

TAREA3

HERRAMIENTAS BIOINFORMÁTICAS PARA LAS CIENCIAS BIOLÓGICAS: INTRODUCCIÓN A PYTHON

Introducción

La acetilación es una modificación postraduccional que ocurre en histonas y proteínas no-histonas. Esta modificación postraduccional juega un papel clave en varios procesos incluyendo vías metabólicas y respuesta a estrés ¹. Walley *et al.*, demostraron que el hongo fitopatógeno *Cochliobolus carbonum* produce la toxina HC (“HC-toxin”), capaz de inhibir la función de la histona desacetilasa de *Zea mays*, alterando la acetilación de histonas y su respuesta inmune ².

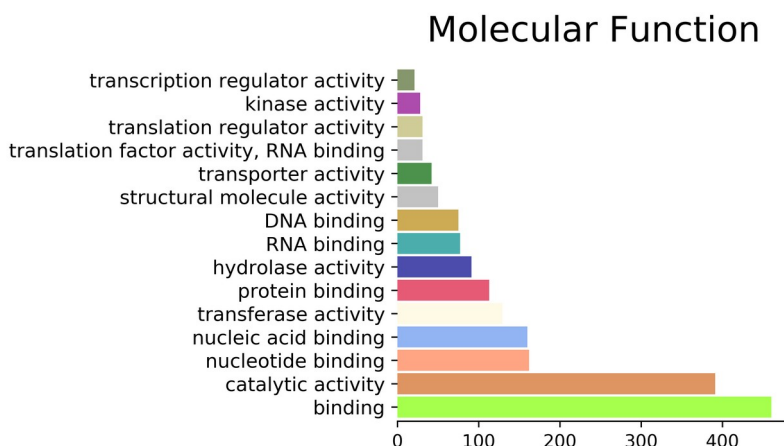
Objetivo: Clasificar por proceso biológico, función molecular y componente celular las proteínas acetiladas identificadas por Walley *et al.* Aplicar algunas funcionalidades del módulo Pandas para la manipulación, depuración y preparación de datos. Aplicar funciones de programación estructurada para resolver un problema relacionado al manejo de información.

Enviar al correo: pcr2.1@hotmail.com

Plazo para entregar la tarea: **hasta las 23:59 del domingo 16, pasado ese tiempo ya no lo calificaré.**

1. Cuántas proteínas acetiladas únicas hay en total en tabla **Zea_mays_acetylome**.
2. A partir de la tabla **Zea_mays_acetylome**, determina los **15 procesos biológicos, funciones moleculares y componentes celulares** más sobrerrepresentados. Crea un gráfico de barras para cada Aspecto.

Recomendación: Para mapear el proceso, función y componente tienes que usar la Ontología Genética, la cual se encuentra en el archivo **Ontology.csv**. Recuerda quitar columnas innecesarias y duplicados a la hora de cuantificar. Ejemplo, el siguiente gráfico es el de función molecular, tu gráfico debe salir igual, no importa si los colores son diferentes. Debes enviar las gráficas que obtuviste.



3. Dentro de los procesos biológicos se identifica el término “**response to stress**”, indica cuántas proteínas acetiladas (solo el número) están implicadas en este proceso.
4. Ahora a partir de este grupo de proteínas (que están en “**response to stress**”) determina cuántas lisinas fueron acetiladas en total.

Recomendación: la lisina acetilada está en minúscula (k) dentro de cada péptido, debes crear un bucle usando los péptidos únicos de tal forma que encuentres y cuentes las lisinas acetiladas, observa la siguiente salida, así es como puedes contar y obtener la cuenta final, la clave está en usar un bucle **for**, una de las funciones de **re** y operadores aumentativos. El número de lisinas acetiladas se encuentra entre 150 y 200.

```
['k'] Lysine in = ELGDkVPGDVK
['k'] Lysine in = qLkELGDKVPGDVK
['k'] Lysine in = kLVDDASK
['k'] Lysine in = TATkdAGR
['k'] Lysine in = LTEAAEkAK
['k'] Lysine in = DkQALQR
['k'] Lysine in = DAKLQFK
['k'] Lysine in = NDEGIDLLkDK
['k'] Lysine in = VGEAEkFAK
['k'] Lysine in = NQAESVIYQTEkQLK
['k'] Lysine in = qAVVNPENTFFSVkR
['k'] Lysine in = VESkLQELK
['k'] Lysine in = LVDDASKFLNDK
['k'] Lysine in = IPAVQELVKK
['k'] Lysine in = TTPSVVAYTkSGDR
['k'] Lysine in = FLNDkVTK
['k'] Lysine in = KQDITITGASTLPkDEVER
['k'] Lysine in = DAIDTkNQAESVIYQTEK
['k'] Lysine in = LQELkDAVAGGSTQTK
```

Los archivos de la tarea los descargas a partir de (revisa su estructura antes de abrirlos):

Ontología Genética:

<https://raw.githubusercontent.com/Bioinformatica2020/Semana6/master/Ontology.csv>

Proteínas acetiladas:

https://raw.githubusercontent.com/Bioinformatica2020/Semana6/master/Zea_mays_acetylome.tsv

Referencias

1. Kim SC, Sprung R, Chen Y, et al. Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Mol Cell*. 2006;23(4):607-618. doi:10.1016/j.molcel.2006.06.026
2. Walley JW, Shen Z, McReynolds MR, Schmelz EA, Briggs SP. Fungal-induced protein hyperacetylation in maize identified by acetylome profiling. *Proc Natl Acad Sci U S A*. 2018;115(1):210-215. doi:10.1073/pnas.1717519115