

Estudio de los patrones de distribución de grupos funcionales de genes en el genoma de cinco organismos modelo

Flavio Pazos Obregón* 1,2 , Pablo Soto 1, Diego Silvera 1,2 , José Luis Lavín 3, Ana Rosa Cortázar 3, Rosa Barrio 4, Ana María Aransay 3,5, Rafael Cantera 1,6.

1: Departamento de Biología del Neurodesarrollo, IIBCE. 2: Instituto de Matemática y Estadística “Rafael Laguarda”, Facultad de Ingeniería, UDELAR. 3: Plataforma de Análisis del Genoma, CIC bioGUNE, Derio, España. 4: Unidad de Genómica Funcional, CIC bioGUNE, Derio, España. 5: CIBERehd, ISCIII, Madrid, España. 6: Departamento de Zoología, Universidad de Estocolmo, Suecia. * fpazos@iibce.edu.uy

En organismos eucariotas, los grupos de genes que comparten función no están distribuidos aleatoriamente en el genoma. Nuestra hipótesis es que los grupos de genes con funciones biológicas relacionadas tienen patrones de distribución en el genoma que son similares. Estamos investigando la posibilidad de entrenar algoritmos de aprendizaje automático para predecir nuevas funciones de genes a partir de sus ubicaciones relativas en el genoma y para ello nos proponemos comprender los patrones de distribución de grupos funcionales de genes en el genoma de cinco organismos modelo. En esta línea, hemos comenzado a desarrollar herramientas de análisis que permitan estudiar sistemáticamente esos patrones.

Aquí presentamos “Cluster Locator”, una herramienta en línea que permite caracterizar la manera en que los genes de una lista están distribuidos a lo largo del genoma al que pertenecen. Dada una lista proporcionada por el usuario, Cluster Locator ubica y cuantifica todos los max-gap clusters formados por los genes de la lista y asocia un p-valor al porcentaje de genes de la lista que forman clusters. Presentamos además los primeros resultados de la caracterización de los patrones de distribución de cientos de grupos funcionales en cinco organismos modelo; *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans* y *Saccharomyces cerevisiae*.

Los métodos tradicionales para determinar funciones de genes no pueden acompañar la velocidad a la que se generan datos de relevancia biológica. Es por ello que han cobrado gran relevancia los métodos computacionales de predicción, siendo aquellos que se basan en principios de aprendizaje automático los que obtienen mejores resultados. Se ha entrenado a este tipo de algoritmos para predecir funciones de genes a partir de distintas características de los mismos, pero hasta donde sabemos, aun no se lo ha hecho con los patrones de distribución de grupos funcionales. Nuestros resultados representan un avance en este sentido.