

PROFILING EPIGENETIK H3K27Ac MENGGUNAKAN DATA ChIP-Seq
SEL KANKER PARU-PARU AKIBAT ASAP ROKOK
PADA PROGRAM R

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Jurusan Statistika



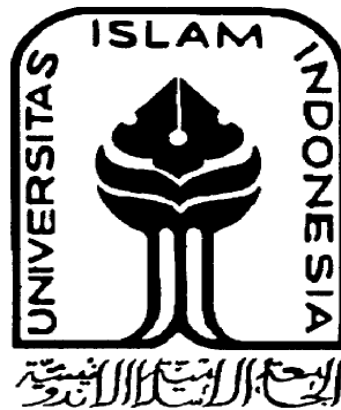
Tri Atmaja Huda
14611213

JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2019

**PROFILING EPIGENETIK H3K27Ac MENGGUNAKAN DATA ChIP-Seq
SEL KANKER PARU-PARU AKIBAT ASAP ROKOK
PADA PROGRAM R**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Jurusan Statistika



Tri Atmaja Huda
14611213

**JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2019**

HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR

Judul : *Profiling* Epigenetik H3K27Ac Menggunakan Data ChIP-Seq Sel Kanker Paru-paru Akibat Asap Rokok Pada Program R

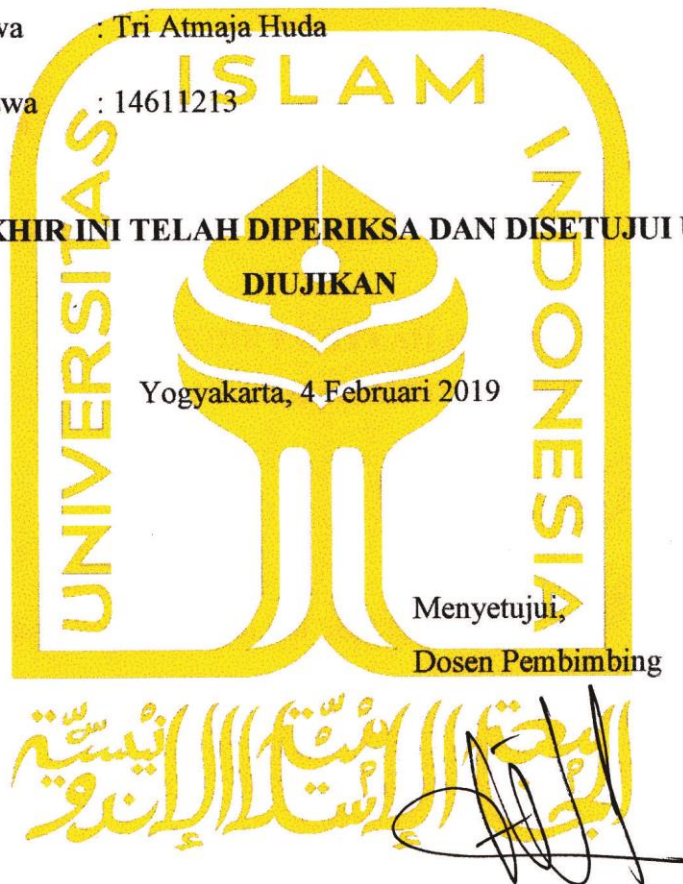
Nama Mahasiswa : Tri Atmaja Huda

Nomor Mahasiswa : 14611213

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK

DIUJIKAN

Yogyakarta, 4 Februari 2019



Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.

**HALAMAN PENGESAHAN
TUGAS AKHIR**

**PROFILING EPIGENETIK H3K27Ac MENGGUNAKAN DATA ChIP-Seq
SEL KANKER PARU-PARU AKIBAT ASAP ROKOK
PADA PROGRAM R**

Nama Mahasiswa : Tri Atmaja Huda

Nomor Mahasiswa : 14611213

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 14 FEBRUARI 2019**

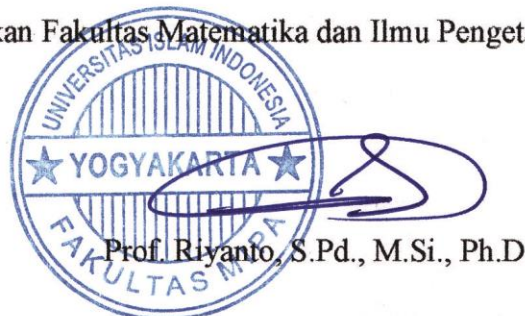
Nama Penguji

1. Saepudin, M.Si., Ph.D., Apt.
2. M. Hasan Sidiq Kurniawan, S.Si., M.Sc.
3. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.

Tanda Tangan

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



KATA PENGANTAR



Assalamu'alaikum Warahmatullaahi Wabarakaatuh

Alhamdulillah segala puji bagi Allah SWT atas segala rahmat dan karunia-Nya yang telah dilimpahkan kepada penulis sehingga penulis dapat menyelesaikan penulisan tugas akhir ini yang berjudul **“Profiling Epigenetik H3K27Ac Menggunakan Data CHIP-Seq Sel Kanker Paru-paru Akibat Asap Rokok Pada Program R”** sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan jenjang strata satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Shalawat serta salam semoga selalu tercurah kepada Nabi Muhammad SAW serta para sahabat dan pengikutnya yang senantiasa menjaga keimanan dan keislamannya hingga akhir hayatnya.

Penyusunan tugas akhir ini tidak terlepas dari dukungan, bantuan, arahan, dan bimbingan dari berbagai pihak. Untuk itu dengan hati yang tulus penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

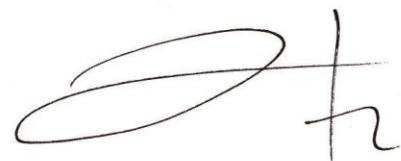
1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Bapak Dr. Edy Widodo, S.Si., M.Si. selaku Ketua Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam yang juga telah banyak membantu dan memberi ilmu dan wawasan baru kepada Penulis.
3. Ibu Dr. techn. Rohmatul Fajriyah, S.Si., M.Si. selaku dosen pembimbing tugas akhir penulis, yang dengan sabar membimbing penulis untuk menyelesaikan tugas akhir ini. Semoga Allah memberikan kesehatan kepada Ibu dan ilmu yang diberikan dapat bermanfaat bagi Penulis.
4. Kedua orang tua tercinta, Ibu dan Bapak serta kedua kakak kandung yang selalu setia mendoakan, memberikan dukungan, kasih sayang, menemani serta memberikan motivasi kepada penulis selama menyelesaikan tugas akhir ini.

5. Seluruh staf pengajar Jurusan Statistika Universitas Islam Indonesia terimakasih atas ilmu yang diberikan semoga senantiasa diberikan kebaikan oleh Allah SWT.
6. Teman-teman: Avitariella, Baist, Janu, Zainul, Sidqi, Rizal dan Agil yang selalu menjadi tempat berbagi, mendukung, memberi semangat dan do'a. Terimakasih atas segalanya, semoga silaturahmi kita tetap terjaga dan selalu di ridhoi Allah SWT.
7. Teman-teman bimbingan TA : Husain, Hawila, Nanda, Gilang, Shoddiq, Lina, Ika, Ummi, Leni, Himelda, Maida, Ari, dan teman-teman bimbingan angkatan 2015 yang sudah sama-sama berjuang, saling mengingatkan dan memberi motivasi serta dorongan untuk menyelesaikan Tugas Akhir ini.
8. Teman-teman seperjuangan Statistika UII Angkatan 2014 sukses dan berjuanglah menuju hari esok yang lebih baik.
9. Semua pihak yang terlibat dalam penyusunan tugas akhir ini yang tidak bisa disebutkan satu per satu yang telah memberikan dukungan dan motivasi selama menyelesaikan tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran dari semua pihak yang bersifat membangun selalu penulis harapkan demi kesempurnaan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat. Akhir kata, semoga Allah SWT senantiasa meridhai segala usaha kita. *Aamiin yaa rabbal'aalamiin.*

Wassalamu'alaikum Warahmatullaahi Wabarakaatuh

Yogyakarta, 4 Februari 2019



Tri Atmaja Huda

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	viii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN.....	x
TAKARIR.....	xi
PERNYATAAN.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
BAB III LANDASAN TEORI.....	9
3.1 Bioinformatika	9
3.2 Kanker Paru-Paru.....	13
3.3 DNA.....	13
3.4 RNA	13
3.5 Histone	14
3.6 Epigenetika	14
3.7 <i>Markov Model</i>	15
3.8 <i>Hidden Markov Model</i>	17
3.8.1 Persoalan Utama HMM.....	18
3.8.2 Solusi Persoalan 1	18
3.8.3 Solusi Persoalan 2	19
3.8.4 Solusi Persoalan 3	20

3.9	ChIP-Seq.....	21
3.10	<i>Trimming & Filtering</i>	22
3.11	<i>Sequence Alignment</i>	22
3.12	<i>Quality Check</i>	23
3.13	<i>Peak Calling</i>	23
3.14	<i>Peak Calling Menggunakan HMM</i>	23
3.14.1	Algoritma <i>Baum-Welch</i>	24
3.14.2	Algoritma Viterbi	26
3.15	<i>Annotation</i>	24
3.16	<i>Functional Enrichment Analysis</i>	26
3.17	<i>Software R</i>	26
BAB IV METODOLOGI PENELITIAN		28
4.1	Populasi dan Sampel Penelitian	28
4.2	Jenis dan Sumber Data	28
4.3	Variabel Penelitian	28
4.4	Metode Analisis Data	28
BAB V ANALISIS DAN PEMBAHASAN		31
5.1	Deskripsi Data	31
5.2	<i>Alignment</i>	32
5.3	<i>Quality Check</i>	33
5.4	<i>Peak Calling</i>	34
5.5	<i>Annotation & Visualization</i>	36
5.6	<i>Functional Enrichment Analysis</i>	37
5.7	Perbandingan Hasil Analisis Dengan Penelitian Terhadulu.....	41
BAB VI KESIMPULAN DAN SARAN		42
6.1	Kesimpulan.....	42
6.2	Saran	43
DAFTAR PUSTAKA		44
LAMPIRAN.....		47

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terdahulu	7
Tabel 5.1 Sampel Data ChIP-Seq.....	31

DAFTAR GAMBAR

Gambar 3.1 <i>Markov Chain</i>	16
Gambar 4.1 Alur Penelitian.....	29
Gambar 5.1 (a) <i>Sequence DNA Sampel Data Control</i>	31
Gambar 5.1 (b) <i>Sequence DNASampel Data Case</i>	32
Gambar 5.2 (a) <i>Mapping Statistics Sampel Data Control</i>	32
Gambar 5.2 (b) <i>Mapping Statistics Sampel Data Case</i>	32
Gambar 5.3 (a) <i>Persentase Bacaan Unik Sampel Data Control</i>	33
Gambar 5.3 (b) <i>Persentase Bacaan Unik Sampel Data Case</i>	33
Gambar 5.4 (a) <i>Quality Score Sampel Data Control</i>	34
Gambar 5.4 (b) <i>Quality Score Sampel Data Case</i>	34
Gambar 5.5 <i>ChIP Peaks</i>	35
Gambar 5.6 <i>Coverage Plot</i>	35
Gambar 5.7 <i>Pie Chart</i>	36
Gambar 5.8 <i>Upset Plot</i>	37
Gambar 5.9 <i>Dot Plot Pathway Analysis</i>	38
Gambar 5.10 <i>Dot Plot Disease Ontology</i>	39
Gambar 5.11 <i>Cnet Plot</i>	40

DAFTAR LAMPIRAN

Lampiran 1 : *Link* Data

Lampiran 2 : *Syntax* R untuk Analisis

TAKARIR

AHRR	: <i>Aryl-Hydrocarbon Receptor Repressor</i>
AT2	: <i>Alveolar Epithelial Type 2</i>
ChIP	: <i>Chromatin Immunoprecipitation</i>
ChIP-Seq	: <i>Chromatin Immunoprecipitation Sequencing</i>
DNA	: <i>Deoxyribonucleic Acid</i>
EMBL-EBI	: <i>The European Bioinformatics Institute</i>
H3K27Ac	: <i>Histone H3 Lysine 27 Acetylation</i>
HMM	: <i>Hidden Markov Model</i>
NGS	: <i>Next Generation Sequencing</i>
IARC	: <i>The International Agency for Research on Cancer</i>
JFK	: <i>Jinfukang</i>
NLT	: <i>Non-Tumor Lung</i>
RNA	: <i>Ribonucleic Acid</i>
WHO	: <i>World Health Organization</i>

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 4 Februari 2019



Tri Atmaja Huda

**PROFILING EPIGENETIK H3K27Ac MENGGUNAKAN DATA ChIP-Seq
SEL KANKER PARU-PARU AKIBAT ASAP ROKOK
PADA PROGRAM R**

Oleh: Tri Atmaja Huda

Program Studi Statistika Fakultas Matematik dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

E-mail: 14611213@students.uui.ac.id

INTISARI

Kanker paru-paru merupakan jenis kanker yang paling banyak diderita dan merupakan penyebab pertama kematian pada kanker, hal ini disebabkan karena prognosis yang buruk terhadap kanker ini di seluruh dunia. Faktor lingkungan seperti asap tembakau dapat memodulasi pembentukan dan pemeliharaan metilasi DNA, dan dengan demikian dapat mempengaruhi penyakit kanker paru-paru melalui mekanisme yang kompleks. Oleh karena itu, pada penelitian ini dilakukan profiling epigenetik untuk melihat gen-gen yang berpengaruh terhadap kanker paru-paru, sehingga diharapkan dapat memberikan treatment yang spesifik terhadap hasil profiling untuk meningkatkan kualitas prognosis, mempercepat proses penyembuhan pasien kanker paru-paru, serta memperkecil tingkat kematian yang disebabkan kanker paru-paru. Metode yang digunakan dalam penelitian ini adalah analisis data ChIP-Seq terhadap modifikasi histone H3K27Ac, penanda epigenomik yang terkait dengan ekspresi gen aktif dalam sel kanker paru yang diakibatkan karena asap rokok, berdasarkan paket-paket pada Bioconductor menggunakan program R. Berdasarkan penelitian ini diperoleh kelompok gen yang memiliki mekanisme, fungsi, dan peranan yang sama dari gen yang mengalami perubahan akibat paparan asap rokok yaitu signaling by receptor tyrosine kinases merupakan yang paling dominan dengan 441 gen dan nilai adjusted P-value adalah 0,015 serta rasio gennya sebesar 0,044. Kemudian asosiasi penyakit yang dihasilkan paling dominan adalah non-small cell lung carcinoma dengan 415 gen dan nilai adjusted P-value adalah 0,00206 serta rasio gennya sebesar 0,057. Gen-gen yang berpengaruh terhadap non-small cell lung carcinoma atau penyakit kanker paru-paru adalah CYP2E1, NAPSA, CYP1A1, EGFR, serta 411 gen lainnya.

Kata Kunci : *Kanker paru-paru, ChIP-Seq, Profiling epigenetik, Gen*

**EPIGENETIC PROFILING H3K27Ac USING DATA ChIP-Seq
LUNG CANCER CELLS CAUSED BY SMOKE CIGARETTE
IN R PROGRAM**

By: Tri Atmaja Huda

Statistics Department Faculty of Mathematics and Natural Sciences

Universitas Islam Indonesia

E-mail: 14611213@students.uii.ac.id

ABSTRACT

Lung cancer is the most common type of cancer and is the first cause of death in cancer, this is due to the poor prognosis of this cancer throughout the world. Environmental factors such as tobacco smoke can modulate the formation and maintenance of DNA methylation, and thus can affect lung cancer through complex mechanisms. Therefore, in this study epigenetic profiling was conducted to look at genes that influence lung cancer, so that it is expected to provide specific treatment for profiling results to improve the quality of prognosis, accelerate the healing process of lung cancer patients, and reduce levels deaths caused by lung cancer. The method used in this study was data analysis of ChIP-Seq against histone modification of H3K27Ac, epigenomic markers related to the expression of active genes in lung cancer cells caused by cigarette smoke, based on packages in Bioconductor using the R program. Based on this study a group was obtained genes that have the mechanism, function, and the same role of genes undergoing changes due to exposure to cigarette smoke, namely signaling by receptor tyrosine kinases are the most dominant with 441 genes and the adjusted P-value value is 0.015 and the gene ratio is 0.044. Then the most dominant disease association was non-small cell lung carcinoma with 415 genes and the adjusted P-value value was 0.00206 and the gene ratio was 0.057. The genes that affect non-small cell lung carcinoma or lung cancer are CYP2E1, NAPSA, CYP1A1, EGFR, and 411 other genes.

Keywords: *Lung Cancer, ChIP-Seq, Epigenetic Profiling, Genes*