

Segmental Duplications Containing Tandem Repeated Genes Encoding Putative Deubiquitinating Enzymes

Hong Liu^{1,2}, Li Li¹, Asher Zilberstein¹, Chang S Hahn^{1,2}

1. Aventis Pharmaceuticals Inc., Route 202/206, Bridgewater, NJ 08807, USA.
2. To whom correspondence should be addressed: hong.liu@aventis.com; or chang.hahn@aventis.com.

Keywords: segmental duplication, deubiquitinating enzyme

Abstracts

Both inter- and intra-chromosomal segmental duplications are known to have occurred in the human genome during evolution. Few cases of such segments involving functional genes have been reported. While searching for the human orthologs of murine hematopoietic deubiquitinating enzymes (DUBs), we identified four clusters of DUB-like genes on chromosome 4p15 and chromosome 8p22-23 that are over 90% identical to each other at the DNA level. These genes are expressed in a cell type- and activation-specific manner, with different clusters possessing potentially distinct expression profiles. Examining the surrounding sequences of these gene duplication events, we have identified previously unreported conserved sequence elements that are as large as 35 to 74 kb encircling the gene clusters. Traces of these elements are also found on chromosome 12p13 and chromosome 11q13. The coding and immediate upstream sequences for DUB-like genes as well as the surrounding conserved elements, are present in the chimpanzee trace database, but not in the rodent genome. We hypothesize that the segments containing these DUB clusters and surrounding elements arose relatively recently in evolution through inter- and intra-chromosomal duplicative transpositions, following the divergence of primates and rodents. Genome wide systematic search of the segmental duplication containing duplicated gene cluster has been performed.

1. Introduction

Duplicative transposition is a segmental duplication event that transfers 1-200 kb blocks of genomic

sequences to other location(s) in the genome during evolution. This may occur either within a chromosome or between chromosomes. Many of the inter-chromosomal duplications are biased toward pericentromeric and subtelomeric regions, while the locations of intra-chromosomal duplications are more variable, and may involve duplication of functional genes [5]. Preliminary analysis of the initial assembled human genome demonstrates that recent segmental duplications cover more than 3.3% of the genome. Intra-chromosomal duplications occupy 2% of the genome and inter-chromosomal duplications occupy 1.5%, with overlaps comprising 0.2% [9]. Segmental duplications are important for restructuring human chromosomes, thus expanding gene diversity [3,11]. While previous studies have described either inter- and intra-chromosomal duplications, we present here an example of a gene cluster that evolved through a series of both types of duplicative events. Moreover, while the arrangement of the genes within these clusters has diverged, a high level of sequence conservation has been preserved within these loci.

2. Results and Discussion

Mouse DUB1, DUB2, and DUB2A are members of a subfamily of deubiquitinating enzymes called hematopoietic DUBs, which are immediate-early response genes induced by cytokines [2,15,16]. Mouse DUB1, DUB2 and DUB2A share high amino acid sequence similarity with more than 88% identity among 1, 2 and 2A, and 95% identity between 2 and 2A. Using mouse DUB1, 2, and 2A sequences to query the mouse genome, there are a total of nine hits distributed on six different contigs on the chromosome 7. In order to identify human orthologs for murine hematopoietic DUBs, mouse DUB1 and 2 sequences were searched against the human genome. Multiple hits were identified in five genomic locations: one at

Figure 1. Distribution of five segments encompassing duplicated pDUBs in the human genome.

Segment 4 is located on 4p15, segments 8.1 and 8.2 are located on 8p23, and segments 8.3 and 8.4 are located on 8p22. Black arrowheads indicate the direction of pDUB coding sequences, which in some cases contain introns, depicted by unfilled boxes (8.1_3, 8.2_1, 8.3_1, 8.4_2). Truncated arrowheads signify pDUBs that undergo premature translational termination (4_8, 4_10, 8.4_1).

4p15, and two segments each at 8p23 and 8p22. In total, 21 putative DUBs (pDUB) were predicted by the GenScan analysis [4] (Figure 1). The pDUBs on chromosome 4 were first identified by Kogi et al. [8] when they sequenced a repeat region in 4p15.1-15.3 and identified 50-70 copies of a putative open reading frame (ORF) of 530 aa. There is a gap immediately up stream of the first ORF (4_1) that is estimated to be 300 kb [6], a feature that limits our ability to precisely elucidate the exact duplications of pDUB on chromosome 4p15. Remarkably, seven of the ten pDUBs on chromosome 4 (4_1, 4_2, 4_3, 4_5, 4_6, 4_7, 4_9) display complete nucleotide sequence identity, as do two of the intron-containing pDUBs on chromosome 8 (8.1_3 and 8.2_1). At the genomic sequence level, these 21 pDUBs share an overall identity of 90%. Some of the pDUBs, such as 8.1_3, 8.2_1, 8.3_1 and 8.4_2, contain a premature stop codon in the first exon, with GenScan analysis predicting extra exon (s). These extra exons lower the overall amino acid homology to as low as 30% (Table 1). The

homology of these human pDUBs to mouse DUB1, 2 and 2A are between 23% and 50%. Sequence alignment suggests that 14 out of 21 (4_1, 4_2, 4_3, 4_4, 4_5, 4_6, 4_7, 4_9, 4_10, 8.1_1, 8.1_2, 8.2_2, 8.2_3, 8.3_3) contain ubiquitin specific protease domains critical for DUB activity.

We examined the expression of these pDUBs by real-time PCR analysis in different cell models. Four primer sets were designed to test the expression of several clusters of these pDUBs. Since these are potential orthologs of murine hematopoietic deubiquitinating enzymes, we first examined their expression in immunocytes (Figure 2A). We found that these pDUBs were expressed in peripheral blood mononuclear cell (PBMC) preparations from several healthy donors. Interestingly, expression of the pDUBs was upregulated upon stimulation with lipopolysaccharides (LPS), Phytohemagglutinin (PHA) or both, with those pDUBs on chromosome 4 (pDUB4) exhibiting faster kinetics and higher levels of upregulation than those on chromosome 8 (pDUB8).

Figure 2. Cell and tissue specific expression of human pDUBs.

A. Expression pattern of pDUBs in PBMC treated with LPS, PHA or LPS+PHA. Freshly isolated human PBMCs were mock treated, treated with either LPS, PHA, or both for 1.5h, 7h and 24h, and pDUBs were quantified by real time PCR. Because of high homology among the pDUBs, primer set pDUB4 targets the eight full-length genes on chromosome 4 (4_1, 4_2, 4_3, 4_4, 4_5, 4_6, 4_7, 4_9), primer set pDUB8-a targets genes 8.1_1, 8.1_2, 8.2_2, and 8.2_3, primer set pDUB8-b targets genes 8.3_2 and 8.3_3, and primer set pDUB8-c targets gene 8.4_2. B. Quantitative PCR-based expression analysis of the pDUB4 cluster in different immunocyte populations in the presence or absence of stimulation as described in the Methods section. C. Expression of the pDUB4 cluster in various human tissues was quantified by real-time PCR.

Table 1. Percentage of amino acid identity among human pDUBs

In addition, the expression levels of different pDUB clusters differ upon stimulation, with the pDUB4 cluster yielding the greatest overall signal. However, the large gene copy number of this cluster may explain this. In contrast to what was observed at 1.5 h and 7 h timepoints, there was no detectable upregulation of any pDUB examined at 24 hours post-stimulation. Next, we examined the expression of the pDUB4 cluster, the highest expressed pDUB set, in individual purified immunocytes in the presence or absence of stimulation appropriate to the particular cell type (Figure 2B). The expression of this gene cluster was high in resting and activated B cells and resting CD8⁺ T cells. The expression levels in CD4⁺ T cells, CD14⁺ monocytes, and granulocytes were very low regardless of stimulation status. Since the expression profile is rather different to that of the murine DUBs, we examined the expression of pDUB4 in various tissues (Figure 2C). Twenty-five different human tissues including two fetal tissues were examined. The expression was highest in CNS tissues (including brain, fetal brain, spinal cord) and testis.

We extended the examination of the upstream sequences of the pDUB clusters and found that they are highly conserved among all four segments on chromosome 8 for up to 74 kb (Figure 3), with pairwise identity greater than 90%. The conserved upstream sequence could not be located on chromosome 4 because of the ~300 kb gap immediately upstream of the first gene (4_1) [6]. Interestingly, this conserved long sequence element (termed LE1) was also located on chromosome 12p13 in a truncated form of 34 kb with deletion at both ends. In addition to a ~3 kb duplication of a small segment near the far end of LE1, there are only sparse and tiny gaps along the 34 kb of five conserved sequences. The overall homology of the various LE1 sequences ranges 90-95%. Analysis of the sequences immediately downstream of the pDUB clusters revealed a similar pattern. All the downstream

sequences (~35 kb) for segments 4, 8.1, 8.2, 8.3, and 8.4 display pairwise sequence identity of greater than 91%. This second conserved long element (LE2) was also found in a truncated version on chromosome 12 (29 kb) juxtaposed to LE1, in the absence of intervening pDUBs. An additional LE2 of 34 kb was also found on chromosome 11q13 with no associated pDUBs or LE1. In contrast to LE1 sequences, the panel of conserved LE2 sequences (29 kb) exhibits slightly more divergence, with larger gaps up to 1.5 kb. The common feature is a ~3 kb palindrome near the far end of the long element. Overall sequence identity among LE2 sequences on chromosome 4 and 8 is between 91 to 95%, similar to that calculated for LE1. The most divergent LE2 sequences are on chromosome 12 and 11 with sequence identity of 81%. As for the chromosomal positions of these segments, 4p15, 8p23, 8p22 and 12p13 are located in subtelomeres, while 11q13 is neither subtelomeric nor pericentromeric. In those segments that contain pDUB clusters, the orientations of LE1 and LE2 with respect to the direction of transcription of pDUBs are conserved (see Figure 3). In order to examine whether there is a potential conserved LE1 in the 300 kb gap on chromosome 4, we searched the human trace database. A number of hits were identified that aligned with LE1, and some of them were different from LE1s for 8.1, 8.2, 8.3, 8.4, or 12 (data not shown). These unmapped fragments may belong to segment 4 and/or elsewhere in the genome.

Further analysis of the LE sequences extending from clusters 8.1 and 8.2, which are juxtaposed as mirror images of one another, revealed a number of interesting features. The LE1 sequences for these two loci extend over a distance of 150 kb, while the LE2 sequences encompass a region of over 280 kb, terminating in a gap of over 100 kb that defines the boundary of these two inverted segments.

Figure 3. Chromosomal distribution of conserved long elements in the presence or absence of pDUB clusters. Shown are boxes depicting pDUBs (), LE1 sequences (), and LE2 sequences (). Arrowheads delineate the direction of individual ORF, while the polarity of LE sequences, which indicates the position on the chromosome, is represented by the fill patterns of the boxes. The length of each LE segment is identified to the left of each box. Empty boxes signify deleted portions of the LE segments; the size of each deletion is indicated to the right of each box.

Between these extended boundaries, the level of overall conservation of 8.1 and 8.2 is greater than 95%. Pairwise sequence alignment of 8.1 and 8.2 revealed the presence of a 240 kb palindromic repeat in the upstream portion of the LE1 associated with 8.1, whereas the 120 kb core sequence of this palindrome was located at a similar position in the LE1 associated with 8.2. Immediately next to the palindromic sequences and the gap, there is a pair of palindromic tandem repeats in both 8.1 and 8.2 (Figure 4).

GenScan prediction suggests there are only 1-2 hypothetical genes located in the ~ 30 kb stretch of LE1 and LE2 that are shared among all segments. Further analysis by RepeatMasker demonstrates that 32-36% of the sequence content of LE1 and LE2 is composed of known repeat sequences. This is higher compared to that of the pDUB region (10%). Nevertheless, it is still lower than that of the whole genome, greater than 50% of which consists of repeated sequences [9]. The GC content of LE1 and LE2 is 40-42%, similar to the average of GC content in human genome [9], while the pDUB region has higher GC content of 46-50%. Moreover, when the

nearest-neighbor base-frequency was analyzed to investigate the local sequence composition, they are very similar not only within, but also between, each of the LE1 and LE2 families. However, the nearest-neighbor base-frequencies for the long elements are quite different from those of the pDUB sequences. All these data support that LE1 and LE2 are gene-poor genomic regions and there is some coherence between the two long elements that is not shared with the pDUB region.

Since we couldn't identify LE sequences in mouse and rat assembled genomes, we examined them in trace databases of mammals. The chimpanzee trace database, which containing 1.8×10^{10} bp of sequences, was searched for the existence of duplicated pDUBs and conserved neighboring long elements. Multiple hits were identified for LE sequences, with more hits for pDUB-containing regions (data not shown). This suggests that duplicated pDUBs may also be conserved in chimpanzees, as are the LE sequences. These sequences could not be found in mouse or rat trace databases.

Figure 4. Dot plot of four-way sequence alignments between extended segments 8.1 and 8.2. Genomic sequence of segments 8.1 and 8.2 were retrieved starting from the central gap and extending 560 kb in each direction. Solid lines define regions of greater than 90% sequence similarity, while arrows identify the positions of the pDUB clusters. Dotted lines extrapolated from these solid lines reveal potential alignments within the gap region. This analysis predicts a gap size of ~150 kb which contains additional palindromic repeat sequences.

It raises the possibility that these duplicative transpositions arose following the divergence of primates and rodents.

In summary, we provide a report of both intra-chromosomal and inter-chromosomal segmental transpositions. The duplicated segments contain highly conserved flanking long elements, many with multiple copies of genes that may have putative deubiquitination function. The four unique structures of the duplication segment or the remnant of the segment are as follows: (1) long conserved elements surrounding two to three copies of pDUB are present in four segments on chromosome 8, (2) long conserved elements with no pDUB cluster (12p13), (3) at least one long conserved element with ten or more copies of pDUB (4p15), and (4) one long conserved element (11q13).

Based on the structure of the duplicated segments we described, we speculate that duplicative transposition happened first by an inter-chromosomal mechanism, then by intra-chromosomal duplication (chromosome 8).

Figure 5. Phylogenetic analysis. (A) internal duplicated pDUB cluster, (B) LE1 sequences, and (C) LE2 sequences. Conserved genomic sequences for LE1 (34 kb), LE2 (29 kb) and internal pDUB cluster (13.5 kb) for each segment were used to build the

phylogenetic trees. Sequences of three genes were retrieved to represent pDUB cluster on chromosome 4.

In order to examine the proximity of duplication events that arose during evolution, phylogenetic analysis of LE1, LE2 and the intervening region containing duplicated pDUBs was conducted (Figure 5). In general the pDUB clusters on chromosome 8 were found to segregate more closely with one another than with the pDUB cluster on chromosome 4. This pattern is consistent with the hypothesis that inter-chromosomal duplication events preceded intra-chromosomal transpositions. Analysis of LE sequences provides additional insights into the evolution of these segments. The phylogenetic trees for chromosomes 4 and 8 LE1 and LE2 sequences exhibit distinct patterns. For example, the LE1 of 8.1 is closer to that of 8.2 while the LE2 of 8.1 is closer to that of 12. In addition, it is telling that the chromosome 4 LE2 interrupts the clustering of chromosome 8 LE2 sequences. These observations raise the notion that several rounds of duplicative transposition may have occurred either simultaneously or within a relatively short time period among these segments. It is worth note that the mechanism of transposition for segments 8.1 and 8.2 is distinct in that it involved an inverted duplication event including larger surrounding sequences (Figure 4), which possibly occurred through secondary chromosomal rearrangements triggered by interactions between the multiple palindromic sequences in those surrounding sequences. It is interesting that the clusters on chromosome 8 all possess either 2 or 3 pDUBs, while that on chromosome 4 contains at least 10 pDUBs and those on chromosomes 11 and 12 lack these inserts altogether. Because the precise sequence of amplification and deletion events that have occurred between the chromosomes are not known, it remains uncertain as to whether pDUBs were present in the progenitor segment or incorporated subsequently.

Segmental duplications are not only important in genome evolution but may introduce chromosomal instability that carries with in the risk of disease induction. For example, 24 out of 169 identified recent segmental duplication regions have been associated with genetic disease [3]. In fact, the genomic regions described here are all reported to have linkage with some genetic diseases. Among them, 8p22-23 has been reported to be linked to schizophrenia [10] and prostate cancer [14] while 4p16 has been linked to bipolar disorder, Huntington's disease and other diseases [6,8]. Both 8p22-23 and 4p16 are segments encompassing duplicated pDUB that showed limited tissue expression with the highest expression in brain. However, whether this correlation is simply by chance

or whether pDUBs and/or the LE sequences may directly contribute to these neurological diseases requires further investigation.

In this study, using *in silico* approach, we have identified segmental duplications that are across four chromosomes (4, 8, 11 and 12), and the duplication has undergone further four rounds in the single chromosome (8). In order to identify segmental duplications containing duplicated gene cluster across the whole genome, we applied the following algorithm:

There are total 792 clusters of duplicated genes (without interruption of other gene) in the human genome. Analyzing those clusters, there are 127 groups that contain duplicated gene clusters either within one chromosome or across multiple chromosomes. Among them, 88 groups contain duplicated gene clusters across different chromosomes. Those 88 groups represent the inter-chromosomal segmental duplications that contain duplicated genes. Genes embedded in those duplicated segments are: seven transmembrane protein receptors including olfactory and taste receptors, zinc finger protein, homeobox protein, keratin, cytochrome P450, protease (ADAM 28), chemokine receptor, and sodium channels, etc. Gene duplication followed by functional specialization has been considered a major evolutionary force for gene diversification, and may contribute to the process of environmental adaptation [3,11]. Many of those duplicated genes have shown function specificity involved in different biological processes. However, whether those segmental duplications involved neighboring long intergenic sequences like DUB cluster need further investigation, and their duplication mechanism or order of duplication (gene duplication vs segment duplication) may not necessary be similar. Nevertheless, the systematical study demonstrates the overall pattern of inter- and intra- chromosomal duplications involved duplicated genes in the human genome and provide resources for future study.

3. Materials and Methods

3.1. Sequence Analysis

Genomic mapping and analysis were performed using NCBI human (NCBI built 34 version 1) and mouse (NCBI built 30 version 1) genome resources. Trace database search was blasted against NCBI Trace Archive databases. Gene prediction was carried out by GenScan [4]. Multiple sequence alignment was accomplished by Clustalw [13]. Genomic sequence alignment and dot plot were built by Pipmaker [12]. Molecular evolution analysis was performed by Phylip package [7]. The reliability of each branch for phylogenetic tree was assessed by 100 bootstrap replications.

Systematical searching of segmental duplication containing duplicated gene cluster were based on data from Ensembl. Human protein sequences and gene positions on chromosomes were downloaded from Ensembl v19 (<http://www.ensembl.org/>). Homolog proteins were defined based on the blastp search [1] with E-value smaller than e-20.

3.2. Sample preparation, and TaqMan real time PCR analysis

PBMC were isolated from fresh blood samples of normal subjects. PBMC band was spun through 5 ml fetal bovine serum cushion and washed extensively with phosphate buffered saline. Subset of PBMC was enriched using MACS negative selection bead cocktails according to the manufacture's specifications. PBMC were mock stimulated or stimulated with either 100 ng/ml LPS, 500 ng/ml PHA or combination for 1.5, 7, or 24 hours. Enriched CD4⁺ or CD8⁺ T cells were stimulated with 10 ng/ml anti-CD3 mAb, 5 ug/ml anti-CD28 mAb and 10 ug/ml protein G at the 10⁶ cells per ml concentration for 8 or 18 hours for CD4⁺ T cells or 4 and 24 hours for CD8⁺ T cells. CD19⁺ enriched B cells were stimulated with either 100 ng/ml LPS or 1.5 ug/ml anti-CD40 mAb with 10 ng/ml IL-4 for 4 or 20 hours. CD14⁺ monocytes were bound to the plate for overnight and stimulated with 100 ng/ml LPS for 1.5, 7 and 24 hours. Total granulocytes were stimulated with 10 ng/ml TNF-alpha for 4 hours. Human tissues were purchased commercially.

RNAs from above cell samples were prepared and Real time PCR was performed following manufacturer's direction. Samples containing GAPDH primer pair and probe were also prepared as control. The primer/probe sets used are as follows: pDUB4: forward primer: 5'-AGCTCCCTGCTAAACCTCTCTTC-3'; reverse

primer: 5'-AGCGAGTGTGCCAGTGTTC-3';
Taqman probe: 5'-TCGACCCCGACACATCAGGAGTCC-3'.
pDUB8_a: forward primer: 5'-CGACGAACCCGACAGATCA-3'; reverse primer: 5'-TGGTCCCTCCCTTGCAGAGAA-3'; Taqman probe: 5'-AGTCCATGAACACTGGCACACTCG-3'.
pDUB8_b: forward primer: 5'-TGGAAAGAGCCACTCAGGAAA-3'; reverse primer: 5'-CAGGCTTCGTTTTGTTTTGCT-3'; Taqman probe: 5'-CACCTTAGACCACTGGAGATTCCCCCAA-3'.
pDUB8_c: forward primer: 5'-TGGAAAGAGCCACTCAGGAAA-3'; reverse primer: 5'-GGCTTCGTTTTGTTTTGCTTTT-3'; Taqman probe: 5'-CACCTTAGACCACTGGAAATTCCCC-3'.

Acknowledgments

We are grateful to Dr. Christopher Arendt for editing the manuscript and his comments. We thank Dr. Kin-Tak Yu for helping designing the PCR primer sets.

References

- [1] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, L.D. J., Basic local alignment search tool., *J. Mol. Biol.* 215 (1990) 403-410.
- [2] K.H. Baek, M.A. Mondoux, R. Jaster, E. Fire-Levin, A.D. D'Andrea, DUB-2A, a new member of the DUB subfamily of hematopoietic deubiquitinating enzymes, *Blood* 98 (2001) 636-642.
- [3] J.A. Bailey, Z. Gu, R.A. Clark, K. Reinert, R.V. Samonte, S. Schwartz, M.D. Adams, E.W. Myers, P.W. Li, E.E. Eichler, Recent segmental duplications in the human genome, *Science* 297 (2002) 1003-1007.
- [4] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, *J Mol Biol* 268 (1997) 78-94.
- [5] E.E. Eichler, Recent duplication, domain accretion and the dynamic mutation of the human genome, *Trends Genet* 17 (2001) 661-669.
- [6] K.L. Evans, S. Le Hellard, S.W. Morris, D. Lawson, C. Whitton, C.A. Semple, J.A. Fantes, H.S. Torrance, M.P. Malloy, J.C. Maule, S.J. Humphray, M.T. Ross, D.R. Bentley, W.J. Muir, D.H. Blackwood, D.J. Porteous, A 6.9-Mb high-resolution BAC/PAC contig of human 4p15.3-p16.1, a candidate region for bipolar affective disorder, *Genomics* 71 (2001) 315-323.
- [7] J. Felsenstein, PHYLIP, <http://evolution.genetics.washington.edu/phylip.html>.
- [8] M. Kogi, S. Fukushige, C. Lefevre, S. Hadano, J.E. Ikeda, A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization, *Genomics* 42 (1997) 278-283.
- [9] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W.

- FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860-921.
- [10] A.E. Pulver, V.K. Lasseter, L. Kasch, P. Wolyniec, G. Nestadt, J.L. Blouin, M. Kimberland, R. Babb, S. Vourlis, H. Chen, et al., Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes, *Am J Med Genet* 60 (1995) 252-260.
- [11] R.V. Samonte, E.E. Eichler, Segmental duplications and the evolution of the primate genome, *Nat Rev Genet* 3 (2002) 65-72.
- [12] S. Schwartz, Z. Zhang, K.A. Frazer, A. Smit, C. Riemer, J. Bouck, R. Gibbs, R. Hardison, W. Miller, PipMaker--a web server for aligning two genomic DNA sequences, *Genome Res* 10 (2000) 577-586.
- [13] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res* 22 (1994) 4673-4680.
- [14] F. Wiklund, B.A. Jonsson, I. Goransson, A. Bergh, H. Gronberg, Linkage analysis of prostate cancer susceptibility: confirmation of linkage at 8p22-23, *Hum Genet* 112 (2003) 414-418.
- [15] Y. Zhu, M. Carroll, F.R. Papa, M. Hochstrasser, A.D. D'Andrea, DUB-1, a deubiquitinating enzyme with growth-suppressing activity, *Proc Natl Acad Sci U S A* 93 (1996) 3275-3279.
- [16] Y. Zhu, K. Lambert, C. Corless, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, A.D. D'Andrea, DUB-2 is a member of a novel family of cytokine-inducible deubiquitinating enzymes, *J Biol Chem* 272 (1997) 51-57.