EST-based detection and analysis of mammalian transcripts

Shobhit Gupta

Nov 2005

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich für Mathematik und Informatik
der Freien Universität Berlin
vorgelegte
Dissertation

1. Reviewer: Prof. Dr. Martin Vingron

2. Reviewer: Prof. Dr. Ulf Leser

Defence Date: 16-Nov-2005

For my Parents

Acknowledgments

My first days in berlin were actually my first days outside India. That had quite an impact on my ability to get adapted to the new environment. But for that matter, the CMB department of Max Planck Institute was a very good choice. The group members were extremely friendly and easily approachable for both personal as well as work-related issues.
To start with the acknowledgments I thank my supervisor, Martin Vingron. I am most grateful to Martin for offering me an opportunity to work under his supervision. His excellence at work as well as clarity of ideas was obvious to me beforehand. Additionally he has proved to be a fantastic supervisor both in terms of project-related and personal guidance.
I express my deepest gratitude to Stefan Haas. He has been my mentor throughout. I sincerely appreciate his patience with me especially during the paper-writing sessions (read cycles). I sincerely thank my collaborators Dorothea Zink and Bernd Korn for providing a close experimental support to my computational work. Bernd has also been a critical and very helpful member of my Phd committee.
I deeply appreciate the company of all past and present CMB group members. The discussions during the weekly group meetings have been very fruitful for the advancement of my project as well as for the broadening of my knowledge-base. I would also like to thank our system administrator, Wilhelm Ruesing. He has been relentless and exceptionally prompt in cracking down on infra-structural problems.
On a personal note, special thanks goes to Birgit Pils and Johanna Holbrook for great help during my initial days. Additionally, I am thankful to my friends in Germany especially Stefanie Koeniger, Shivendra Kishore, Harindar Keer and Abha Singh. They have been very motivating during the inevitable low periods of graduate research.
Lastly but perhaps the most significantly, I appreciate the support of my mother before and during my graduate studies. I would have been nowhere without her enormous struggle for my education.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviations in alphabetical order

A	adenine
AEDB	alternative exon database
ASAP	alternative splicing annotation database
ASD	alternative splicing database
BLAST	Basic local alignment search tool
bp	base pair(s)
°C	degree celsius
C	cytosine
CAGE	cap analysis gene expression
cDNA	complementary DNA; DNA synthesized from mRNA by RT
CDS	coding sequence
CGH	Comparative genomic Hybridization
ChIP	chromatin immunoprecipitation
DNA	deoxyribonucleic acid
EST	expressed sequence tag
G	guanine
Kb	kilobase(s); 1,000 nt
Mb	megabase(s); 1,000,000 nt
mRNA	messenger RNA
nt	nucleotide
PCR	polymerase chain reaction
polyA	polyadenylation signal
RNA	ribonucleic acid
RT	reverse transcription
RTase	reverse transcriptase; an enzyme
RT-PCR	reverse transcription-polymerase chain reaction
SAGE	serial analysis of gene expression
snRNP	small nuclear ribonucleoprotein
T	thymine
U	uracil
UTR	untranslated region; part of mRNA transcripts

For a compilation of IUPAC symbols for nucleotide nomenclature see Appendix C.