9-16. doi: 10.1016/j.mib.2015.03.004. PubMed PMID: 25818841.
5. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The Epigenomic Landscape of Prokaryotes. PLoS Genet. 2016;12(2):e1005854. doi: 10.1371/journal.pgen.1005854. PubMed PMID: 26870957; PubMed Central PMCID: PMC4752239.
6. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev. 2013;77(1):53-72. Epub 2013/03/09. doi: 10.1128/MMBR.00044-12. PubMed PMID: 23471617; PubMed Central PMCID: PMC4021643.

- 1011 7. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing
1012 from single polymerase molecules. *Science*. 2009;323(5910):133-8. doi:
1013 10.1126/science.1162986. PubMed PMID: 19023044.
- 1014 8. McCarthy A. Third generation DNA sequencing: pacific biosciences' single
1015 molecule real time technology. *Chem Biol*. 2010;17(7):675-6. Epub 2010/07/28. doi:
1016 10.1016/j.chembiol.2010.07.004. PubMed PMID: 20659677.
- 1017 9. Kelleher P, Murphy J, Mahony J, van Sinderen D. Identification of DNA Base
1018 Modifications by Means of Pacific Biosciences RS Sequencing Technology. *Methods*
1019 *Mol Biol*. 2018;1681:127-37. Epub 2017/11/15. doi: 10.1007/978-1-4939-7343-9_10.
1020 PubMed PMID: 29134592.
- 1021 10. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al.
1022 Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat*
1023 *Methods*. 2010;7(6):461-5. Epub 2010/05/11. doi: 10.1038/nmeth.1459. PubMed PMID:
1024 20453866; PubMed Central PMCID: PMCPMC2879396.
- 1025 11. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, et al.
1026 Characterization of DNA methyltransferase specificities using single-molecule, real-time
1027 DNA sequencing. *Nucleic Acids Res*. 2012;40(4):e29. Epub 2011/12/14. doi:
1028 10.1093/nar/gkr1146. PubMed PMID: 22156058; PubMed Central PMCID:
1029 PMCPMC3287169.
- 1030 12. Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. Complete Genome
1031 Sequence of *emm28* Type *Streptococcus pyogenes* MEW123, a Streptomycin-
1032 Resistant Derivative of a Clinical Throat Isolate Suitable for Investigation of

1033 Pathogenesis. *Genome Announc.* 2016;4(2). doi: 10.1128/genomeA.00136-16. PubMed
 1034 PMID: 26988051; PubMed Central PMCID: PMC4796130.

1035 13. Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. Complete Genome
 1036 Sequence of *emm4 Streptococcus pyogenes* MEW427, a Throat Isolate from a Child
 1037 Meeting Clinical Criteria for Pediatric Autoimmune Neuropsychiatric Disorders
 1038 Associated with *Streptococcus* (PANDAS). *Genome Announc.* 2016;4(2). doi:
 1039 10.1128/genomeA.00127-16. PubMed PMID: 26988046; PubMed Central PMCID:
 1040 PMC4796125.

1041 14. Cunningham MW. Pathogenesis of group A streptococcal infections. *Clin*
 1042 *Microbiol Rev.* 2000;13(3):470-511. PubMed PMID: 10885988.

1043 15. Areschoug T, Carlsson F, Stalhammar-Carlemalm M, Lindahl G. Host-pathogen
 1044 interactions in *Streptococcus pyogenes* infections, with special reference to puerperal
 1045 fever and a comment on vaccine development. *Vaccine.* 2004;22 Suppl 1:S9-S14. doi:
 1046 10.1016/j.vaccine.2004.08.010. PubMed PMID: 15576204.

1047 16. Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, et al.
 1048 Disease manifestations and pathogenic mechanisms of Group A *Streptococcus*. *Clin*
 1049 *Microbiol Rev.* 2014;27(2):264-301. doi: 10.1128/CMR.00101-13. PubMed PMID:
 1050 24696436; PubMed Central PMCID: PMC3993104.

1051 17. Kreikemeyer B, McIver KS, Podbielski A. Virulence factor regulation and
 1052 regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host
 1053 interactions. *Trends Microbiol.* 2003;11(5):224-32. PubMed PMID: 12781526.

1054 18. Vega LA, Malke H, McIver KS. Virulence-Related Transcriptional Regulators of
 1055 *Streptococcus pyogenes*. In: Ferretti JJ, Stevens DL, Fischetti VA, editors.

1056 Streptococcus pyogenes : Basic Biology to Clinical Manifestations. Oklahoma City
 1057 (OK)2016.

1058 19. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A
 1059 nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases
 1060 and their genes. Nucleic Acids Res. 2003;31(7):1805-12. Epub 2003/03/26. PubMed
 1061 PMID: 12654995; PubMed Central PMCID: PMCPMC152790.

1062 20. Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, et al. Complete
 1063 genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc Natl Acad Sci U S
 1064 A. 2001;98(8):4658-63. Epub 2001/04/11. doi: 10.1073/pnas.071559398. PubMed
 1065 PMID: 11296296; PubMed Central PMCID: PMC31890.

1066 21. Okada R, Matsumoto M, Zhang Y, Isaka M, Tatsuno I, Hasegawa T. Emergence
 1067 of type I restriction modification system-negative *emm1* type *Streptococcus pyogenes*
 1068 clinical isolates in Japan. APMIS. 2014;122(10):914-21. Epub 2014/10/31. PubMed
 1069 PMID: 25356467.

1070 22. Euler CW, Ryan PA, Martin JM, Fischetti VA. M.SpyI, a DNA methyltransferase
 1071 encoded on a *mefA* chimeric element, modifies the genome of *Streptococcus pyogenes*.
 1072 J Bacteriol. 2007;189(3):1044-54. Epub 2006/11/07. doi: 10.1128/JB.01411-06.
 1073 PubMed PMID: 17085578; PubMed Central PMCID: PMCPMC1797290.

1074 23. Hanski E, Horwitz PA, Caparon MG. Expression of protein F, the fibronectin-
 1075 binding protein of *Streptococcus pyogenes* JRS4, in heterologous streptococcal and
 1076 enterococcal strains promotes their adherence to respiratory epithelial cells. Infect
 1077 Immun. 1992;60(12):5119-25. Epub 1992/12/01. PubMed PMID: 1452345; PubMed
 1078 Central PMCID: PMC258286.

- 1079 24. Cho KH, Caparon MG. Patterns of virulence gene expression differ between
1080 biofilm and tissue communities of *Streptococcus pyogenes*. Mol Microbiol.
1081 2005;57(6):1545-56. Epub 2005/09/02. doi: 10.1111/j.1365-2958.2005.04786.x.
1082 PubMed PMID: 16135223.
- 1083 25. Perez-Casal J, Price JA, Maguin E, Scott JR. An M protein with a single C repeat
1084 prevents phagocytosis of *Streptococcus pyogenes*: use of a temperature-sensitive
1085 shuttle vector to deliver homologous sequences to the chromosome of *S. pyogenes*.
1086 Mol Microbiol. 1993;8(5):809-19. Epub 1993/05/01. PubMed PMID: 8355608.
- 1087 26. Nielsen HV, Guiton PS, Kline KA, Port GC, Pinkner JS, Neiers F, et al. The metal
1088 ion-dependent adhesion site motif of the *Enterococcus faecalis* EbpA pilin mediates
1089 pilus function in catheter-associated urinary tract infection. MBio. 2012;3(4):e00177-12.
1090 doi: 10.1128/mBio.00177-12. PubMed PMID: 22829678; PubMed Central PMCID:
1091 PMC3419518.
- 1092 27. Ghosh J, Caparon MG. Specificity of *Streptococcus pyogenes* NAD(+)
1093 glycohydrolase in cytolysin-mediated translocation. Mol Microbiol. 2006;62(4):1203-14.
1094 Epub 2006/10/18. doi: 10.1111/j.1365-2958.2006.05430.x. PubMed PMID: 17042787.
- 1095 28. Lyon WR, Madden JC, Levin JC, Stein JL, Caparon MG. Mutation of luxS affects
1096 growth and virulence factor expression in *Streptococcus pyogenes*. Mol Microbiol.
1097 2001;42(1):145-57. Epub 2001/10/27. PubMed PMID: 11679074.
- 1098 29. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes and
1099 their relatives. Nucleic Acids Res. 2014;42(1):20-44. Epub 2013/09/27. doi:
1100 10.1093/nar/gkt847. PubMed PMID: 24068554; PubMed Central PMCID:
1101 PMCPMC3874165.

- 1102 30. Dryden DT, Cooper LP, Murray NE. Purification and characterization of the
1103 methyltransferase from the type 1 restriction and modification system of *Escherichia coli*
1104 K12. J Biol Chem. 1993;268(18):13228-36. Epub 1993/06/25. PubMed PMID: 8514761.
- 1105 31. Suri B, Shepherd JC, Bickle TA. The EcoA restriction and modification system of
1106 *Escherichia coli* 15T-: enzyme structure and DNA recognition sequence. EMBO J.
1107 1984;3(3):575-9. Epub 1984/03/01. PubMed PMID: 6325176; PubMed Central PMCID:
1108 PMCPMC557390.
- 1109 32. Makovets S, Titheradge AJ, Murray NE. ClpX and ClpP are essential for the
1110 efficient acquisition of genes specifying type IA and IB restriction systems. Mol
1111 Microbiol. 1998;28(1):25-35. Epub 1998/05/21. PubMed PMID: 9593294.
- 1112 33. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, et al.
1113 Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli*
1114 using single-molecule real-time sequencing. Nat Biotechnol. 2012;30(12):1232-9. Epub
1115 2012/11/10. doi: 10.1038/nbt.2432. PubMed PMID: 23138224; PubMed Central PMCID:
1116 PMCPMC3879109.
- 1117 34. Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, Parkhill J, et al.
1118 Phase variation of a Type IIG restriction-modification enzyme alters site-specific
1119 methylation patterns and gene expression in *Campylobacter jejuni* strain NCTC11168.
1120 Nucleic Acids Res. 2016;44(10):4581-94. Epub 2016/01/21. doi: 10.1093/nar/gkw019.
1121 PubMed PMID: 26786317; PubMed Central PMCID: PMCPMC4889913.
- 1122 35. Casselli T, Tourand Y, Scheidegger A, Arnold WK, Proulx A, Stevenson B, et al.
1123 DNA Methylation by Restriction Modification Systems Affects the Global Transcriptome

1124 Profile in *Borrelia burgdorferi*. J Bacteriol. 2018. Epub 2018/09/27. doi:
 1125 10.1128/JB.00395-18. PubMed PMID: 30249703.

1126 36. Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. Mismatch repair in
 1127 Gram-positive bacteria. Res Microbiol. 2016;167(1):4-12. Epub 2015/09/08. doi:
 1128 10.1016/j.resmic.2015.08.006. PubMed PMID: 26343983.

1129 37. Mclver KS, Scott JR. Role of mga in growth phase regulation of virulence genes
 1130 of the group A streptococcus. J Bacteriol. 1997;179(16):5178-87. PubMed PMID:
 1131 9260962; PubMed Central PMCID: PMCPMC179378.

1132 38. Hondorp ER, Mclver KS. The Mga virulence regulon: infection where the grass is
 1133 greener. Mol Microbiol. 2007;66(5):1056-65. PubMed PMID: 18001346.

1134 39. Vega LA, Malke H, Mclver KS. Virulence-Related Transcriptional Regulators of
 1135 *Streptococcus pyogenes*. In: Ferretti JJ, Stevens DL, Fischetti VA, editors.
 1136 *Streptococcus pyogenes: Basic Biology to Clinical Manifestations*. Oklahoma City
 1137 (OK)2016.

1138 40. Watson ME, Jr., Neely MN, Caparon MG. Animal Models of *Streptococcus*
 1139 *pyogenes* Infection. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. *Streptococcus*
 1140 *pyogenes: Basic Biology to Clinical Manifestations*. Oklahoma City (OK)2016.

1141 41. Bunce C, Wheeler L, Reed G, Musser J, Barg N. Murine model of cutaneous
 1142 infection with gram-positive cocci. Infect Immun. 1992;60(7):2636-40. Epub 1992/07/01.
 1143 PubMed PMID: 1612733; PubMed Central PMCID: PMC257214.

1144 42. Fielding CA, McLoughlin RM, McLeod L, Colmont CS, Najdovska M, Grail D, et
 1145 al. IL-6 regulates neutrophil trafficking during acute inflammation via STAT3. J Immunol.
 1146 2008;181(3):2189-95. Epub 2008/07/22. PubMed PMID: 18641358.

- 1147 43. Wright HL, Cross AL, Edwards SW, Moots RJ. Effects of IL-6 and IL-6 blockade
1148 on neutrophil function in vitro and in vivo. *Rheumatology*. 2014;53(7):1321-31. Epub
1149 2014/03/13. doi: 10.1093/rheumatology/keu035. PubMed PMID: 24609058.
- 1150 44. Cho JS, Pietras EM, Garcia NC, Ramos RI, Farzam DM, Monroe HR, et al. IL-17
1151 is essential for host defense against cutaneous *Staphylococcus aureus* infection in
1152 mice. *J Clin Invest*. 2010;120(5):1762-73. Epub 2010/04/07. doi: 10.1172/JCI40891.
1153 PubMed PMID: 20364087; PubMed Central PMCID: PMC2860944.
- 1154 45. Staali L, Morgelin M, Bjorck L, Tapper H. *Streptococcus pyogenes* expressing M
1155 and M-like surface proteins are phagocytosed but survive inside human neutrophils.
1156 *Cell Microbiol*. 2003;5(4):253-65. PubMed PMID: 12675683.
- 1157 46. Kihlberg BM, Cooney J, Caparon MG, Olsen A, Bjorck L. Biological properties of
1158 a *Streptococcus pyogenes* mutant generated by Tn916 insertion in *mga*. *Microb Pathog*.
1159 1995;19(5):299-315. PubMed PMID: 8778565.
- 1160 47. Caparon MG, Scott JR. Identification of a gene that regulates expression of M
1161 protein, the major virulence determinant of group A streptococci. *Proc Natl Acad Sci U S*
1162 *A*. 1987;84(23):8677-81. Epub 1987/12/01. PubMed PMID: 2446327; PubMed Central
1163 PMCID: PMC299609.
- 1164 48. Caswell CC, Lukomska E, Seo NS, Hook M, Lukomski S. Scl1-dependent
1165 internalization of group A *Streptococcus* via direct interactions with the alpha2beta(1)
1166 integrin enhances pathogen survival and re-emergence. *Mol Microbiol*.
1167 2007;64(5):1319-31. Epub 2007/06/05. doi: 10.1111/j.1365-2958.2007.05741.x.
1168 PubMed PMID: 17542923.

- 1169 49. Terao Y, Kawabata S, Kunitomo E, Murakami J, Nakagawa I, Hamada S. Fba, a
1170 novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry
1171 into epithelial cells, and the fba gene is positively transcribed under the Mga regulator.
1172 Mol Microbiol. 2001;42(1):75-86. Epub 2001/10/27. PubMed PMID: 11679068.
- 1173 50. Green NM, Beres SB, Graviss EA, Allison JE, McGeer AJ, Vuopio-Varkila J, et
1174 al. Genetic diversity among type emm28 group A streptococcus strains causing invasive
1175 infections and pharyngitis. J Clin Microbiol. 2005;43(8):4083-91. Epub 2005/08/06. doi:
1176 10.1128/JCM.43.8.4083-4091.2005. PubMed PMID: 16081955; PubMed Central
1177 PMCID: PMC1233891.
- 1178 51. Green NM, Zhang S, Porcella SF, Nagiec MJ, Barbian KD, Beres SB, et al.
1179 Genome sequence of a serotype M28 strain of group A streptococcus: potential new
1180 insights into puerperal sepsis and bacterial disease specificity. J Infect Dis.
1181 2005;192(5):760-70. Epub 2005/08/10. doi: 10.1086/430618. PubMed PMID: 16088825.
- 1182 52. Bruins MJ, Damoiseaux RA, Ruijs GJ. Association between group A beta-
1183 haemolytic streptococci and vulvovaginitis in adult women: a case-control study. Eur J
1184 Clin Microbiol Infect Dis. 2009;28(8):1019-21. PubMed PMID: 19343383.
- 1185 53. Anteby EY, Yagel S, Hanoch J, Shapiro M, Moses AE. Puerperal and intrapartum
1186 group A streptococcal infection. Infect Dis Obstet Gynecol. 1999;7(6):276-82. Epub
1187 1999/12/22. doi: 10.1155/S1064744999000514. PubMed PMID: 10598916; PubMed
1188 Central PMCID: PMC1784762.
- 1189 54. Stalhammar-Carlemalm M, Areschoug T, Larsson C, Lindahl G. The R28 protein
1190 of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins,

1191 confers protective immunity and promotes binding to human epithelial cells. Mol
 1192 Microbiol. 1999;33(1):208-19. PubMed PMID: 10411737.
 1193 55. Watson ME, Jr., Nielsen HV, Hultgren SJ, Caparon MG. Murine Vaginal
 1194 Colonization Model for Investigating Asymptomatic Mucosal Carriage of *Streptococcus*
 1195 *pyogenes*. Infect Immun. 2013;81(5):1606-17. Epub 2013/03/06. doi:
 1196 10.1128/IAI.00021-13. PubMed PMID: 23460515.
 1197 56. Balbontin R, Rowley G, Pucciarelli MG, Lopez-Garrido J, Wormstone Y, Lucchini
 1198 S, et al. DNA adenine methylation regulates virulence gene expression in *Salmonella*
 1199 *enterica* serovar Typhimurium. J Bacteriol. 2006;188(23):8160-8. Epub 2006/09/26. doi:
 1200 10.1128/JB.00847-06. PubMed PMID: 16997949; PubMed Central PMCID:
 1201 PMCPMC1698197.
 1202 57. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, et al. The
 1203 methylomes of six bacteria. Nucleic Acids Res. 2012;40(22):11450-62. Epub
 1204 2012/10/05. doi: 10.1093/nar/gks891. PubMed PMID: 23034806; PubMed Central
 1205 PMCID: PMCPMC3526280.
 1206 58. Estibariz I, Overmann A, Ailloud F, Krebs J, Josenhans C, Suerbaum S. The
 1207 core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in
 1208 orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res. 2019. Epub
 1209 2019/01/10. doi: 10.1093/nar/gky1307. PubMed PMID: 30624738.
 1210 59. Watson ME, Jr., Jarisch J, Smith AL. Inactivation of deoxyadenosine
 1211 methyltransferase (*dam*) attenuates *Haemophilus influenzae* virulence. Mol Microbiol.
 1212 2004;53(2):651-64. doi: 10.1111/j.1365-2958.2004.04140.x. PubMed PMID: 15228541.

- 1213 60. Hu C, Zhao Y, Sun H, Yang Y. Synergism of Dam, MutH, and MutS in
1214 methylation-directed mismatch repair in *Escherichia coli*. *Mutat Res*. 2017;795:31-3.
1215 Epub 2017/01/21. doi: 10.1016/j.mrfmmm.2016.12.002. PubMed PMID: 28107644;
1216 PubMed Central PMCID: PMCPMC5299848.
- 1217 61. Morgan RD, Bhatia TK, Lovasco L, Davis TB. Mmel: a minimal Type II restriction-
1218 modification system that only modifies one DNA strand for host protection. *Nucleic*
1219 *Acids Res*. 2008;36(20):6558-70. Epub 2008/10/22. doi: 10.1093/nar/gkn711. PubMed
1220 PMID: 18931376; PubMed Central PMCID: PMCPMC2582602.
- 1221 62. Willemse N, Schultsz C. Distribution of Type I Restriction-Modification Systems in
1222 *Streptococcus suis*: An Outlook. *Pathogens*. 2016;5(4). Epub 2016/11/22. doi:
1223 10.3390/pathogens5040062. PubMed PMID: 27869755; PubMed Central PMCID:
1224 PMCPMC5198162.
- 1225 63. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, et
1226 al. Methylome diversification through changes in DNA methyltransferase sequence
1227 specificity. *PLoS Genet*. 2014;10(4):e1004272. Epub 2014/04/12. doi:
1228 10.1371/journal.pgen.1004272. PubMed PMID: 24722038; PubMed Central PMCID:
1229 PMCPMC3983042.
- 1230 64. Sitaraman R, Dybvig K. The *hsd* loci of *Mycoplasma pulmonis*: organization,
1231 rearrangements and expression of genes. *Mol Microbiol*. 1997;26(1):109-20. Epub
1232 1998/01/31. PubMed PMID: 9383194.
- 1233 65. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, et al. Epigenetic Switch Driven by DNA
1234 Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. *PLoS Pathog*.

1235 2016;12(7):e1005762. Epub 2016/07/20. doi: 10.1371/journal.ppat.1005762. PubMed
 1236 PMID: 27427949; PubMed Central PMCID: PMCPMC4948785.

1237 66. McIver KS, Myles RL. Two DNA-binding domains of Mga are required for
 1238 virulence gene activation in the group A streptococcus. *Mol Microbiol.* 2002;43(6):1591-
 1239 601. Epub 2002/04/16. PubMed PMID: 11952907.

1240 67. Almengor AC, McIver KS. Transcriptional activation of *sclA* by Mga requires a
 1241 distal binding site in *Streptococcus pyogenes*. *J Bacteriol.* 2004;186(23):7847-57. Epub
 1242 2004/11/18. doi: 10.1128/JB.186.23.7847-7857.2004. PubMed PMID: 15547255;
 1243 PubMed Central PMCID: PMCPMC529090.

1244 68. Wexler DE, Cleary PP. Purification and characteristics of the streptococcal
 1245 chemotactic factor inactivator. *Infect Immun.* 1985;50(3):757-64. Epub 1985/12/01.
 1246 PubMed PMID: 3905613; PubMed Central PMCID: PMCPMC261145.

1247 69. Ji Y, McLandsborough L, Kondagunta A, Cleary PP. C5a peptidase alters
 1248 clearance and trafficking of group A streptococci by infected mice. *Infect Immun.*
 1249 1996;64(2):503-10. Epub 1996/02/01. PubMed PMID: 8550199; PubMed Central
 1250 PMCID: PMC173793.

1251 70. Li J, Zhu H, Feng W, Liu M, Song Y, Zhang X, et al. Regulation of inhibition of
 1252 neutrophil infiltration by the two-component regulatory system CovRS in subcutaneous
 1253 murine infection with group A streptococcus. *Infect Immun.* 2013;81(3):974-83. Epub
 1254 2013/01/16. doi: 10.1128/IAI.01218-12. PubMed PMID: 23319556; PubMed Central
 1255 PMCID: PMCPMC3584857.

1256 71. Lynskey NN, Reglinski M, Calay D, Siggins MK, Mason JC, Botto M, et al. Multi-
 1257 functional mechanisms of immune evasion by the streptococcal complement inhibitor

1258 C5a peptidase. PLoS Pathog. 2017;13(8):e1006493. Epub 2017/08/15. doi:
 1259 10.1371/journal.ppat.1006493. PubMed PMID: 28806402; PubMed Central PMCID:
 1260 PMCPMC5555575.

1261 72. National Research Council (U.S.). Committee for the Update of the Guide for the
 1262 Care and Use of Laboratory Animals. Guide for the care and use of laboratory animals.
 1263 Washington, D.C.: National Academies Press; 2011. Available from:
 1264 <http://lib.myilibrary.com?id=297592>.

1265 73. Lyon WR, Gibson CM, Caparon MG. A role for trigger factor and an *rgg*-like
 1266 regulator in the transcription, secretion and processing of the cysteine proteinase of
 1267 *Streptococcus pyogenes*. Embo J. 1998;17(21):6263-75. Epub 1998/11/03. doi:
 1268 10.1093/emboj/17.21.6263. PubMed PMID: 9799235; PubMed Central PMCID:
 1269 PMC1170952.

1270 74. Kietzman CC, Caparon MG. CcpA and LacD.1 affect temporal regulation of
 1271 *Streptococcus pyogenes* virulence genes. Infect Immun. 2010;78(1):241-52. Epub
 1272 2009/10/21. doi: 10.1128/IAI.00746-09. PubMed PMID: 19841076; PubMed Central
 1273 PMCID: PMC2798178.

1274 75. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable
 1275 way to create recombinant plasmids. Biotechniques. 2010;48(6):463-5. Epub
 1276 2010/06/24. doi: 10.2144/000113418. PubMed PMID: 20569222; PubMed Central
 1277 PMCID: PMC3121328.

1278 76. Caparon MG, Stephens DS, Olsen A, Scott JR. Role of M protein in adherence of
 1279 group A streptococci. Infect Immun. 1991;59(5):1811-7. Epub 1991/05/01. PubMed
 1280 PMID: 2019444; PubMed Central PMCID: PMC257920.

1281 77. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene
1282 expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207-
1283 10. Epub 2001/12/26. PubMed PMID: 11752295; PubMed Central PMCID:
1284 PMCPMC99122.

1285 78. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
1286 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic*
1287 *Acids Res.* 2015;43(7):e47. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792;
1288 PubMed Central PMCID: PMC4402510.

1289 79. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for
1290 differential expression analysis of digital gene expression data. *Bioinformatics.*
1291 2010;26(1):139-40. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308;
1292 PubMed Central PMCID: PMC2796818.

1293 80. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization.
1294 *BMC bioinformatics.* 2010;11:367. doi: 10.1186/1471-2105-11-367. PubMed PMID:
1295 20598126; PubMed Central PMCID: PMC2912887.

1296 81. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T)
1297 method. *Nat Protoc.* 2008;3(6):1101-8. Epub 2008/06/13. PubMed PMID: 18546601.

1298 82. Brenot A, King KY, Janowiak B, Griffith O, Caparon MG. Contribution of
1299 glutathione peroxidase to the virulence of *Streptococcus pyogenes*. *Infect Immun.*
1300 2004;72(1):408-13. Epub 2003/12/23. PubMed PMID: 14688122; PubMed Central
1301 PMCID: PMC344014.

1302 83. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of
1303 image analysis. *Nat Methods.* 2012;9(7):671-5. PubMed PMID: 22930834.

1304 84. Sheen TR, Jimenez A, Wang NY, Banerjee A, van Sorge NM, Doran KS. Serine-
 1305 rich repeat proteins and pili promote *Streptococcus agalactiae* colonization of the
 1306 vaginal tract. J Bacteriol. 2011;193(24):6834-42. Epub 2011/10/11. doi:
 1307 10.1128/JB.00094-11. PubMed PMID: 21984789; PubMed Central PMCID:
 1308 PMC3232834.

1309 85. Fichorova RN, Rheinwald JG, Anderson DJ. Generation of papillomavirus-
 1310 immortalized cell lines from normal human ectocervical, endocervical, and vaginal
 1311 epithelium that maintain expression of tissue-specific differentiation proteins. Biol
 1312 Reprod. 1997;57(4):847-55. Epub 1997/10/07. PubMed PMID: 9314589.

1313 86. Okada N, Tatsuno I, Hanski E, Caparon M, Sasakawa C. *Streptococcus*
 1314 *pyogenes* protein F promotes invasion of HeLa cells. Microbiology. 1998;144 (Pt
 1315 11):3079-86. Epub 1998/12/10. PubMed PMID: 9846743.

1316 87. Decleva E, Menegazzi R, Busetto S, Patriarca P, Dri P. Common methodology is
 1317 inadequate for studies on the microbicidal activity of neutrophils. J Leukoc Biol.
 1318 2006;79(1):87-94. Epub 2005/10/26. doi: 10.1189/jlb.0605338. PubMed PMID:
 1319 16244110.

1320
 1321
 1322

Figure captions

Fig 1. M.SpyMEW123I dependent m6A modifications in the *S. pyogenes* genome.

A) Genomic organization of the SpyMEW123I Type I RM system gene cluster, *hsdRSM*, in strain MEW123 (WT) and the MEW513 in-frame deletion mutant (Δ RSM). B) Detection of genomic m6A base modifications (red line) in the MEW123 genome or the Δ RSM genome via PacBio SMRT sequencing. Modification quality values (modQVs) are indicated on the x-axis and the number of bases is indicated on the y-axis. ModQVs indicate if the polymerase kinetics at a position differs from the expected background, where a modQV of 30 corresponds to a p-value of 0.001. C) Dot blot with α -m6A antibody on genomic DNA isolated from the following strains: MEW123 (WT), restriction endonuclease *hsdR* antibiotic cassette-disruption mutant MEW489 (Ω RE), in-frame deletion of the *hsdRSM* gene cluster (Δ RSM), the Δ RSM strain complemented with plasmid-encoded *hsdRSM* (Δ RSM/pRSM), and unmodified DNA oligonucleotides serving as a negative control (Neg cont.) (500 ng DNA per spot).

Fig 2. The Type I RM system, SpyMEW123I, is functional for endonuclease activity and influences transformation efficiency.

A) Transformation efficiencies of WT (MEW123), the *hsdRSM* in-frame deletion (Δ RSM, MEW513), the *hsdR* single-gene disrupted mutant (Ω RE, MEW489), and the *scpA* insertional inactivation mutant (Ω *scpA*, MEW380). A Mann-Whitney U test was used to identify significant differences between groups for transformation efficiency; ** $P < 0.01$, *** $P < 0.001$, n = 3-7 replicates per point. B) Each SpyMEW123I RM site is represented on the plot where the position that the site occurs in the genome (from 0

Mb -1.8 Mb) is represented on the x-axis. The fraction of reads called as methylated from PacBio SMRT sequencing for each RM site is represented on the y-axis. RM sites that occur on the + and – strand of the DNA duplex are represented by green triangles and red dots, respectively.

Fig 3. Gene transcripts involved in immune evasion and adherence are down regulated in the Δ RSM strain compared to WT.

A) Heatmap of differentially expressed genes between WT and Δ RSM strains. Replicates in triplicate are represented on the x-axis and published/putative gene functions are on the y-axis with gene references to the MEW123 genome. Relative expression is compared row-wise with more highly expressed replicates in red. B) Verification of RNA-Seq data using qRT-PCR with individual primer sets shown on the x-axis for the Δ RSM mutant (blue), complemented mutant strain Δ RSM/pRSM (green), *hsdR* single gene mutant Ω RE (orange), and the Δ *mga* mutant (purple). The y-axis indicates Relative Transcript Levels for individual transcripts compared to *recA* reference transcript. Each gene transcript was analyzed in triplicate. Shown is fold change compared to WT expression, with genes showing greater than two-fold change as significant. Genes extending lower than the x-axis were down-regulated several hundred-fold.

Fig 4. Deletion of the MEW123 RSM gene cluster is associated with larger skin lesion formation in a murine subcutaneous infection model.

C57BL/6J mice were inoculated subcutaneously at the shaved flank with 1×10^7 CFUs MEW123 (WT, red squares) or the Δ RSM mutant (blue squares) on Day 0. Lesions

were photographed daily and lesion area was calculated using ImageJ software. A) Shown are individual mouse lesion area sizes with mean values (black bars) over Days Post-Infection (DPI). A Mann-Whitney U test was used to identify significant differences in lesion size between strains at each time point; ns, not significant, * $P < 0.05$, *** $P < 0.001$. B) Representative images of mice skin lesions over time at days 2-, 4-, and 6-days post-infection. Shown are skin lesions following infection with MEW123 (WT), MEW513 (Δ RSM), and the complemented strain MEW552 (Δ RSM/pRSM). Black bars are 1 cm for reference. Black dashed lines highlight the area of tissue injury. C) Shown are individual mouse lesion area sizes from a representative experiment with mean values (black bars) over Days Post-Infection. Shown are lesions from mice infected with MEW123 (WT), MEW513 (Δ RSM), the complemented strain MEW552 (Δ RSM/pRSM), the in-frame deletion of *Mga*, MEW480 (Δ *mga*), the cassette-insertion disruption mutant of *emm28*, MEW409 (Ω *emm28*), and the cassette-insertion disruption mutant of *scpA*, MEW380 (Ω *scpA*). A Mann-Whitney U test was used to identify significant differences in lesion size between strains at each time point; * $P < 0.05$.

Fig 5. Deletion of the MEW123 RSM gene cluster increases the host inflammatory cytokine response in the murine subcutaneous infection model.

C57BL/6J mice were inoculated subcutaneously at the shaved flank and resulting lesions were dissected on day 6 post-infection for cytokine analysis (panel A) and histology (panel B). A) Shown are ELISA results of homogenized murine skin biopsy specimens from mice previously inoculated with sterile phosphate buffered saline (PBS, black bars), 1×10^7 CFUs MEW123 (WT, red bars), or 1×10^7 CFUs of the Δ RSM mutant (Δ RSM, blue bars). Results are pooled from biopsies of 3 mice per group with mean and

SEM cytokine concentrations. A Mann-Whitney U test was used to identify significant differences in cytokine concentration; ns, not significant, * $P < 0.05$, ** $P < 0.01$. B) Representative images at 10X magnification of skin biopsies from mice inoculated with MEW123 (WT, upper row), or the Δ RSM mutant (Δ RSM, bottom row). Slides were stained with hematoxylin and eosin (H&E) for general neutrophil and overall inflammatory response, or specifically by immunohistochemistry for T cells (α -CD3) or macrophages (α -F4/80). Focal areas of intense inflammation were outlined with dashed lines for comparison.

Fig 6. Deletion of MEW123 RSM gene cluster impairs resistance to human neutrophil bactericidal activity.

Streptococcal resistance to human neutrophil bactericidal activity was examined using an *in vitro* assay. Human neutrophils were incubated with *S. pyogenes* strains in a 1:10 ratio for 10 min. Extracellular streptococci were eliminated with gentamicin and penicillin for 20 min. Neutrophil lysis and release of viable intracellular streptococci was performed with 2% saponin in pH 11 water for 20 min. Surviving CFUs were plated onto THY agar. Shown are streptococcal strains HSC5 (M14) as a positive control, the HSC5 Ω emm (M14 Ω emm) as a negative control, MEW123 (M28), MEW123 Ω emm (M28 Ω emm), and the MEW123 RSM deletion mutant (M28 Δ RSM). Shown are mean $\pm$ SEM CFU counts performed in triplicate from a representative experiment. An non-paired t-test was used for statistical significance, comparing mutant to parent strain; * $P < 0.05$, *** $P < 0.001$.

Fig 7. Deletion of MEW123 RSM gene cluster impairs streptococcal adherence to human vaginal epithelial cells, but does not impact carriage duration in a murine vaginal colonization model.

A) Confluent wells of VK E6E7 human vaginal epithelial cells were inoculated with 5×10^6 CFUs of *S. pyogenes* MEW123 (WT, red), the *emm28* gene-disrupted mutant ($\Delta emm28$, gray), or the RSM deletion mutant (ΔRSM , blue) (multiplicity of infection 5-10:1). Following 1h incubation, non-adherent streptococci were washed away with sterile PBS followed by lysis of the epithelial cells with pH 11 water, serial dilution, and determination of viable streptococci remaining. The % Cell Associated (mostly adherent plus few intracellular) is the percentage of the inoculum CFUs remaining detectable at 1h. Significant differences between groups were calculated by non-paired t-test (** $P < 0.01$, $n = 6$ replicates per point). B) Estrogenized female C57BL/6J mice were intravaginally inoculated with 1×10^6 CFU of *S. pyogenes* MEW123 (WT, red), or the RSM deletion mutant (ΔRSM , blue). Mice were cultured by intravaginal washes with sterile PBS and plated onto selective media (THY supplemented with streptomycin 1000 $\mu\text{g/ml}$) for quantification. Shown are mean $\pm$ SEM Log_{10} CFU counts over time post-infection of 10 mice per group pooled from two separate experiments. A repeated-measures ANOVA was used for statistical significance; at no time point tested over the 28-day experiment were the two groups statistically significantly different.

Supporting information

S1 Fig. Growth curves of WT and Δ RSM mutant are similar in THY broth and C-media, and in THY-broth containing penicillin, gentamicin, or erythromycin.

Growth was monitored using a Synergy HTX plate reader (BioTek) in 96 well plates (Greiner Bio-One). Briefly, 4 μ l of overnight culture grown in THY broth was inoculated into 200 μ l of the described fresh media, with identical strains and conditions measured in at least triplicate. Growth was at 37°C, room air, in static conditions for 12 hours (time on X-axis) and OD_{620nm} was measured every 3 seconds (Y-axis). Where indicated, antibiotic was added to THY broth at concentrations ranging from 0.05 μ g/mL to 100 μ g/mL, with concentrations shown in the color-coded key on the right [μ g/mL].

S2 Fig. Streptococcal CFU counts in skin lesions at day#4 post-infection. Skin lesions from infected mice were dissected on day#4 post-infection and homogenized in 1 ml sterile PBS. The homogenate was serially diluted in sterile PBS and plated onto THY agar plates containing streptomycin (1000 μ g/mL) to select for *S. pyogenes* strain MEW123 and its mutants. Each point represents the CFU counts from one mouse lesion, with black bars indicating mean CFU values for that group. Groups are MEW123 (WT), MEW513 (Δ RSM), MEW552 (Δ RSM/pRSM), and MEW480 (Δ *mga*). The Mann-Whitney U-test was used to test for statistical significance, though none were statistically different from the WT strain.

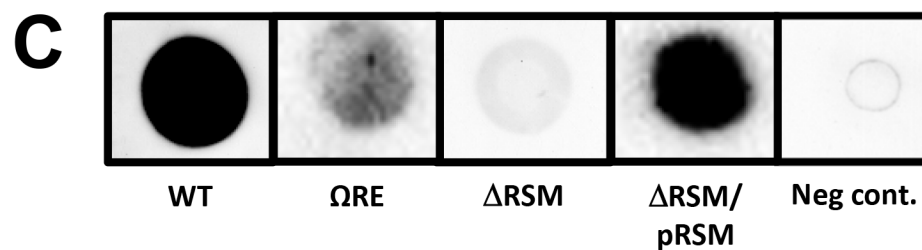
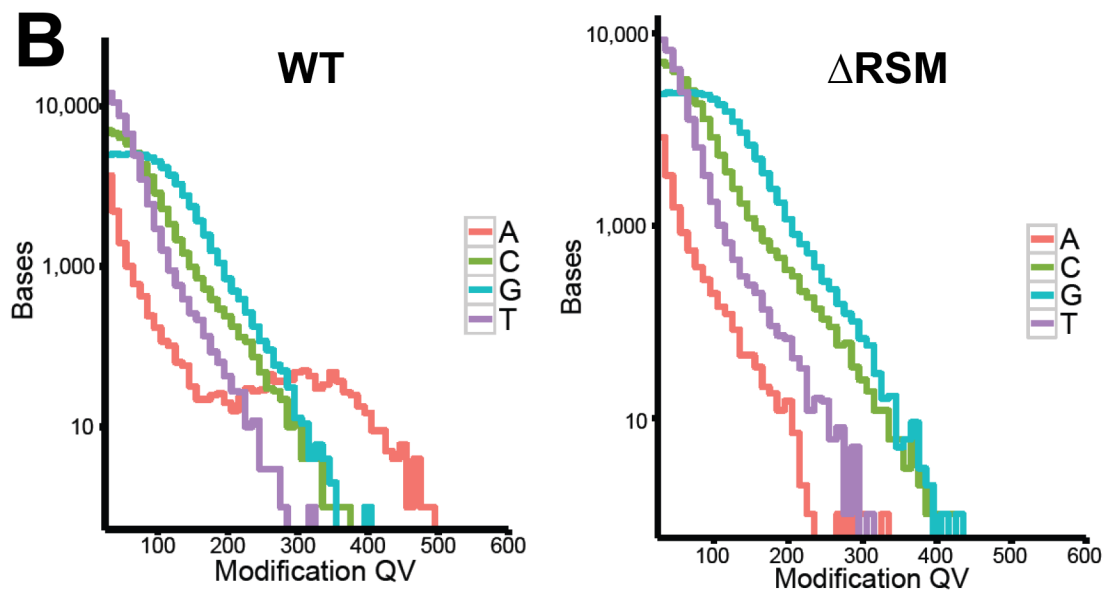
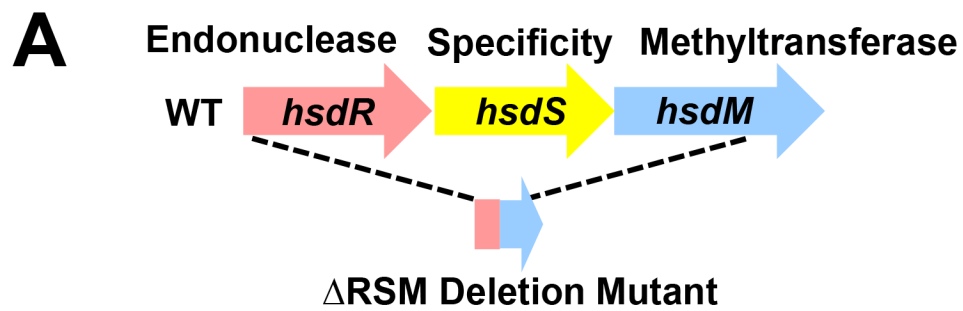


Figure 1.

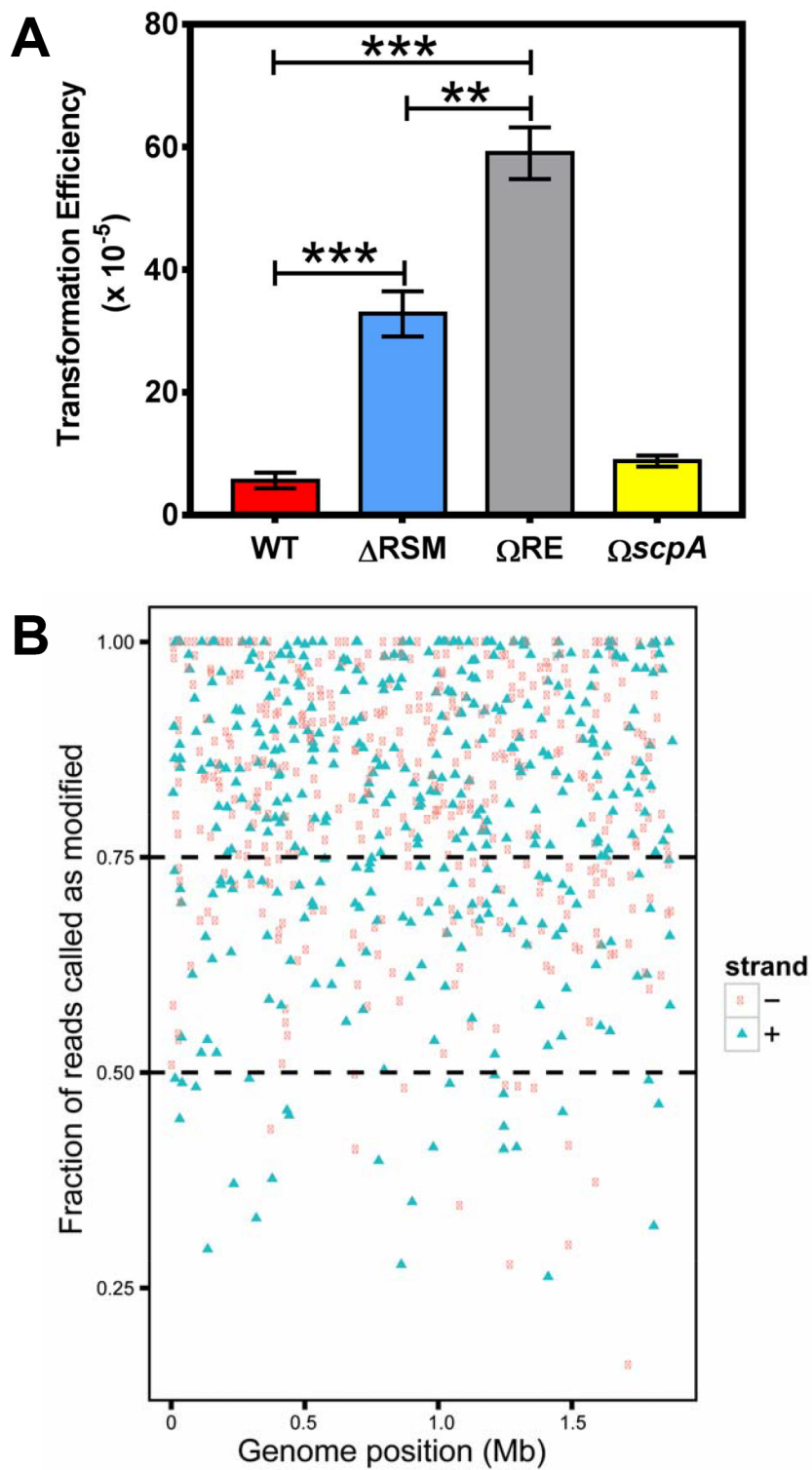


Figure 2.

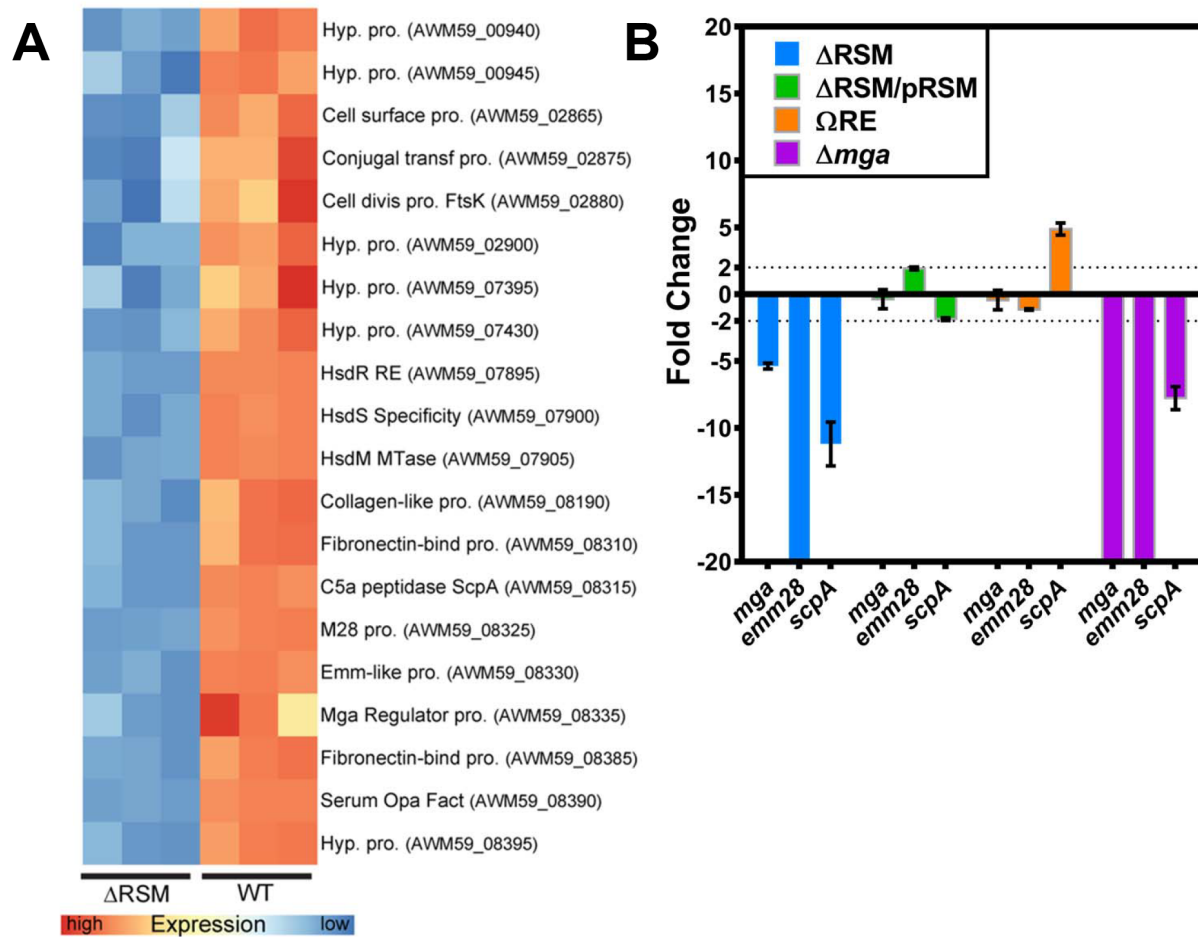


Figure 3.

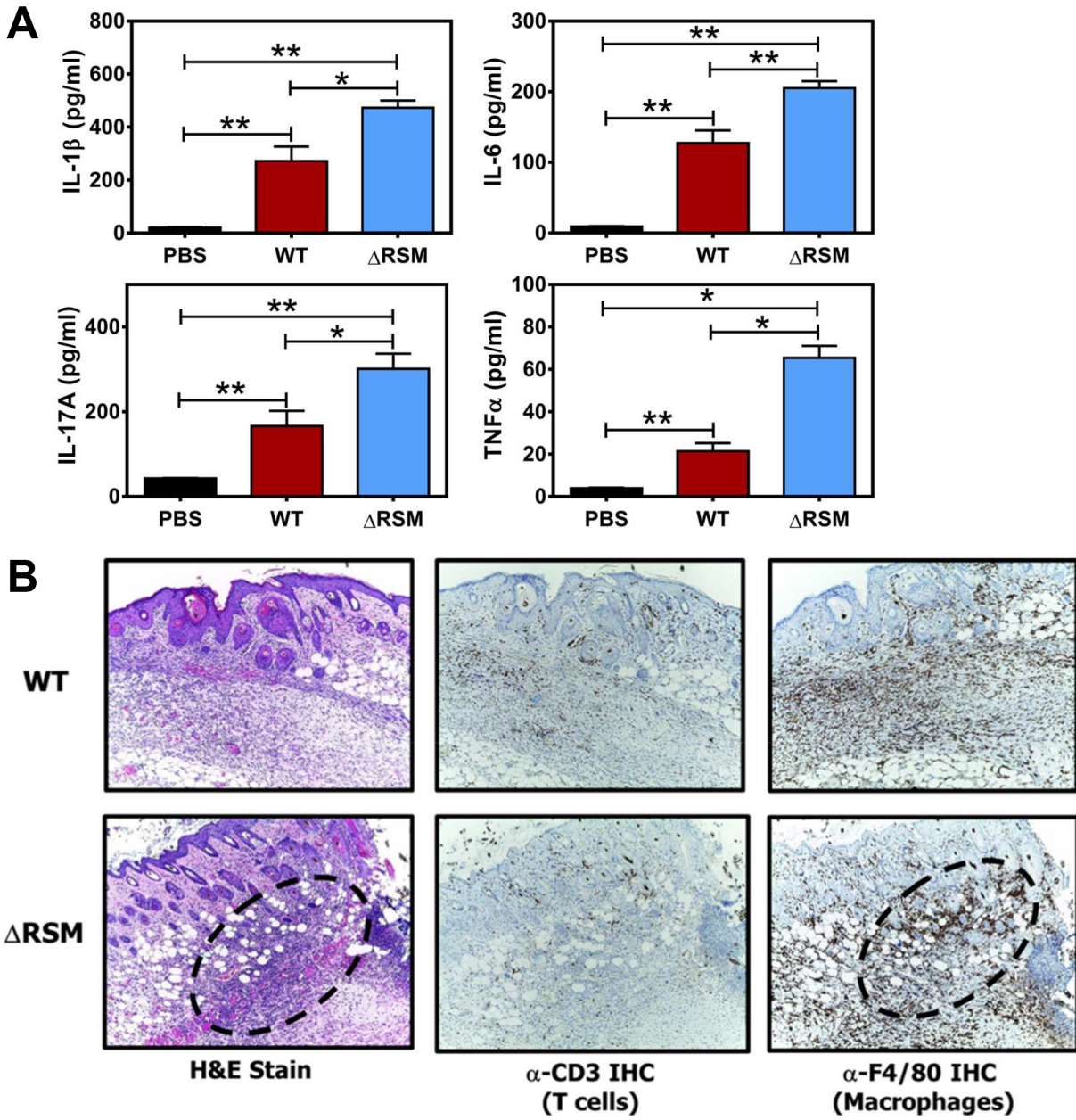
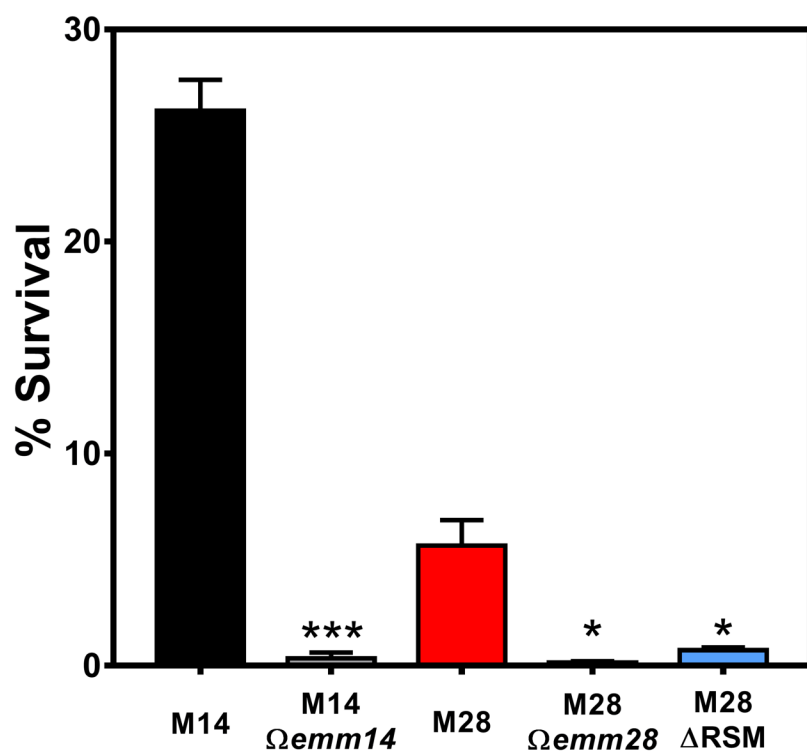


Figure 5.



1466

1467

1468

1469 Figure 6.

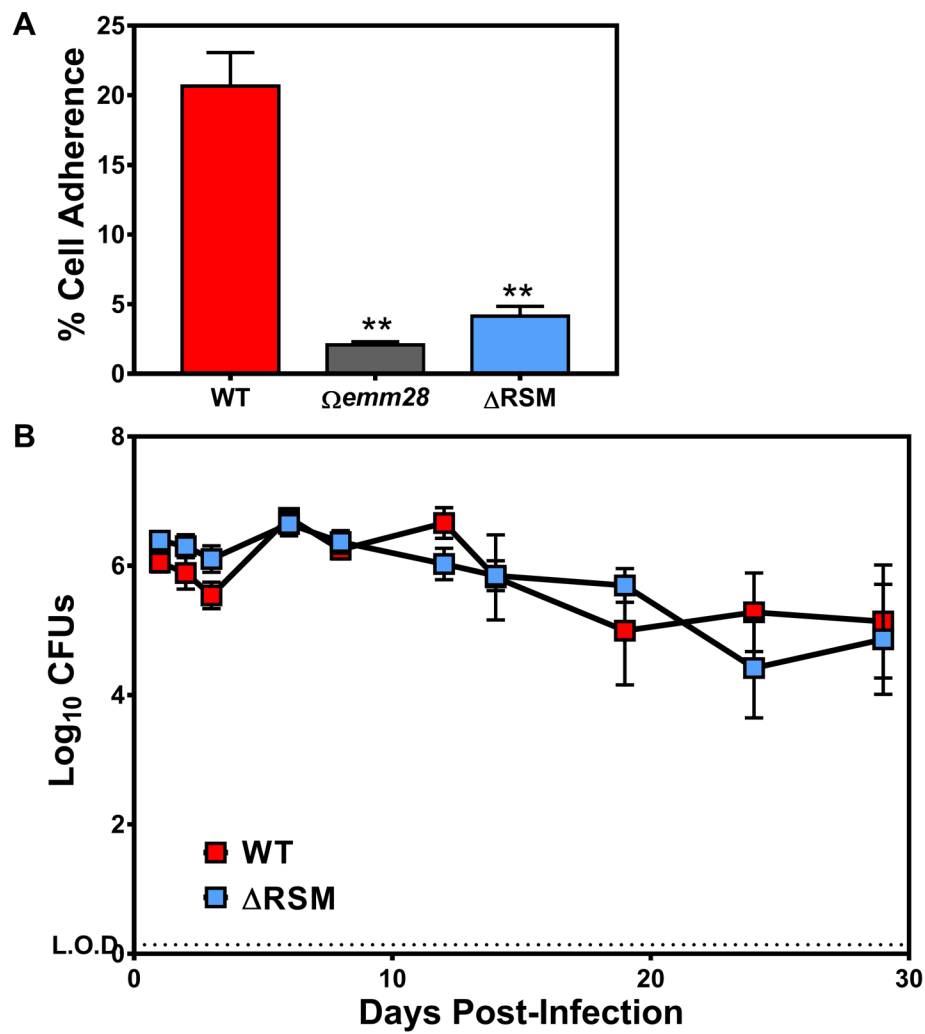


Figure 7.

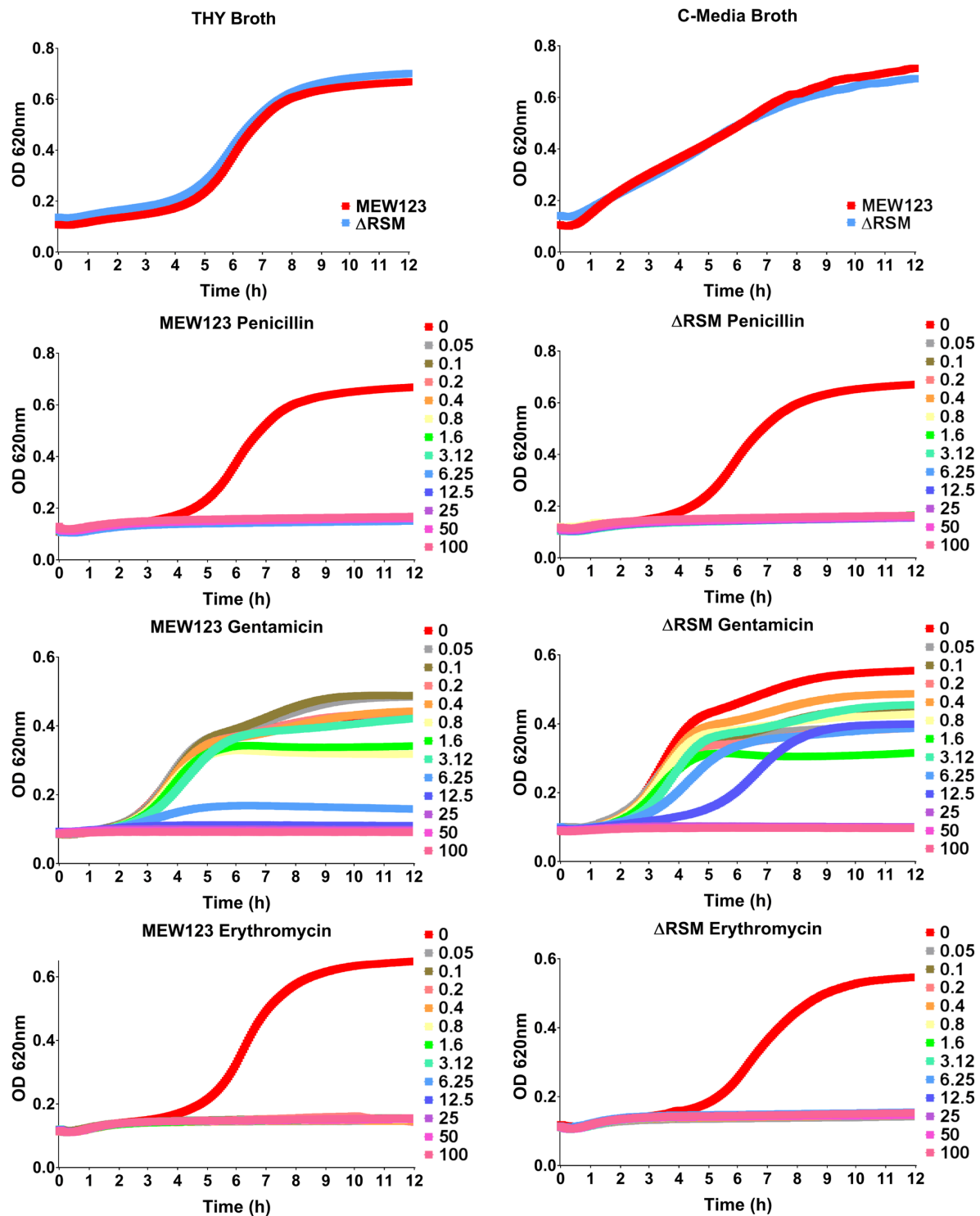
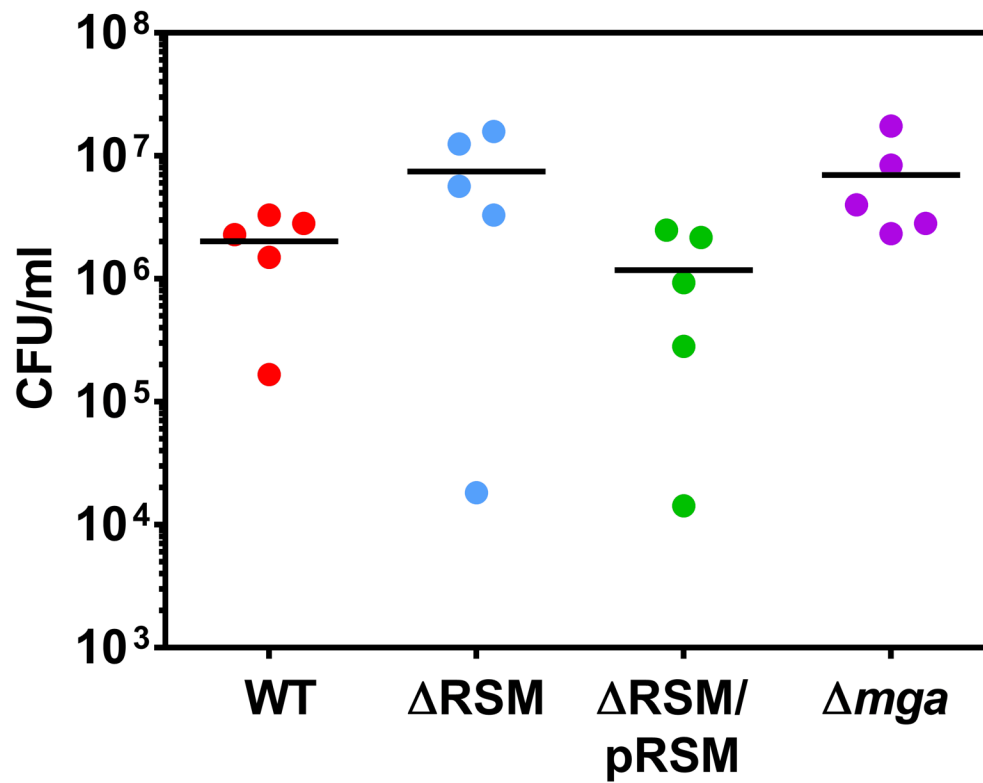


Figure S1.

Skin Lesion CFUs Day#4 Post-Infection



1477

1478 Figure S2.