

PRUEBA AUTOMÁTICA DE HIPÓTESIS SOBRE BASES DE DATOS
BIOLÓGICAS HETEROGÉNEAS UTILIZANDO REDES DE
CONOCIMIENTO ABIERTAS

Automatic Hypotheses Testing over Heterogeneous Biological Databases using Open
Knowledge Networks

Jamil Hasan¹

jamil@uidaho.edu

(1) Associate professor of the Department of Computer Science at the University of IDAHO, Estados Unidos.

RESUMEN

Es necesario comprender las bases moleculares del dolor musculoesquelético para el desarrollo de terapias, su manejo y posible personalización. Uno de cada tres estadounidenses usa analgésicos de venta libre y una décima parte usa medicamentos recetados para controlar el dolor. Los CDC también estiman que alrededor del 20\% de los estadounidenses sufren dolor crónico. Dado que la experiencia del dolor agudo o crónico varía debido a la genética y la fisiología individual, es imperativo que los investigadores continúen encontrando nuevas terapias para tratar o controlar los síntomas. En este artículo, nuestro objetivo es desarrollar una plataforma computacional basada en el conocimiento semilla, llamada BioNursery, que permitirá a los biólogos formular hipótesis, definir y probar computacionalmente los mecanismos moleculares subyacentes al dolor. En nuestro ecosistema de conocimiento, acumulamos información seleccionada de los usuarios sobre las relaciones entre bases de datos biológicas, herramientas de análisis y contenidos de bases de datos para generar módulos de análisis biológicos, llamados pi-graphs o gráficos de procesos. Proponemos una función de mapeo desde una descripción en lenguaje natural de un modelo molecular hipotético hasta un flujo de trabajo computacional para realizar pruebas en BioNursery. Utilizamos un

sistema de curación y retroalimentación de computación colectiva, llamado Explorer, para mejorar los modelos computacionales propuestos para el descubrimiento de mecanismos moleculares y el crecimiento del ecosistema de conocimiento.

Palabras claves: Ecosistema de conocimiento, BioNursery, terapéutica, base molecular.

ABSTRACT

An understanding of the molecular basis of musculoskeletal pain is necessary for the development of therapeutics, their management, and possible personalization. One-in-three Americans use OTC pain killers, and one tenth use prescription drugs to manage pain. The CDC also estimates that about 20\% Americans suffer from chronic pain. As the experience of acute or chronic pain varies due to individual genetics and physiology, it is imperative that researchers continue to find novel therapeutics to treat or manage symptoms. In this paper, our goal is to develop a seed knowledgebased computational platform, called BioNursery, that will allow biologists to computationally hypothesize, define and test molecular mechanisms underlying pain. In our knowledge ecosystem, we accumulate curated information from users about the relationships among biological databases, analysis tools, and database contents to generate biological analyses modules, called pi-graphs, or process graphs. We propose a mapping function from a natural language description of a hypothesized molecular model to a computational workflow for testing in BioNursery. We use a crowd computing feedback and curation system, called Explorer, to improve proposed computational models for molecular mechanism discovery, and growing the knowledge ecosystem.

Palabras claves: knowledge ecosystem, BioNursery, therapeutics, molecular basic

REFERENCIAS

1. Clark, C.A.C., Helikar, T., Dauer, J.: Simulating a computational biological model, rather than reading, elicits changes in brain activity during biological reasoning. *CBE Life Sci. Educ.* 19(3), ar45 (2020)
 2. Jamil, H., Naha, K.: Mapping strategies for declarative queries over online heterogeneous biological databases for intelligent responses. In: SAC 2023, Tallinn, 27–31 March 2023. ACM (2023)
 3. Jamil, H.M.: Knowledge rich natural language queries over structured biological databases. In: BCB 2017, Boston, 20–23 August 2017, pp. 352–361 (2017)
 4. Jamil, H.M., Sadri, F.: Crowd enabled curation and querying of large and noisy text mined protein interaction data. *Distribut. Parall. Datab.* 36(1), 9–45 (2018)
 5. Liu, G., et al.: Aging atlas: a multi-omics database for aging biology. *Nucl. Acids Res.* 49(Database-Issue), D825–D830 (2021)
 6. Medlock, L., Sekiguchi, K., Hong, S., Dura-Bernal, S., Lytton, W.W., Prescott, S.A.: Multiscale computer model of the spinal dorsal horn reveals changes in network processing associated with chronic pain. *J. Neurosci.* 42(15), 3133–3149 (2022)
 7. Mou, X., Jamil, H.M.: Visual life sciences workflow design using distributed and heterogeneous resources. *IEEE/ACM TCBB* 17(4), 1459–1473 (2020)
 8. Rasu, R., et al.: Cost of pain medication to treat adult patients with nonmalignant chronic pain in the United States. *J. Manag. Care Spec. Pharm.* 20(9), 921–928 (2014)
-