

LAPORAN AKHIR
Penelitian Unggulan Perguruan Tinggi (P)



**Sistem Deteksi dan Pengobatan Penyakit Kanker
Berdasarkan Sekuens DNA**

Tahun ke 1 dari rencana 2 tahun

Laili Muffikhab, S.Kom., M.Sc.	0013117402
Edy Santoso, S.Si, M.Kom.	0014047403
Imam Cholissodin, S.Si., M.Kom.	0019078505

Dibiayai oleh :
Direktorat Jenderal Pendidikan Tinggi,
Kementerian Pendidikan dan Kebudayaan, Melalui DIPA Universitas Brawijaya
Nomor : DIPA-023.04.2.414989/2013, Tanggal 5 Desember 2012,
dan berdasarkan SK Rektor Universitas Brawijaya
Nomor : 407/SK/2013 tanggal 2 September 2013

UNIVERSITAS BRAWIJAYA
Desember 2013

HALAMAN PENGESAHAN

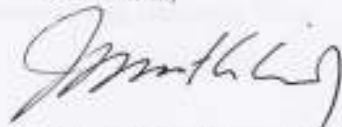
Judul : **Sistem Deteksi dan Pengobatan Berdasarkan Sekuens DNA**
Peneliti / Pelaksana : **DNA**
Nama Lengkap : **Laili Muflikah, S.Kom, M.Sc.**
NIDN : **0013117402**
Jabatan Fungsional : **Lektor**
Program Studi : **Teknik Informatika/ Ilmu Komputer**
Nomor HP : **081 232 20198**
Alamat surel (e-mail) : **laili@ub.ac.id**
Anggota (1)
Nama Lengkap : **Edy Santoso, S.Si., M.Kom.**
NIDN : **0014047403**
Perguruan Tinggi : **Universitas Brawijaya**
Anggota (2)
Nama Lengkap : **Imam Cholissodin, S.Si., M.Kom.**
NIDN : **00190780505**
Perguruan Tinggi : **Universitas brawijaya**
Tahun Pelaksanaan : **Tahun ke 1 dari rencana 2 tahun**
Biaya Tahun Berjalan : **Rp. 50.000.000,00**
Biaya Keseluruhan : **Rp. 100.000.0000,00**

Malang, 23 Desember 2013

Mengetahui,
Ketua PPTK/UB

Ir. Sutrisno, M.T.
NIP. 19570525 198701 1 001

Ketua Peneliti,



Laili Muflikah, S.Kom., M.Sc.
NIP. 19741113 2005 012 001

Menyetujui,
Pjs. Ketua LPPM UB



Prof. Dr. Ir. Siti Chuzaimi, MS.
NIP. 19530514 198002 2 001

ABSTRAK

Penyakit kanker merupakan penyakit yang sangat berbahaya, salah satu penyebabnya adalah adanya sel yang mengalami pertumbuhan tidak terkendali secara abnormal sehingga terjadi mutasi gen, dimana salah satunya adalah p53. Mutasi pada gen tersebut mengakibatkan perbedaan urutan asam amino protein p53. Mutasi p53 atau TP53 adalah perubahan genetik yang paling sering banyak ditemukan pada kanker manusia. Analisis mutasi menyebabkan banyak penelitian dan informasi tentang protein TP53 dan fungsinya. Kemajuan terbaru dalam metodologi *sequencing genom* kanker berdampak pada informasi penyembuhan dan manajemen data

Tingginya tingkat kematian akibat adanya penyakit kanker mengakibatkan banyaknya penelitian yang dilakukan untuk deteksi dini. Kajian dan analisa dalam level molekul dilakukan dengan memadukan biologi molekuler, matematika dan teknik informatika yang dikenal dengan bioinformatika. Salah satu proses di bidang informatika yang dapat dimanfaatkan untuk mencari pola DNA (sebagai penyusun asam amino) dari gen p53 penyebab kanker adalah *Sequential Pattern Mining*. *Sequential pattern mining* merupakan proses penggalian data (*data mining*) yang menghasilkan pengetahuan mengenai serangkaian kejadian-kejadian yang memiliki frekuensi kemunculan yang melebihi nilai *threshold* yang ditentukan. Pencarian pola DNA penyebab kanker ini diharapkan nantinya membantu dalam proses diagnosa dini penyakit kanker. Algoritma *PrefixSpan* merupakan algoritma yang memiliki performa yang bagus dalam hal waktu komputasi digunakan dalam penelitian ini.

Uji coba dilakukan dengan memberikan *threshold* berupa panjang sekuens (*length*) dan minimum kemunculan (*min_sup*) terhadap sejumlah database sekuens DNA dari beberapa penderita kanker pada setiap ekson yang sama. Adapun keluaran dari sistem yang dibuat diperoleh pola atau motif dari sekuens DNA bagi si penderita. Kemudian dari pola-pola tersebut diambil pola yang sama atau memiliki kemiripan yang tinggi untuk digunakan mendeteksi bagi penderita atau orang sehat (diambil data yang berbeda). Performansi dari sistem yang terlihat melalui nilai *support* dan *lift ratio*. Hasil evaluasi uji coba menunjukkan nilai *support* rata-rata yang cenderung stabil pada kisaran 0.8 memiliki nilai *lift ratio* rata-rata ≥ 1 .

Kata kunci: kanker, *sequencing genom*, *sequential pattern mining*, *prefixspan*, DNA

RINGKASAN

Penyakit kanker merupakan penyakit yang sangat berbahaya, salah satu penyebabnya adalah adanya sel yang mengalami pertumbuhan tidak terkendali secara abnormal sehingga terjadi mutasi gen. Mutasi tersebut mengakibatkan adanya perbedaan urutan susunan asam amino yang terbentuk dari sekuens DNA. Bioinformatika merupakan suatu kajian ilmu pengetahuan dengan memadukan biologi molekuler, matematika dan teknik informatika sebagai upaya untuk menganalisa dan menginterpretasikan kehidupan dalam level molekuler. Salah satu metoda dalam bidang informatika yang digunakan adalah *sequential pattern mining discovery*, untuk mendapatkan pola (motif) dari sekuens DNA sebagai deteksi dini bagi penderita penyakit kanker. Adapun data set yang digunakan dalam penelitian ini adalah beberapa gen yang terkena penyakit kanker payudara (*breast cancer*) yang diambil dari situs gen bank (NCBI) dengan alamat URL: www.ncbi.nlm.nih.gov. Algoritma *Prefix Span* digunakan dalam pencarian pola beberapa gen yang terjangkit. Dengan memberikan threshold berupa panjang sekuens database DNA, minimum kemunculan didapatkan pola-pola yang memiliki performansi yang tinggi. Pola tersebut juga digunakan untuk deteksi terhadap data baru (baik bagi penderita maupun bukan). Hasil pendeteksian didapatkan terlihat memiliki tingkat akurasi yang tinggi.

DAFTAR PUSTAKA

- Bai, L. & Zhu, W.G. (2006). *Journal of Cancer Molecules* 2 (4): 141-153. Guilin, China
- Barnes, M. R., & Gray, I. C. (2003). *Bioinformatics For Genetics*. Chichester, England: John Wiley & Sons Ltd
- Brown, T., & Brown, D. (2005). *Mutagenesis and DNA repair*. Dipetik Mei 17, 2012, dari <http://www.atdbio.com>
- Cesario, A., & Frederick, M. B. (2011). *Cancer Systems Biology, Bioinformatics and Medicine : Research and Clinical Applications*. New York: Springer Dordrecht Heidelberg London.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers
- D. Benson, D.J. Lipman, and J. Ostel I (1993) GenBank. *Nuclei Acid Resc* 21, 2963-2965.
- C. Sander (2001) Bioinformatics challenges in 2001. *Bioinformatics* 17, 1-2.
- Hen, J. Y., & Lonardi, S. (2010). *Biological Data Mining*. Chapman & Hall/CRC Broken Sound Parkway NW
- Jones, N. C., & Pevzner, P. A. (2004). *An Introduction to Bioinformatics Algorithms*. Massachusetts London: The MIT Press Cambridge
- L. Hjortsberg, J.M. Rubio-Navado, D. Hamroun, M. Claustre, C. Bérout and T. Soussi. The p53 Mutation HandBook (dipetik Juni 2, 2012) dari http://p53.fr/TP53_database_download/TP53%20handbook/TP53_handbook.html
- Pang-Ning Tan, Michael Steinbach and Vipin Kumar. (2006). *Introduction to Data Mining*, Pearson Education.
- Parthasarathy, S. (2007). *Building Genetic Medicine : technology, breast cancer, and the comparative politics of health care*. Massachusetts London: The MIT Press Cambridge.
- Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, et al. (2001) "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," Proc. 2001 Int'l Conf. Data Eng. (ICDE '01), pp. 215-224.
- Pustai L., Lewis, C., & Yap, E. (1996). *Cell Proliferation in Cancer- Regulation Mechanisms of Neoplastic Cell Growth*. Oxford: Oxford University Press.
- Thierry Soussi. 2011. *TP53 Mutations in Human Cancer: Database Reassessment and Prospects for the Next Decade*. *Adv Cancer Res* 110: 107-139
- Wiseman SM, Griffith O, Deen S, Rajput A, Masoudi H, Gilks B, Goldstein L, Gown A, Jones S. (2007). *Identification of Molecular Markers Altered During Transformation Of Differentiated Into Anaplastic Thyroid Carcinoma*. *Archives of Surgery* (2007) 142:717-727, discussion 727-9.