Global Analysis of Transcription Start Sites and Transcription Units in Bacterial Genomes

Collado-Torres L.1,2, Reyes-Quiroz A.1, Cuellar-Partida G., Moreno-Mayar V., Vargas-Chávez C.1,2, Taboada B.1, Vega-Alvarado L.1, Jimenez-Jacinto V.1, Mendoza-Vargas A.1, Grande R.1, Olvera L.1, Olvera M.1, Juárez K.1, Collado-Vides J., Morett E.

1Winter Genomics, 2LCG-UNAM, 3CCADET-UNAM, 4UUSMD-UNAM, 5IBT-UNAM, 6CCG-UNAM

Summary
With high throughput sequencing it is possible to identify at the genomic scale the transcription start sites (TSSs) and transcription units (TUs) in bacterial genomes. Due to the biological and data complexity, these analyses are challenging and require the development of custom algorithms. Critical steps involve 1) maximizing the number of reads that can be used without introducing false alignments, 2) removing biological noise: random transcription and degradation products, 3) identifying the TSSs, 4) visualizing global TSSs patterns, and 5) identifying TUs. Condensing the analyses tools into a Bioconductor package will guarantee the reproducibility of the work. The methods have been developed with data from *Escherichia coli* and *Geobacter sulfurreducens*.

Methods

**TSSs experimental methods**

- Choosing the minimum read length
- Number of putative TSSs positions
- Read coverage (frequency) of the putative TSSs
- Number of TSSs related to the stringency
- TSSs in genomic regions be removing low frequency TSSs
- TSSs in genomic regions be removing high frequency TSSs

**Transcription Units**

- Heads minus tail method
- Left position gaps method
- Differentiation method

**Transcription Start Sites**

Distribution of the distance to gene start

TSsgram sample

Conclusions
These preliminary analyses will lead to the improvement of the accuracy of promoter prediction, operon structure and regulatory networks and support a new understanding from a genome perspective of the complex regulatory network that governs transcription and regulation in bacterial genomes such as *E. coli* and *G. sulfurreducens*.

References

- Collado-Torres, L. et al. manuscript in preparation.


We acknowledge support from NIGMS-NH grant R01 GM071962-05 and from CONACYT México (83688 G.I.).