

Ovation Prokaryotic RNA-Seq System : Whole Transcriptome Profiling in Bacteria

Chris Armour, Himanshu Sethi, Victor Sementchenko, Matt Biery and Chris Raymond, NuGEN Technologies, Inc.

Overview

박테리아는 최근까지 과학적으로 쉽게 근접할 수 없었던 수 많은 생물 다양성을 갖고 있다. 차세대 시퀀싱(Next-Generation Sequencing, NGS)의 개발로 미생물 분야의 연구에 변화가 나타나고 있는데, 그 예로 박테리아 계놈의 전체 시퀀싱은 이제 비교적 쉬운 일이 되었다. 박테리아 연구를 위한 다음 과정은 다양한 표현형 정보, 박테리아 계놈 서열과 whole transcriptome 프로파일의 조합을 포함하게 될 것이다 (그림 1). 이렇게 통합된 결과는 종간의 유전적인 다양성이 유전자 발현의 차이를 통해 궁극적으로는 어떻게 표현형의 특성으로 나타나는지 알려줄 것이다.

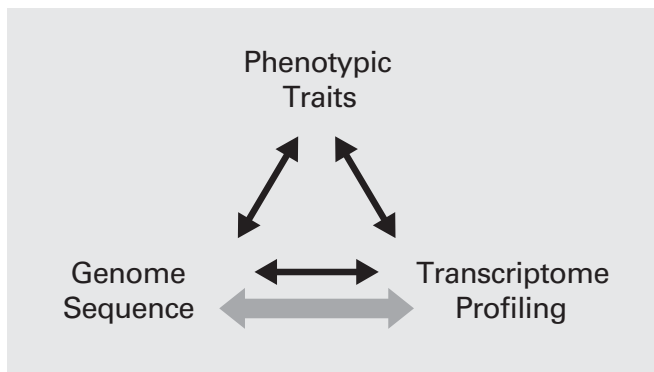


그림 1. Bacterial genomics. NGS technologies afford the opportunity to integrate phenotypes, genome sequences and global transcriptome profiling. Genome sequencing and transcriptome profiling are tightly coupled (heavy arrow) because transcriptome analysis requires the genome sequence for sequence alignments, and conversely, the interpretation of unique genomic features is guided by global transcriptome analysis.

박테리아의 전반적인 유전자 발현 패턴 연구에는 여러가지 기술적인 어려움이 있다. 특히, 대부분의 박테리아 RNA는 ribosomal subunit RNA(rRNAs)로 구성되어 있고 박테리아의 전사, 번역 그리고 RNA의 분해가 견고하게 연결되어 있다는 사실은 생물학적으로 꼭 필요한 정보를 포함하고 있는 RNA들이 종종 small RNA와 분명하지 않은 RNA 단편에 의해 대변된다는 것을 의미한다. 결론적으로 다양한 표현형을 나타내지만 동일한 서열을 가진 이 다양성은 발현된 유전자 서열을 밝혀내는데 있어서 불확실성으로 작용하게 된다. NGS에서 나오는 방대한 결과를 이용하는 이상적인 과정은 계놈 서열분석과 유전자 발현분석의 결합이 될 것이다 (그림 1). 이런 관점에서 계놈 서열분석은 transcriptome 서열의 alignment를 위한 비교대상은 물론 한 조직에서 완전하게 유전적인 분석이 가능하다는 청사진을 제공한다. 반면 transcriptome 데이터를 보면 표현형의 변화를 주도하는 전사 반응 네트워크를 알 수 있게 된다.

Ovation Prokaryotic RNA-Seq System 특징

- 뛰어난 rRNA의 감소
- 모든 박테리아 종에 적용할 수 있는 보편적 제품
- 초기 필요한 샘플량이 매우 적음
- 빠르고, 간편하고, RNA-to-NGS에 최적화
- 비용 절감이 가능한 multiplex 분석 가능
- 분해된 total RNA 샘플에 효과적
- 기존에 사용하고 있는 rRNA depletion 방법과 병행 가능

NuGEN Technologies Inc.(이하 NuGEN)에서는 모든 박테리아 종에 적용할 수 있고, NGS 기술에 대응 가능하도록 최적화된 prokaryotic transcriptome 프로파일링 방법의 필요성을 인식하고, total RNA에서 시퀀싱까지 가능한 end-to-end workflow를 만들었다. 독자적으로 개발한 프라이머를 사용하여 double strand cDNA를 합성하며 (그림 2), 이렇게 합성한 cDNA는 NuGEN의 Encore NGS Library System은 물론 타사의 시퀀싱용 library 제작에도 적용 가능하다.

Performance

Ovation Prokaryotic RNA-Seq System의 특징은 위에 묘사한 바와 같다. NuGEN은 rRNA read 감소, 전반적인 유전자 발현 분석, 다양한 박테리아 종에서의 성능과 균일한 sequence coverage를 포함한 몇 가지의 성능을 평가하였다. rRNA reduction은 유효성의 핵심 지표로 사용된다 (그림 3). 전형적인 박테리아의 total RNA는 95-99%가 rRNA이고, 나머지 1-5%의 RNA가 생물학적으로 유용한 발현패턴을 만들어내고, 모니터링하는데 사용하는 unique RNA(mRNA)이다. Total RNA내 unique RNA의 비중은 종이나 균집, 샘플의 성장단계에 따라서 매우 다양하게 변화할 수 있다. 일반적으로 빠르게 성장하는 "log phase"에서 준비한 샘플은 unique RNA의 비율이 낮다. 일반적인 박테리아 샘플(95-99% rRNA)에서 rRNA의 90%가 성공적으로 제거되면 (Ovation Prokaryotic System의 일반적인 performance) unique RNA의 read의 비율이 total read의 10-35% 수준까지 올라갈 수 있다. 즉, rRNA 양이 매우 많아 total RNA 중 rRNA의 90%를 제거해도, rRNA는 여전히 read count의 대다수를 나타낸다.

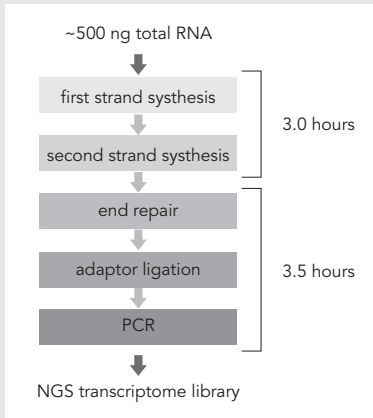


그림 2. Construction of NGS transcriptome libraries. First- and second-strand cDNA Synthesis with selective priming to enrich non-rRNA transcripts from bacterial and archaeal total RNA inputs. Following end repair, the double strand cDNA is compatible with NuGEN's Encore NGS Library Systems that allows easy sample multiplexing within a single-end, transcriptome sequencing format.

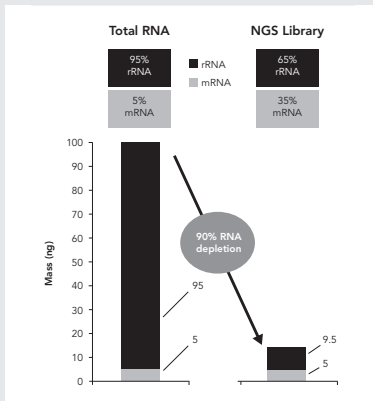


그림 3. Ribosomal RNA depletion from an NGS transcriptome library. In a total RNA sample that is 95% rRNA and 5% uniquely mapping mRNA, elimination of 90% of the rRNA results in an NGS library that is 65% rRNA and 35% unique RNAs. The key feature of this library is a 7-fold enrichment of unique RNA species.

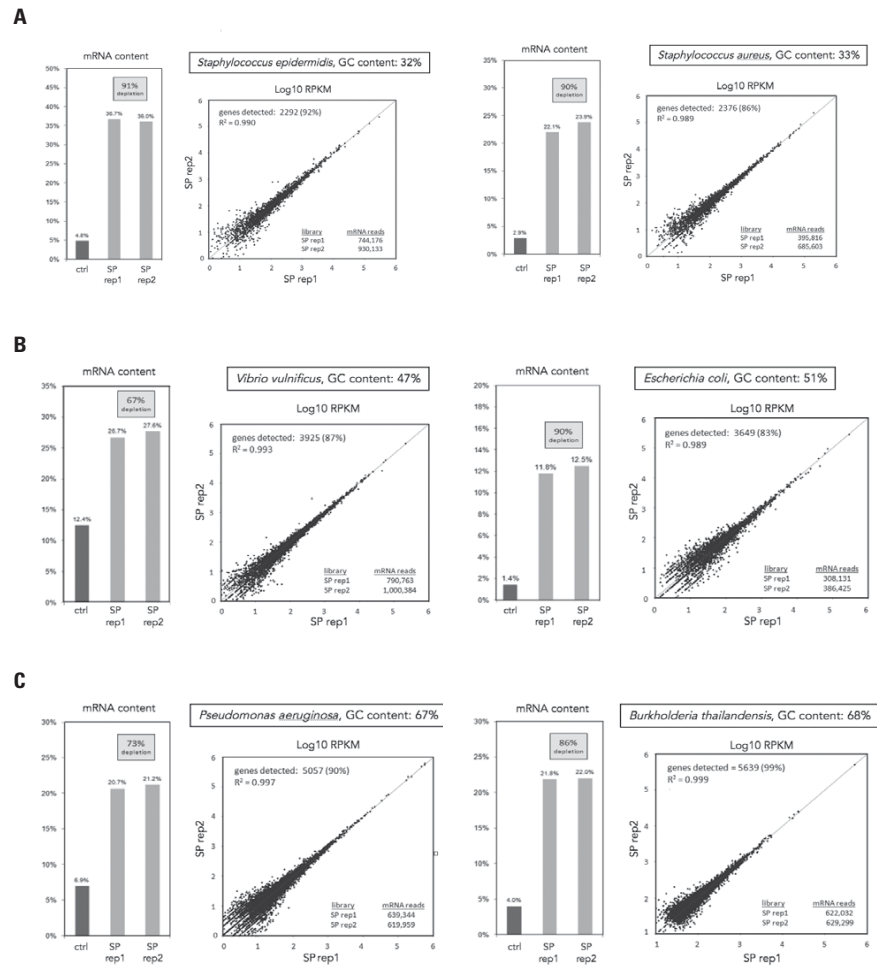


그림 4. The Ovation Prokaryotic RNA-Seq System. (A) Performance in bacteria with low GC content in their genomes. (B) Performance in bacteria with ~50% GC content in their genomes. (C) Performance in bacteria with high GC content in their genomes. Samples labeled with SP were processed with the selective priming technology in the Ovation Prokaryotic RNA-Seq System.

NuGEN에서는 박테리아 연구에 이와 같은 접근이 중대한 공헌을 할 것이라고 본다. 예를 들어 전형적인 박테리아에서 유전자 발현 프로파일링 실험을 한다고 가정 해보자. 만약 4,500,000 bp 게놈의 90%가 전사된다고 가정하면, 이것은 4,000 kbp의 transcriptome이 된다. 발현 프로파일링 제작을 위하여 sequencing depth 200 reads/kbp로 읽는다고 하면 이 transcriptome을 커버하기 위해서는 약 800,000 mapping read가 필요하다. 이 sequencing depth 기준으로, 평균 발현 수준의 400 bp의 평균 ORF-encoding gene은 약 80 reads에 의해 커버 가능하다. 실제 NuGEN에서 실험한 바에 의하면 여러 종의 박

테리아에서 400,000~800,000 mapping read의 sequencing depth로 유전자의 90% 혹은 그 이상을 검출하였다. 박테리아 배양액에서 추출한 total RNA 중 95%가 rRNA라고 추정하면, rRNA reduction 없이 이 transcriptome을 프로파일링하려면 약 2천만 sequencing read가 필요하다. 그러나 NuGEN의 rRNA reduction 방법을 사용하여 90%를 감소시킬 경우 2,800,000 read면 transcriptome 분석이 가능하고, NuGEN Encore Multiplex System으로 multiplex library를 제작한 후, Illumina Genome Analyzer II에 적용하면 multiple sample의 시퀀싱도 가능하다.

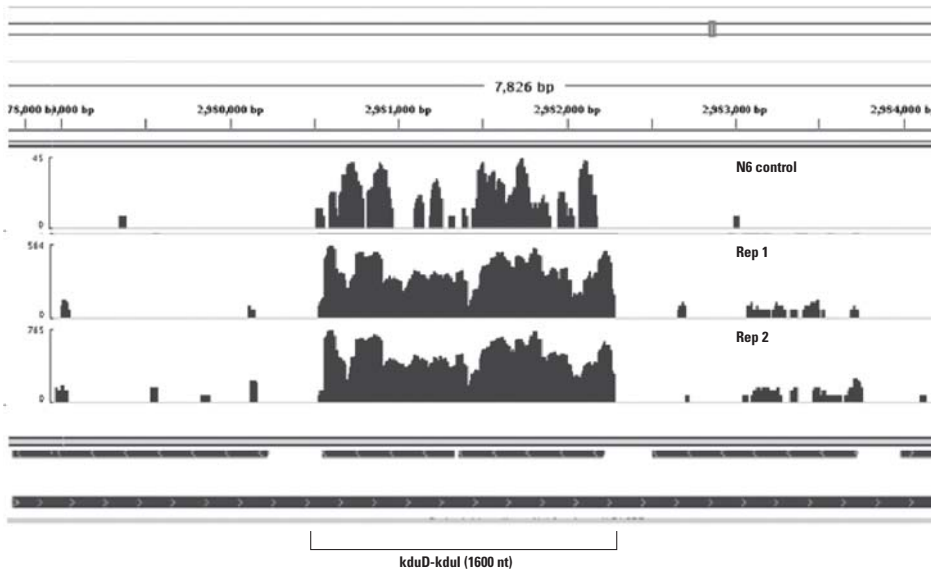


그림 5. Uniformity of transcript coverage. The horizontal axis shows a portion of the *E. coli* genome that includes the *kduD-kdul* bicistronic operon. The vertical axis shows the density of tag reads on a logarithmic scale. The sequencing results of three independent transcriptome libraries are shown; one is the random primer control and the other two are replicates of libraries produced with the Ovation Prokaryotic RNA-Seq methodology.

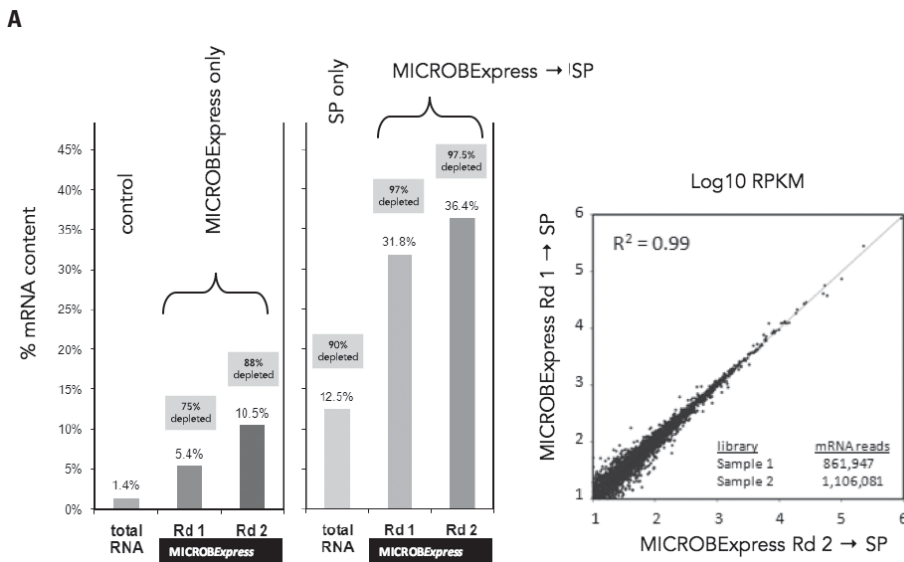
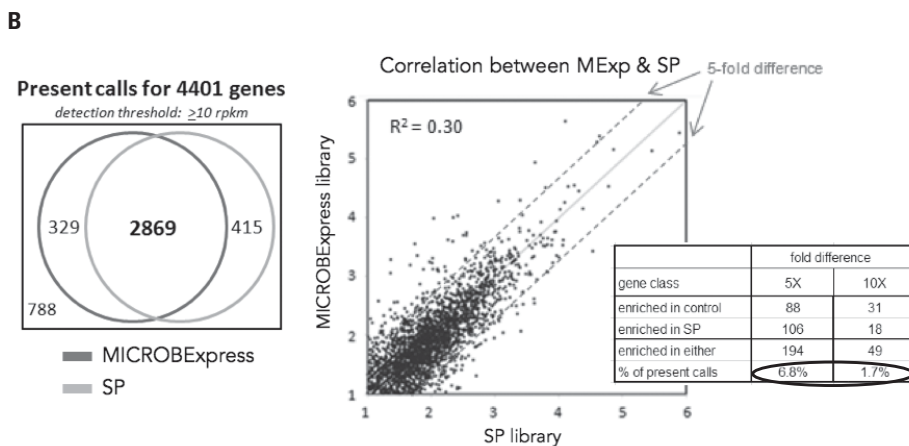


그림 6. Comparison of Ovation Prokaryotic RNA-Seq System selective priming (SP) approach with MICROExpress. (A) Left: Depletion of rRNA reads with MICROExpress alone, SP alone, or sequential treatment of samples with both methods. Right: Correlation between independent samples exposed to both depletion methods. (B) Left: Venn diagram of all 4401 annotated ORFs in *E. coli* showing genes detected by both methods, one method alone, or neither method. Right: Correlation of RPKM values between the two rRNA depletion methods. The inset shows 7% of genes differ by greater than 5-fold in RPKM values and less than 2% of genes differ by greater than 10-fold.



Ovation Prokaryotic RNA-Seq System은 고전적인 진정세균계 Eubacteria와 고세균계Archaeobacteria내 모든 원핵생물에서 재현성 있는 transcriptome 프로파일링을 만들 수 있도록 고안되었다. NuGEN에서는 이 시스템을 G:C 함량이 다양한 6종의 박테리아에 적용하여 확인하였다(그림 4). 중요한 점은 이 반복적인 실험에서 독특한 mapping reads가 매우 재현성이 있었다는 점이다. 더욱이 유전자발현 수준(reads-per-kbp/1 million, RPKM convention에 따라 정의됨, Mortazavi *et al.*, 2008)은 매우 재현성이 있고 10^4 의 범위를 커버했다. 위에서 설명한 것과 같이, 이 실험에서는 일반적으로 200,000~400,000 mapping reads의 coverage depth에서 알려진 유전자의 90% 이상을 검출하였다.

대부분의 원핵생물에서는 전사된 게놈의 전체 서열은 특징이 정확히 구분되어 있지 않았다. 일반적으로 게놈 서열의 annotation은 open reading frame 검출과 structural RNA 사이의 서열 보존(sequence conservation)에 맞춰 계산적으로 진행된다. 원론적으로 실제로 얻은 RNA-Seq 데이터는 게놈 annotation을 늘리는데 사용할 수 있다. 이렇게 하기 위해서 sequencing reads는 transcriptional unit의 5'과 3' 경계boundaries 범위를 정확하게 정해야 하고, read depth는 전사된 단편 transcribed segment이 매우 균일해야 한다. 그림 5에서는 보통수준의 전사 활성을 갖고 있는 게놈 영역에서 높게 발현된 유전자 쌍을 보여준다.

최근에 많이 사용되는 prokaryotic transcriptome library 제작방법에는 hybridization(He *et al.*, 2010)이나 nuclease를 이용해 rRNA를 물리적으로 제거하는 방법이 포함되어 있다. NuGEN은 Ovation Prokaryotic RNA-Seq System과 Ambion의 MICROBExpress Bacterial mRNA Enrichment Kit(Austin, TX)과의 성능을 비교하였다. 또한 NuGEN의 선택적인 priming 방법과 hybridization으로 rRNA를 감소시키는 방법을 결합하여 사용하였다(그림 6). Ambion에서 구매할 수 있는 *E. coli* RNA에도 동일한 처리를 적용하였는데, 이 total RNA는 99%의 rRNA로 구성되어 있다. MICROBExpress 처리는 실제로 rRNA의 양을 상당히 감소시켰고, 특히 두번째 농축enrichment 이후에 효과가 컸다. 이와 동일한 실험에서 Ovation Prokaryotic RNA-Seq

System을 사용한 경우 rRNA 제거가 보다 효과적이었다. 그러나 만약 rRNA 감소시키는 방법이 유전자 발현의 특징을 왜곡한다면, 장점이라고 할 수 없을 것이다. MICROBExpress는 최소의 편향성을 생성한다고 알려져 있으므로 MICROBExpress를 이용해 얻은 결과와 Ovation Prokaryotic RNA-Seq System의 결과를 비교하였다(그림 6). 이 벤다 이아그램에는 두 가지 방법에서 모두 검출된 유전자들과 한가지 방법에서만 발견된 유전자들, 또 다른 쪽에선 일반적으로 낮게 발현되는 유전자를 나타내었다(데이터 표기 안함). Correlation plot은 두 방법 모두에서 검출된 유전자의 관계를 보여준다. 놀랍게도 두가지 방법에 의해 검출된 유전자의 98%가 서로의 RPKM value의 10배 이내 였다. 그림 6B는 이 두 가지 매우 다른 방법 사이에 절대적인 비교와 프로토콜의 재현성이 매우 높음을 나타내었다(그 예로 그림 5 참고). 흥미롭게도 두 가지 방법을 동시에 적용할 때 즉, MICROBExpress hybridization-based depletion 다음에 Ovation Prokaryotic RNA-Seq System 적용했을 때 가장 높은 독특한 mapping RNA(97% rRNA depletion, 1% 에서 37%로 uniquely mapping reads가 증가)를 검출하였다.

Encore NGS Library System과 Ovation Prokaryotic RNA-Seq System를 조합하여 사용하면 효과적이고 경제적으로 박테리아 transcriptome 서열을 분석할 수 있다. Ovation Prokaryotic RNA-Seq System은 rRNA reduction 기술을 바탕으로 박테리아의 상보적인 whole transcriptome 프로파일링을 위한 완전한 솔루션을 제공한다. 또한 미량의 초기 시료 작용과 total RNA에서 NGS Library 구축까지 심플한 시스템으로 미생물 샘플을 다양하게 적용할 수 있게 되었다.

References

- 1 He S., Wurtzel O., Singh K., Froula J.L., Yilmaz S., Tringe S.G., Wang Z., Chen F., Lindquist E.A., Sorek R., Hugenholtz P. (2010) Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature Methods* 7: 807-812.
- 2 Levin J.Z., Yassour M., Adiconis X., Nusbaum C., Thompson D.A., Friedman N., Gnirke A., Regev A. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods* 7: 709-715
- 3 Mortazavi A., Williams B.A., McCue K., Schaeffer L., and Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5: 621-628