BMC Genomics

RESEARCH ARTICLE

Open Access

Differential expression of small RNAs under chemical stress and fed-batch fermentation in *E. coli*

Martin Holm Rau, Klara Bojanovič, Alex Toftgaard Nielsen and Katherine S. Long[*]

Abstract

Background: Bacterial small RNAs (sRNAs) are recognized as posttranscriptional regulators involved in the control of bacterial lifestyle and adaptation to stressful conditions. Although chemical stress due to the toxicity of precursor and product compounds is frequently encountered in microbial bioprocessing applications, the involvement of sRNAs in this process is not well understood. We have used RNA sequencing to map sRNA expression in *E. coli* under chemical stress and high cell density fermentation conditions with the aim of identifying sRNAs involved in the transcriptional response and those with potential roles in stress tolerance.

Results: RNA sequencing libraries were prepared from RNA isolated from *E. coli* K-12 MG1655 cells grown under high cell density fermentation conditions or subjected to chemical stress with twelve compounds including four organic solvent-like compounds, four organic acids, two amino acids, geraniol and decanoic acid. We have discovered 253 novel intergenic transcripts with this approach, adding to the roughly 200 intergenic sRNAs previously reported in *E. coli*. There are eighty-four differentially expressed sRNAs during fermentation, of which the majority are novel, supporting possible regulatory roles for these transcripts in adaptation during different fermentation stages. There are a total of 139 differentially expressed sRNAs under chemical stress conditions, where twenty-nine exhibit significant expression changes in multiple tested conditions, suggesting that they may be involved in a more general chemical stress response. Among those with known functions are sRNAs involved in regulation of outer membrane proteins, iron availability, maintaining envelope homeostasis, as well as sRNAs incorporated into complex networks controlling motility and biofilm formation.

Conclusions: This study has used deep sequencing to reveal a wealth of hitherto undescribed sRNAs in *E. coli* and provides an atlas of sRNA expression during seventeen different growth and stress conditions. Although the number of novel sRNAs with regulatory functions is unknown, several exhibit specific expression patterns during high cell density fermentation and are differentially expressed in the presence of multiple chemicals, suggesting they may play regulatory roles during these stress conditions. These novel sRNAs, together with specific known sRNAs, are candidates for improving stress tolerance and our understanding of the *E. coli* regulatory network during fed-batch fermentation.

Keywords: sRNA, RNA-seq, Fermentation, Chemical stress, MicF, RybB, OmrB, CyaR, RyhB

* Correspondence: kalon@biosustain.dtu.dk
Novo Nordisk Foundation Center for Biosustainability, Technical University of
Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark

Rau *et al. BMC Genomics* (2015) 16:1051

Page 2 of 16

Background

Bacteria encode hundreds of small regulatory RNA molecules with typical sizes from 50 to 300 nucleotides that regulate gene expression [1, 2]. Although some sRNAs function by binding to protein targets and sequestering their activities, the majority act via base pairing interactions with target mRNA molecules, thereby affecting their translation and/or stability. The base pairing sRNAs can be divided into *cis*- and *trans*-encoded sRNAs that are encoded just opposite or at a different chromosomal location relative to their targets, respectively. The latter group of intergenic sRNAs is characterized by short and imperfect target base-pairing interactions, multiple mRNA targets, and in some bacteria, the requirement of the RNA chaperone Hfq.

The bacterium *Escherichia coli* has been the model system for the study of sRNAs and it is therefore the organism with the most comprehensive information on sRNA function available [3]. Early approaches for sRNA identification were largely based on high abundance, sequence conservation, and protein co-purification, in particular with Hfq [4]. Systematic screens for sRNAs have focused mostly on intergenic regions and utilized computational methods [5–10], shotgun cloning strategies [11, 12] and high-density oligonucleotide probe arrays [13–15]. In recent years the application of RNA-sequencing (RNA-seq) has led to the identification of hundreds of novel transcripts in diverse bacteria. In *E. coli*, two studies have used RNA-seq approaches to identify novel sRNAs and detect sRNAs predicted previously with computational methods [16, 17].

Bacterial sRNAs are often expressed in response to changing environmental conditions and function to modulate gene expression. There are a plethora of documented connections between sRNAs and stress, where sRNAs regulate important processes in response to metabolite/nutrient, envelope/outer membrane, oxidative, iron deficiency, anaerobic and pH stress [18, 19]. Chemical stress is routinely encountered during microbial bioprocessing applications due to product toxicity [20–24] because high titers above 100 g per liter are usually required for economically viable production. Despite this, the cellular response and involvement of sRNAs in this process are poorly understood. Biobased production of chemicals is usually performed by high cell density fed-batch fermentation and complete knowledge of regulation of cellular metabolism is critical for achieving improved production. This includes sRNA-derived regulation and its effect on metabolism during the different fermentation phases, such as the fed-batch phase and the transition between exponential and stationary phases. We have used an RNA-seq approach to map the *E. coli* sRNome during chemical stress and high cell density fermentation with the aim of gaining insights into the chemical stress response and identifying sRNAs with roles in stress tolerance that have potential applications in the design and optimization of future production strains. As significant differences in growth physiology are observed between batch and fed-batch fermentation conditions, we have also studied the expression of sRNAs during these two conditions. Moreover, we have investigated the expression of 462 small RNAs, comprised of previously annotated and 253 novel transcripts, and show that a significant fraction of them are differentially expressed under chemical stress and during high cell-density fermentations.

Results

Experimental approach

An initial list of future building block and precursor biochemicals was compiled with inspiration from a study that aimed to identify the top value added chemicals that can be produced from biomass [25]. This analysis considered the necessary transformations to convert sugars into the building block chemicals and the further conversion of these into secondary chemicals and derivatives, as well as economic parameters including the known and potential market data for the compounds. Several other targets of commercial interest as well as some inhibitors commonly found in biomass hydrolysate were also included. A series of growth inhibition tests were performed on the *E. coli* K-12 MG1655 strain with a range of compound concentrations. The initial list was reduced, where compounds with low inhibition of bacterial growth within solubility limits, similar chemical structures and low commercial potential were excluded.

The final list of twelve compounds (Table 1) includes organic acids (acetate, succinic acid, itaconic acid, and levulinic acid), amino acids (serine and threonine), organic

Table 1 The chemical compounds and growth-inhibiting concentrations used in this study

Compound	Concentration	
	g/L or % (v/v)	mM
Acetate (sodium)	7.5 g/L	55
1,4-Butanediol	4 %	452
Butanol	0.75 %	82
γ-Hydroxy-butyrolactone	1.10 %	144
Decanoic acid	0.13 %	7.3
Furfural	0.10 %	12
Geraniol	0.03 %	1.7
Itaconic acid	29 g/L	223
Levulinic acid	0.70 %	69
Serine	3 g/L	29
Succinic acid	32 g/L	271
Threonine	7 g/L	59

Rau *et al. BMC Genomics* (2015) 16:1051

Page 3 of 16

solvent-like compounds (butanol, 3-hydroxybutyrolactone, 1,4-butanediol, and furfural), the isoprenoid precursor geraniol and the fatty acid decanoic acid. In order to investigate the response to growth inhibiting concentrations of the chemicals and to detect compound-specific responses in *E. coli* K-12 MG1655, we chose to use the concentration of compounds that reduced the exponential-phase growth rate by 33 %. Growth inhibition experiments were performed to determine this concentration for each compound (Table 1). A wide concentration range of growth inhibition was observed, including compounds with low inhibitory effects such as succinic acid and 1,4-butanediol (inhibiting concentrations > 200 mM) and high effects such as geraniol and decanoic acid (inhibiting concentrations in the low millimolar range).

In order to detect sRNAs, cDNA sequencing libraries were generated using RNA isolated from *E. coli* cells subjected to chemical stress or grown under high cell density fermentation conditions. For the chemical stress experiments, cells were initially grown to exponential phase, followed by compound addition, one hour of growth, and cell harvest (Fig. 1a). High cell density fermentations were performed with *E. coli* K-12 MG1655 including glucose batch and fed-batch phases, where OD_{600} values greater than 100 were reached (Fig. 1b). Cells were harvested at several points along the growth curve including exponential batch (point 1), glucose-limited exponential fed-batch (points 2–5), transition (points 6–7) and stationary (points 8–9) phases (Fig. 1b). All steps including cell growth, RNA isolation and library preparation for deep sequencing were performed in triplicate for the chemical stress samples (except those with threonine, itaconic acid and decanoic acid that were performed once) and in duplicate for the fermentation samples.

Identification of sRNAs

A total of 217 million sequencing reads were obtained for all samples, with an average of 4.3 million reads per sample (Additional file 1). The average percentage of reads mapping to intergenic sRNAs was 18.7 % per sample, corresponding to 800,000 reads (40 million in total), while the majority of reads mapped to rRNAs. For sRNA identification, trimmed sequence reads for each sample were mapped to the *E. coli* K-12 MG1655 genome and the resulting mapping files for all samples were merged into one file containing information on all sequence reads for each library. This procedure was performed for both the fermentation and stress libraries. Expression values containing the number of reads per base pair were extracted from the mapping files, and transcript regions were identified using a cutoff of 100 reads. The transcript file was reduced to contain only transcripts residing entirely within intergenic regions using a previously designed script [26]. The expression profiles of adjacent regions abutting the intergenic transcripts were further manually curated to evaluate whether each transcript corresponded to an sRNA or untranslated region (UTR). In cases where similar expression levels and/or the lack of a gap between the transcript and adjacent gene were observed, the transcript was classified as an UTR. A total of 253 novel intergenic sRNAs were identified by this approach, adding to the 92 previously annotated sRNAs in RegulonDB 8.6 [27], and the 117 sRNAs previously detected by RNA-seq [17]. Of these groups of previously reported sRNAs, 86 % and 79 %, respectively, could be detected here. Although there was evidence of transcription for most of the other annotated sRNAs, the level was low and these were therefore classified as undetected. The novel sRNAs are denoted ES001 to

Fig. 1 Representative growth curves of *E. coli* K12 MG1655 for the investigated conditions. **a** Chemical stress experiments were performed by addition of compound in exponential phase, followed by cell harvest after approximately one hour of growth. The stress samples were compared to a control where no compound was added and harvest performed at same OD_{600}. **b** During high cell density fed batch fermentation cells were harvested at different time points along the growth curve. Measurement time points are indicated (blue circles), as well as sample time points (red circles) in exponential batch (point 1), glucose-limited exponential fed-batch (points 2–5), transition (points 6–7) and stationary (points 8–9) phases

Rau *et al. BMC Genomics* (2015) 16:1051

Page 4 of 16

ES253 based on their genomic coordinates, whereas the previously detected sRNAs are referred to either by their annotated names or the ECS designations used in an earlier RNA-seq study [17].

Some representative examples of expression profiles for annotated known and novel small RNA transcripts are shown in Fig. 2. The first five plots display expression profiles in the regions containing the annotated and experimentally verified sRNAs GlmZ, CyaR, MicF, McaS, and RyhB. They demonstrate a very good correlation between actual coordinates and transcript boundaries as determined from read coverage. Although the expression profile of McaS seems to diverge from the annotated coordinates, it is actually consistent with a shorter transcript previously described in the literature [28]. The other plots illustrate expression profiles of novel sRNAs, where their coordinates can be estimated from the expression boundaries. For some sRNAs such as ES043 the boundaries are very distinct, facilitating easy estimation of sRNA coordinates, while others have less pronounced boundaries such as ES067 and ES116. The coordinates were defined based on the cutoff of 100 reads for the majority of sRNAs, whereas a subset of sRNAs chosen for further study was manually curated by visual inspection.

As the actual coordinates for a few small RNAs including ES116 are not necessarily those determined through the implementation of the cutoff value, the stated coordinates for these transcripts may differ to a minor degree. The RNA-seq derived coordinates and direction for all novel and previously detected sRNAs are presented in Additional files 2 and 3, respectively.

The chromosomal positions of the novel sRNA transcripts are depicted in Fig. 3a. Together with the positions of previously detected transcripts, the localization of sRNAs is evenly distributed on the genome. The lengths of the novel sRNAs were compared to previously identified transcripts as shown in Fig. 3b. Over 75 % of the novel transcripts have lengths of 50–150 nucleotides and over 95 % are shorter than 300 nucleotides. The lengths of 90 % of the annotated known sRNAs are between 50 and 300 nucleotides, where over one-third of these are in the range of 100–150 nucleotides. The other previously described sRNAs [17] are generally shorter, with over two-thirds between 50 and 75 nucleotides long, possibly due to selection of transcripts below 200 nucleotides in that study. The length distribution of annotated sRNAs is slightly skewed towards longer transcripts compared to both RNA-seq studies. This could
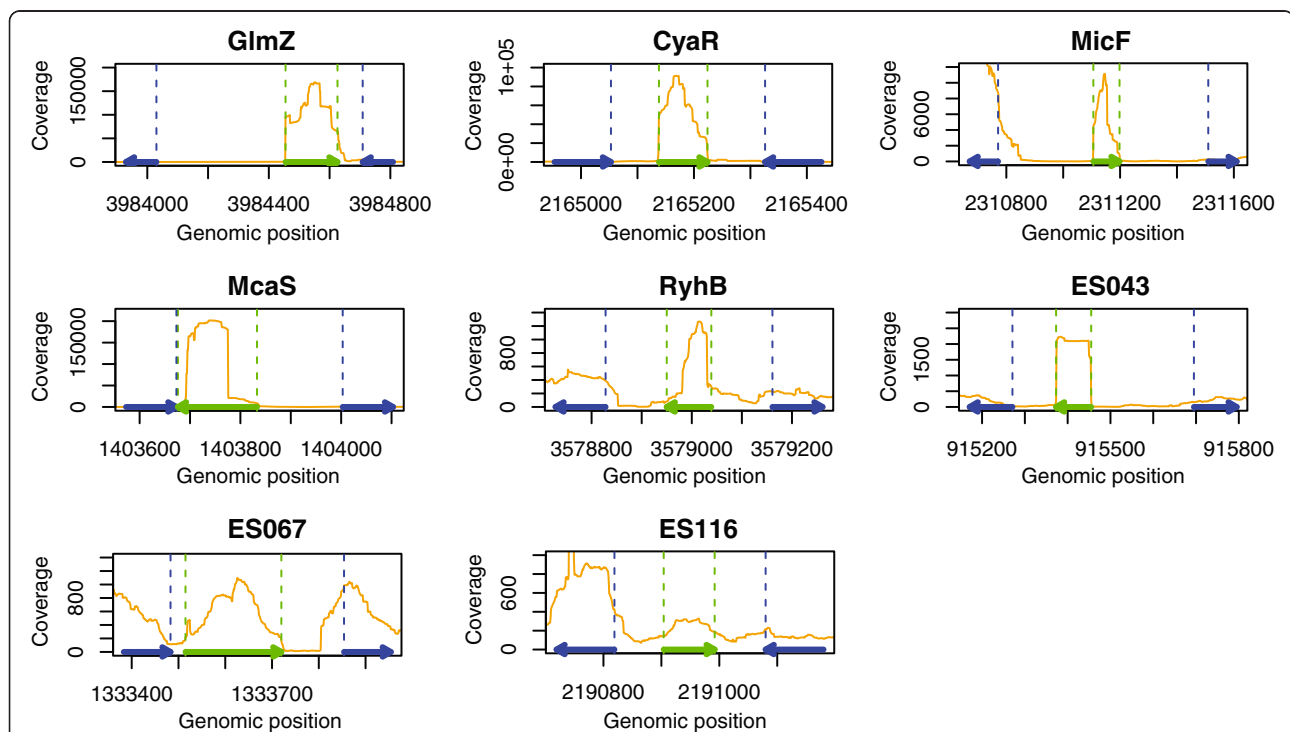

Fig. 2 Small RNA expression profiles detected with RNA-seq. The panels include the known sRNAs GlmZ, CyaR, MicF, McaS and RyhB, a novel sRNA with clearly defined coordinates, ES043, a novel sRNA with less defined coordinates, ES067, and a novel sRNA with vaguely defined coordinates and low expression level, ES116. The y-axes of the plots denote read coverage at each nucleotide position for the pooled fermentation and stress samples. The x-axes denote the genomic positions according to the *E. coli* K-12 MG1655 genome coordinates (NC_000913.2). Legend: orange lines, expression signal; dashed green lines, sRNA coordinates; dashed blue lines, intergenic region coordinates; arrows denote directions of flanking genes (blue) and sRNAs (green)
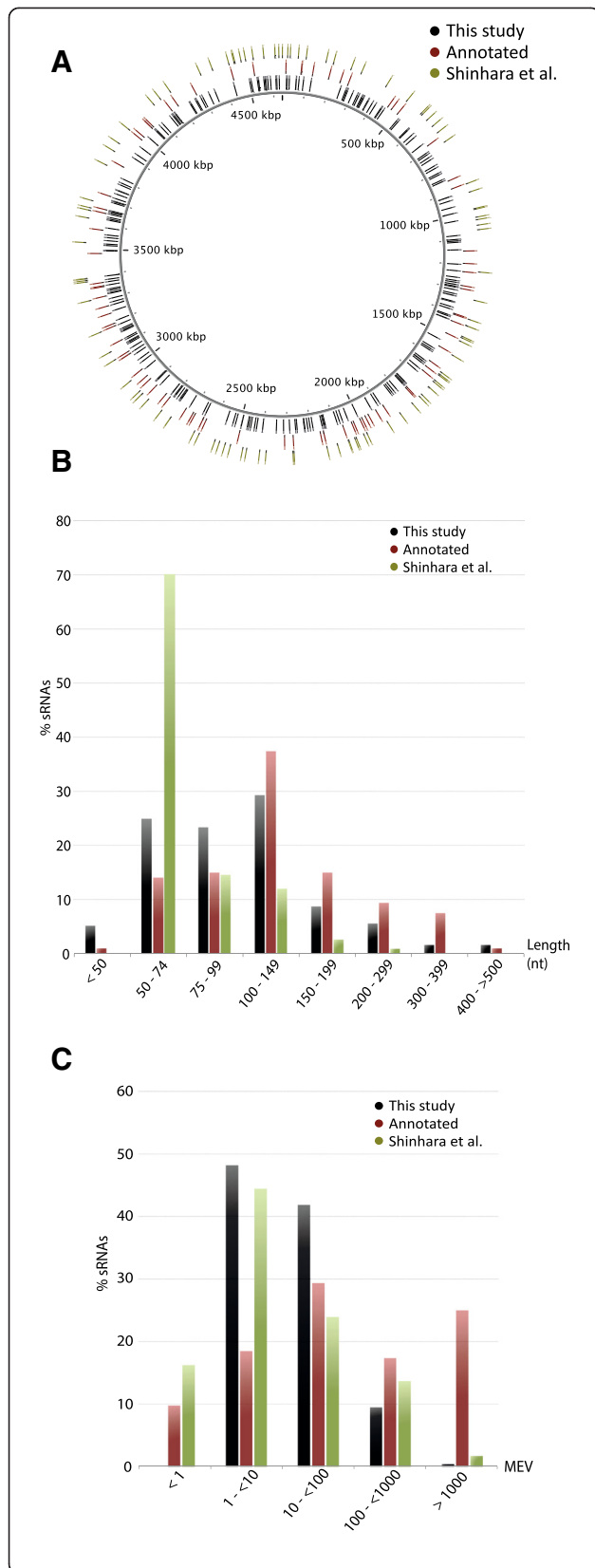
Rau *et al. BMC Genomics* (2015) 16:1051

Page 5 of 16

Fig. 3 Features of the 253 novel small RNA transcripts detected in this study in relation to previously described sRNAs. Distribution of small RNAs on the *E. coli* MG1655 genome (**a**), where the novel sRNAs detected in this study are depicted on the inner ring, the known annotated sRNAs on the middle ring and the sRNAs detected in [17] on the outer ring. The transcript lengths and expression levels for the novel small RNAs compared to the known annotated and those predicted by Shinhara *et al.* [17], are shown in (**b**) and (**c**), respectively. Expression levels are presented as cumulative Mean Expression Values (MEV) defined as total stress and fermentation library reads for each sRNA divided by sRNA length

be due to a bias in detecting longer sRNAs experimentally or that some transcripts predicted by RNA-seq are actually longer *in vivo*. The expression levels of all known transcripts, including novel transcripts and previously reported sRNAs, are represented as cumulative mean expression values (MEVs) in Fig. 3c (MEVs of each sRNA are provided in Additional files 2 and 3). Approximately 42 % of the novel transcripts have MEVs between 10 and 100 and 10 % have MEVs > 100. For the known annotated and other previously described sRNAs [17], roughly 70 % and 40 %, respectively, have MEVs of >10. The majority of highly expressed sRNAs belong to the group of annotated sRNAs, which is sensible as highly expressed sRNAs are more likely to be discovered experimentally. However a significant number also have quite low expression. This could be constitutive or limited to the particular experimental conditions investigated here.

Differential expression of sRNAs during high cell density fermentation

For differential expression analysis, read counting was performed for each sRNA, followed by normalization and statistical analysis. In the high cell density fermentation, cells were harvested in four different phases of growth, including batch exponential (glucose excess), fed-batch exponential (glucose limitation), transition and stationary phases. A heat map including only the eighty-four sRNAs with significant differential expression reveals four different expression profile types according to the harvest point (Fig. 4a). Among these the batch exponential expression profile is especially distinct. In comparison, the fed-batch exponential profiles show some level of similarity but are much less pronounced, likely due to the approximately five-fold slower growth rate in this phase. There is a marked shift in samples 6 and 7 and these are similar to the stationary phase samples 8 and 9. In the hierarchical clustering, samples 6 and 7 cluster separately from samples 8 and 9 and are therefore likely capturing the transition into stationary phase during high cell density fed-batch fermentation. Two major clusters of sRNAs are evident from the heat map with high expression in either exponential or stationary phase (Fig. 4a). The substantial changes in growth

Rau *et al. BMC Genomics* (2015) 16:1051

Page 6 of 16

conditions during the fermentation are underscored by the fact that 60 % of the differentially expressed sRNAs show changes that are four-fold or higher.

The sRNAs with significant expression changes during the fermentation are mainly a result of differential expression between batch exponential and stationary phases (Table 2), where seventy-six out of 462 sRNAs (16.5 %) show differential expression between these two growth phases. The largest changes are observed for the SraL (30-fold) and novel ES217 (33-fold) RNAs that are upregulated and the novel ES220 (177-fold) and ES003 (133-fold) RNAs that are downregulated (Fig. 4b). Several known sRNAs show differential expression in stationary phase that is consistent with expression patterns documented in the literature. The CyaR, MicF, MicL, RybA (MntS), RybB, SraL, and SroC RNAs exhibit increased expression [12, 14, 29–32], whereas the IS092 (IsrB), MgrR, Spf, SraG (PsrO), SroG and SroA RNAs exhibit decreased expression [5, 6, 12, 15]. Most of the differentially expressed sRNAs are novel and the specific expression patterns observed suggest that they may play important regulatory roles during growth (Fig. 4b). Further investigation of these could provide valuable information on cellular regulation and their effects on metabolism during fed-batch and high cell density fermentation. The nine sRNAs with differential expression in fedbatch versus batch exponential phases are novel except for SroA and SroC and may have potential utility in the regulation of metabolism in the fed-batch phase. All nine are changed at greater magnitude in the corresponding direction in stationary phase and their differential expression could consequently be an effect of overlapping conditions, such as slow growth or high cell density. The downregulation of SroA and upregulation of SroC in stationary relative to exponential growth phases is consistent with previous data from Northern blots [12]. A total of fifty-four differentially expressed sRNAs are observed between transition and batch exponential phases, where forty-six are also differentially expressed in the same direction in stationary versus batch exponential phases. Eight transcripts are specifically changed in transition phase compared to batch exponential phase, including RyhB and RprA that are downregulated and SroD and C0719 that are upregulated. There are no sRNAs with differential expression between stationary and transition phases, possibly due to the limited number of samples and the more highly variable physiology encountered during the high cell density fed batch fermentation when compared to the exponential growth phases.
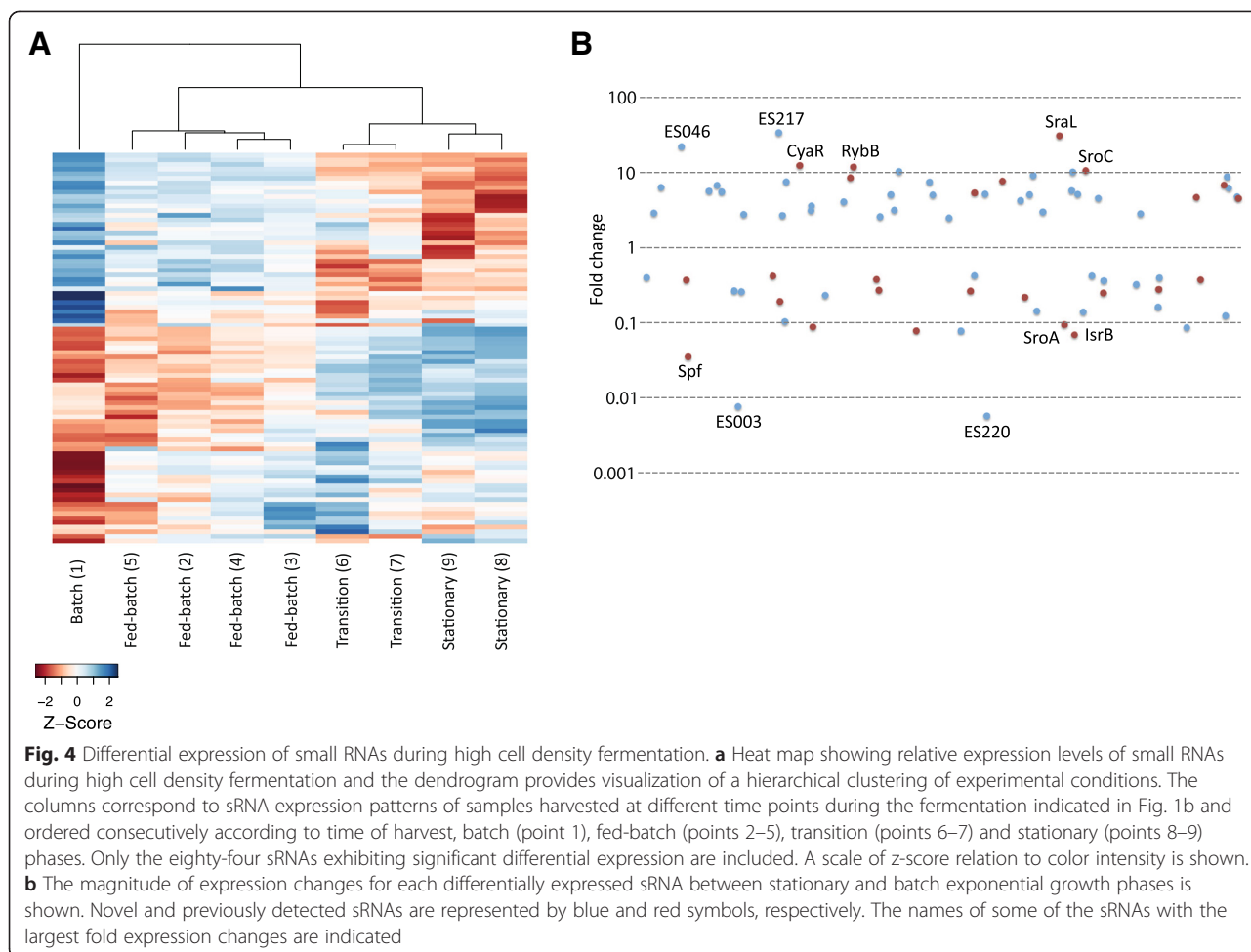
Differential expression of sRNAs during chemical stress

For the 462 identified sRNAs, 138 (29.9 %) are differentially expressed during chemical stress, eighty-four (18.2 %) are differentially expressed during high cell density fermentation, and 177 (38.3 %) are differentially expressed in at least one condition. There are forty-four sRNAs (9.5 %) that exhibit differential expression under both chemical stress and fermentation conditions. The number of differentially expressed sRNAs for each chemical stress condition is shown in Fig. 5a. The numbers of sRNAs with significantly changed expression during chemical stress vary according to the different compound types, where the organic solvents produce the most widespread changes, followed by the acids, and then relatively few changes with the amino acids, decanoic acid and geraniol. Most of the chemicals elicit similar percentages of up- and downregulated sRNAs. However, the differentially expressed sRNAs are mostly upregulated in the case of itaconic acid, decanoic acid and geraniol stress, and downregulated under serine stress. Approximately 33 % of the differentially expressed sRNAs under these conditions show changes that are four-fold or higher.

For the samples exposed to chemical stress, a principal component analysis plot provides an overview of the expression profiles (Fig. 5b). The samples exposed to the organic solvents cluster together, whereas those treated with acids are more spread. A possible reason for the more separate location of the itaconic and succinic acid samples is the high sodium concentration that was introduced as sodium hydroxide in order to neutralize their acidity, thereby adding an osmotic effect. Other samples such as those treated with serine, geraniol and decanoic acid resemble the untreated control sample and also show minor differential expression.

In order to compare the changes in sRNA expression between different conditions, the number of common differentially expressed sRNAs between each pair of stress conditions was scored (Fig. 5c). There is a significant overlap among the differentially expressed sRNAs between the datasets for samples subjected to stress with the organic solvents and on average an organic solvent condition has an overlap of 45 % of its differentially expressed sRNAs with the other organic solvent conditions but only 23 % with the acid conditions. The overlap of differentially expressed sRNAs in the samples treated with acids is smaller in absolute values. However, in relative values (Additional file 4) an acid condition on average shares 43 % and 42 % with other acid and organic solvent conditions, respectively. The relative overlap with stationary phase is for most chemical stress conditions above 33 %. Consequently a great deal of similarity in sRNA expression exists between acid, organic solvent and stationary phase conditions and especially within the group of organic solvent conditions where the overlap is high in both absolute and relative terms. The extent of sRNA differential expression is a possible

Rau *et al. BMC Genomics* (2015) 16:1051

Page 7 of 16

Fig. 4 Differential expression of small RNAs during high cell density fermentation. **a** Heat map showing relative expression levels of small RNAs during high cell density fermentation and the dendrogram provides visualization of a hierarchical clustering of experimental conditions. The columns correspond to sRNA expression patterns of samples harvested at different time points during the fermentation indicated in Fig. 1b and ordered consecutively according to time of harvest, batch (point 1), fed-batch (points 2–5), transition (points 6–7) and stationary (points 8–9) phases. Only the eighty-four sRNAs exhibiting significant differential expression are included. A scale of z-score relation to color intensity is shown. **b** The magnitude of expression changes for each differentially expressed sRNA between stationary and batch exponential growth phases is shown. Novel and previously detected sRNAs are represented by blue and red symbols, respectively. The names of some of the sRNAs with the largest fold expression changes are indicated

reflection of the overall cellular effects of the stress and hence the effects of e.g., the organic solvents on the cell could be both extensive and similar.

Differential sRNA expression in multiple stress conditions

A group of sRNAs including nineteen known and ten novel transcripts show differential expression in at least four and up to twelve out of thirteen tested stress conditions (Fig. 6). For most of these transcripts, the individual differential expression is unidirectional, especially for conditions within the same chemical group. This is also observed for sRNAs exhibiting changes in both directions, where similar chemicals generally yield expression changes in the same direction. Three overall patterns of differential expression are observed. The first group of transcripts exhibits increased expression in multiple chemical stress conditions (Fig. 6, columns with blue highlighting), where roughly half of these also show increased expression in stationary phase. A second group has decreased expression in chemical stresses, with some also having decreased expression in stationary phase (Fig. 6, columns with red highlighting). A single transcript, CyaR, exhibits decreased expression in chemical stress conditions and increased expression in

Table 2 Summary of differentially expressed sRNAs during high cell density fermentation

Growth phases compared	Number of differentially expressed sRNAs			Known upregulated sRNAs	Known downregulated sRNAs
	Total	Up	Down		
Fed-batch exponential vs. Batch exponential	9	5	4	SroC	SroA
Transition vs. Batch exponential	54	29	25	CyaR, C0719, RNA0-365, RybA, RybB, SraL, SroC, SroD	IsrB, IS128, RprA, RyhB, Spf, SroA, Tff
Stationary vs. Batch exponential	76	43	33	CyaR, MicF, MicL, RNA0-365, RybA, RybB, SraL, SroC	IsrB, IS128, MgrR, PsrN, PsrO, Spf, SroA, SroG, Tff

Rau *et al. BMC Genomics* (2015) 16:1051
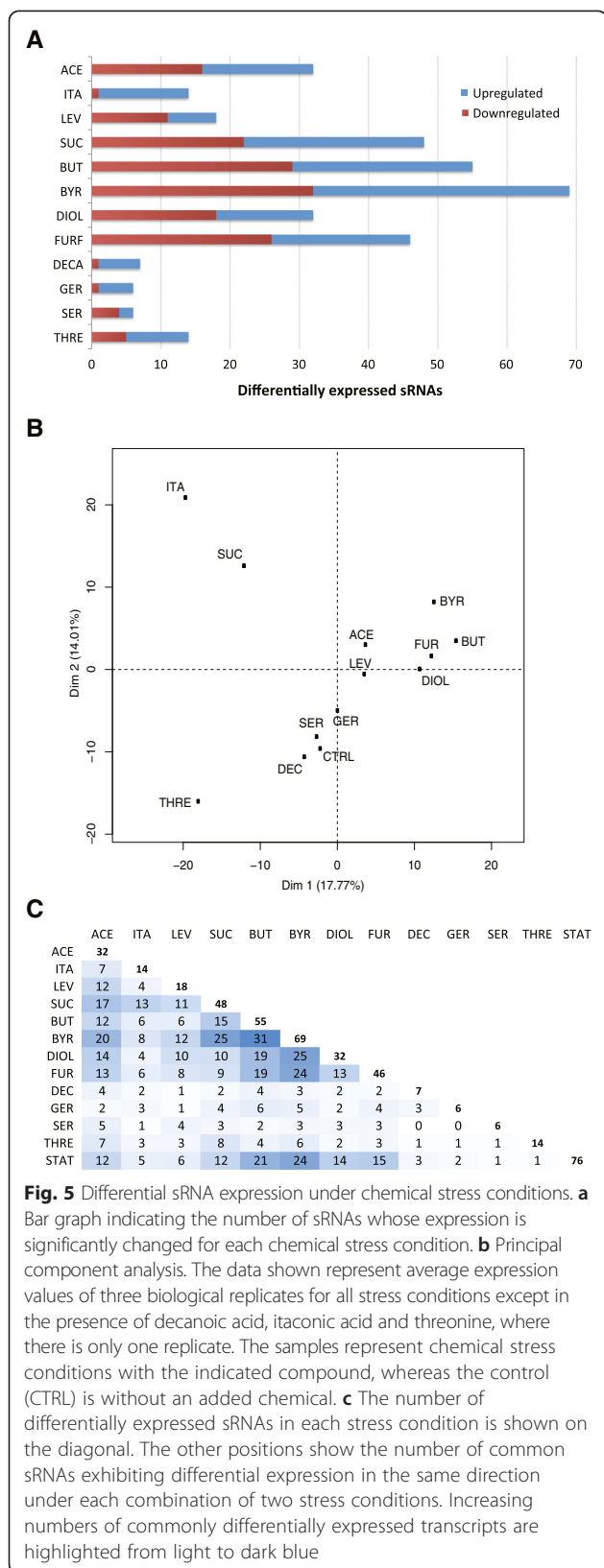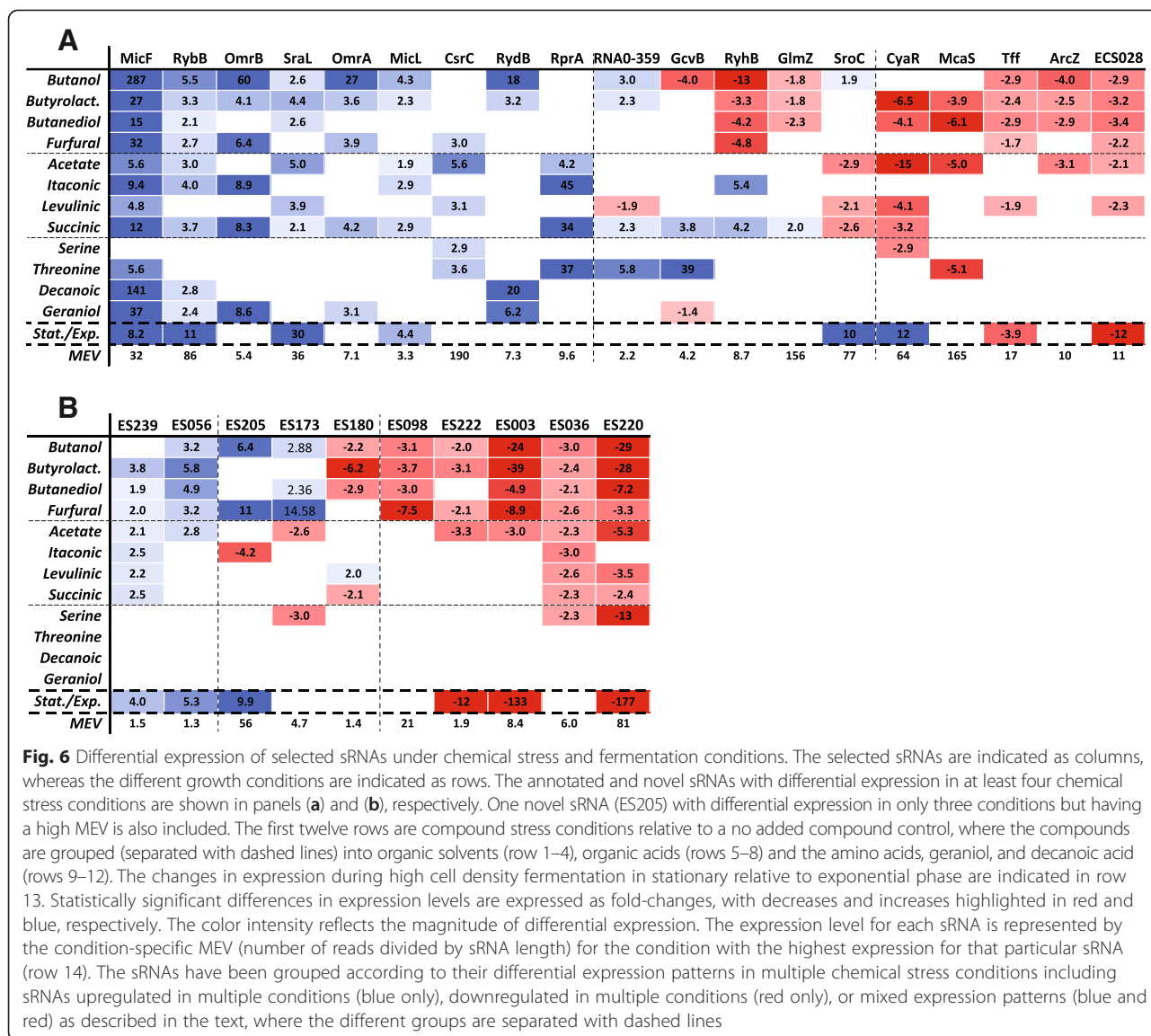
Page 8 of 16

Fig. 5 Differential sRNA expression under chemical stress conditions. **a** Bar graph indicating the number of sRNAs whose expression is significantly changed for each chemical stress condition. **b** Principal component analysis. The data shown represent average expression values of three biological replicates for all stress conditions except in the presence of decanoic acid, itaconic acid and threonine, where there is only one replicate. The samples represent chemical stress conditions with the indicated compound, whereas the control (CTRL) is without an added chemical. **c** The number of differentially expressed sRNAs in each stress condition is shown on the diagonal. The other positions show the number of common sRNAs exhibiting differential expression in the same direction under each combination of two stress conditions. Increasing numbers of commonly differentially expressed transcripts are highlighted from light to dark blue

stationary phase. The third group of transcripts shows unique mixed differential expression patterns under the tested conditions (Fig. 6, columns with blue and red highlighting). In addition to these selected sRNAs, grouping of all differentially expressed sRNAs by hierarchical clustering reveals two overall clusters based on expression in fermentation conditions and six overall clusters based on expression in chemical stress conditions (Additional file 5). The two fermentation-related clusters consist of sRNAs with increased expression in either batch exponential or stationary phases, respectively, while the six chemical stress-related clusters each have their own signature of condition-specific sRNA differential expression.

Several small RNAs of unknown function show differential expression under the tested conditions. Ten hitherto undescribed transcripts show differential expression in four to nine out of thirteen tested stress conditions (Fig. 6b and Table 3). These transcripts are expressed at a level over an MEV of 1 in at least one condition (corresponding to just under 100 reads) and a subset are expressed at higher levels with MEVs between 20 and 82 in at least one growth condition. The expression profiles of these transcripts for the pooled fermentation and stress samples are included in Additional file 6. Most of the sRNA profiles show clearly defined signals compared to the signals of the adjacent genes. For a few transcripts including ES003 the expression profiles from the separate fermentation or stress samples showed a more sharp separation from adjacent genes and these transcripts were therefore assigned as intergenic sRNAs rather than UTRs. All transcripts except two have experimentally determined transcription start sites [27, 33–35] that are consistent with the coordinates defined in this study (Table 3). The ES036 and ES098 transcripts lacking promoter evidence could come from longer processed transcripts and be part of the growing group of sRNAs that are derived from within coding regions, in particular 3′-UTRs [36]. The novel transcripts do not contain predicted Rho-independent terminator sequences. For most of the novel transcripts, sequence conservation is limited to the *Escherichia* and *Shigella* genera, but the ES056 and ES098 transcripts are also conserved in *Citrobacter* and *Salmonella*, respectively (Table 3).

Growth experiments with sRNA mutant strains

Strains overexpressing two of the novel (ES205, ES220) and seven of the previously annotated or predicted (GcvB, McaS, RprA, RydB, RyhB, SraL, SroC) differentially expressed sRNAs, as well as strains with deleted sRNA (for all except GcvB and SraL) were constructed and used in chemical stress growth experiments to measure growth rates compared to wild type controls

Rau *et al. BMC Genomics* (2015) 16:1051

Page 9 of 16

A

	MicF	RybB	OmrB	SraL	OmrA	MicL	CsrC	RydB	RprA	RNA0-359	GcvB	RyhB	GlmZ	SroC	CyaR	McaS	Tff	ArcZ	ECS028
Butanol	287	5.5	60	2.6	27	4.3		18		3.0	-4.0	-13	-1.8	1.9			-2.9	-4.0	-2.9
Butyrolact.	27	3.3	4.1	4.4	3.6	2.3		3.2		2.3		-3.3	-1.8		-6.5	-3.9	-2.4	-2.5	-3.2
Butanediol	15	2.1		2.6								-4.2	-2.3		-4.1	-6.1	-2.9	-2.9	-3.4
Furfural	32	2.7	6.4		3.9		3.0					-4.8					-1.7		-2.2
Acetate	5.6	3.0		5.0		1.9	5.6		4.2					-2.9	-15	-5.0		-3.1	-2.1
Itaconic	9.4	4.0	8.9			2.9		45			5.4				-4.1		-1.9		-2.3
Levulinic	4.8			3.9			3.1					-1.9		-2.1	-4.1		-1.9		-2.3
Succinic	12	3.7	8.3	2.1	4.2	2.9			34	2.3	3.8	4.2	2.0	-2.6	-3.2				
Serine							2.9							-2.9					
Threonine	5.6						3.6		37	5.8	39					-5.1			
Decanoic	141	2.8						20											
Geraniol	37	2.4	8.6		3.1			6.2			-1.4								
Stat./Exp.	8.2	11		30		4.4								10	12		-3.9		-12
MEV	32	86	5.4	36	7.1	3.3	190	7.3	9.6	2.2	4.2	8.7	156	77	64	165	17	10	11

B

	ES239	ES056	ES205	ES173	ES180	ES098	ES222	ES003	ES036	ES220
Butanol		3.2	6.4	2.88	-2.2	-3.1	-2.0	-24	-3.0	-29
Butyrolact.	3.8	5.8			-6.2	-3.7	-3.1	-39	-2.4	-28
Butanediol	1.9	4.9		2.36	-2.9	-3.0		-4.9	-2.1	-7.2
Furfural	2.0	3.2	11	14.58		-7.5	-2.1	-8.9	-2.6	-3.3
Acetate	2.1	2.8		-2.6			-3.3	-3.0	-2.3	-5.3
Itaconic	2.5		-4.2						-3.0	
Levulinic	2.2				2.0				-2.6	-3.5
Succinic	2.5				-2.1				-2.3	-2.4
Serine				-3.0					-2.3	-13
Threonine										
Decanoic										
Geraniol										
Stat./Exp.	4.0	5.3	9.9				-12	-133		-177
MEV	1.5	1.3	56	4.7	1.4	21	1.9	8.4	6.0	81

Fig. 6 Differential expression of selected sRNAs under chemical stress and fermentation conditions. The selected sRNAs are indicated as columns, whereas the different growth conditions are indicated as rows. The annotated and novel sRNAs with differential expression in at least four chemical stress conditions are shown in panels (**a**) and (**b**), respectively. One novel sRNA (ES205) with differential expression in only three conditions but having a high MEV is also included. The first twelve rows are compound stress conditions relative to a no added compound control, where the compounds are grouped (separated with dashed lines) into organic solvents (row 1–4), organic acids (rows 5–8) and the amino acids, geraniol, and decanoic acid (rows 9–12). The changes in expression during high cell density fermentation in stationary relative to exponential phase are indicated in row 13. Statistically significant differences in expression levels are expressed as fold-changes, with decreases and increases highlighted in red and blue, respectively. The color intensity reflects the magnitude of differential expression. The expression level for each sRNA is represented by the condition-specific MEV (number of reads divided by sRNA length) for the condition with the highest expression for that particular sRNA (row 14). The sRNAs have been grouped according to their differential expression patterns in multiple chemical stress conditions including sRNAs upregulated in multiple conditions (blue only), downregulated in multiple conditions (red only), or mixed expression patterns (blue and red) as described in the text, where the different groups are separated with dashed lines

(Additional file 7). Overexpression was achieved by high copy number plasmid using rhamnose-inducible expression. Three or four chemical concentrations for each compound were employed ranging from the relatively low concentration used for RNA sequencing and up to ten-fold higher concentrations, with the latter often inhibiting growth completely (Additional file 7). However, no significant differences in growth rate were detected in the mutant strains (data not shown). These results indicate that increasing the expression of these sRNAs or deleting them does not result in measurable changes in growth and as a result does not appear to increase the stress tolerance in the conditions tested. It is possible that increased chemical tolerance would be present in a more sensitive assay such as competition experiments or by assaying e.g., survival of chemical stress in stationary phase, a relevant production phase. Redundancy in regulatory mechanisms compensating for engineered changes in expression of the selected sRNA is another possibility.

Discussion

The expression of many hitherto undescribed small RNA transcripts is reported in this study. In relation to the two other RNA-seq studies in *E. coli*, there is a much greater sequencing depth relative to one study [17] and a comparable sequencing depth to the other study [16]. The former study, with 89395 reads mapping to known sRNAs and 30831 reads mapping to novel sRNAs, employed computational and experimental evidence of transcription initiation combined with RNA-seq data to identify 117 novel intergenic transcripts [17]. The latter study had 3 million reads for annotated sRNAs, 4.3 million reads for novel intergenic regions and reported

Rau *et al. BMC Genomics* (2015) 16:1051

Page 10 of 16

Table 3 Selected novel intergenic small RNA transcripts

Name	Start[a]	End[a]	Length[b]	Direction	TSS coordinates[c]	Conservation[d]
ES003	29551	29603	53	+	I	Esc,Shi
ES036	740211	740253	43	+	-	Esc,Shi
ES056	1102475	1102566	92	-	+1	Esc,Shi,Cit
ES098	1814204	1814245	42	+	-	Esc,Shi,Sal
ES173	3154543	3154606	64	+	+8	Esc,Shi
ES180	3295030	3295119	90	-	+11	Esc,Shi
ES205	3657034	3657105	72	+	I	Esc,Shi
ES220	4010909	4011013	105	+	I	Esc,Shi
ES222	4056241	4056348	108	+	−12	Esc,Shi
ES239	4434589	4434711	123	+	I	Esc,Shi

[a]The transcript coordinates determined by RNA seq data
[b]The transcript length determined by RNA seq data
[c]The presence of experimentally determined transcription start sites (TSS) from previous studies is indicated, as well as any differences in TSS coordinates between previous studies and present study. For example '+1' denotes a previously determined TSS one nucleotide downstream of that determined in this study. Identical coordinates are indicated with 'I' and cases with no previously determined TSSs or predicted promoters are marked with '-'
[d]The sequence conservation of the transcripts in other bacterial genera is indicated: Esc *Escherichia*, Shi *Shigella*, Sal *Salmonella*, Cit *Citrobacter*

only 10 novel sRNAs [16]. Although a number of transcripts mapping to intergenic regions with high expression levels were also detected, direct comparison with these is precluded by the fact that only coordinates of entire transcript-containing intergenic regions rather than transcript coordinates are reported [16]. Considering that 117 novel intergenic transcripts were found with low sequencing depth our finding of 253 novel sRNAs with high sequencing depth seems proportionate.

There are many examples of sRNAs with expression changes in the same direction under the studied acid and organic solvent stress conditions (Fig. 6). Consequently the differential expression of these sRNAs could mediate similar functional outcomes within but also between the two chemical groups (Fig. 7). For the sRNAs exhibiting differential expression in multiple chemical stress conditions (Fig. 6) three overall groups emerged. These include groups of sRNAs exhibiting solely upregulation, solely downregulation or mixed directions of differential expression in multiple conditions, respectively.

For the group of sRNAs with increased differential expression in multiple stress conditions, several of these are involved in modulating the expression of outer membrane proteins. The MicF and RybB sRNAs are upregulated in nearly all of the 13 tested conditions and modulate the bacterial cell surface by regulating expression of the major outer membrane proteins (OMPs), OmpF and OmpC, which control the transport of molecules into and out of the cell [37]. In particular, RybB is a global regulator that is part of the σ^E envelope stress response activation network [38–42] and functions to impart repressor functions to the σ^E regulon to maintain envelope homeostasis [43]. As porins comprise only a subset of its targets, the scope of control by RybB

extends well beyond the cell envelope and is connected to other global regulatory systems [43]. The MicF sRNA is also known to bind to other targets including the global transcription factor Lrp, where it indirectly activates genes in the Lrp regulon via antisense base-pairing with the translation initiation region of *lrp* mRNA [44]. As Lrp is considered to be an activator of genes required during nutrient-poor conditions and a repressor of genes required in nutrient-rich conditions [44], its potential repression by MicF suggests an increase in energy consumption during chemical stress. The OmrA and OmrB sRNAs are upregulated in five and six chemical stress conditions, respectively, and regulate expression of β-barrel family OMPs including gated channels for iron-siderophore complexes (CirA, FecA, FepA) and the protease OmpT [45]. The newly characterized MicL sRNA is upregulated in five chemical stress conditions and is one of the increasing numbers of sRNAs that originate from within protein-coding rather than intergenic regions only [36, 46]. It is part of the σ^E regulon and regulates expression of the abundant outer membrane lipoprotein Lpp by binding to the *lpp* mRNA and inhibiting translation [31]. Taken together, the data indicate that a major physiological outcome mediated by sRNAs under chemical stress is the repression of outer membrane protein expression, likely as a defense mechanism potentially limiting the entry of harmful extracellular molecules.

Other sRNAs with increased expression under multiple chemical stress conditions include the RpoS-dependent SraL RNA, the RpoS-targeting RNA RprA, and the protein-binding CsrC RNA (Figs. 6a and 7). SraL targets the ribosome-associated protein folding chaperone, trigger factor, via binding to the translation initiation region of *tig* mRNA [32]. Upregulation of SraL could

Rau *et al. BMC Genomics* (2015) 16:1051

Page 11 of 16

Fig. 7 Model of sRNA involvement in the chemical stress response in *E. coli*. An overview of differentially expressed annotated sRNAs during chemical stress with acids and organic solvents. Only sRNAs with significant expression changes in the presence of three or four out of the four acids or four organic solvents are included (filled orange squares, acids; filled green squares, organic solvents). Small RNAs with differential expression in multiple chemical stress conditions are shown with small vertical arrows to indicate increased or decreased expression. The documented action of sRNAs on selected targets or processes is indicated either with blocked lines (repression) or arrows (activation). The potential physiological effects of sRNA differential expression under chemical stress are shown with highlighting, where red indicates inhibition and blue indicates activation. Arrows from the sigma factors RpoE and RpoS indicate the sRNAs with sigma factor dependent expression. The blocked line and arrows from sRNAs to RpoS indicate sRNAs that target the *rpoS* mRNA and result in repression or activation of expression, respectively

function to adjust the level of trigger factor to reduced ribosome and protein synthesis in stationary phase and under chemical stress. RprA activates expression of RpoS by binding to the 5′-UTR of *rpoS* mRNA in an Hfq-dependent manner [47, 48]. The strong increases in RprA expression observed in the presence of acetate, itaconic acid and succinic acid is supported by documented activation of RprA under conditions of cell envelope stress and acidification [49, 50]. Furthermore, expression of RprA has been shown to confer acid resistance when overexpressed together with other RpoS-activating sRNAs, DsrA and ArcZ [49, 51]. The CsrC sRNA binds to and sequesters the mRNA binding protein CsrA via multiple GGA motifs in the loops of predicted stem-loop structures and thereby competes with mRNAs for CsrA binding. Increased CsrC expression in multiple chemical stresses would serve to relieve the effects of CsrA,

which represses stationary phase gene expression and activates genes needed for growth [52].

The sRNAs with decreased expression under multiple chemical stress conditions include the CRP-dependent McaS and CyaR RNAs and the RpoS-targeting RNA ArcZ (Figs. 6a and 7). CyaR represses expression of LuxS, NadE, and OmpX by base-pairing with the translation initiation region of target mRNAs [30]. Downregulation of CyaR and possible de-repression of LuxS suggests an increase in quorum sensing during acid stress. McaS is a dual-function RNA that regulates motility and biofilm formation through base-pairing with target mRNAs and via protein binding to the CsrA protein, a negative regulator of e.g., the expression of exopolysaccharide β-1,6 N-acetyl-D-glucosamine (PGA) biosynthesis genes [28, 53]. ArcZ activates RpoS expression by base pairing to the *rpoS* mRNA at overlapping positions

Rau *et al. BMC Genomics* (2015) 16:1051

Page 12 of 16

targeted by the RprA and DsrA sRNAs [54]. In addition to repression of several specific mRNA targets, overexpression of ArcZ results in differential expression of over 750 genes and loss of motility [55].

Other sRNAs show a mixed response, with increased expression in the presence of some chemicals and decreased expression with others. The RyhB and GlmZ sRNAs are involved in iron and glucosamine-6-phosphate homeostasis [56, 57], respectively, and show similar patterns of mixed differential expression with decreases in the presence of organic solvents and increases in acids (Fig. 7). The combination of upregulation of OmrA/B and downregulation of RyhB suggests a reduction of iron availability during organic solvent stress. The strong increase observed in GcvB expression in the presence of threonine is consistent with its role in regulating expression of amino acid transporters and the *thrL* leader peptide upstream of the threonine biosynthetic operon [58, 59]. A recent investigation has revealed that the SroC RNA functions as an RNA sponge by base pairing with GcvB, triggering its degradation by RNase E, and thereby relieving repression of genes in the GcvB regulon [60]. Their expression is also inversely correlated in the two conditions in which both are differentially expressed.

A number of the sRNAs exhibiting differential expression under chemical stress including OmrA, OmrB, RprA, CsrC, McaS, ArcZ and GcvB are incorporated into complex networks controlling motility and biofilm formation [61, 62]. These regulatory circuits allow sRNAs to integrate extracellular signals to control transcription factor expression and generate appropriate physiological outputs. The mRNAs encoding the RpoS sigma subunit of RNA polymerase and the master regulators of flagella expression (*flhDC*) and curli synthesis (CsgD) have been shown to be targeted by multiple sRNAs and are thus major hubs for sRNA regulation. The above sRNAs regulate the expression of at least one of these three mRNAs via base pairing interactions with the exception of CsrC that acts indirectly via the RNA-binding protein CsrA [54, 63–66].

Conclusions
In this study, RNA sequencing was used to monitor expression of intergenic small RNAs in the bacterium *E. coli* during high cell density fermentation and chemical stress conditions. A total of 462 transcripts have been investigated, including 253 hitherto unknown transcripts, thereby more than doubling the number of intergenic small RNAs described in this organism. Differential expression analysis of fermentation samples showed that three-quarters of the eighty-four transcripts with significantly changed expression levels are novel, supporting possible regulatory roles for these transcripts in adaptation during different fermentation stages. Further investigation

of these could provide new insight into *E. coli* regulation during fed-batch fermentation, which is important for a complete understanding of metabolism that often affects biochemical production. The study has identified twenty-nine novel and previously annotated small RNAs that exhibit differential expression in multiple chemical stress conditions. The effects of differentially expressed sRNAs with known function indicate a general downregulation of many outer membrane proteins, of which many are porins, suggesting a decreased influx of molecules into the cell (Fig. 7). Other effects appear to be decreased availability of iron, RpoS upregulation, as well as an increase in quorum sensing. Overall the differentially expressed sRNAs include several that regulate envelope homeostasis and control bacterial lifestyle, underscoring the involvement of specific sRNAs in coordinating the necessary physiological changes to respond to chemical stress. This work provides insights into sRNAs involved in the chemical stress response and their relevance for engineering in biotechnological applications.

Methods
Bacterial strains & growth conditions
The *Escherichia coli* K-12 MG1655 strain was employed in this study. For strain engineering LB medium was used. For chemical stress experiments M9 medium [67] with 0.2 % glucose was used with added trace elements (per liter: 5 mg $FeCl_3 \times 6H_2O$, 1 mg $ZnSO_4 \times 7H_2O$, 0.2 mg $CuCl_2 \times 2H_2O$, 0.5 mg $MnSO_4 \times H_2O$, 0.3 mg $CoCl_2 \times 6H_2O$, 0.3 mg $Na_2EDTA \times 2H_2O$) and vitamins (per liter: 10 µg pyridoxine-HCl, 5 µg thiamine-HCl, 5 µg riboflavin, 5 µg nicotinic acid, 5 µg calcium D-(+)-pantothenate, 5 µg p-aminobenzoic acid, 2 µg biotin, 2 µg lipoic acid, 0.1 µg vitamin B12). For high-cell density fermentation a custom medium was employed containing (per liter): 10 g glucose, 14 g KH_2PO_4, 5 g $(NH_4)_2SO_4$, 2 g citric acid x H_2O, 1 g $MgSO_4 \times 7$ H_2O, 100 µL antifoam (Antifoam A, Sigma), trace elements (per liter: 0.2 g citric acid x H_2O, 25 mg $MnCl_2 \times 4H_2O$, 50 mg NaCl, 75 mg $FeCl_3 \times 6H_2O$, 5 mg $CoCl_2 \times 6$ H_2O, 25 mg $ZnSO_4 \times 7H_2O$, 0.5 mg $CuCl_2 \times 2H_2O$, 0.5 mg boric acid, 0.5 mg $NaMoO_4 \times 2H_2O$, 5 mg $CaCl_2$), vitamins (per liter: 30 mg pyridoxine-HCl, 15 mg thiamine-HCl, 15 mg riboflavin, 15 mg nicotinic acid, 15 mg calcium D-(+)-pantothenate, 15 mg p-aminobenzoic acid, 15 mg lipoic acid, 15 mg biotin, 15 mg vitamin B12). The feed medium for the fed-batch fermentation consisted of 70 % glucose, 10 g/L $MgSO_4$, trace elements (per liter: 250 mg citric acid x $2H_2O$, 31 mg $MnCl_2 \times 4H_2O$, 62.5 mg, NaCl, 94 mg $FeCl_3 \times 6$ H_2O, 62.5 mg $CoCl_2 \times 6$ H_2O, 31 mg $ZnSO_4 \times 7H_2O$, 0.6 mg $CuCl_2 \times 2H_2O$, 0.6 mg boric acid, 0.6 mg $NaMoO_4 \times 2H_2O$, 62.5 mg $CaCl_2$), vitamins (per liter: 25 mg pyridoxine-HCl, 12.5 mg thiamine-HCl, 12.5 mg riboflavin, 12.5 mg

Rau *et al. BMC Genomics* (2015) 16:1051

Page 13 of 16

nicotinic acid, 12.5 mg calcium D-(+)-pantothenate, 12.5 mg p-aminobenzoic acid, 12.5 mg lipoic acid, 12.5 mg biotin, 12.5 mg vitamin B12).

For chemical stress experiments cells were grown overnight in M9 medium and diluted to an OD_{600} of 0.05 in 25 mL M9 medium in 250 mL baffled shake flasks. After growth to around OD_{600} 0.9, 25 mL M9 medium with chemical stressor was added and cells were grown for 1 h before harvest. Control cells without chemical stressor in the added 25 mL of M9 medium were grown for approximately 0.5 h to reach the same OD_{600}, 0.6-0.7, as the chemically stressed conditions. The employed chemicals were sodium acetate (Sigma, S8750), butanol (Sigma, 281549), 3-hydroxy-butyrolactone (TCI Chemicals, H0939), 1,4-butanediol (Merck, 801534), decanoic acid (Sigma, W236403), furfural (Sigma, 185914), geraniol (TCI Chemicals G0027), itaconic acid (Sigma, I29204), levulinic acid (Sigma, L2009), L-serine (Sigma, S4311), succinic acid (Sigma, S9512) and L-threonine (Sigma, T8441). Triplicates were performed for most compounds except three (decanoic acid, itaconic acid and threonine) to reduce the total number of samples added to the Illumina flow-cell lane. Compounds were selected based on importance and similarity to other compounds.

For high cell density fermentation, cells grown overnight were diluted to an OD of 0.1 in 0.5 L fermentation medium within a 1 L fermentor. At an OD_{600} of approximately 11, glucose was consumed and feed was turned on. During fed-batch fermentation feed was added to sustain an average growth rate of 0.14 h^{-1} corresponding to a doubling time of 5 h. The feed was increased at a steady state while avoiding excess glucose in the medium, thereby keeping cells glucose limited. The pH was monitored and maintained at 6.8 and oxygen saturation was measured and oxygen sparged into the medium during later stages of fed-batch growth to maintain saturation above 50 %.

Isolation and processing of RNA
RNA isolation and processing was performed as previously described with minor modifications [26]. Briefly, cells were harvested by adding 10 mL of cell culture to 2 mL ethanol containing 5 % phenol, followed by pelleting of cells by centrifugation and freezing at –80 °C. Subsequently cell pellets were lysed with 1 mg/mL lysozyme for 5 min and RNA extracted using Trizol and chloroform. For DNA removal each sample was treated with 40 units DNase I for 30 min. The purity and quality of RNA was verified with spectrophotometer and Bioanalyzer (Agilent Technologies). RNA molecules of length 50 – 500 nucleotides, containing sRNAs, were selected on polyacrylamide-urea gels (Bio-Rad), followed by depletion of 5S rRNA employing the MICROBExpress kit (Life Technologies) using an HPLC purified custom

Capture Oligo specific for 5S rRNA (5′-AAAAAAAA AAAAAAAAAAGCGTTTCACTTCTGAGTTCGGCA- 3′). RNA was then treated with Tobacco Acid Pyrophosphatase (Epibio) (10 U per sample, containing up to 10 µg RNA) and fragmented using RNase III (Fermentas) for 10 min. Purity and size distribution of RNA was assessed with spectrophotometer and Bioanalyzer after this and the preceding steps. Finally sRNA libraries were constructed using the TruSeq Small RNA Sample Preparation Kit (Illumina) with a few modifications. The protocol was initiated at the section of *Ligate 3′ and 5′ adapters* using 100–400 ng of prepared RNA and was followed up to and including the section *Reverse Transcribe and Amplify*. Subsequently, resulting libraries were purified using AMPure XP beads (Beckmann Coulter). The concentration of the resulting DNA libraries was measured using a Qubit Fluorometer and size distribution assessed using a Bioanalyzer. The libraries were then subjected to paired-end sequencing on an Illumina Hi-Seq 2000 at Beckmann Coulter Genomics.

Sequencing data analysis
The obtained reads were trimmed using Trimmomatic (standard settings) for removal of low-quality regions of reads. The reads were then mapped to the *E. coli* K-12 MG1655 genome (NC_000913.2) using Bowtie 2 [68] and the resulting sam files were converted to bam files using Samtools [69]. For sRNA identification, bam files for stress samples were merged together, while bam files for fermentation samples were merged together. Using a custom script [26] the number of mapped reads for each nucleotide position in the genome was obtained and intergenic regions containing above 100 reads per nucleotide were selected as potential transcripts. These regions were subsequently manually curated to e.g., join adjacent regions with gaps of few base pairs with fewer than 100 reads per nucleotide or filter out short regions. Potential transcripts were further manually curated by visual inspection using IGV Viewer [70] to remove likely UTRs. In cases where similar expression levels and/or the lack of a gap between the transcript and adjacent gene were observed, the transcript was classified as a UTR and excluded. For differential expression analysis the list of newly identified small RNAs were joined with already annotated sRNAs and previously predicted sRNAs [17]. Read counts for each sRNA in each sample were obtained using CLC Genomics software. Subsequently TMM normalization [71], a scaling normalization method employing weighted trimmed mean of the log expression ratios, was utilized. Differential expression analysis was performed with the edgeR R statistics package [72], regarding counts as a negative binomial distribution and data fitted to generalized linear models. Default

Rau *et al. BMC Genomics* (2015) 16:1051

Page 14 of 16

parameters were employed and genes with FDR-values below 0.05 were defined as significantly differentially expressed. The mean expression value (MEV) is defined as the number of reads divided by sRNA length. Cumulative MEVs include all reads from all samples, while condition-specific MEVs include only reads from a certain experimental condition. Genome coordinate plot was performed with cgview [73], heatmap and hierarchical clustering employing the Pearson correlation as distance measure with the R packages bioDist and gplots using normalized log-transformed expression values. Principal component analysis was performed with the R package FactoMineR.

Secondary information on selected novel sRNAs

Promoters of novel sRNAs were either identified from experimental sources [27, 33–35] or predicted using BProm that predicts σ^{70} promoters [74]. Transcriptional terminators were predicted using ARNold [75, 76]. Conservation was estimated by blast analysis towards organisms of the following genera: *Citrobacter, Cronobacter, Escherichia, Edwardsiella, Enterobacter, Erwinia, Klebsiella, Pantoea, Salmonella, Serratia, Shigella, Sodalis, Vibrio* and *Yersinia*. If the identity over the entire length of the sRNA was on average above 90 %, then it was classified as conserved in the respective organism.

sRNA mutant strain engineering

Briefly, sRNA deletion mutants were constructed using the λ red recombineering technique [77] while sRNA overexpression was performed by insertion of sRNAs downstream of a rhamnose inducible promoter on a pRSFDuet-1 high-copy plasmid. The selected sRNAs were ES205, ES220, GcvB, McaS, RprA, RydB, RyhB, SraL and SroC. For further details see Additional file 7.

Growth rate inhibition experiments

Growth rates of sRNA deletion mutants were tested using three or four chemical concentrations for each chemical (see Additional file 7) and three biological replicates. Overexpression strains were tested in three or four chemical concentrations using three different rhamnose inducer concentrations in each, but without replicates. Cells were grown overnight in M9 with 0.2 % glucose and diluted into M9 medium with 0.2 % glucose, trace elements, vitamins (the same concentrations as for chemical stress experiments) and relevant chemical added. For overexpression rhamnose was present at transfer. Growth was performed in microtiter 96 square well plates (Enzyscreen B.V.) and these were incubated at 37 °C with 225 rpm shaking in a Growth Profiler 1152 (Enzyscreen B.V.). Growth rates of mutants compared to wild type were tested for statistically significant

differences. For sRNA overexpression the control was WT with pRSF-Duet-1 without sRNA insertion.

Additional files

Additional file 1: RNA sequencing information. (XLSX 11 kb)

Additional file 2: Novel intergenic small RNAs detected in this study. (XLSX 130 kb)

Additional file 3: Previously identified intergenic small RNAs. (XLSX 107 kb)

Additional file 4: Relative overlap of differentially expressed sRNAs between conditions. (EPS 1956 kb)

Additional file 5: Clustering of differentially expressed sRNAs. (EPS 7438 kb)

Additional file 6: Expression profiles of selected novel sRNAs. (EPS 2014 kb)

Additional file 7: Supplementary Materials and Methods. (DOCX 31 kb)

References
1. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell. 2011;43(6):880–91.
2. Waters LS, Storz G. Regulatory RNAs in bacteria. Cell. 2009;136(4):615–28.
3. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol. 2011;3(12):a003798.
4. Sharma CM, Vogel J. Experimental approaches for the discovery and characterization of regulatory small RNA. Curr Opin Microbiol. 2009;12(5):536–46.
5. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, et al. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr Biol. 2001;11(12):941–50.
6. Chen S, Lesnik EA, Hall TA, Sampath R, Griffey RH, Ecker DJ, et al. A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. Biosystems. 2002;65(2–3):157–77.
7. Rivas E, Klein RJ, Jones TA, Eddy SR. Computational identification of noncoding RNAs in E. coli by comparative genomics. Curr Biol. 2001;11(17):1369–73.
8. Saetrom P, Sneve R, Kristiansen KI, Snove Jr O, Grunfeld T, Rognes T, et al. Predicting non-coding RNA genes in Escherichia coli with boosted genetic programming. Nucleic Acids Res. 2005;33(10):3263–70.
9. Tran TT, Zhou F, Marshburn S, Stead M, Kushner SR, Xu Y. De novo computational prediction of non-coding RNA genes in prokaryotic genomes. Bioinformatics. 2009;25(22):2897–905.
10. Yachie N, Numata K, Saito R, Kanai A, Tomita M. Prediction of non-coding and antisense RNA genes in Escherichia coli with Gapped Markov Model. Gene. 2006;372:171–81.

Rau *et al. BMC Genomics* (2015) 16:1051

Page 15 of 16

11. Kawano M, Reynolds AA, Miranda-Rios J, Storz G. Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. Nucleic Acids Res. 2005;33(3):1040–50.

12. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, et al. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. 2003;31(22):6435–43.

13. Tjaden B, Saxena RM, Stolyar S, Haynor DR, Kolker E, Rosenow C. Transcriptome analysis of Escherichia coli using high-density oligonucleotide probe arrays. Nucleic Acids Res. 2002;30(17):3732–8.

14. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 2001;15(13):1637–51.

15. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol. 2003; 50(4):1111–24.

16. Raghavan R, Groisman EA, Ochman H. Genome-wide detection of novel regulatory RNAs in E. coli. Genome Res. 2011;21(9):1487–97.

17. Shinhara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, et al. Deep sequencing reveals as-yet-undiscovered small RNAs in Escherichia coli. BMC Genomics. 2011;12:428.

18. Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, et al. Small RNA regulators and the bacterial response to stress. Cold Spring Harb Symp Quant Biol. 2006;71:1–11.

19. Hoe CH, Raabe CA, Rozhdestvensky TS, Tang TH. Bacterial sRNAs: regulation in stress. Int J Med Microbiol. 2013;303(5):217–29.

20. Heer D, Sauer U. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. Microb Biotechnol. 2008;1(6):497–506.

21. Palmqvist E, Hahn-Hägerdal B. Fermentation of lignocellulosic hydrolysates. I. inhibition and detoxification. Bioresour Technol. 2000;74:17–24.

22. Roe AJ, O'Byrne C, McLaggan D, Booth IR. Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology. 2002;148(Pt 7):2215–22.

23. Royce LA, Liu P, Stebbins MJ, Hanson BC, Jarboe LR. The damaging effects of short chain fatty acids on Escherichia coli membranes. Appl Microbiol Biotechnol. 2013;97(18):8317–27.

24. Rutherford BJ, Dahl RH, Price RE, Szmidt HL, Benke PI, Mukhopadhyay A, et al. Functional genomic study of exogenous n-butanol stress in Escherichia coli. Appl Environ Microbiol. 2010;76(6):1935–45.

25. Werpy T, Petersen G. Top value added chemicals from biomass, Volume I: Results of screeening potential candidates from sugars and synthesis gas. Oak Ridge, TN: US Department of Energy; 2004.

26. Gomez-Lozano M, Marvig RL, Molin S, Long KS. Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa. Environ Microbiol. 2012; 14(8):2006–16.

27. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. 2013;41(Database issue):D203–213.

28. Thomason MK, Fontaine F, De Lay N, Storz G. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in Escherichia coli. Mol Microbiol. 2012;84(1):17–35.

29. Chen S, Zhang A, Blyn LB, Storz G. MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli. J Bacteriol. 2004;186(20):6689–97.

30. De Lay N, Gottesman S. The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. J Bacteriol. 2009; 191(2):461–76.

31. Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G. MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev. 2014; 28(14):1620–34.

32. Silva IJ, Ortega AD, Viegas SC, Garcia-Del Portillo F, Arraiano CM. An RpoS-dependent sRNA regulates the expression of a chaperone involved in protein folding. RNA. 2013;19(9):1253–65.

33. Cho BK, Zengler K, Qiu Y, Park YS, Knight EM, Barrett CL, et al. The transcription unit architecture of the Escherichia coli genome. Nat Biotechnol. 2009;27(11): 1043–9.

34. Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, et al. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. J Bacteriol. 2015;197(1):18–28.

35. Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, et al. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. MBio. 2014;5(4):e01442–01414.

36. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. 2012;31(20):4005–19.

37. Guillier M, Gottesman S, Storz G. Modulating the outer membrane with small RNAs. Genes Dev. 2006;20(17):2338–48.

38. Balbontin R, Fiorini F, Figueroa-Bossi N, Casadesus J, Bossi L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in Salmonella enterica. Mol Microbiol. 2010;78(2):380–94.

39. Johansen J, Rasmussen AA, Overgaard M, Valentin-Hansen P. Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. J Mol Biol. 2006;364(1):1–8.

40. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J. Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA. Proc Natl Acad Sci U S A. 2010;107(47):20435–40.

41. Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, Vogel J. SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. Mol Microbiol. 2006;62(6):1674–88.

42. Thompson KM, Rhodius VA, Gottesman S. SigmaE regulates and is regulated by a small RNA in Escherichia coli. J Bacteriol. 2007;189(11):4243–56.

43. Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. Proc Natl Acad Sci U S A. 2011;108(31): 12875–80.

44. Holmqvist E, Unoson C, Reimegard J, Wagner EG. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. Mol Microbiol. 2012;84(3):414–27.

45. Guillier M, Gottesman S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. Mol Microbiol. 2006;59(1):231–47.

46. Miyakoshi M, Chao Y, Vogel J. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. Curr Opin Microbiol. 2015;24:132–9.

47. Majdalani N, Hernandez D, Gottesman S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol Microbiol. 2002;46(3):813–26.

48. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA. Positive regulation by small RNAs and the role of Hfq. Proc Natl Acad Sci U S A. 2010;107(21):9602–7.

49. Bak G, Han K, Kim D, Lee Y. Roles of rpoS-activating small RNAs in pathways leading to acid resistance of Escherichia coli. Microbiologyopen. 2014;3(1): 15–28.

50. Majdalani N, Chen S, Murrow J, St John K, Gottesman S. Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol Microbiol. 2001; 39(5):1382–94.

51. Gaida SM, Al-Hinai MA, Indurthi DC, Nicolaou SA, Papoutsakis ET. Synthetic tolerance: three noncoding small RNAs, DsrA, ArcZ and RprA, acting supra-additively against acid stress. Nucleic Acids Res. 2013;41(18):8726–37.

52. Romeo T, Vakulskas CA, Babitzke P. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. Environ Microbiol. 2013; 15(2):313–24.

53. Jorgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G. Dual function of the McaS small RNA in controlling biofilm formation. Genes Dev. 2013;27(10):1132–45.

54. Mandin P, Gottesman S. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. 2010;29(18): 3094–107.

55. Papenfort K, Said N, Welsink T, Lucchini S, Hinton JC, Vogel J. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol Microbiol. 2009;74(1):139–58.

56. Gopel Y, Khan MA, Gorke B. Menage a trois: post-transcriptional control of the key enzyme for cell envelope synthesis by a base-pairing small RNA, an RNase adaptor protein, and a small RNA mimic. RNA Biol. 2014;11(5):433–42.

57. Salvail H, Masse E. Regulating iron storage and metabolism with RNA: an overview of posttranscriptional controls of intracellular iron homeostasis. Wiley Interdiscip Rev RNA. 2012;3(1):26–36.

58. Sharma CM, Darfeuille F, Plantinga TH, Vogel J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev. 2007;21(21):2804–17.

59. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. Pervasive post-transcriptional control of genes involved in amino acid

Rau *et al. BMC Genomics* (2015) 16:1051

Page 16 of 16

metabolism by the Hfq-dependent GcvB small RNA. Mol Microbiol. 2011; 81(5):1144–65.

60. Miyakoshi M, Chao Y, Vogel J. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. EMBO J. 2015; 34(11):1478–92.

61. Mika F, Hengge R. Small Regulatory RNAs in the Control of Motility and Biofilm Formation in E. coli and Salmonella. Int J Mol Sci. 2013;14(3):4560–79.

62. Mika F, Hengge R. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of Escherichia coli. RNA Biol. 2014;11(5):494–507.

63. De Lay N, Gottesman S. A complex network of small non-coding RNAs regulate motility in Escherichia coli. Mol Microbiol. 2012;86(3):524–38.

64. Holmqvist E, Reimegard J, Sterk M, Grantcharova N, Romling U, Wagner EG. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. EMBO J. 2010;29(11):1840–50.

65. Jorgensen MG, Nielsen JS, Boysen A, Franch T, Moller-Jensen J, Valentin-Hansen P. Small regulatory RNAs control the multi-cellular adhesive lifestyle of Escherichia coli. Mol Microbiol. 2012;84(1):36–50.

66. Mika F, Busse S, Possling A, Berkholz J, Tschowri N, Sommerfeldt N, et al. Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in Escherichia coli. Mol Microbiol. 2012;84(1):51–65.

67. Miller JH. Experiments in molecular genetics. Cold Spring Harbor, NY: Cold Spring Laboratory; 1972.

68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16): 2078–9.

70. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.

71. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.

72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

73. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. Bioinformatics. 2005;21(4):537–9.

74. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In: Metagenomics and its applications in agriculture, biomedicine and environmental studies. Li RW, editor. Hauppauge, NY: Nova Science Publishers; 2011: 61–78.

75. Gautheret D, Lambert A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol. 2001;313(5):1003–11.

76. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 2001;29(22):4724–35.

77. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000; 97(12):6640–5.