

SCIENTIFIC REPORTS



OPEN

Long noncoding RNA expression profiles in gut tissues constitute molecular signatures that reflect the types of microbes

Received: 21 April 2015

Accepted: 03 June 2015

Published: 30 June 2015

Lunxi Liang^{1,2,*}, Luoyan Ai^{1,2,*}, Jin Qian^{1,2}, Jing-Yuan Fang^{1,2} & Jie Xu^{1,2}

The gut microbiota is commonly referred to as a hidden organ due to its pivotal effects on host physiology, metabolism, nutrition and immunity. The gut microbes may be shaped by environmental and host genetic factors, and previous studies have focused on the roles of protein-coding genes. Here we show a link between long non-coding RNA (lncRNA) expression and gut microbes. By repurposing exon microarrays and comparing the lncRNA expression profiles between germ-free, conventional and different gnotobiotic mice, we revealed subgroups of lncRNAs that were specifically enriched in each condition. A nearest shrunken centroid methodology was applied to obtain lncRNA-based signatures to identify mice in different conditions. The lncRNA-based prediction model successfully identified different gnotobiotic mice from conventional and germ-free mice, and also discriminated mice harboring transplanted microbes from fecal samples of mice or zebra fishes. To achieve optimal prediction accuracy, fewer lncRNAs were required in the prediction model than protein-coding genes. Taken together, our study demonstrated the efficacy of lncRNA expression profiles in discriminating the types of microbes in the gut. These results also provide a resource of gut microbe-associated lncRNAs for the development of lncRNA biomarkers and the identification of functional lncRNAs in host-microbes interactions.

The intestinal tract harbors trillions of commensal bacteria representing over a thousand species and encoding over one hundred and fifty fold more genes than the human genome. The human intestinal microbiota has been shown to participate in epithelium maturation and proliferation, host nutrition and metabolism, as well as immune responses and protection against pathogens^{1,2}. It is increasingly likely that specific compositional patterns of gut microbiota may associate with different human diseases, such as colorectal cancer^{3–6} and inflammatory bowel disease (IBD)^{7–9}. The gut microbiota composition is shaped by multiple factors such as food intake, colonization history, and host genetic factors^{10,11}. Our current knowledge about the relationship between host genetic background and microbiota composition is still limited, and most previous studies have focused on the potential roles of protein-coding genes^{12–14}.

Recent genomic studies have revealed tens of thousands of long non-coding RNAs (lncRNAs) in the mammalian genomes¹⁵. lncRNAs may participate in many essential biological processes, such as genomic imprinting, maintenance of pluripotency, immune response and development. Moreover, lncRNAs have also been linked to different human diseases, such as neurodegenerative disorders, cardiovascular diseases and cancers^{16–18}. While the roles of protein-coding genes in host-microbiota interactions

¹State Key Laboratory for Oncogenes and Related Genes; Key Laboratory of Gastroenterology & Hepatology, Ministry of Health; Division of Gastroenterology and Hepatology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University. ²Shanghai Cancer Institute, Shanghai Institute of Digestive Disease, 145 Middle Shandong Rd, Shanghai 200001, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.X. (email: jhexu@yahoo.com) or J.-Y.F. (email: jingyuanfang@sjtu.edu.cn)

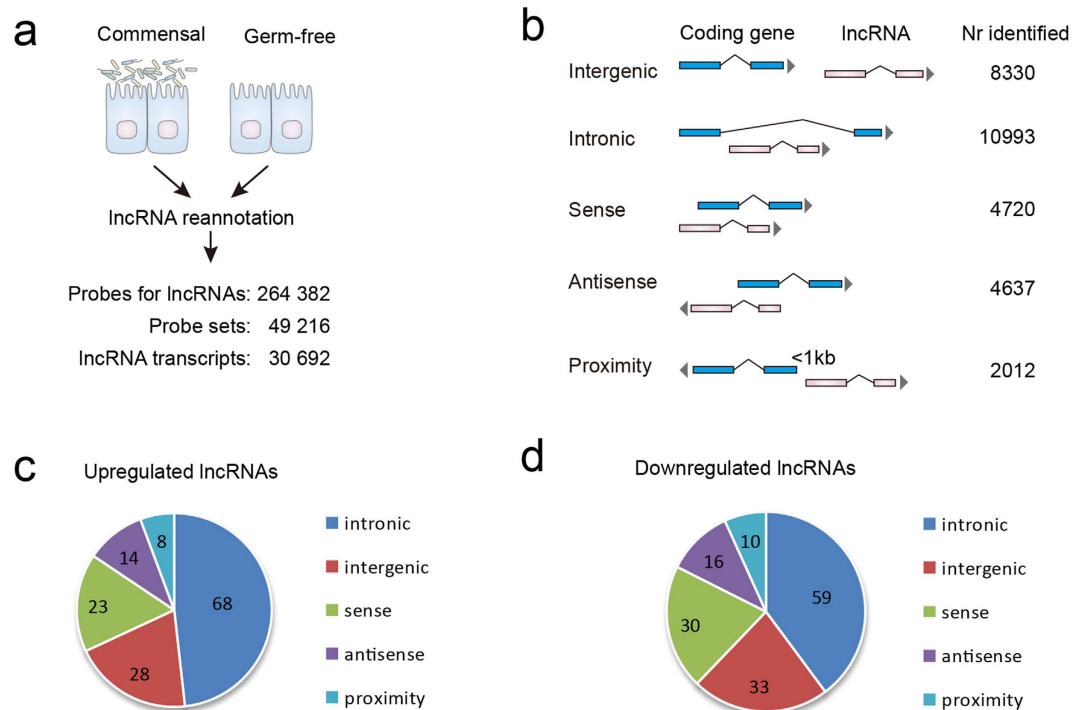


Figure 1. Identification of lncRNAs regulated by commensal microbiota in mice. (A) Schematic representation of the lncRNA reannotation process. The mouse exon array contained 264 382 probes mapping to 30692 lncRNA transcripts. Probes that overlap with coding genes have been removed. The dataset was based on intestinal epithelium cells of conventionalized mice as compared to germ-free mice. (B) The classification of lncRNAs into five categories (intergenic, intronic, sense, antisense and proximity) according to their relationships with protein-coding genes. The numbers of identified lncRNAs in different categories are shown on the right. (C,D) Pie charts showing the number of microbiota-regulated lncRNAs in each category. When a criteria of fold change >2 and $P < 0.05$ was accepted, 141 and 148 lncRNAs were respectively up- and down-regulated.

have been subjected to intensive investigation, it is largely unclear if lncRNAs may participate in the responses of intestinal epithelial cells to gut flora.

Recent studies have suggested the involvement of lncRNAs in inflammatory signaling. As an example, lncRNA-Cox2 has been reported as a downstream target of TLR signaling that serves as a transcriptional cofactor through interactions with various regulatory complexes¹⁹. In addition, the lncRNA THRIL was found to regulate TNF α by binding to hnRNPL during innate activation of macrophages²⁰. However, it is unknown whether and to what extent lncRNAs may be regulated by gut microbiota. It is also unclear if lncRNA expression profiles may reflect certain features of microbes in the gut.

In this study, we characterize lncRNAs that are regulated by gut microbiota (in conventional or gnotobiotic mice), which may be useful for further functional investigations. We also present a proof-of-concept study for the efficacy of lncRNA-based signatures in discriminating conventional and gnotobiotic mice.

Results

Identification of commensal microbiota-regulated lncRNAs. To systematically identify microbiota-regulated lncRNAs, we focused on published microbiota re-colonization studies that utilized the Affymetrix mouse exon microarray platform, which has many more probes mapping to lncRNA genes^{15,21}. We were specifically interested in the lncRNA expression profiles in gut epithelial tissues that interact with gut microbes, thus the datasets concerning other tissues or cell types (e.g., liver, macrophages, etc) were excluded from further investigation. Since the gut microbiota of laboratory mice is variable due to both genetic and environmental factors^{22,23}, we focused on data from one laboratory to avoid potential inconsistency caused by microbial variation. These criteria have selected the GSE46952 dataset²⁴, which included conventional, germ-free and gnotobiotic mice (re-colonized with either E.coli or E.coli expressing bile salt hydrolase) with at least 4 biological replicates for each condition. A comprehensive computational pipeline¹⁵ was used to re-annotate the probes that uniquely map lncRNA transcripts (overall design shown in Fig. 1a). The reliability of this method has been supported by RT-PCR validation and the high consistency with RNA-seq data¹⁵. As indicated in Fig. 1b, five categories of

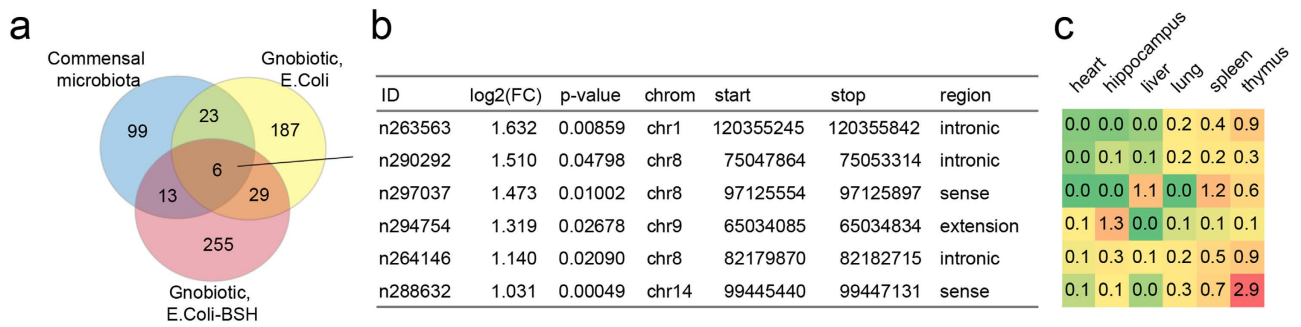


Figure 2. Venn's diagram showing the overlapping lncRNAs that are induced by microbiota, E.coli or E.coli-BSH. The six commonly upregulated lncRNAs are described in the table, and their expression levels in different tissues (heart, hippocampus, liver, lung, spleen and thymus) are indicated in the heat map on the right. It can be seen that most lncRNAs are highly expressed in immune organs (spleen and thymus).

lncRNAs have been identified according to their relationships with protein-coding genes, including intergenic, intronic, sense, antisense and proximity (a full list of lncRNAs is provided in Supplementary Table 1).

According to the findings of the MicroArray Quality Control (MAQC) project, fold-change based selection criteria can significantly improve the agreement of the biological interpretation of the data²⁵. In contrast, when a t-statistic (P-value) ranking is used as the primary criterion the reproducibility would be substantially lower²⁶. Therefore, we determined differentially expressed lncRNAs based on fold-change (>2 or <0.5) plus a nonstringent P-value cutoff (0.05). This criteria has been also been accepted in previous studies^{27,28}. While intronic lncRNAs represented the largest group (35.8%) in all identified lncRNAs, we found even higher rates of intronic lncRNAs in both upregulated (48.2%, $P < 0.001$, chi-test) and downregulated lncRNAs (39.9%, $P < 0.001$, chi-test) caused by re-colonization of commensal microbiota in germ-free mice (Fig. 1c,d, altered lncRNAs listed in Supplementary Table 2).

Limited overlap between lncRNAs associated with distinct gut microbes. Since it has been suggested that gnotobiotic mice may have specific expression patterns of protein-coding genes²⁹, we tested whether lncRNAs are also differentially expressed in mice that were re-colonized by different types of microorganisms. To this end, we analyzed the expression of lncRNAs in germ-free (GF) mice in comparison to mice that were re-colonized with either mouse microbiota (mouse), E.coli (EC) strain, or E.coli expressing bile salt hydrolase (EC-BSH). Interestingly, only low level of overlap was found between these conditions, with most altered lncRNAs being type-specific (Fig. 2a, listed in Supplementary Table 3). In the six commonly upregulated lncRNAs (Fig. 2b), most were also highly expressed in immune organs such as spleen and thymus (Fig. 2c), suggesting potential involvement of these lncRNAs in host immune responses.

Previous studies have demonstrated the crucial role of NF- κ B in transactivating a large number of protein-coding genes in response to gut microbiota^{1,30,31}, thus we explored the potential relationship between NF- κ B and upregulated lncRNAs. The genomic binding sites of NF- κ B was extracted from GSM611117 dataset, and were compared to the transcription starting sites (TSS) of upregulated lncRNAs. As a result, only 72 of 612 (11.7%) upregulated lncRNAs (in any condition) were found with NF- κ B binding sites <10 kb upstream of their TSS (Fig. 3, listed in Supplementary Table 4), indicating that most upregulated lncRNAs may not be direct transcriptional targets of NF- κ B.

lncRNA-based signatures correctly identify gnotobiotic mice from conventional mice.

Although the exact mechanisms underlying microbiota-affected lncRNA expression are largely unknown, it is likely that lncRNA expression may partially result from host-microbe interactions and therefore constitute a signature that reflects the status in the gut. Based on this hypothesis, we questioned if lncRNA expression profiles may provide sufficient information for discriminating gnotobiotic and conventional mice. Using an established method for feature extraction and sample classification named PAM algorithm³², we classified the germ-free (GF), re-conventionalized (RC), and gnotobiotic mice (EC or EC-BSH) based on lncRNA expression profiles. The PAM algorithm uses the "nearest shrunken centroids" model to identify gene signatures that best characterize each class, and its effectiveness on lncRNA-based sample classification has been demonstrated in our previous study¹⁶. As expected, the PAM algorithm identified lncRNAs with microbe-specific expression patterns (Fig. 4a), and the final classification model successfully discriminated these mice (GF, RC, EC and EC-BSH) with an overall error rate of 0.114 (Fig. 4b). Notably, the mice in EC and EC-BSH groups were classified without an error (Fig. 4c), suggesting that lncRNA expression profiles may discriminate gnotobiotic mice more efficiently from other types.

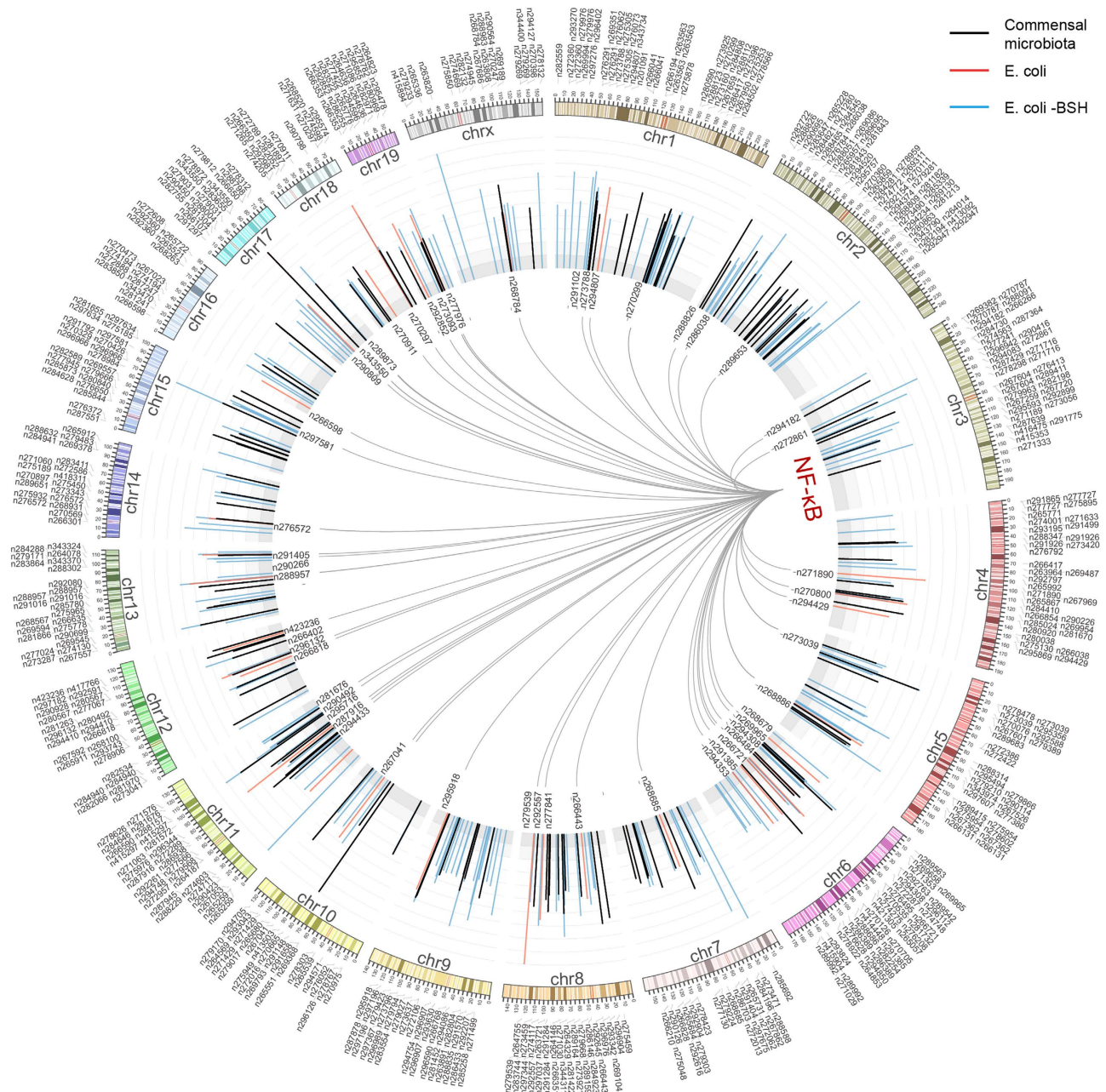


Figure 3. Circos plot showing the lncRNAs that are upregulated by commensal microbiota, *Escherichia coli* (E.coli), or *Escherichia coli* expressing bile salt hydrolase (E.coli-BSH). The histogram indicates expression levels of lncRNAs that are upregulated by microbiota (black), E.coli (blue) or E.coli-BSH (red). The links indicate potential binding of NF- κ B to the promoters of these upregulated lncRNAs. The Genbank accession numbers of lncRNA transcripts are indicated in the plot.

LncRNA signatures discriminate mice with different transplanted microbiota. The higher accuracy for gnotobiotic mice identification was not surprising, since the colony of single bacterial strain may represent a status of extremely imbalanced microbiota. Therefore, it is meaningful to test if lncRNAs-based signatures could efficiently discriminate mice bearing composite microbes. The GSE5198 dataset included germ-free mice that were re-colonized with fecal-derived microbiota from mice or zebra fish. Interestingly, the PAM algorithm identified a considerable number of lncRNAs with type-specific expression (Fig. 5a), and perfectly discriminated all mouse groups (Fig. 5b,c). Therefore, it seems that lncRNA expression profiles could identify not only gnotobiotic mice, but also mice with different composite microbes.

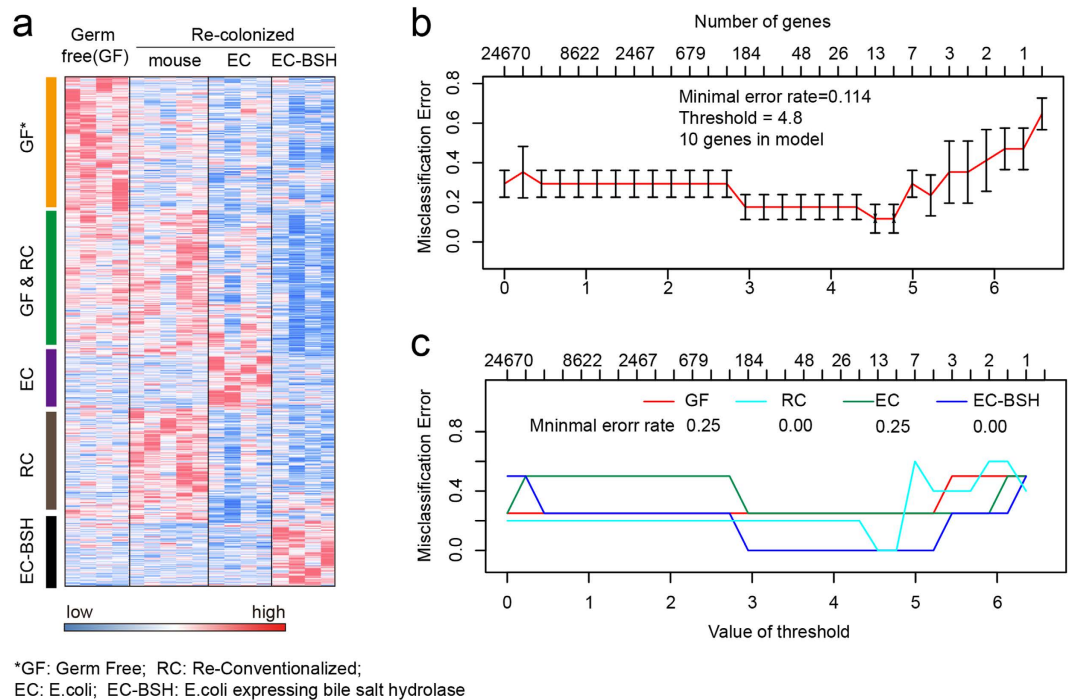


Figure 4. LncRNA expression profiles discriminate conventional and gnotobiotic mice. Four types of gut microorganism conditions in dataset GSE46952 (germ free, re-conventionalized with mouse microbiota, E.coli, or E.coli expressing bile salt hydrolase) were classified by PAM algorithm based on the expression profiles of lncRNAs. (A) The expression profiles of lncRNAs that have microbiota-specific expression in intestinal epithelium cells. The lncRNAs with greater centroids (>3) were selected by the PAM algorithm, which uses the nearest shrunken centroid methodology to identify sample types. These lncRNAs associate with specific microorganism status within the gut, such as germ free (GF), re-conventionalized mouse microbiota (RC), E.coli (EC), and E.coli expressing bile salt hydrolase (EC-BSH). (B) The misclassification error rates by PAM as a function of threshold for centroid shrinkage. The error rates firstly decreased at threshold of 3, and further reached minimal value (0.114) when threshold was set to 4.5. (C) Type-specific error rates for classification as a function of threshold for centroid shrinkage.

Discussions

Previous investigations have mainly focused on the potential roles of protein-coding genes in host-microbe interactions, but our results suggest a link between lncRNA expression and gut microbes. To probe the expression of lncRNAs in re-conventionalized and gnotobiotic mice, we used a comprehensive bioinformatics pipeline to reannotate probes that uniquely map to lncRNAs from public expression microarray datasets.

The comparisons between re-conventionalized (RC) mice and gnotobiotic mice (EC and EC-BSH) suggested considerable type-specific expression patterns of lncRNAs. Only six lncRNAs were commonly upregulated in RC, EC and EC-BSH mice, although 613 lncRNAs were found upregulated in at least one condition. These 6 lncRNAs were also highly expressed in immune organs such as spleen and thymus, suggesting their potential involvement in host immune responses. Since immune cells may be recruited and activated upon re-colonization of microbes in the gut, it still requires clarification whether these commonly upregulated lncRNAs may accurately reflect the change in epithelial cells alone.

Our classification models based on lncRNA expression profiles have successfully discriminated mice that were re-colonized with different E. coli strains or fecal-derived microbiota. These findings further support a more generalized hypothesis that lncRNAs may be as important as protein-coding genes for the purposes of indicating biological status. As we have discussed previously, the expression level of a non-coding gene may better represent its activity than a protein-coding gene (PCG), because the function of a PCG may be affected by more factors such as translation, posttranslational modification, conformational regulation, and proteasomal degradation^{17,18}.

To avoid potential inconsistency caused by microbial variation, our study was based on microarray data from the same laboratory. Future studies should test the cross-platform (e.g. RNA-seq vs microarray), cross-strain (BALB/c vs C57BL/6) and cross-laboratory overlapping of differentially expressed lncRNAs. It would also be worthwhile to further clarify the exact roles of lncRNAs in host-microbe interactions. Another important direction would be discovering disease-associated lncRNA signatures, which may be useful for developing novel biomarkers and therapeutic targets.

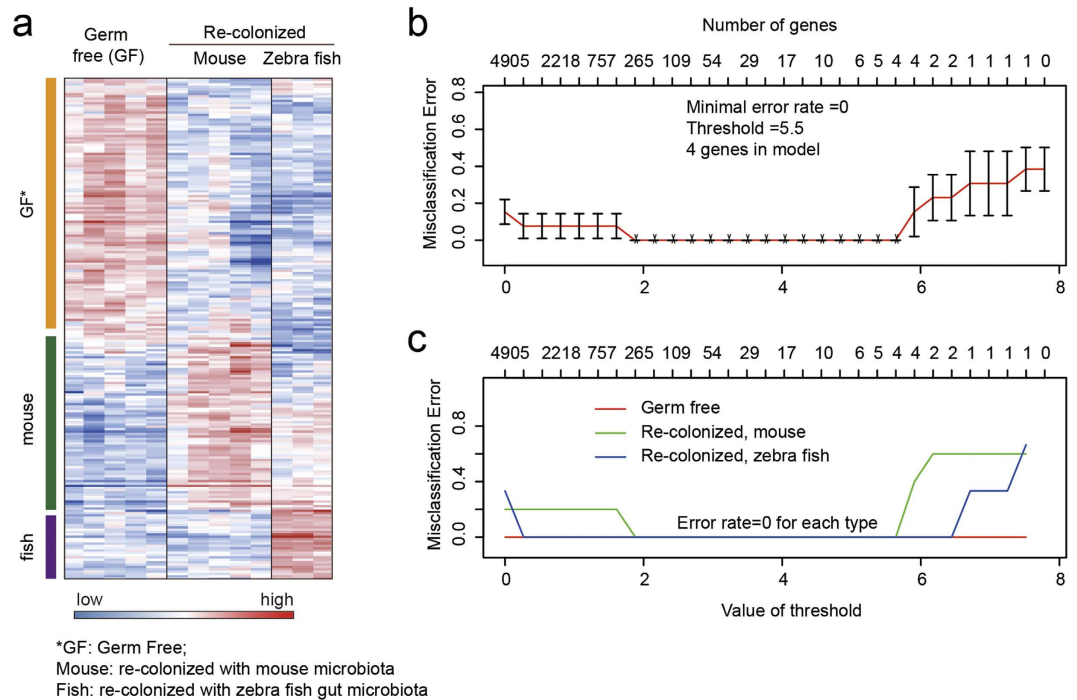


Figure 5. Discrimination of mice with gut microbiota transplanted from conventional mice or zebra fish. The lncRNA expression profiles were extracted from GSE5198 dataset, which included gene expression in the small intestine of mice that were either germ-free (GF), colonized with a conventional mouse cecal microbiota (Re-colonized, mouse) or colonized with a conventional zebrafish gut microbiota (re-colonized, zebrafish). (A) Heat map showing the sample-specific expression profiles of lncRNAs in intestinal epithelium cells. The genes were screened according to a centroid shrinkage threshold of 2, as determined by overall prediction error. (B) Overall prediction error rate as a function of centroid shrinkage threshold in the PAM classification model. An minimal error of 0.00 was reached when the threshold increased to 2. (C) Type-specific error rates for classification as a function of threshold for centroid shrinkage.

Methods

Re-annotation of microarray probes. The raw microarray data of mouse intestinal tissues were downloaded from Gene Expression Omnibus (GEO). The dataset GSE46952 included conventional mice ($n = 5$), germ-free (GF, $n = 4$), and gnotobiotic mice that were colonized with *E. coli* (EC, $n = 4$) or *E. coli* expressing bacterial bile salt hydrolase (EC-BSH, $n = 5$). The pipeline for annotating probes that uniquely map to lncRNAs has been described previously¹⁵. Briefly, the 4.7 million probes in the Affymetrix GeneChip Exon 1.0 ST arrays were filtered to discard those mapping to none or multiple locations, and probes overlapping with protein-coding genes were also excluded for further processing. The remaining probes were aligned with lncRNA genes (>200 bp) that were included in the NONCODE3 database³⁵. After ambiguous hits were removed, the probes mapped to 30692 lncRNAs in the mouse genome.

Identification of differentially expressed lncRNAs. The expression levels of lncRNAs were compared between different conditions using Linear Models for Microarray Data (LIMMA)³⁴. A widely accepted criteria of fold change > 2 and $P < 0.05$ was used to identify differentially expressed genes. According to the MicroArray Quality Control (MAQC) project^{25,26}, gene lists generated by fold-change ranking plus a nonstringent P-value cutoff were more reproducible than those obtained by significance analysis. The genomic locations and expression levels of altered lncRNAs were visualized using the circo program³⁵.

Mapping NF- κ B binding sites on lncRNA promoters. The NF- κ B ChIP-seq data (on p65 subunit) were obtained from Expression Omnibus (GEO) with the accession number GSM611117. The distances between these peaks and the transcription starting sites (TSSs) of lncRNAs were calculated with the ChIPpeakAnno package of the Bioconductor program³⁶. When the binding site was located within 10 kb upstream the TSS of lncRNA, a putative binding was considered. This criteria has been adopted by many previous studies^{37–39}.

Sample classification based on gene expression profiles. We used lncRNA expression profiles to predict the types of mice, based on the PAM algorithm that shrinks the prototypes and hence obtains a

classifier³². PAM applies the “nearest shrunken centroids” method to identify subsets of genes that best characterize each class. The shrinkage consists of moving the centroid towards zero by a threshold, which is determined according to the prediction error of the model. As the threshold increases, the number of genes left in the model decreases. In the present study, the prediction model was based on minimal sets of genes at a shrinkage threshold immediately before the error rates escalate. The prediction based on protein-coding genes used the same method as lncRNAs, and the threshold for centroid shrinkage was determined independently. Moreover, genes left in the model displayed strong type-specific expression feature when the increase of shrinkage caused an initial decrease in the misclassification error (note the threshold level was lower than the final value). In this context, genes left in the model displayed significant type-specific distribution patterns. The heat maps were generated using the TM4 software package⁴⁰.

References

- Lakhdari, O. *et al.* Functional metagenomics: a high throughput screening method to decipher microbiota-driven NF- κ B modulation in the human gut. *PLoS One* **5**, doi: 10.1371/journal.pone.0013092 (2010).
- Oliveira, M. R. *et al.* Germ-free mice produce high levels of interferon-gamma in response to infection with *Leishmania major* but fail to heal lesions. *Parasitology* **131**, 477–488, doi: 10.1017/S0031182005008073 (2005).
- Klimesova, K. *et al.* Altered gut microbiota promotes colitis-associated cancer in IL-1 receptor-associated kinase M-deficient mice. *Inflamm Bowel Dis* **19**, 1266–1277, doi: 10.1097/MIB.0b013e318281330a (2013).
- Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* **12**, 661–672, doi: 10.1038/nrmicro3344 (2014).
- Smith, K. Microbiota: Manipulating the microbiota could affect colorectal cancer development. *Nat Rev Gastroenterol Hepatol* **11**, 4, doi: 10.1038/nrgastro.2013.225 (2014).
- Jobin, C. Colorectal cancer: looking for answers in the microbiota. *Cancer Discov* **3**, 384–387, doi: 10.1158/2159-8290.CD-13-0042 (2013).
- Bellaguarda, E. & Chang, E. B. IBD and the Gut Microbiota—from Bench to Personalized Medicine. *Curr Gastroenterol Rep* **17**, 439, doi: 10.1007/s11894-015-0439-z (2015).
- Ray, K. IBD. Understanding gut microbiota in new-onset Crohn’s disease. *Nat Rev Gastroenterol Hepatol* **11**, 268, doi: 10.1038/nrgastro.2014.45 (2014).
- Dessein, R., Rosenstiel, P. & Chamillard, M. Debugging the intestinal microbiota in IBD. *Gastroenterol Clin Biol* **33 Suppl 3**, S131–136, doi: 10.1016/S0399-8320(09)73148-3 (2009).
- Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**, 18933–18938, doi: 10.1073/pnas.1007028107 (2010).
- Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest* **124**, 3617–3633, doi: 10.1172/JCI75436 (2014).
- Manichanh, C., Borruel, N., Casellas, F. & Guarner, F. The gut microbiota in IBD. *Nat Rev Gastroenterol Hepatol* **9**, 599–608, doi: 10.1038/nrgastro.2012.152 (2012).
- Penders, J. *et al.* Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**, 511–521, doi: 10.1542/peds.2005-2824 (2006).
- Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab* **17**, 141–152, doi: 10.1016/j.cmet.2012.12.007 (2013).
- Gellert, P., Ponomareva, Y., Braun, T. & Uchida, S. Noncoder: a web interface for exon array-based detection of long non-coding RNAs. *Nucleic Acids Res* **41**, e20, doi: 10.1093/nar/gks877 (2013).
- Hu, Y. *et al.* Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer. *Cancer Res* **74**, 6890–6902, doi: 10.1158/0008-5472.CAN-14-0686 (2014).
- Ciacci, C. *et al.* Immunomodulation in *Mytilus galloprovincialis* by non-toxic doses of hexavalent chromium. *Fish Shellfish Immunol* **31**, 1026–1033, doi: 10.1016/j.fsi.2011.09.002 (2011).
- Hu, Y. *et al.* A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* **5**, 2230–2242 (2014).
- Carpenter, S. *et al.* A long noncoding RNA mediates both activation and repression of immune response genes. *Science* **341**, 789–792, doi: 10.1126/science.1240925 (2013).
- Li, Z. *et al.* The long noncoding RNA THRIL regulates TNF α expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A* **111**, 1002–1007, doi: 10.1073/pnas.1313768111 (2014).
- Du, Z. *et al.* Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* **20**, 908–913, doi: 10.1038/nsmb.2591 (2013).
- Hildebrand, F. *et al.* Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol* **14**, R4, doi: 10.1186/gb-2013-14-1-r4 (2013).
- Hufeldt, M. R., Nielsen, D. S., Vogensen, F. K., Midtvedt, T. & Hansen, A. K. Variation in the gut microbiota of laboratory mice is related to both genetic and environmental factors. *Comp Med* **60**, 336–347 (2010).
- Joyce, S. A. *et al.* Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *Proc Natl Acad Sci U S A* **111**, 7421–7426, doi: 10.1073/pnas.1323599111 (2014).
- Consortium, M. *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24**, 1151–1161, doi: 10.1038/nbt1239 (2006).
- Patterson, T. A. *et al.* Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* **24**, 1140–1150, doi: 10.1038/nbt1242 (2006).
- Tschopp, P. *et al.* A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* **516**, 391–394, doi: 10.1038/nature13819 (2014).
- Miyazaki, M. *et al.* Id2 and Id3 maintain the regulatory T cell pool to suppress inflammatory disease. *Nat Immunol* **15**, 767–776, doi: 10.1038/ni.2928 (2014).
- Yamamoto, M. *et al.* A microarray analysis of gnotobiotic mice indicating that microbial exposure during the neonatal period plays an essential role in immune system development. *BMC Genomics* **13**, 335, doi: 10.1186/1471-2164-13-335 (2012).
- Erkosar Combe, B. *et al.* *Drosophila* microbiota modulates host metabolic gene expression via IMD/NF- κ B signaling. *PLoS One* **9**, e94729, doi: 10.1371/journal.pone.0094729 (2014).
- Dantoft, W. *et al.* The Oct1 homolog Nubbin is a repressor of NF- κ B-dependent immune gene expression that increases the tolerance to gut microbiota. *BMC Biol* **11**, 99, doi: 10.1186/1741-7007-11-99 (2013).
- Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567–6572, doi: 10.1073/pnas.082099299 (2002).

33. Bu, D. *et al.* NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res* **40**, D210–215, doi:10.1093/nar/gkr1175 (2012).
34. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi: 10.1093/nar/gkv007 (2015).
35. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645, doi: 10.1101/gr.092759.109 (2009).
36. Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237, doi: 10.1186/1471-2105-11-237 (2010).
37. Deschenes, J., Bourdeau, V., White, J. H. & Mader, S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J Biol Chem* **282**, 17335–17339, doi: 10.1074/jbc.C700030200 (2007).
38. Feng, D. *et al.* A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **331**, 1315–1319, doi: 10.1126/science.1198125 (2011).
39. Khan, A. H., Lin, A. & Smith, D. J. Discovery and characterization of human exonic transcriptional regulatory elements. *PLoS One* **7**, e46098, doi: 10.1371/journal.pone.0046098 (2012).
40. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).

Acknowledgements

This project was supported by grants from National Natural Science Foundation of China (30971330, 31371420, 81320108024, 81000861, 81322036, and 81272383); Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 81421001), the Program for Innovative Research Team of Shanghai Municipal Education Commission; Shanghai “Oriental Scholars” project (2013XJ); Shanghai Science and Technology Commission “Pujiang Project” (13PJ1405900) and Shanghai Natural Science Foundation (12ZR1417900). The sponsors of this study had no role in the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, or the writing of the manuscript.

Author Contributions

J.X. and L.L. wrote the manuscript. L.L., L.A., J.Q., J.X. and J.-Y.F. collected data and performed analysis. J.X. conceived this work.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liang, L. *et al.* Long noncoding RNA expression profiles in gut tissues constitute molecular signatures that reflect the types of microbes. *Sci. Rep.* **5**, 11763; doi: 10.1038/srep11763 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>