

A STRUCTURAL GENOMICS APPROACH TO THE REGULATION OF APOPTOSIS: CHIMP VS. HUMAN

JESSICA AHMED, STEFAN GÜNTHER, FRIEDRICH MÖLLER, ROBERT PREISSNER

Structural Bioinformatics Group, Institute of Molecular Biology and Bioinformatics, Charité-University Medicine Berlin, Arnimallee 22, 14195 Berlin, Germany

Keywords: chimpanzee, cancer, apoptosis, DNA-repair

Abstract

After the sequencing of the human genome, the publication of the genome of our nearest relative, the chimpanzee (*Pan troglodytes*) provided groundbreaking data improving the understanding of the recent human evolution. There are about forty million changes, most of them single nucleotide substitutions, which teach us about ourselves, both in terms of similarities and differences with chimpanzees. From a medical point of view differences in incidence and severity of diseases are of special importance to pinpoint novel targets and to develop innovative therapies.

This analysis focuses on the cognition that chimpanzees rarely suffer from cancer. To elucidate possible reasons for this finding, we compare differences regarding apoptosis and DNA-repair on different levels of chromosome organization, gene structure, post-transcriptional and post-translational modifications to functional changes in protein structures. The result is a complex pattern of subtle variances and a few large-scale changes.

1. Introduction

Today, one of the most frequent causes of death is cancer. Among the malignant neoplasms, the most prominent types of cancer in Europe in 2006 are breast (30.9%), prostate (24.1%), lung (11.2%), colon, and rectum (13%) cancer [1]. Possible reasons for cancer are determined by environmental or intrinsic factors.

Nutritional differences and other ecological causes are responsible for disparity in cancer incidence, but the tenfold increase, as observed for carcinomas of breast, ovary, lung, stomach, colon and, rectum cancer, between chimpanzee and human [2] cannot be explained coherently by such arguments. This directs the focus on intrinsic factors like susceptibility, and tolerance, which depend on genetic factors. Genetic variability leads to a diversity of cancer incidence in different populations. The Japanese and non-Hispanic white men show the highest cancer risk, whereas the Alaska's men show the lowest one [3]. Especially, the prostate [4] and breast [5] types of cancer do rarely occur in the Eskimo population. However, the Indians and Aleuts show a significantly increased rate of prostate cancer compared to the native Alaska's men [5]. Well-known genetic abnormality responsible for different cancer incidence exhibits e.g. the breast cancer gene BRCA1, which possesses mutations in breast cancer patients. This protein acts as a tumor suppressor and is involved in processes of the DNA-repair. In most instances a DNA-damage leads either to DNA-repair or apoptosis. However, if these programs fail, cancer may occur. It is generally accepted that genetic variants lead to disruption of the apoptotic and/or DNA-repair pathways and therewith to a preferential cancer development [6, 7]. It was disclosed that a recurrent mutation of the BRCA1-gene occurs in cancer patients of the Chinese population [8]. Furthermore, in case of the Philadelphia

chromosome a short variant of chromosome 22 exists. This short chromosome 22 leads to a transcription of a new gene, named BCR-ABL, which is involved in the Leukaemia development [9]. Currently, several Single Nucleic Polymorphisms (SNPs), which are involved in a higher cancer susceptibility, are collected in databases [10-12].

Summing up, it can be concluded, that distinct genetic differences between chimp and human exist, which are responsible for the increased cancer incidence in human, and might be associated with apoptosis or DNA-repair. An important question, which may cause even therapeutic consequences, is: How can these genetic factors, which affect cancer susceptibility, be detected? One possibility is to compare the machinery and regulation of DNA-repair and apoptosis between human and chimpanzee.

The chimpanzee, our closest living relative, has a genome wide identity of more than 98 percent [13, 14]. Nevertheless, chimpanzees rarely suffer from cancer. Especially prostate [15], breast, and lung cancer and spontaneous neoplasms [16] do rarely occur in chimpanzees population [17-19]. The sequencing of the chimpanzee genome elucidates that human and chimps diverged in about 35 million single nucleotide changes and about 5 million insertion/deletion events (indels) [13]. Previous works have shown that differences on nucleotide- and amino acid levels between chimp and human proteins, which are involved in cancer development, are existent [18]. These findings support the hypothesis that the low cancer incidence in chimpanzees has genetic reasons.

Puente et al. selected 333 cancer relevant genes and found about 1,500 changes in the amino acid sequence of the proteins [18], but their detailed analysis remains restricted to a fistful of proteins. For instance, the coding sequence for the tumor suppressor p53 has an arginine at position 73, which does only occur in the human lineage, whereas other primates like the chimpanzee, gorilla and mandrill have a proline at this position [14]. Furthermore, in the work of Nielsen et al. about 13,700 annotated genes of chimps and human were analyzed and their selectivity was calculated [20]. In this comparison, it was pointed out that several cancer related proteins, which are involved in apoptosis, tumor suppression and in the cell cycle control are positively selected. This is supported by the results of the Nature consortium [13], which determined the selectivity of groups of proteins in the GeneOntology [21]. In this case, the apoptotic pathway and the proteins, which belong to the cell cycle control, are also positively selected [13].

To pinpoint the reasons for the low cancer incidence in chimps and hopefully to deduce novel possibilities for cancer treatment, we consider an enlarged dataset of about 500 proteins which are either involved in the apoptotic or in the DNA-repair pathway. In contrast to previous studies, we will analyze these proteins on different levels: chromosomal organization, gene structure, posttranscriptional and posttranslational modifications as well as structural and functional changes.

2. Data and Methods

Based on the assumption, that a change in the regulation of DNA-repair or apoptosis might be the reason for the decreased cancer susceptibility in chimpanzee, 493 proteins involved in these processes were selected. The protein and DNA-alignments were

extracted from the Ensembl database vol. 42-44 [22]. Instead of just analyzing the number of mutations in the gene and protein sequences of both species, other topics were considered in this work as well.

These proteins were analyzed on different levels of biological organization:

- Chromosome and gene structure
- Post-transcriptional modifications and post-translational modifications
- SNPs and indels leading to structural and functional changes.

First, the chromosome organization, which means e.g. the protein distribution over the chromosomes of the 493 selected proteins, was considered for both species. Moreover, the positive selection of the group of genes and also of the single genes was calculated according the work of Nielson et al. [20]. To analyze the chromosome organization, the tool AutoGraph was used [23] to compare the positions of the genes on the chromosomes.

For the next level, the gene structure of the 493 proteins was examined and differences were recognized by changes of the number and size of introns and exons. These differences could, on the one hand, lead to different numbers of splice variants, and on the other hand to missing regulatory RNAs, which would have been transcribed from deleted intron sequences [24]. Therefore, the human proteins were selected from the Swissprot database [25] and compared with the chimpanzee and human genome of the GenBanks RefSeq [26] to find their coding region. This was carried out by use of GenomeThreader [27]. The matches for chimp and human were compared. The GenomeThreader tool marked the exons and introns to facilitate the comparison of the number of exons and the length of the introns of both species [28, 29].

Moreover, pseudogenozation, also analyzed in this work, results in a loss of genes and paralogous gene copies [14] and, therefore, leads to a decline of protein-families, which is an important event during evolution. In a previous work of Wang et al. [30], several pseudogenozations of chimp and human genes have been described. In this work the pseudogenes of the RefSeq annotation were considered. The annotation of the chimp genome is an ongoing process and will possibly lead to new results in the future. One reason for the potential of improvement of the annotation is the lack of ESTs, which usually take an important part in the annotation process. While the dbEST section of NCBI (state: July 2007) lists more than 8 million human ESTs, there are currently less than 5000 chimp ESTs available. Most annotations on *Pan troglodytes* were derived by their similarity to nucleotide and protein sequences from other species and lack support from original chimp EST and full length cDNA data.

On the RNA-level, we analyzed post-transcriptional modifications. Alternative splicing is a prerequisite after the transcription of a pre-mRNA to yield various new mRNAs. Disruptions of enhancers, which are located upstream of a gene are indispensable for alternative splicing. A single mutation in an enhancer region leads in most cases to the loss of an alternative splice site and could also cause either a new splice variant or even the loss of a splice variant. To compare the alternative splice variants, the annotated

splice variants per gene, which are stored in the GenBank [26] of the NCBI database [31] were selected and compared .

Another important point we focused on in this work, is the post-translational modification after the biosynthesis. These modifications are important for the controlling of the localisation, enzyme activity and regulation of the final, native protein. For instance, the phosphorylation is a necessary step for the ubiquitination and, therefore, for the degradation of a protein. By loss or change of modification sites the biological activity of proteins could be up- or down-regulated, respectively [32]. To determine the modification sites in the protein sequences of both species, the PROSITE pattern [33] for seven different modification types was used. The number of occurrences of the seven modification types in both species was compared. The measurement of positive selection aims in the detection of genes that evolve faster than others on the basis of a selection pressure for novel forms. Changes in the nucleotide sequence that produce synonyms in the triplet code (Ks) and cause no changes in the protein sequence represent a kind of steady background noise. Non-synonymous nucleotide changes (Ka) are either advantageous for the organism or does not become accepted during evolution. A high Ka/Ks ratio indicates a strong selection pressure, while a low ratio means selection has been working to conserve the sequence. In the comparison of two species, the Ka/Ks ratio does not state which of the two species has changed the most from their common ancestor but genes that diverged the most are identified. The Ka/Ks ratio was estimated as $(dA/dS)/(NA/NS)$ where dA and dS are the number of amino acid mutations and synonymous mutations observed in each coding sequence. NA and NS is the number of possible non-synonymous and synonymous mutations [34].

Finally, the proteins were analyzed on the structural and functional level. Structural information was used to find changes in protein-protein and protein-compound interaction sites between the chimpanzee and human lineage. Already a single amino acid change of a conserved region of a binding site could yield to a disruption of a compound binding site. The available crystal structures of 209 proteins were collected from the PDB [35] to check whether a single amino acid change between human and chimp changes a protein-protein or a protein-compound binding site. Beyond that, an occurring change of a protein-protein or protein-compound binding site, resulting of an indel, was analyzed separately, because for these cases an increased assembly error rate has to be expected [36]. For this reason 48 proteins that were excluded from the Ensembl database vol. 44 were not considered in the gene related part of this analysis.

3. Results and discussion

3.1. Chromosome organization

The genome of the chimpanzee consists of 24 different chromosomes, whereas the human genome has 23 different chromosomes [13]. During evolution two chromosomes, which are designated 2a and 2b in chimpanzee, were merged to one larger chromosome 2 in the human genome [13]. Furthermore, nine pericentric inversions occur , in which the centromere is included [13]. One of these inversions exists on chromosome 18 and is shown in Fig. 1.

first step matched the protein sequence of the Swissprot database against the human and chimpanzee genomes and selected the annotated CDSs of GenBank.

For the RING finger protein 7 exists an annotated CDS in the GenBank for the chimpanzee and the human with 100 percent identity to the protein sequence of the Swissprot database. Nevertheless, the annotated CDS belongs to chromosome 3 in the human and to chromosome 7 in the chimpanzee genome. Moreover, a region on chromosome 3, which could also transcribe the RNA for this protein, is found with no existing annotated CDS. However, we found no matching coding region on human chromosome 7 for this protein. Similarly, for the proteins Replication protein C and the Nucleophosmin no coding regions could be detected on the corresponding chromosomes using the GenomeThreader.

Tab. 1: Proteins, which have its best matching coding sequence on different chromosomes.

Protein ID	Acc. number	Best match on human chromosome	Best match on chimp chromosome
RING finger protein 7	RBX2	3	7
Replication protein C	PP2AA	5	X
Nucleophosmin	NPM	5	16

3.2. Gene structure

To analyze the gene structure, the exon/intron structure has to be considered. The first step of analyzing the gene structure is to compare the numbers of exons, which are transcribed into the mRNAs in chimpanzee and human. Overall, an annotated CDS for human and chimpanzee is available for about 300 proteins with an identity of nearly 100 percent. For these CDSs the exon/intron structure was examined. About 5% of the proteins are transcribed from different numbers of exons during the human and chimpanzee biosynthesis. In one third of them the human mRNAs are transcribed of at least one more exon than the orthologous mRNA of the chimpanzee and in the two third it is conversely. These differences in the exon numbers can lead to different splice variants. To have an overview of the intron structure, the intron length of human and chimpanzee was determined and compared. Altogether, 84 proteins, which have a sequence identity of 95-100 percent and the same number of introns between human and chimpanzees have either at least one intron which is at least 50 base pairs longer than the orthologous introns. The genes of the chimpanzee have longer introns than the human genes (Fig. 3). Because introns are coding for regulatory RNAs, it might be possible that in chimpanzee more or different such RNAs exist, which might be also involved in the regulation of the apoptosis or the replication of apoptotic relevant genes.

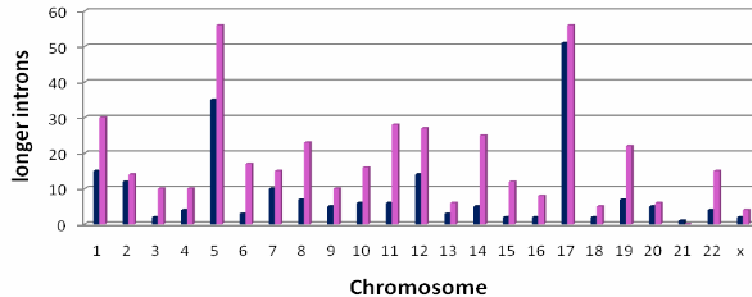


Figure 3: Number of longer introns per human and chimpanzee chromosome for the 84 proteins. On average, the genes of the chimpanzee have longer introns than the human genes. Blue bars: number of longer introns per human chromosome, magenta: number of longer introns per chimpanzee chromosome.

On average, the apoptotic and DNA-repair proteins have about 9 mutations per gene. Nielsen et al. [20] found a very large proportion of cancer-related genes that are affected by positive selection. To detect candidates that evolve faster than others the Ka/Ks ratios of the 493 genes were calculated. To determine whether a single protein is positively selected, the Ka/Ks value has to be higher than one. Altogether, 13 proteins, including the breast cancer protein BRCA1 (Tab. 2) are positively selected. The positive selection and the meaning of BRCA1 for cancer development have been shown earlier [13, 20].

Tab. 2: Proteins with Ka/Ks value higher than 1, indicating positive selection.

Protein	Acc. number	Ka/Ks-value
Thioredoxin-like protein p46	TXND5	3,13
Immediate early protein GLY96	IEX1	2,13
Poly [ADP-ribose] polymerase 3	PARP3	2,06
dUTP pyrophosphatase	DUT	1,70
Tumor necrosis factor receptor superfamily member 10D	TR10D	1,67
Tumor necrosis factor receptor superfamily member 18	TNR18	1,43
Breast cancer type 1 susceptibility protein	BRCA1	1,28
Tumor necrosis factor receptor superfamily member 10A	TR10A	1,21
Caspase-5	CASP5	1,21
Vascular endothelial growth factor A	VEGFA	1,17
Myc proto-oncogene protein	MYC	1,14
Serpin B9	SPB9	1,02
Transcription factor E2F1	E2F1	1,00

A process, which is likely to be more important than single mutations, is the pseudogenization [30, 37]. Pseudogenization is a gene inactivation, which is either caused by nonsense/frameshift mutations or by the loss of paralogous gene copies that were duplicated during hominoid evolution. In the 493 considered proteins about 49 proteins are possible pseudogenes in the chimpanzee, but active genes in human. One of these proteins is the transcription factor E2F3, which is involved in the cell cycle control [38]. This protein is described as a pseudogene in the RefSeq annotation of the NCBI. In contrast, the Ensembl annotation lists this gene locus as a functional gene, but achieves

this by the declaration of a cytosine as single nucleotide intron, which might be an artefact of the Ensembl annotation process. Further experimental validation is necessary to prove if one of these annotations is right. Moreover, also an error in the existing sequence could be the cause for these annotations. The loss of the transcription factor E2F3 could lead to a decreased proliferation rate. Surprisingly, many differences in the gene structure between chimpanzee and human were found. Especially, the great number of genes, which have larger introns, could be important for the cancer development. In this process, the regulatory mRNAs as well as the great number of different splice variants play a central role. Selective siRNAs will offer the possibility to analyze their occurrence in tumor cells experimentally. Corresponding experiments are in preparation with a cooperating lab regarding Bcl-2 proteins.

3.3. Post-transcriptional and post-translational modifications

An important post-transcriptional modification is alternative splicing, which yields to a varying number of mRNAs. More than 60 percent of human genes employ alternative splicing [39]. Altogether, about 1.5 annotated splice variants per human gene and 2.5 annotated variants per chimpanzee gene exist in the 493 proteins. The increased number of alternative splice variants can be explained by the increased number of exons in chimpanzee. This could lead to a more complex regulation of the DNA-repair and apoptotic pathway.

After the translation, many proteins have to be modified. These modifications regulate for example the proteins function, activity and localisation. For the seven types of modification cAMP-, PKC- and CK2-dependent phosphorylation, myristylation, ASN-glycosylation, TYR-phosphorylation, and amidation the number of proteins, which have at least one more modification site than the orthologue are considered. Therefore, the human patterns of these modification sites were used. The result is that the human proteins have more modification sites for all seven modification types than the chimpanzee (Fig. 4). From the analysis of the frequency of post-translational modifications, it can be hypothesized that either fewer modifications in the chimpanzees proceed or, more likely, that new patterns for these modifications have evolved. Hence, the development of new PROSITE patterns for the chimpanzee will be required.

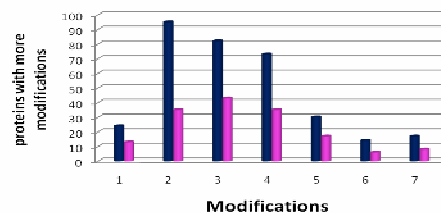


Figure 4: Number of proteins with one minimum modification site.

Number of proteins with at least one more cAMP-phosphorylation site (1), PKC-phosphorylation site (2), CK2-phosphorylation site (3), myristyl site (4), ASN-glycosylation site (5), TYR-phosphorylation site (6) and/or amidation site (7). For all cases the human has more modification sites. Blue: human, magenta: chimpanzee.

3.4. SNPs and indels leading to structural and functional changes

On average, the human-chimp lineage shows two amino acid mutations per protein [40], whereas in the reduced dataset of apoptotic and DNA-repair proteins eight amino acid mutations per protein can be found. Altogether, 4,111 single amino acid changes occur and 288 insertion/deletion events. However, Puente et al. just identified 1,542 amino acid changes in 333 cancer relevant proteins [18]. Moreover, the total protein sequence identity of these 493 proteins amounts 96%. In the schedule below, recognition patterns of important domains, which are destructed by a mutation, are listed (Tab. 3). For instance, the BH2-motif of the Bcl-2-related protein B2La1, which retards apoptosis, is lost. To understand the genetic changes between human and chimpanzee in terms of functional differences, the analysis of the structure and function is necessary. Altogether, 209 protein structures of the 493 proteins exist in the PDB, where 59 proteins have protein-protein interaction site changes and 10 proteins have compound binding site changes.

Tab. 3: Proteins, in which a motif is destructed by a mutation.

Protein ID	Motif	Human	Chimp
IFNA2	Interferon α, β, δ	X*	
MIF	Macrophage migration inhibitory factor	X	
SSR3	Microbodies C-terminal targeting signal		X
TXND5	Thioredoxin family active site	X	
VEFFA	Platelet-derived growth factor family signature	X	
B2LA1	BH2 motif	X	
BRCA2	ATP-binding motif	X	
ERCC3	Helicase ATP-binding type domain profile	X	
ERN1	SER/THR protein kinases active-site signature	X	
TBB2C	Tubulin β mRNA autoregulation signal	X	

*The cross in the last two columns indicates whether the human or the chimpanzee has this motif in the protein.

An example for a change in the compound binding site is the Topoisomerase II alpha.

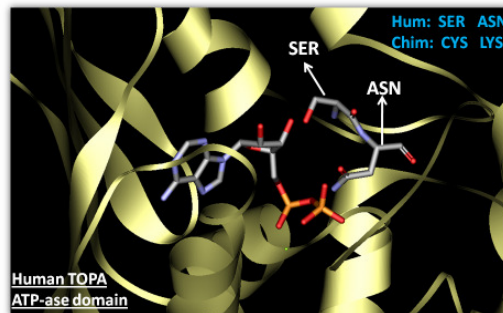


Figure 5: ATPase domain of human Topoisomerase II alpha of pdb code: 1ZXM.

The ATPase domain with a bound ATP molecule is shown. Two residues differ in the protein sequence of the human and chimps. SER and ASN are changed to CYS and LYS. These amino acids are very important for the binding of the ATP and a change could influence the strength of the binding. In this case, the negative charge of the phosphate group and the positive charge of LYS, which is in the chimpanzee sequence, could result in a better binding of the ATP.

The crystal structure of the ATPase-domain was determined. In this case the ATP-binding site will be changed, because of a mutation of two important amino acids (Fig. 5; PDB-code: 1ZXM) [41]. The Topoisomerase II alpha is a key enzyme, which is involved in DNA replication and is frequently amplified in breast cancer, which emphasizes the meaning of this finding.

We report that the transcription factor E2F3 might be a pseudogene in the chimpanzees. A crystal structure of the E2F3 domain, which might be a pseudogene in the chimpanzee genome, but an active gene in the humans, is available in the PDB (Fig. 6; PDB-code: 1CF7) [38].

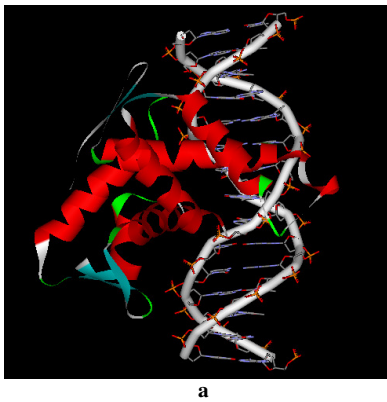


Figure 6: Possible pseudogenosation of E2F3.

a) DNA-binding domain of a transcription factor E2F-family member; PDB-code: 1CF7. The E2F transcription factor is shown as ribbon and the bound DNA in stick representation. The binding of E2F family members to the promoter region of cell-cycle controlling proteins enhances their expression.

This protein belongs to the E2F-family and acts as a transcription activator, by binding the DNA at the recognition site 5'-TTTC[CG]CGC-3', which is found in the promoter region of a number of genes, which are involved in the cell cycle regulation or in the DNA replication pathway. It specifically binds the protein Rb1. Moreover, E2F3 controls the cell-cycle progression from G1 to S phase and its high concentration results in a high proliferation rate.

Further works show that this protein is highly expressed in human prostate and lung cancer type and is actually a target for cancer therapy [42, 43]. One drug which inhibits the protein E2F3 is the silybin [44]. The recent use of E2F3 as cancer target emphasizes its meaning as key regulator of apoptosis. This supports our finding that its pseudogenosation might be an important reason for the low cancer incidence in chimpanzees. Of the 159 amino acid changes occurring in the 493 proteins could be mapped onto the 3D structure. A detailed analysis regarding structural changes and their influence on the protein function will be topic of a separate publication.

Tab. 4: Summary.

Level	Observation
Gene structure	<ul style="list-style-type: none"> - 49 putative pseudogenes in the chimp genome, which are active genes in human (e.g. transcription factor E2F3) - Chimps have exceeding longer introns, which may contain regulatory parts - Chimps have more exons, which transcribe for the same mRNA, which potentially leads to more splice variants - Group of proteins is positively selected ($Ka/Ks = 0.4$)
Post-transcriptional modification (alternative splicing)	<ul style="list-style-type: none"> - 1.5 annotated alternative splice variants per gene in human (738 proteins) - 2.5 annotated alternative splice variants per gene in chimps (1204 proteins)
Post-translational modification	<ul style="list-style-type: none"> - Humans have exceeding post-translational modifications or different patterns
Structure and function	<ul style="list-style-type: none"> - 6 amino acid changes per protein - 209 crystal structures of 493 proteins available - 59 proteins with possible changes in protein-protein binding sites - 10 proteins with changes in compound-protein binding sites (e.g. Topoisomerase II alpha)

4. Conclusion and perspective

A major outcome of this analysis represents the pseudogenozation of a number of proteins. The increased number of exons, which leads to more splice variants, is of outstanding importance and is a new finding. Beyond that, various amino acid changes can be mapped onto 3D structures and will be analyzed elsewhere. In a next step, their role in the apoptotic and DNA-repair pathways will be deconstructed. Longer introns are found in the chimpanzee. Transposable elements like ALUs may be responsible for that and should be investigated circumstantially, because they add regulatory sequences, like steroid binding sites and influence the methylation status of promoters and thus, the activity of genes. This epigenetic level should be included in a further analysis.

Acknowledgements

This work was supported by the International Research Training Group Boston-Kyoto-Berlin, funded by the DFG.

References

1. Ferlay, J., et al., *Estimates of the cancer incidence and mortality in Europe in 2006*. Ann Oncol, 2007. **18**(3): p. 581-92.
2. Varki, A., *A chimpanzee genome project is a biomedical imperative*. Genome Res, 2000. **10**(8): p. 1065-70.
3. Miller, B.A., S.M. Scoppa, and E.J. Feuer, *Racial/ethnic patterns in lifetime and age-conditional risk estimates for selected cancers*. Cancer, 2006. **106**(3): p. 670-82.
4. Wampler, N.S., et al., *Breast cancer survival of American Indian/Alaska Native women, 1973-1996*. Soz Präventivmed, 2005. **50**(4): p. 230-7.
5. Snyder, O.B., J.J. Kelly, and A.P. Lanier, *Prostate cancer in Alaska Native men, 1969-2003*. Int J Circumpolar Health, 2006. **65**(1): p. 8-17.

6. Houtgraaf, J.H., J. Versmissen, and W.J. van der Giessen, *A concise review of DNA damage checkpoints and repair in mammalian cells*. *Cardiovasc Revasc Med*, 2006. **7**(3): p. 165-72.
7. Nakanishi, M., M. Shimada, and H. Niida, *Genetic instability in cancer cells by impaired cell cycle checkpoints*. *Cancer Sci*, 2006. **97**(10): p. 984-9.
8. Li, W.F., et al., [*BRCA1 1100delAT is a recurrent mutation in Chinese women with familial breast cancer*]. *Zhonghua Yi Xue Za Zhi*, 2007. **87**(2): p. 76-80.
9. Balatzenko, G., et al., *Philadelphia variant, t(5;9;22)(q13;q34;q11), in a case with chronic myeloid leukemia*. *J Buon*, 2003. **8**(1): p. 65-7.
10. McKusick, V.A., *Mendelian Inheritance in Man and its online version, OMIM*. *Am J Hum Genet*, 2007. **80**(4): p. 588-604.
11. Shimizu, N., M. Ohtsubo, and S. Minoshima, *MutationView/KMcancerDB: a database for cancer gene mutations*. *Cancer Sci*, 2007. **98**(3): p. 259-67.
12. Bajdik, C.D., et al., *CGMIM: automated text-mining of Online Mendelian Inheritance in Man (OMIM) to identify genetically-associated cancers and candidate genes*. *BMC Bioinformatics*, 2005. **6**: p. 78.
13. *Initial sequence of the chimpanzee genome and comparison with the human genome*. *Nature*, 2005. **437**(7055): p. 69-87.
14. Kehrer-Sawatzki, H. and D.N. Cooper, *Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons*. *Hum Mutat*, 2007. **28**(2): p. 99-130.
15. Waters, D.J., et al., *Workgroup 4: spontaneous prostate carcinoma in dogs and nonhuman primates*. *Prostate*, 1998. **36**(1): p. 64-7.
16. Beniashvili, D.S., *An overview of the world literature on spontaneous tumors in nonhuman primates*. *J Med Primatol*, 1989. **18**(6): p. 423-37.
17. McClure, H.M., *Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony*. *Am J Phys Anthropol*, 1973. **38**(2): p. 425-9.
18. Puente, X.S., et al., *Comparative analysis of cancer genes in the human and chimpanzee genomes*. *BMC Genomics*, 2006. **7**(1): p. 15.
19. Seibold, H.R. and R.H. Wolf, *Neoplasms and proliferative lesions in 1065 nonhuman primate necropsies*. *Lab Anim Sci*, 1973. **23**(4): p. 533-9.
20. Nielsen, R., et al., *A scan for positively selected genes in the genomes of humans and chimpanzees*. *PLoS Biol*, 2005. **3**(6): p. e170.
21. Camon, E., et al., *The Gene Ontology Annotation (GOA) Database--an integrated resource of GO annotations to the UniProt Knowledgebase*. *In Silico Biol*, 2004. **4**(1): p. 5-6.
22. Hubbard, T.J., et al., *Ensembl 2007*. *Nucleic Acids Res*, 2007. **35**(Database issue): p. D610-7.
23. Derrien, T., et al., *AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps*. *Bioinformatics*, 2007. **23**(4): p. 498-9.
24. Mattick, J.S. and I.V. Makunin, *Small regulatory RNAs in mammals*. *Hum Mol Genet*, 2005. **14 Spec No 1**: p. R121-32.
25. *The Universal Protein Resource (UniProt)*. *Nucleic Acids Res*, 2007. **35**(Database issue): p. D193-7.
26. Benson, D.A., et al., *GenBank*. *Nucleic Acids Res*, 2007. **35**(Database issue): p. D21-5.

27. Gremme, G., et al., *Engineering a software tool for gene structure prediction in higher organisms*. Information and Software Technology, 2005. **47(15)**: p. 965-978.
28. Usuka, J., W. Zhu, and V. Brendel, *Optimal spliced alignment of homologous cDNA to a genomic DNA template*. Bioinformatics, 2000. **16(3)**: p. 203-11.
29. Usuka, J. and V. Brendel, *Gene structure prediction by spliced alignment of genomic DNA with protein sequences: increased accuracy by differential splice site scoring*. J Mol Biol, 2000. **297(5)**: p. 1075-85.
30. Wang, X., W.E. Grus, and J. Zhang, *Gene losses during human origins*. PLoS Biol, 2006. **4(3)**: p. e52.
31. Maglott, D., et al., *Entrez Gene: gene-centered information at NCBI*. Nucleic Acids Res, 2007. **35(Database issue)**: p. D26-31.
32. Basu, A., G. DuBois, and S. Haldar, *Posttranslational modifications of Bcl2 family members--a potential therapeutic target for human malignancy*. Front Biosci, 2006. **11**: p. 1508-21.
33. de Castro, E., et al., *ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins*. Nucleic Acids Res, 2006. **34(Web Server issue)**: p. W362-5.
34. Yang, Z., D. Balding, M. Bishop, and C. Cannings., *Adaptive molecular evolution, in Handbook of statistical genetics*. Wiley, London, 2001. **Chapter 12**: p. pp. 327-350,.
35. Berman, H., et al., *The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data*. Nucleic Acids Res, 2007. **35(Database issue)**: p. D301-3.
36. Overduin, B., ENSEMBL, Editor. 2007.
37. Fairbanks, D.J. and P.J. Maughan, *Evolution of the NANOG pseudogene family in the human and chimpanzee genomes*. BMC Evol Biol, 2006. **6**: p. 12.
38. Zheng, N., et al., *Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP*. Genes Dev, 1999. **13(6)**: p. 666-74.
39. Mironov, A.A., J.W. Fickett, and M.S. Gelfand, *Frequent alternative splicing of human genes*. Genome Res, 1999. **9(12)**: p. 1288-93.
40. Glazko, G., et al., *Eighty percent of proteins are different between humans and chimpanzees*. Gene, 2005. **346**: p. 215-9.
41. Wei, H., et al., *Nucleotide-dependent domain movement in the ATPase domain of a human type IIA DNA topoisomerase*. J Biol Chem, 2005. **280(44)**: p. 37041-7.
42. Grasmann, C., et al., *Gains and overexpression identify DEK and E2F3 as targets of chromosome 6p gains in retinoblastoma*. Oncogene, 2005. **24(42)**: p. 6441-9.
43. Oeggerli, M., et al., *E2F3 is the main target gene of the 6p22 amplicon with high specificity for human bladder cancer*. Oncogene, 2006. **25(49)**: p. 6538-43.
44. Tyagi, A., C. Agarwal, and R. Agarwal, *Inhibition of retinoblastoma protein (Rb) phosphorylation at serine sites and an increase in Rb-E2F complex formation by silibinin in androgen-dependent human prostate carcinoma LNCaP cells: role in prostate cancer prevention*. Mol Cancer Ther, 2002. **1(7)**: p. 525-32.