Appendix

A The Genetic code

The genetic code consists of 64 triplets of nucleotides. These triplets are called *codons*. With three exceptions, each codon encodes for one of the 20 amino acids used for the synthesis of proteins. That produces some redundancy in the code: most of the amino acids being encoded by more than one codon. The three codons which do not code for any amino acid are called *stop* codons. Additionally, the codon, AUG signals the start of translation.

The genetic code can be expressed as either RNA codons or DNA codons. RNA codons occur in messenger RNA (mRNA) and are the codons that are actually *read* during the translation procedure. But each mRNA molecule acquires its sequence of nucleotides by transcription from the corresponding gene, therefore the RNA codons can the converted into DNA based codons. Figure A.1 lists all the codon-amino acid relationships.

First	Second				Third
	U	**C**	**A**	**G**	
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine	**U**
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine	**C**
	UUA Leucine	UCA Serine	UAA STOP	UGA STOP	**A**
	UUG Leucine	UCG Serine	UAG STOP	UGG Tryptophan	**G**
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine	**U**
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine	**C**
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine	**A**
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine	**G**
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	**U**
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine	**C**
	AUA Isoleucine>	ACA Threonine	AAA Lysine	AGA Arginine	**A**
	AUG Methionine	ACG Threonine	AAG Lysine	AGG Arginine	**G**
G	GUU Valine	GCU Alanine	GAU Aspartic acid	GGU Glycine	**U**
	GUC Valine	Valine Alanine	GAC Aspartic acid	GGC Glycine	**C**
	GUA Valine	GCA Alanine	GAA Glutamic acid	GGA Glycine	**A**
	GUG Valine	GCG Alanine	GAG Glutamic acid	GGG Glycine	**G**

Figure A.1: The Universal Genetic Code. There are 64 different codons out of which 61 correspond to one of the 20 amino acids. The remaining 3 codons signal the termination of the growing poly-peptide chain. Image reproduced from http://www.scienceden.com/mbiology/concepts/translation.

B The RT-PCR Experiments

The RT-PCR experiments were performed by Dr. Dorothea Zink in the lab of Dr. Bernhard Korn at German Cancer Research Center, Heidelberg, Germany.

B.1 Experimental Protocol

Total RNA was prepared using the single-step guanidinum method according to the manufacturer's instructions (TRIZOL, Gibco BRL). First strand cDNA synthesis was carried out in 20μl reaction using the Omniscript Reverse transcriptase (Qiagen) and the oligo(dT) primers with 2μg of total RNA. RT-PCR was carried out in a 20μl reaction in 1 x buffer [1.5 mM Mg2+, 0.2 mM dNTPs, 0,4 μM primers each, 1 Unit of Taq polymerase (Roche)] and 1 μl of cDNA. Amplification steps were as follows: 95°C for 90 sec; 9 cycles of 94°C for 20 sec, 64°C for 10 sec decreasing the annealing temp for 1°C with each cycle (touchdown), 72°C for 20 sec; followed by 30 cycles of 94°C for 20 sec, 55°C for 10 sec, 72°C for 20 sec, followed by an extension at 72°C for 10 min. For the refined PCR, the amplification step was repeated with identical PCR conditions but with 2 μl of PCR product instead of 1 μl of cDNA. All PCR products were resolved on 2% agarose gels run at 90V/20 cm for 1.5 h in TAE buffer. As a control, a fraction of variants were sequenced using the ABI Prism BigDye Terminators and the ABI Prism 3100 sequencer (Applied Biosystems).

B.2 List of tissues

Tissue
Adrenal gland
Antrum
Bladder
Brain, Brain fetal
Breast, Breast tumor
Cervix
Coecum
Colon, Colon tumor
Corpus
Duodenum
Endometrium
Heart, Heart fetal
Kidney, Kidney fetal, Kidney tumor
Large Intestine
Liver, Liver fetal
Lung, Lung tumor
Lymph nodes tumor
Marrow
Muscle
Oesophagus
Ovary, Ovary tumor
Pancreas
Placenta
Prostate gland
Rectum
Salivary gland
Spinal cord
Spleen, Spleen fetal
Small Intestine
Stomach, Stomach tumor
Testis, Testis tumor
Thymus, Thymus fetal
Thyroid gland
Tonsil
Retina
Uterus, Uterus tumor

Table B.1: Tissues for which RT-PCR experiments were performed

C IUPAC nucleotide ambiguity codes

Symbol	Meaning	Nucleic Acid
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
U	U	Uracil
M	A or C	
R	A or G	
W	A or T	
S	C or G	
Y	C or T	
K	G or T	
V	A or C or G	
H	A or C or T	
D	A or G or T	
B	C or G or T	
X	G or A or T or C	
N	G or A or T or C	

Reference: IUPAC-IUB SYMBOLS FOR NUCLEOTIDE NOMENCLATURE: Cornish-Bowden (1985) Nucl. Acids Res. 13: 3021-3030.

D Availability

D.1 Quality computation software

All software related to the computation of quality values is available as a composite perl/C pipeline upon request.

D.2 T-STAG software

All software related to the build up, maintenance, visualization and querying of T-STAG is available on request. Please contact `shobhit.gupta@molgen.mpg.de`.

D.3 T-STAG Database

All the gene/transcript expression data is made available via a web-queryable relational database system. Visit `http://tstag.molgen.mpg.de` for an overview of the available resources.

E List of related publications

Gupta S., Vingron M., Haas S.A. (2005) *T-STAG: Resource and web-interface for Tissue-Specific Transcripts And Genes.* Nucleic Acids Research, in press.

Gupta S., Zink D., Korn B., Vingron M., Haas S.A. (2004) *Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing.* BMC Genomics, 5:72.

Gupta S., Zink D., Korn B., Vingron M., Haas S.A. (2004) *Genome wide identification and classification of Alternative Splicing based on EST data.* Bioinformatics, 20(16): 2579-2585.

F Curriculum vitae

M. Tech. Shobhit Gupta
Goerzallee 135
D-12207 Berlin, Germany

Tel.: $++49 - 30 - 99\,21\,22\,76$
E-mail: Shobhit.Gupta@molgen.mpg.de

Date of birth: July 11, 1978
Place of birth: Ghaziabad, India
Citizenship: Indian
Marital status: Single

Degrees

M. Tech in Biochemical Engineering and Biotechnology, Indian Institute of Technology, India.
Degree with honors

- One year research experience in implementing a bayesian gene prediction program.

B. Tech in Biochemical Engineering and Biotechnology, Indian Institute of Technology, India.
Degree with honors.

- Research experience in metabolic control analysis for detection of weak drug targets.

Education

2002 - present PhD student in the department of Computational Molecular Biology, MPI for Molecular Genetics, Berlin

1997 - 2002 Engineering student at Indian Institute of Technology, Delhi, India

Prizes and Awards

2001 - 2002 Scholarship for masters project from IBM, India Research Lab.

2001 - 2002 Institute scholarship for clearing GATE Examination.

2001 All India Rank *0017* in GATE examination.

1998 - 2001 Brij Mohan Gupta scholarship (awarded annually to three outstanding students of IIT, Delhi).

1997 - 1998 Merit cum Means institute scholarship.

1997 Ranked among top *0.5%* students in Joint Entrance Exam(JEE) which included more than 0.2 million people.

1994 - 1996 Merit based fees exemption scholarship in school.

1992 All India Rank *0482* in NSTC examination.