Genome-scale Analysis of *Escherichia coli* FNR Reveals Complex Features of Transcription Factor Binding

Kevin S. Myers, Huihuang Yan, Irene M. Ong, Dongjun Chung, Kun Liang, Frances Tran, Su¨ ndu¨ z Keles¸, Robert Landick, Patricia J. Kiley

Investigation

- To investigate the roles of TF action and chromosome structure in a prototypical bacterial regulon and study the regulon of the anaerobic TF FNR.

- Regulon--

  A regulon is a group of genes that are regulated as a unit, generally controlled by the same regulatory gene that expresses a protein acting as a repressor or activator.

Background and Introduction

- FNR

FNR is a well-studied global regulator of anaerobiosis, which is widely conserved across bacteria.

- transcription factors (TFs)

Regulation of transcription initiation by transcription factors (TFs) is a key step in controlling gene expression in all domains of life.

Nucleoid-associated proteins (NAPs), such as e NAPs H-NS, IHF and Fis.

Regulated promoters are controlled by multiple TFs .

Interaction between FNR and TFs

- FNR

FNR is widely conserved throughout the bacterial domain, where it evolved to allow facultative anaerobes to adjust to O2 deprivation under anaerobic conditions.

FNR controls expression of a large number of genes under anaerobic growth conditions.

From studies, FNR binding sites can have only a partial match to the consensus sequence of TTGATnnnnATCAA, and be located at variable positions within promoter regions.

FNR has either a positive or negative affect on transcription controlling promoter as depressed or activated.

- Transcription factors (TFs)

Many FNR regulated promoters are controlled by multiple TFs (for example CRP, NarL, NarP, and NAPs, which can have either positive or negative effects on FNR function depending on the promoter architecture.

Finding

- FNR occupancy at many target sites is strongly influenced by TF nucleoid-associated proteins (NAPs) that restrict access to many FNR binding sites.

- In cells lacking H-NS and its paralog StpA showed increased FNR occupancy at sites bound by H-NS in WT strains, indicating that large regions of the genome are not readily accessible for FNR binding.

- Genome-wide FNR occupancy did not correlate with the match to consensus at binding sites, suggesting that significant variation in ChIP signal was attributable to cross-linking or immunoprecipitation efficiency rather than differences in binding affinities for FNR sites.

Materials and Methods

Strains and growth conditions

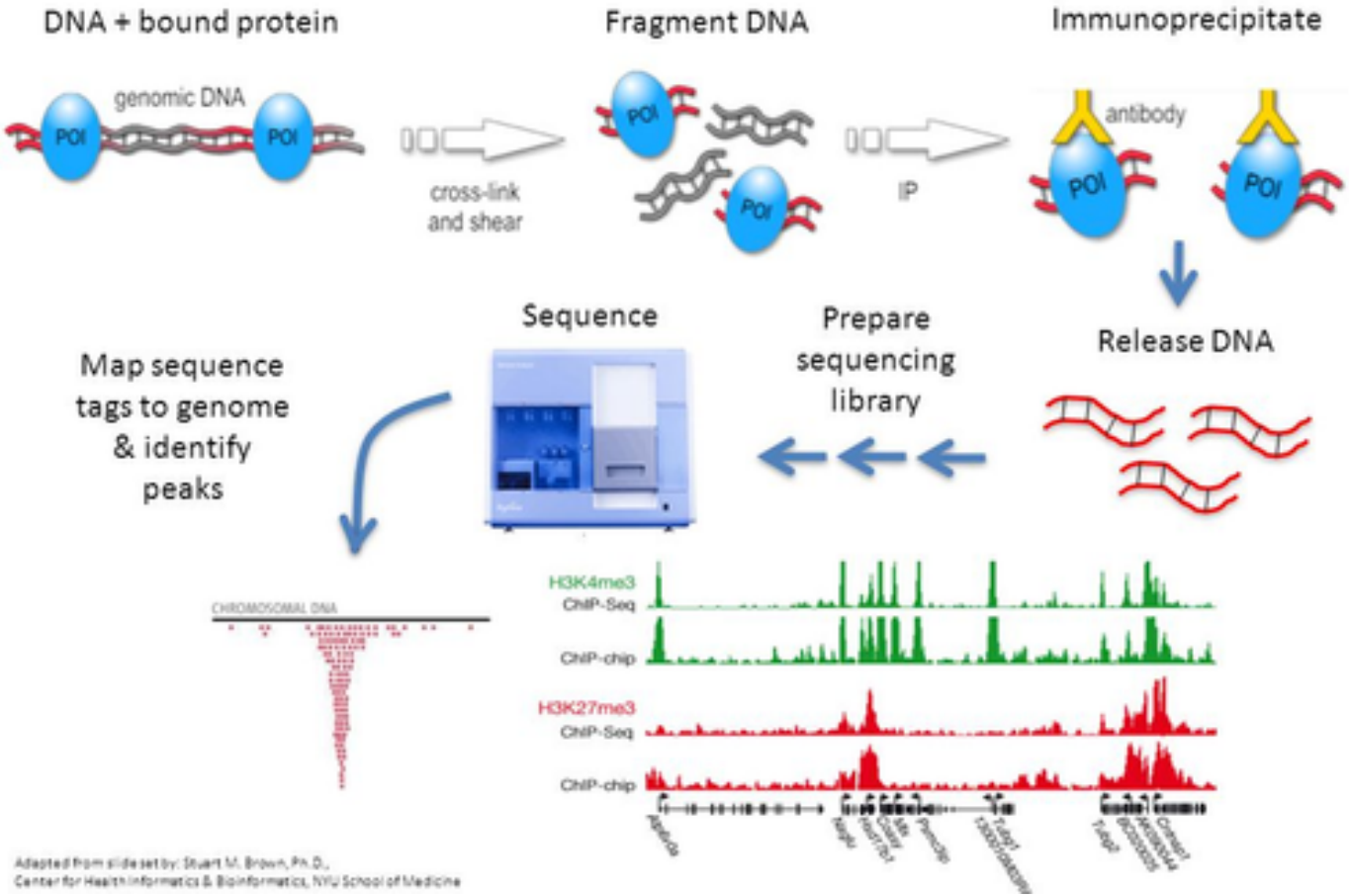
RNA isolation and DNA fragment purification

Whole genome transcriptomic microarray analysis and High-throughput RNA sequencing (RNA-seq) analysis

Chromatin immunoprecipitation followed by hybridization to a microarray chip

or high-throughput sequencing

ChIP-seq overview



Materials and Methods

- To systematically investigate FNR binding genome-wide, performed chromatin immunoprecipitation followed by micro-array hybridization (ChIP-chip) and high-throughput sequencing (ChIP-seq) for WT FNR from *E Coli.*

- Computational and bioinformatic analyses were used to refine a FNR position weight matrix (PWM) to determine the relationship between ChIP-seq/ChIP-chip enrichment and match to the PWM, and to identify predicted FNR binding sites not detected by ChIP-seq.

- To examine the subset of high quality predicted FNR binding sites lacking a FNR ChIPseq peak, obtaine and analyze aerobic and/or anaerobic ChIP-chip data for NAPs H-NS and IHF along with analysis of previously published aerobic ChIP-seq data.

Materials and Methods

- The effect of H-NS on FNR occupancy was examined directly using ChIP-chip analysis of FNR as well as on O2 dependent changes in expression in the absence of H-NS and its paralog StpA

- After identifying FNR binding sites genome-wide, performed whole genome transcription profiling experiments using expression microarrays and highthroughput RNA sequencing (RNA-seq) to compare a WT and ∆fnr strain grown in the same medium used for the DNA binding studies.

- The transcriptional impact of FNR binding genome-wide was investigated by correlating the occupancy data with the transcriptomic data to determine which binding events led to changes in transcription, to identify the direct and indirect regulons of FNR, and to define categories of FNR regulatory mechanisms.

- The aerobic and anaerobic ChIP-chip and ChIP-seq distributions of the s70 and ß subunits of RNAP throughout the genome were analyzed to determine the role of O2 and FNR regulation on RNAP occupancy and transcription

ChIP-Seq analysis

1. Sequences preparation

Experiment sequence fastq file

Platforms (1) GPL16109 Illumina Genome Analyzer IIx (Escherichia coli str. K-12 substr. MG1655star)

Samples (9)
More...

GSM1010219 FNR IP ChIP-seq Anaerobic A

GSM1010220 FNR IP ChIP-seq Anaerobic B

GSM1010221 σ70 IP ChIP-seq Aerobic A

Platform ID GPL16109

Series (2)

GSE41187 Genome-wide analysis of FNR and σ70 in E. coli under aerobic and anaerobic growth conditions.

GSE41195 Genome-scale Analysis of E. coli FNR Reveals Complex Features of Transcription Factor Binding

Relations

SRA SRX189773

BioSample SAMN01731116

Download report: JSON TSV

Download Files as ZIP

Download selected files

Download All

Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Scientific Name	Generated FASTQ files: FTP
PRJNA176149	SAMN01731116	SRX189773	SRR576933	879462	Escherichia coli str. K-12 substr. MG1655star	SRR576933.fastq.gz

Control sequence fastq file

Samples (9)
Less...

GSM1010219 FNR IP ChIP-seq Anaerobic A
GSM1010220 FNR IP ChIP-seq Anaerobic B
GSM1010221 σ70 IP ChIP-seq Aerobic A
GSM1010222 σ70 IP ChIP-seq Anaerobic A
GSM1010223 aerobic INPUT DNA
GSM1010224 anaerobic INPUT DNA

Platform ID GPL16109
Series (2) GSE41187 Genome-wide analysis of FNR and σ70 in E. coli under aerobic and anaerobic growth conditions.
GSE41195 Genome-scale Analysis of E. coli FNR Reveals Complex Features of Transcription Factor Binding

Relations
SRA SRX189778
BioSample SAMN01731121

Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Scientific Name	Generated FASTQ files: FTP
PRJNA176149	SAMN01731121	SRX189778	SRR576938	879462	Escherichia coli str. K-12 substr. MG1655star	☐ SRR576938.fastq.gz

Reference genome sequence fasta file

Escherichia coli str. K-12 substr. MG1655

Download sequences in FASTA format for **genome**, **protein**

Download genome annotation in **GFF**, **GenBank** or **tabular** format

BLAST against Escherichia coli **genome**, **protein**

All 28128 genomes for species:

Browse the **list**

Download sequence and annotation from **RefSeq** or **GenBank**

NEW Try **NCBI Datasets** - a new way to download genome sequence and annotation we're testing in NCBI Labs

Download sequence and annotation from **RefSeq** or **GenBank**

NEW Try **NCBI Datasets** - a new way to download genome sequence and annotation we're testing in NCBI Labs

2. Prepare the index file

3. Quality Control of the reads

4. Mapping the reads with Bowtie

chingyao@sclogin2:~/chipseqanalysis

```
#!/bin/bash

#SBATCH --workdir=/home/c/chingyao/chipseqanalysis/
#SBATCH --job-name=ChIPseqMapping
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --mem=10000
#SBATCH -t 00:10:00
#SBATCH -o run.out
#SBATCH -e run.err
#SBATCH --mail-user=chingyao@mail.usf.edu
#SBATCH --mail-type=BEGIN
#SBATCH --mail-type=END
#SBATCH --mail-type=FAIL

module purge
module add apps/bowtie/2.3.2
module add apps/samtools/1.3.1
bowtie2 -x refgenome/GCF_000005845.2_ASM584v2_genomic -3 1 -q SRR576933/SRR576933.fastq.gz -S SRR576933.sam 2> SRR576933.out
bowtie2 -x refgenome/GCF_000005845.2_ASM584v2_genomic -3 1 -q SRR576938/SRR576938.fastq.gz -S SRR576938.sam 2> SRR576938.out
samtools view -u SRR576933.sam | samtools sort -o SRR576933.bam
samtools view -u SRR576938.sam | samtools sort -o SRR576938.bam
samtools index SRR576933.bam
samtools index SRR576938.bam
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"mapping_Bowtie.sh" 24L, 852C                                                    24,1          All
```

5. Peak calling with MACS2

chingyao@sclogin2:~/macs2

```
#!/bin/bash

#SBATCH --workdir=/home/c/chingyao/chipseqanalysis/
#SBATCH --job-name=PeakCalling
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --mem=10000
#SBATCH -t 00:10:00
#SBATCH -o run.out
#SBATCH -e run.err
#SBATCH --mail-user=chingyao@usf.edu
#SBATCH --mail-type=BEGIN
#SBATCH --mail-type=END
#SBATCH --mail-type=FAIL

module purge
module add apps/miniconda/3.6.1-intel
conda activate macs2
unset PYTHONPATH
macs2 callpeak -t SRR576933.sam -c SRR576938.sam -n MACSpeaks -q 0.05 --gsize 4639675 --keep-dup 1 --nomodel --extsize 400
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"peakCalling.sh" 20L, 545C                                                    3,1           All
```

6. Result

chingyao@sclogin2:~/chipseqanalysis

```
drwxr-x---  8 chingyao usfuser  6 Feb 23 15:23 Genomics_Training
drwxr-x---  7 chingyao usfuser  5 Feb 23 14:02 Genomics_Training_Ref
drwxr-x---  5 chingyao usfuser  3 Feb 23 13:46 Process_RNA-seq_Data__Project
-rw-r-----  1 chingyao usfuser 23 Feb 19 11:03 aaa
-rw-r-----  1 chingyao usfuser  8 Feb 19 10:40 aaa~
drwxr-x---  6 chingyao usfuser 25 Feb 28 21:20 chipseqanalysis
drwxr-x---  2 chingyao usfuser  1 Feb 19 10:28 do
-rw-r-----  1 chingyao usfuser 11 Feb 19 10:26 doing.txt
-rw-r-----  1 chingyao usfuser  1 Feb 19 10:18 doo
drwxr-x---  3 chingyao usfuser 11 Feb  6 19:36 genomeTrain
drwxr-x---  2 chingyao usfuser  2 Jan 21 17:20 human_GRCH38_genomeFile
drwxr-x---  4 chingyao usfuser 21 Feb 28 21:25 macs2
drwxr-x--- 27 chingyao usfuser 26 Feb 23 14:02 miniconda2
drwxr-x---  2 chingyao usfuser  1 Feb  6 20:08 test
drwxr-x---  2 chingyao usfuser  3 Jan 31 20:13 test_GSE77565
[chingyao@sclogin2 ~]$ cd chipseqanalysis/
[chingyao@sclogin2 chipseqanalysis]$ ls -lh
total 4.6G
-rw-r----- 1 chingyao usfuser  16K Feb 27 20:51 MACSpeaks_peaks.narrowPeak
-rw-r----- 1 chingyao usfuser  19K Feb 27 20:51 MACSpeaks_peaks.xls
-rw-r----- 1 chingyao usfuser  11K Feb 27 20:51 MACSpeaks_summits.bed
drwxr-x--- 2 chingyao usfuser    3 Feb 27 18:44 SRR576933
-rw-r----- 1 chingyao usfuser  91M Feb 27 19:40 SRR576933.bam
-rw-r----- 1 chingyao usfuser  14K Feb 27 19:41 SRR576933.bam.bai
-rw-r----- 1 chingyao usfuser  567 Feb 27 19:37 SRR576933.out
-rw-r----- 1 chingyao usfuser 550M Feb 27 19:37 SRR576933.sam
-rw-r----- 1 chingyao usfuser 119M Feb 27 18:01 SRR576933.zip
drwxr-x--- 2 chingyao usfuser    3 Feb  9 13:48 SRR576934
-rw-r----- 1 chingyao usfuser 255M Feb 11 17:00 SRR576934.bam
-rw-r----- 1 chingyao usfuser  14K Feb 11 17:02 SRR576934.bam.bai
-rw-r----- 1 chingyao usfuser  565 Feb 11 16:56 SRR576934.out
-rw-r----- 1 chingyao usfuser 1.7G Feb 11 16:56 SRR576934.sam
-rw-r----- 1 chingyao usfuser 339M Feb  9 13:09 SRR576934.zip
drwxr-x--- 2 chingyao usfuser    3 Feb  9 13:52 SRR576938
-rw-r----- 1 chingyao usfuser 190M Feb 27 19:41 SRR576938.bam
-rw-r----- 1 chingyao usfuser  14K Feb 27 19:41 SRR576938.bam.bai
-rw-r----- 1 chingyao usfuser  564 Feb 27 19:39 SRR576938.out
-rw-r----- 1 chingyao usfuser 1.2G Feb 27 19:39 SRR576938.sam
-rw-r----- 1 chingyao usfuser 243M Feb  9 13:09 SRR576938.zip
-rw-r----- 1 chingyao usfuser  852 Feb 27 19:34 mapping_Bowtie.sh
drwxr-x--- 2 chingyao usfuser   13 Feb 27 18:30 refgenome
-rw-r----- 1 chingyao usfuser 4.1K Feb 27 20:51 run.err
-rw-r----- 1 chingyao usfuser    0 Feb  9 15:38 run.out
[chingyao@sclogin2 chipseqanalysis]$
```

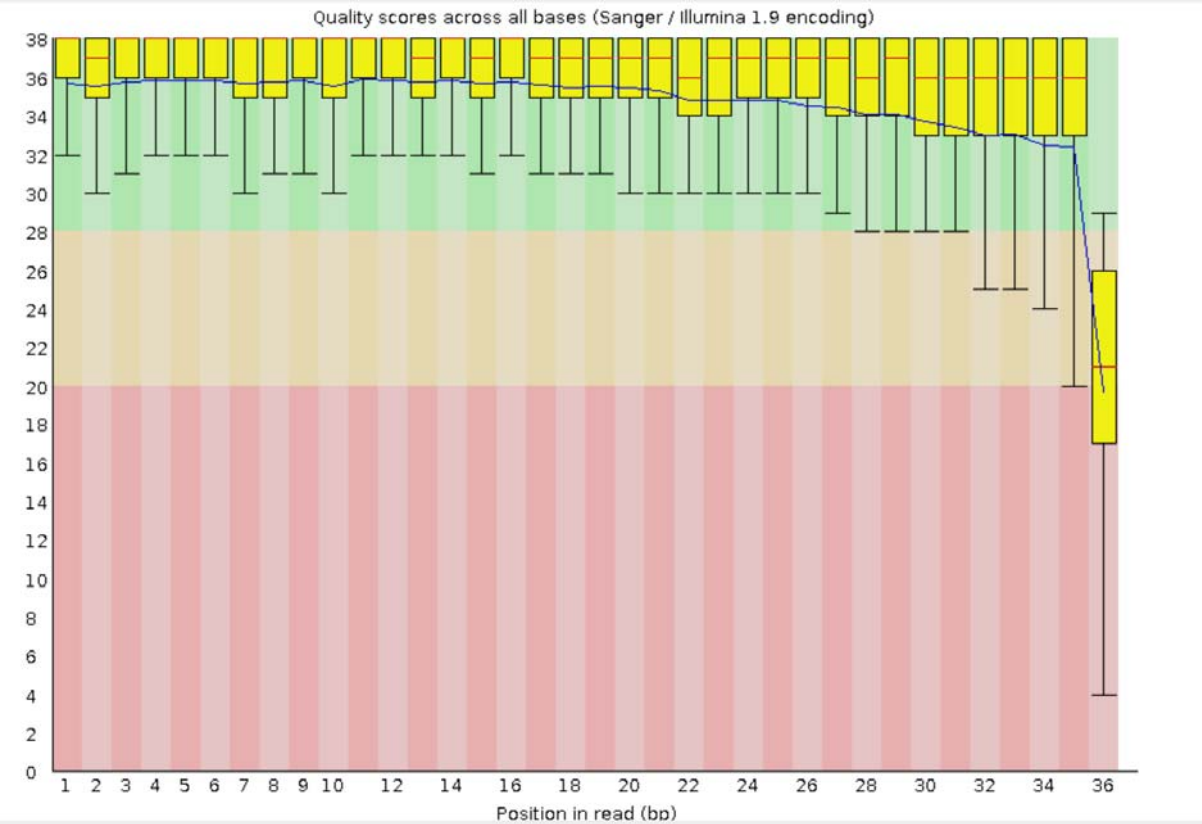
Quality control treatment sample

There are 3,603,544 reads of 36bp in the file.
The overall quality of reads is good.

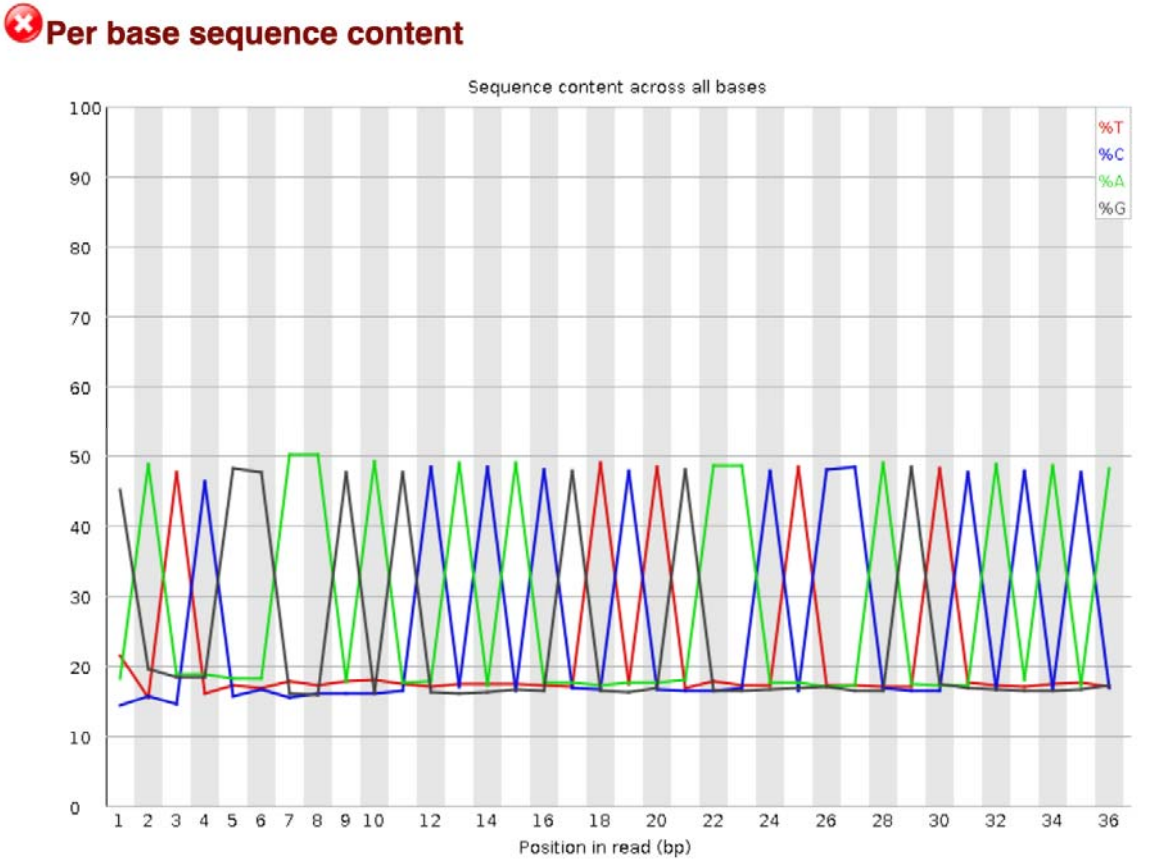
Basic Statistics

Measure	Value
Filename	SRR576933.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3603544
Sequences flagged as poor quality	0
Sequence length	36
%GC	49

Quality control treatment sample

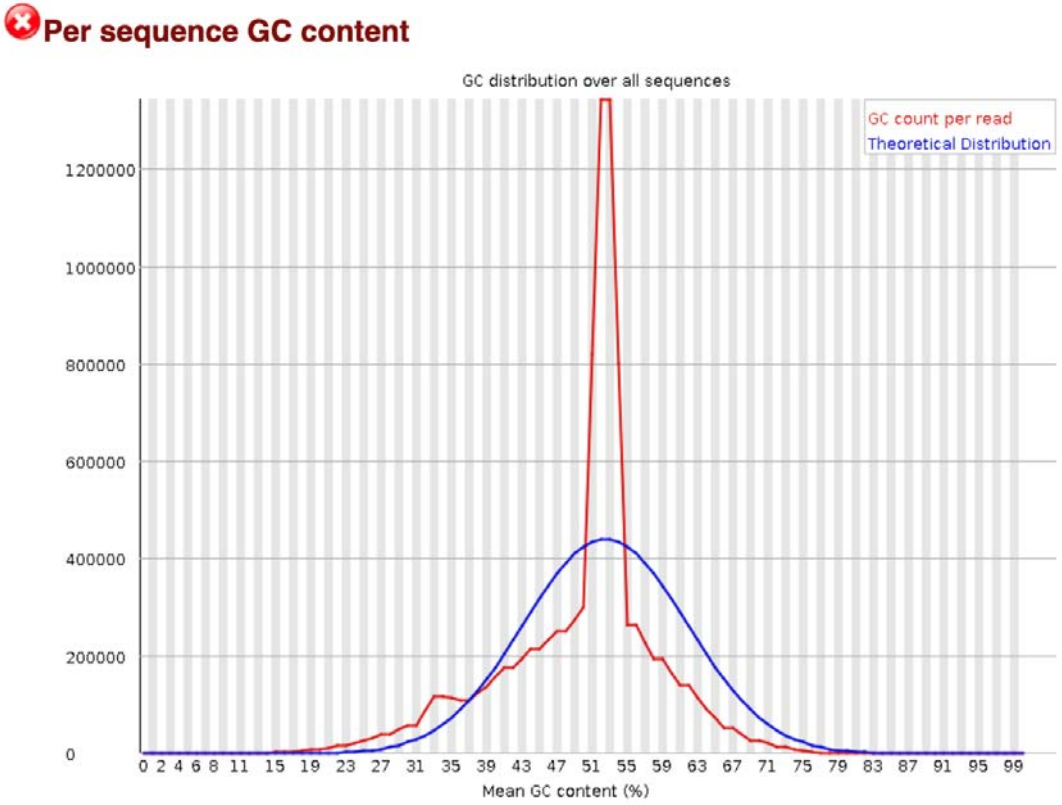


Quality control treatment sample



The base sequence content is not stable. The reason for these bias are the sequence we detected were the sites of FNR binding and they have common patterns.

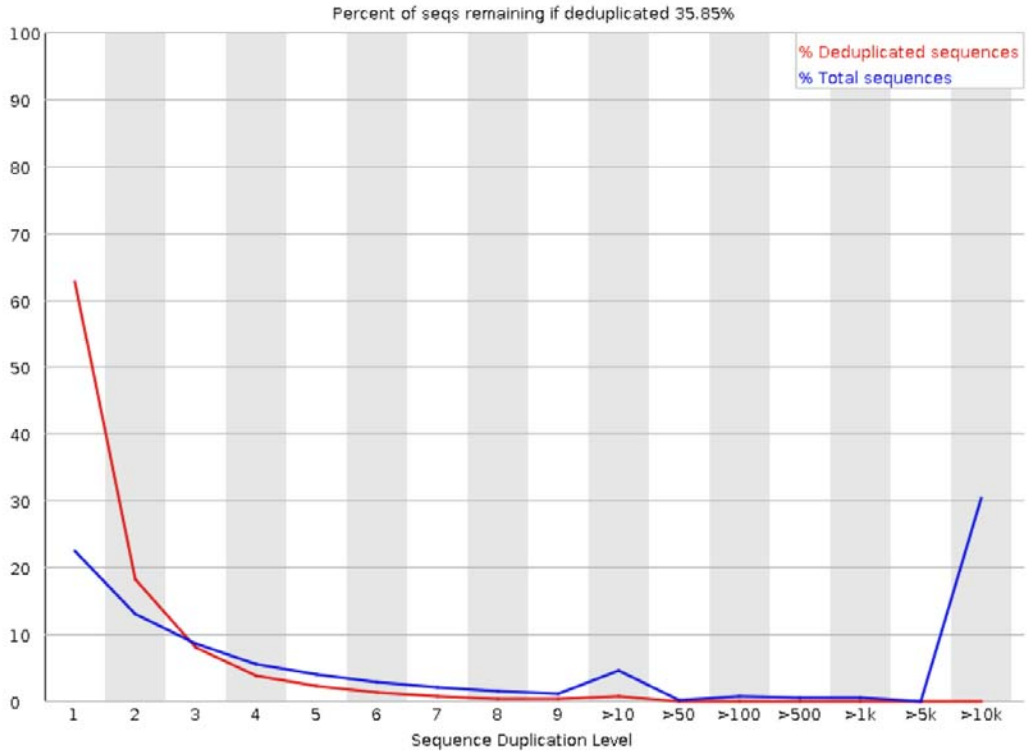
Quality control treatment sample



The GC content dose not obey normal distribution. The kurtosis is too high.

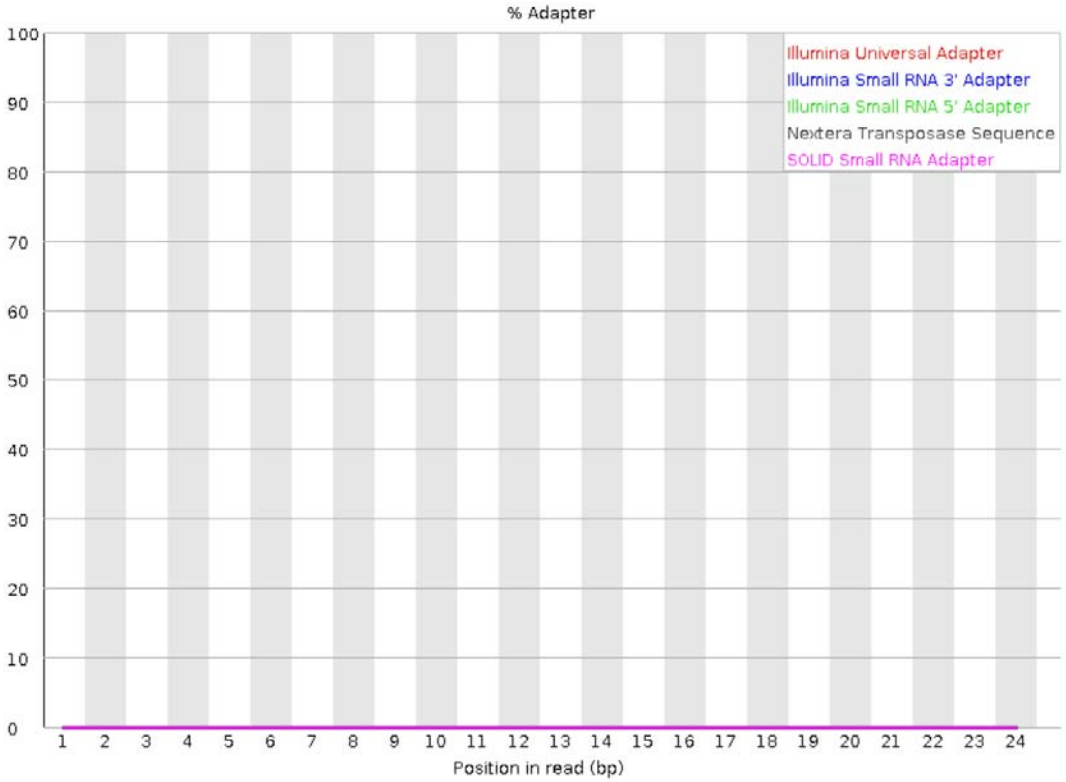
Quality control treatment sample

Sequence Duplication Levels



Quality control treatment sample

Adapter Content



Quality control control sample

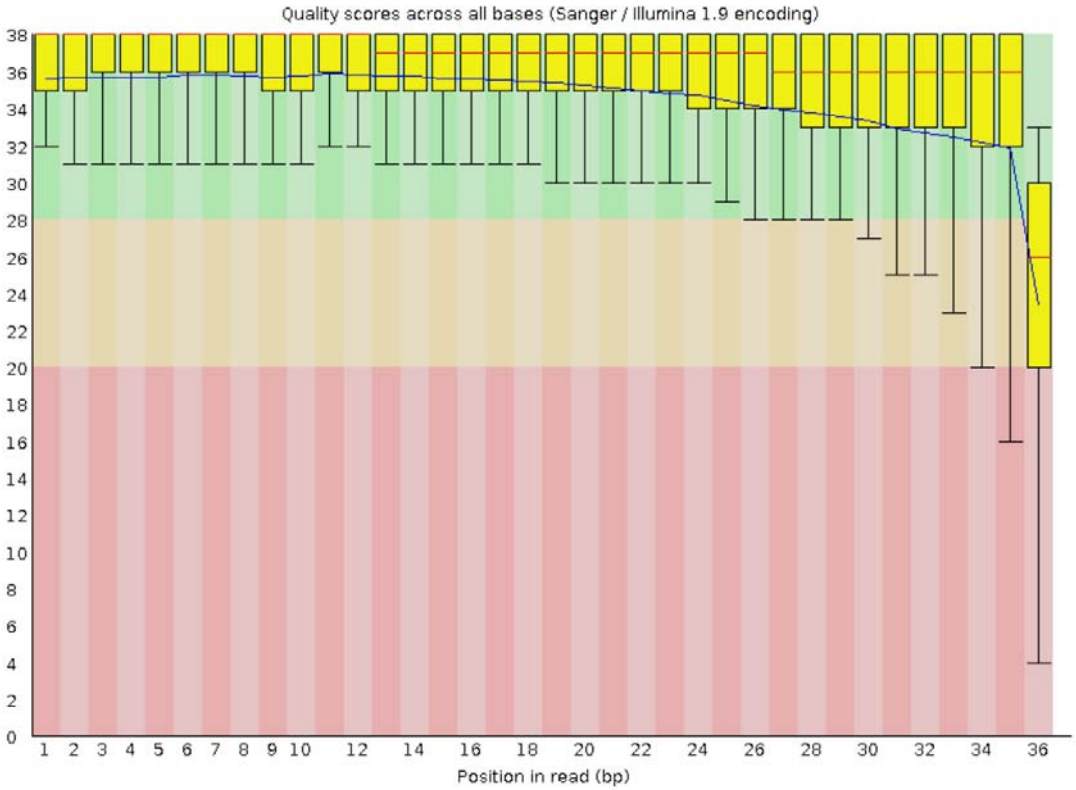
There are 6,717,074 reads of 36bp in the control file. The overall quality of control sample file is good.

Basic Statistics

Measure	Value
Filename	SRR576938.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	6717074
Sequences flagged as poor quality	0
Sequence length	36
%GC	49

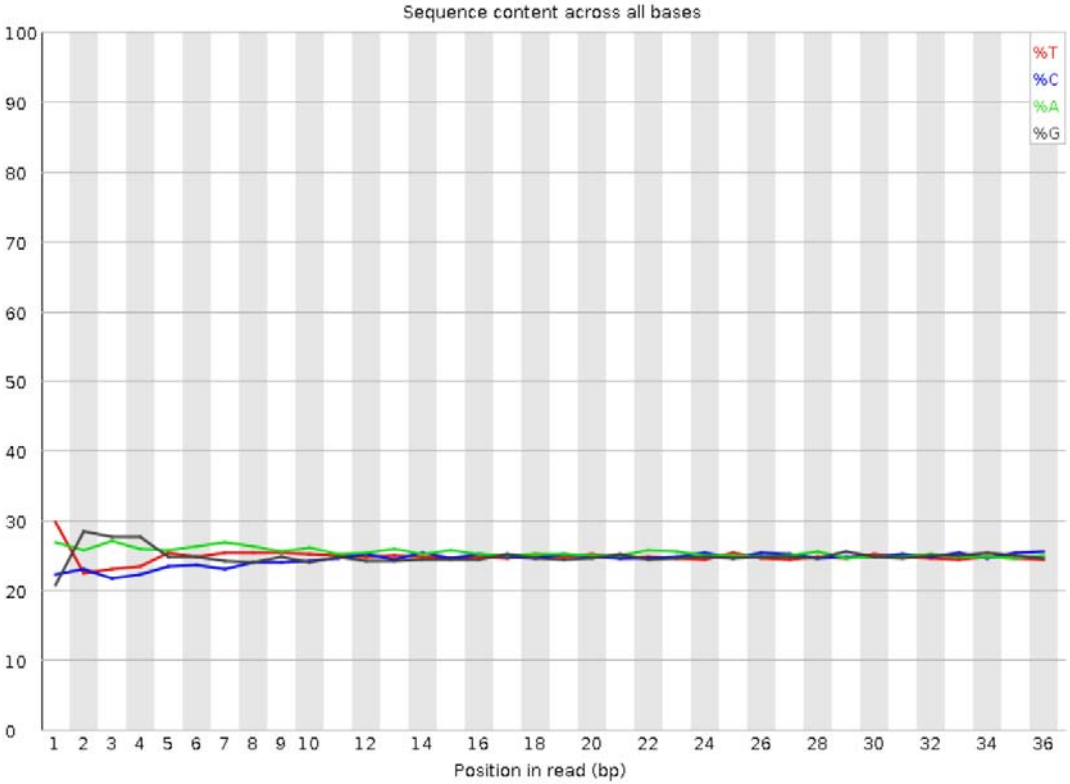
Quality control control sample

Per base sequence quality



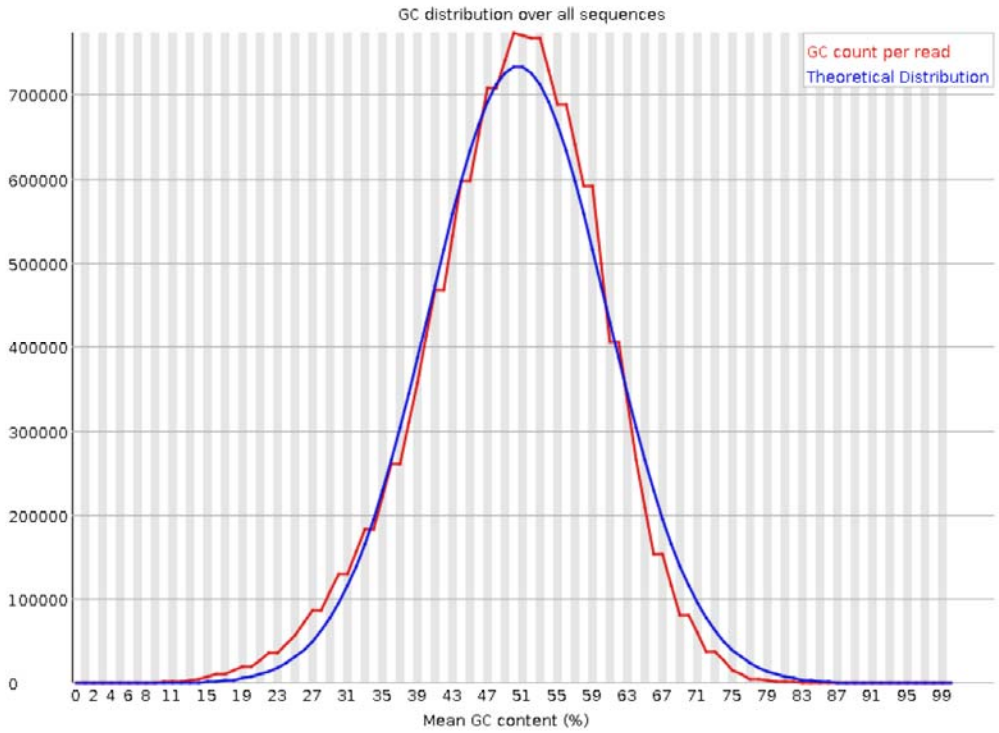
Quality control control sample

Per base sequence content



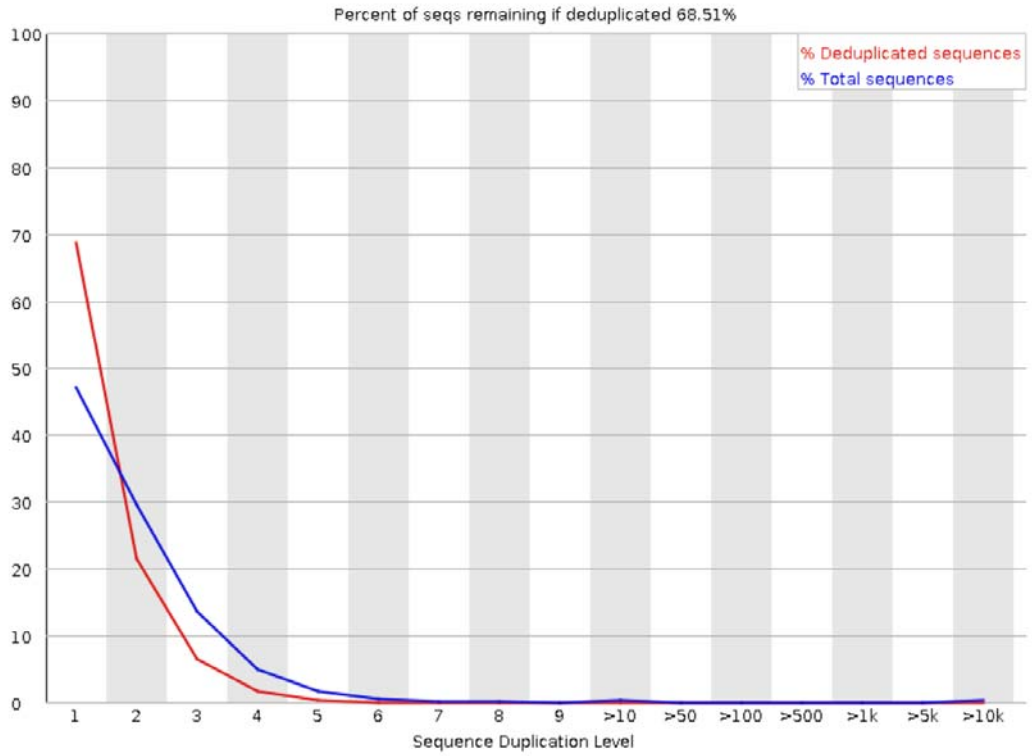
Quality control control sample

Per sequence GC content



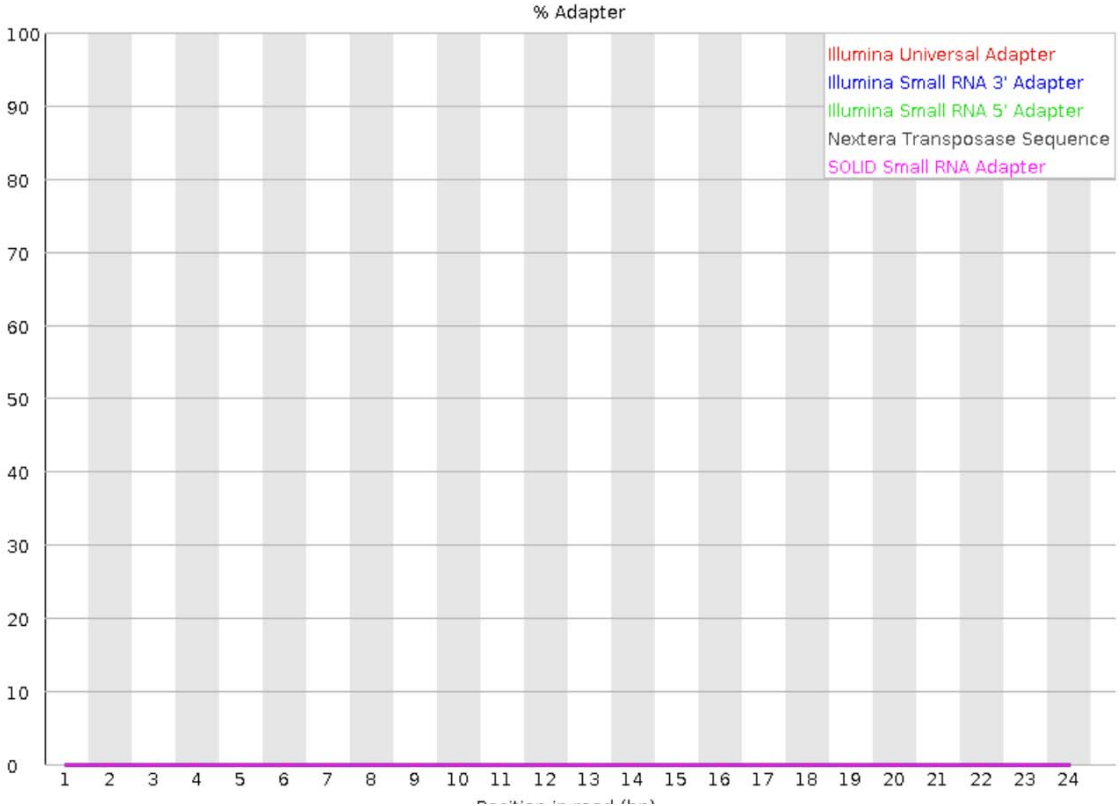
Quality control control sample

Sequence Duplication Levels



Quality control control sample

Adapter Content



Mapping result

- TF binding sites were mapped genome-wide in E. coli K-12 MG1655 using ChIP-chip and/or ChIP-seq for FNR under anaerobic growth conditions.

FNR sample

SRR576933.out

```
3603544 reads; of these:
  3603544 (100.00%) were unpaired; of these:
    1220111 (33.86%) aligned 0 times
    2280575 (63.29%) aligned exactly 1 time
    102858 (2.85%) aligned >1 times
66.14% overall alignment rate
```

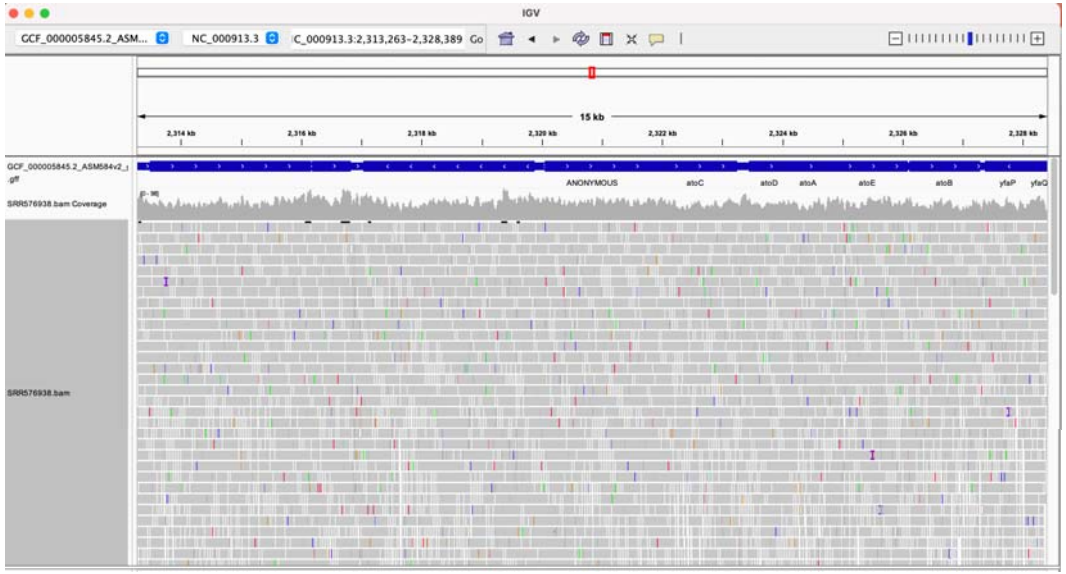
Control sample

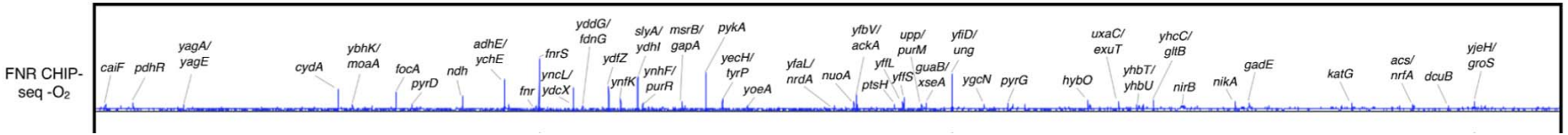
SRR576938.out

```
6717074 reads; of these:
  6717074 (100.00%) were unpaired; of these:
    68441 (1.02%) aligned 0 times
    6433269 (95.77%) aligned exactly 1 time
    215364 (3.21%) aligned >1 times
98.98% overall alignment rate
```

Mapping result

- TF binding sites were mapped genome-wide in E. coli K-12 MG1655 using ChIP-chip and/or ChIP-seq for FNR under aerobic or anaerobic growth conditions, as indicated.





Call peaks

Parameters

```
# This file is generated by MACS version 2.2.6
# Command line: callpeak -t SRR576933.sam -c SRR576938.sam -n MACSpeaks -q 0.05 --gsize 4639675 --keep-dup 1 --nomodel --extsize 400
# ARGUMENTS LIST:
# name = MACSpeaks
# format = AUTO
# ChIP-seq file = ['SRR576933.sam']
# control file = ['SRR576938.sam']
# effective genome size = 4.64e+06
# band width = 300
# model fold = [5, 50]
# qvalue cutoff = 5.00e-02
# The maximum gap between significant sites is assigned as the read length/tag size.
# The minimum length of peaks is assigned as the predicted fragment length "d".
# Larger dataset will be scaled towards smaller dataset.
# Range for calculating regional lambda is: 1000 bps and 10000 bps
# Broad region calling is off
# Paired-End mode is off

# tag size is determined as 35 bps
# total tags in treatment: 2383433
# tags after filtering in treatment: 1166926
# maximum duplicate tags at the same position in treatment = 1
# Redundant rate in treatment: 0.51
# total tags in control: 6648633
# tags after filtering in control: 4400464
# maximum duplicate tags at the same position in control = 1
# Redundant rate in control: 0.34
# d = 400
```

191 MACS peaks were called in analysis.

NC_000913.3	4107189	4107756	568	4107422	186	5.92461	1.43395	4.39349	MACSpeaks_peak_168
NC_000913.3	4133370	4134033	664	4133713	286	32.77546	2.21881	30.90141	MACSpeaks_peak_169
NC_000913.3	4166134	4172106	5973	4168288	374	66.60288	2.83521	64.08263	MACSpeaks_peak_170
NC_000913.3	4175220	4176117	898	4175628	310	40.39673	2.36747	38.42407	MACSpeaks_peak_171
NC_000913.3	4176518	4177519	1002	4177108	193	9.16145	1.5884	7.54299	MACSpeaks_peak_172
NC_000913.3	4178114	4178762	649	4178391	222	15.12597	1.78628	13.41946	MACSpeaks_peak_173
NC_000913.3	4179064	4180559	1496	4179864	276	24.82382	1.99431	23.01505	MACSpeaks_peak_174
NC_000913.3	4180743	4181245	503	4181022	225	11.16394	1.61725	9.50735	MACSpeaks_peak_175
NC_000913.3	4207700	4213395	5696	4208219	359	58.68242	2.68945	56.30525	MACSpeaks_peak_176
NC_000913.3	4287293	4287953	661	4287661	267	26.24872	2.06769	24.4299	MACSpeaks_peak_177
NC_000913.3	4295696	4296403	708	4296092	205	5.95549	1.41043	4.42308	MACSpeaks_peak_178
NC_000913.3	4325959	4326391	433	4326279	189	7.74185	1.52519	6.1567	MACSpeaks_peak_179
NC_000913.3	4348897	4349483	587	4349197	265	23.00013	1.96386	21.20836	MACSpeaks_peak_180
NC_000913.3	4370179	4370885	707	4370513	422	106.20103	3.62289	103.35722	MACSpeaks_peak_181
NC_000913.3	4382059	4382733	675	4382404	241	21.97788	1.99789	20.19629	MACSpeaks_peak_182
NC_000913.3	4392074	4392933	860	4392497	298	42.11026	2.46739	40.11016	MACSpeaks_peak_183
NC_000913.3	4404373	4405384	1012	4404771	310	42.46044	2.42926	40.45461	MACSpeaks_peak_184
NC_000913.3	4462532	4463238	707	4462891	279	40.44931	2.49751	38.47591	MACSpeaks_peak_185
NC_000913.3	4568445	4569015	571	4568777	232	19.50879	1.93205	17.75381	MACSpeaks_peak_186
NC_000913.3	4586614	4587056	443	4586812	191	8.94204	1.58164	7.32829	MACSpeaks_peak_187
NC_000913.3	4602573	4603053	481	4602894	202	10.09154	1.61442	8.45446	MACSpeaks_peak_188
NC_000913.3	4605595	4606734	1140	4606132	280	33.49582	2.26338	31.61395	MACSpeaks_peak_189
NC_000913.3	4616927	4617416	490	4617111	185	9.17807	1.60588	7.55934	MACSpeaks_peak_190
NC_000913.3	4640168	4640944	777	4640560	331	51.90048	2.62044	49.67869	MACSpeaks_peak_191

Discussion

Future direction

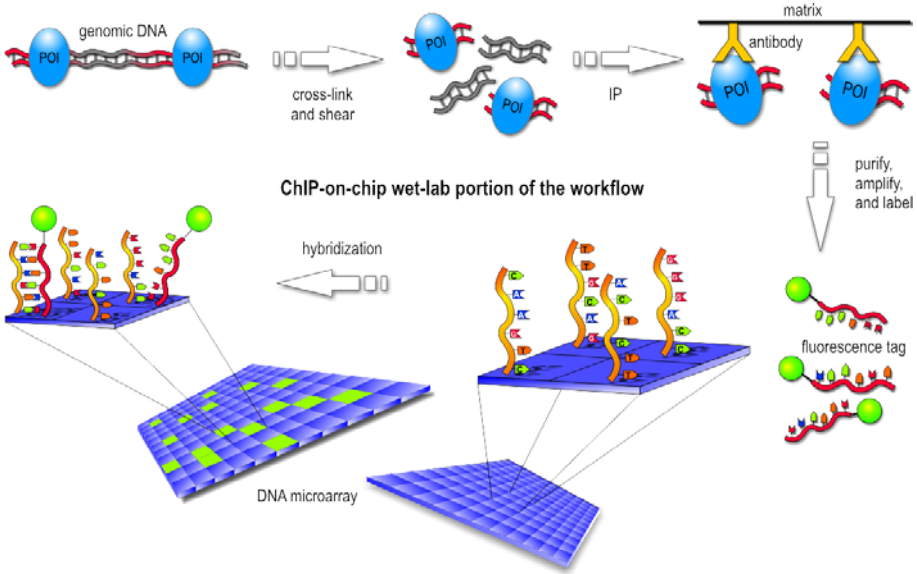
- 1. Find other potential binding sites;
- 2. Predict the binding sites;
- 3. Study the FNR function by examine the function of binding sites;

Chip-on-chip

We identified 191 MACS peaks in the procedure, but is there any other binding sites that were not identified in the analysis?

Computational and bioinformatic analyses were used to refine a FNR position weight matrix (PWM). The PWM was used to determine the relationship between ChIP-seq/ChIP-chip enrichment and match to the PWM, and to identify predicted FNR binding sites not detected by ChIP-seq.

It allows the identification of the cistrome, the sum of binding sites, for DNA-binding proteins on a genome-wide basis.



ChIP-seq generally produces profiles with a better signal-to-noise ratio, and allows detection of more peaks and narrower peaks.

Chip-on-chip

Comparison of ChIP-chip and ChIP-Seq

	ChIP-chip	ChIP-Seq
Resolution	Array-specific, generally 30–100bp	Single nucleotide
Coverage	Limited by sequences on the array; repetitive regions usually masked out	Limited only by alignability of reads to the genome; increases with read length; many repetitive regions can be covered
Cost	$400–$800 per array (1–6 million probes); multiple arrays may be needed for large genomes	$1000–$2000 per Illumina lane (6–15 million readsprior to alignment)
Source of platform noise	Cross-hybridization between probes and non-specific targets	Some GC-biasmay be present
Experimental design	Single- or double-channel, depending on platform	Single channel
Cost-effective cases	Large fraction enriched (broad binding), profiling of selected regions	Small fraction enriched (sharp binding), large genomes
Required amount of ChIP DNA	High (few μg)	Low (10–50 ng)
Dynamic range	Lower detection limit, saturation at high signal	Not limited
Amplification	More required	Less required; single molecule sequencing without amplification is available
Multiplexing	Not possible	Possible

Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009;10(10):669-680. doi:10.1038/nrg2641

Position weight matrix (PWM)

An example of PWM

A

Base probability matrix

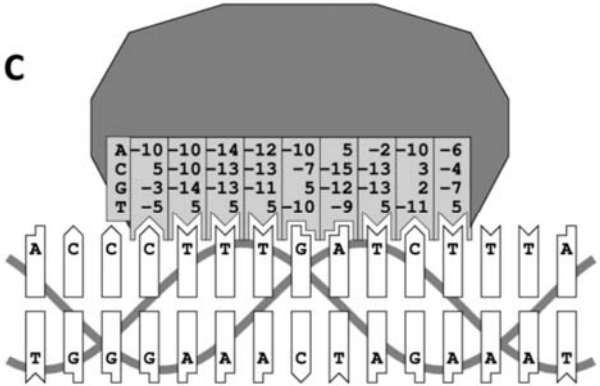
Pos.	1	2	3	4	5	6	7	8	9
A	0.025	0.029	0.012	0.019	0.028	0.935	0.162	0.027	0.063
C	0.775	0.029	0.015	0.015	0.056	0.009	0.013	0.531	0.099
G	0.123	0.012	0.015	0.024	0.888	0.019	0.013	0.422	0.050
T	0.078	0.930	0.958	0.943	0.028	0.037	0.812	0.021	0.788

Log-odds position weight matrix (PWM):

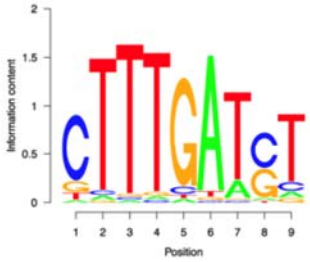
$$w(i,b) = \text{integer}\,(10 * \log10\,(\,p(i,b) / 0.25\,)\,)$$

-10	-10	-14	-12	-10	5	-2	-10	-6
5	-10	-13	-13	-7	-15	-13	3	-4
-3	-14	-13	-11	5	-12	-13	2	-7
-5	5	5	5	-10	-9	5	-11	5

C



B Sequence Logo



PWM score = 5 + 5 + 5 + 5 + 5 + 5 + 5 + 3 + 5 = 43

G. Ambrosini, I. Vorontsov, D. Penzar, R. Groux, O. Fornés, D. Nikolaeva, B. Ballester, J. Grau, I. Grosse, V. Makeev, I. Kulakovskiy and P. Bucher, Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study, *Genome Biol.*, 2020, **21**.

Predict binding sites

Can we find out the common pattern of the binding sites and predict them?

To investigate the usefulness of the PWM generated from our set of ChIP binding sites for predicting FNR sites genome-wide, we initially used a PatSer threshold low enough that a FNR motif was identified in each FNR ChIP-seq peak.

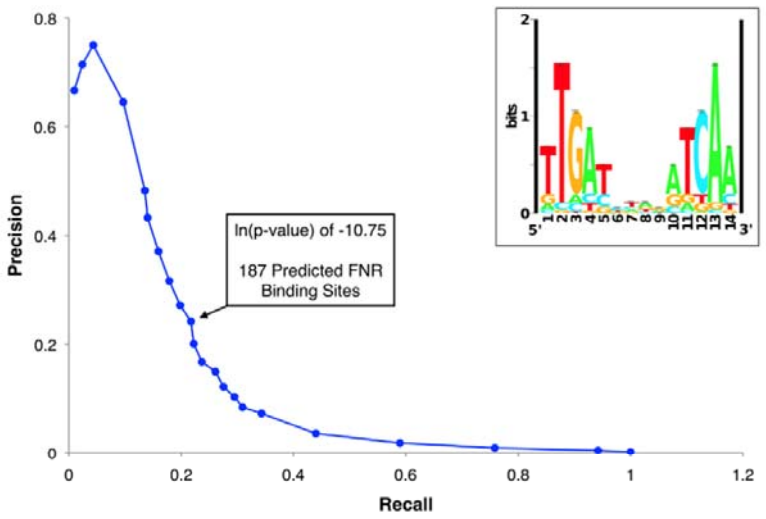
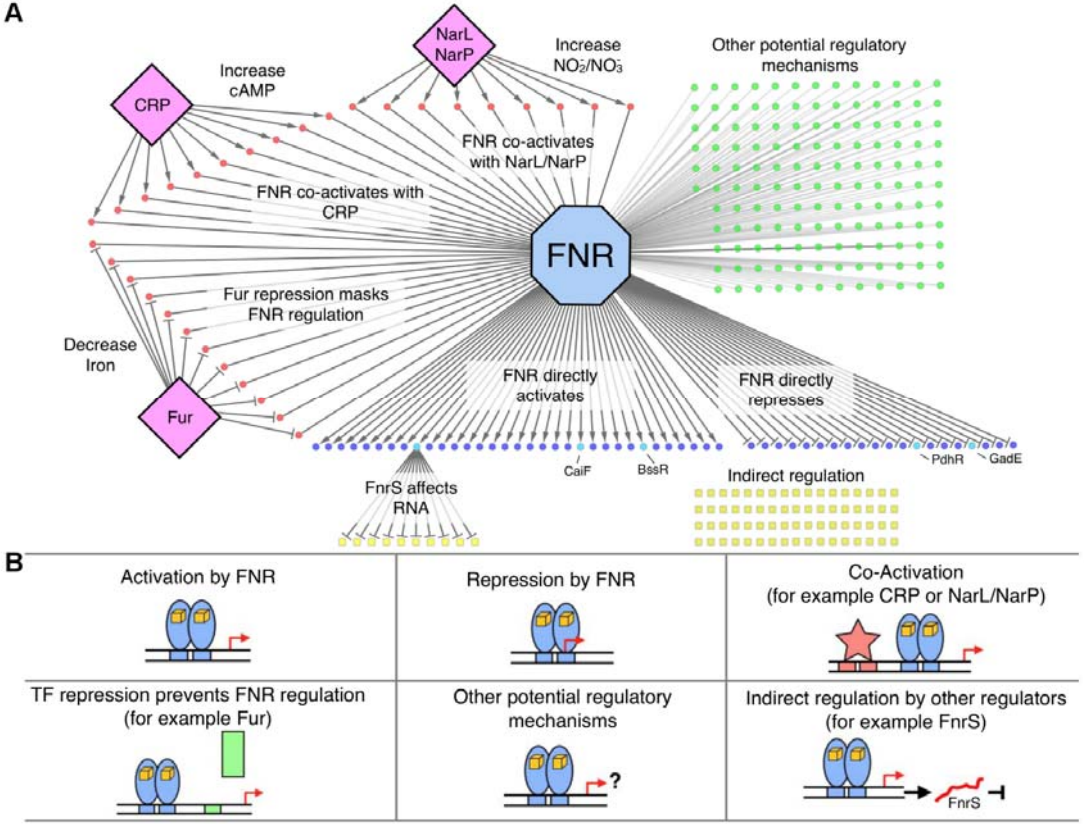


Figure 2. Precision-recall curve used to determine the prediction threshold of FNR binding sites and updated FNR PWM. The precision-recall curve used to determine the optimal threshold for predicting high quality FNR binding sites throughout the genome. The precision and recall values were determined for many ln(p-value) thresholds using the PatSer algorithm and the optimal value is identified by the arrow. The inset shows the FNR position weight matrix (PWM) constructed from the FNR ChIP-seq peak sequences. The height (y-axis) of the letters represents the degree of conservation at that position within the aligned sequence set (in bits), with perfect conservation being 2 bits. The x-axis shows the position of each base (1–14) starting at the 5′ end of the motif.
doi:10.1371/journal.pgen.1003565.g002

FNR function

We identified the binding sites, but the functions of each sites is different. What is the potential pathways of FNR regulation?



FNR function

Table 1. Operons with an upstream FNR ChIP-seq peak and a FNR-dependent change in expression under GMM.

Peak Center (nt)[a]	Operon[b]	B-number of first gene[c]	Function of Operon Product[d]	Number of FNR binding sites[e]	Location top scoring FNR binding site[f]	σ^{70} occupancy -O_2 relative to +O_2[g]	WT -O_2 expression relative to WT +O_2 expression[h]	Previous Experimental Evidence of FNR Binding[i]	Previous Evidence of FNR Regulated Expression[j]
Operons directly activated by FNR (Category 1)									
1,003,976	*pyrD*	b0945	Dihydroorotate Dehydrogenase	1	−38.5	+	o	[29]	[17]
1,656,036	*ynfEFGH- dmsD*	b1587	Putative Selenate Reductase (*ynfEFGH*); DMS Reductase Maturation Protein (*dmsD*)	1	−40.5	+	+	None	[18,19]
1,935,550	*pykA*	b1854	Pyruvate Kinase II	1	−40.5	+	+	[29]	[19]
3,611,605	*nikABCDE*	b3476	Nickel Transporter	1	−40.5	+	+	None	[126]
953,741	*focA-pflB*	b0904	Formate Transporter (*focA*); Pyruvate Formate-Lyase (*pflB*)	2	−40.5	+	+	[127]	[128]
2,714,605	*yfiD*	b2579	Stress-Induced Alternative Pyruvate Formate-Lyase	1	−40.5	+	+	[129]	[129]
940,035	*dmsABC*	b0894	Dimethyl Sulfoxide Reductase	1	−41.5	+	+	[41]	[41]
1,279,003	*narGHJI*	b1224	Nitrate Reductase	1	−41.5	+	+	[41]	[130]
1,627,208	*ydfZ*	b1541	Unknown Function	2	−41.5	+	+	[29]	[18,19]
1,837,412	*ynjE*	b1757	Molybdopterin Synthase Sulfurtransferase	1	−41.5	+	+	None	[18]
3,491,947	*nirBDC- cysG*	b3365	Nitrite Reductase (*nirBDC*); Uroporphyrin III C-Methyltransferase (*cysG*)	1	−41.5	+	+	[21]	[130]
4,285,670	*nrfABCDEFG*	b4070	Periplasmic Nitrite Reductase	1	−41.5	+	+	[42]	[42]
34,059	*caiF*	b0034	Carnitine Transcriptional Activator	1	−41.5	+	o	[29]	[47]
1,277,082	*narK*	b1223	Nitrate/Nitrite Antiporter	1	−41.5	+	+	[41]	[131]
877,441	*bssR*	b0836	Regulator of Biofilm Formation	1	−41.5	+	+	None	None
1,752,688	*ydhYVWXUT*	b1674	Predicted Oxidoreductase System	1	−42.5	+	+	[132]	[18,19]

References:

G. Ambrosini, I. Vorontsov, D. Penzar, R. Groux, O. Fornés, D. Nikolaeva, B. Ballester, J. Grau, I. Grosse, V. Makeev, I. Kulakovskiy and P. Bucher, Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study, Genome Biol., 2020, 21.

J. W. K. Ho, E. Bishop, P. V. Karchenko, N. Nègre, K. P. White and P. J. Park, ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis, *BMC Genomics*, 2011, **12**, 134.

K. S. Myers, H. Yan, I. M. Ong, D. Chung, K. Liang, F. Tran, S. Keleș, R. Landick and P. J. Kiley, Genome-scale analysis of Escherichia coli FNR reveals complex features of transcription factor binding, *PLoS Genet.*, 2013, **9**, e1003565.

Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009;10(10):669-680. doi:10.1038/nrg2641