Université de Montréal

"Evolution of C2H2-Zinc finger genes in mammalian genomes"

par

"Hamsa Dhwani Tadepally"

"Département de Biochimie"

"Faculté de Médecine"

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Maitrise

En Biochimie

"July 2007"

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

"Evolution of C2H2-Zinc finger genes in mammalian genomes"

Présentée par :

"Hamsa Dhwani Tadepally"

a été évaluée par un jury composé des personnes suivantes :

"Martine Raymond"

Président-rapporteur

"Muriel Aubry"

Directrice de recherche

"Gertraud Burger"

Co-directrice

"Nicolas Lartillot"

Membre du jury

Résumé

Les gènes de doigt de zinc de C2H2/Kruppel (C2H2-ZNF) encodent la plus grande classe des facteurs de transcription chez l'homme. Ces gènes constituent une des plus grandes familles de gène chez les mammifères et sont souvent trouvés sous forme de regroupements de gènes juxtaposés sur les chromosomes. Par une recherche extensive basée sur des similitudes de séquences visant à d'identifier l'ensemble des gènes C2H2-ZNF du génome humain, nous avons assemblé un répertoire complet de 718 gènes C2H2-ZNF humains. Les gènes C2H2-ZNF ont été classifiés en sous-familles en fonction des domaines effecteurs N-terminaux aux quels ils sont associés. Nous avons constaté que la sous-famille encodant un domaine KRAB comprend ~ 45% de tous les gènes C2H2-ZNF et est par conséquent la plus grande sous-famille de gènes à motifs doigt de zinc. De plus, nous avons identifié 81 regroupements de gènes C2H2-ZNF qui correspondent à ~ 70% de tous les gènes C2H2-ZNF. Presque 90% des gènes C2H2-ZNF appartenant aux sous-familles KRAB et SCAN sont trouvés sous forme de regroupements. Pour mieux comprendre l'évolution des gènes C2H2-ZNF, nous avons par la suite assemblé un répertoire complet de tous les regroupements de gènes C2H2-ZNF humains ainsi que de leurs contre-parties dans les régions synténiques des génomes de chimpanzé, de souris, de rat et de chien. Une analyse systématique de ce répertoire chez ces mammifères a révélé qu'il existe une variation dans le nombre de regroupements et de gènes faisant partie de ces regroupements parmi les primates, les rongeurs et les canins. Cette variation suggère que ces gènes ont évolué de façon différentielle chez les mammifères. Des études phylogénétiques de plusieurs regroupements de gènes C2H2-ZNF choisis indiquent qu'outre une duplication

différentielle, la perte de gènes dans certaines espèces a conduit à des répertoire différents de gènes C2H2-ZNF chez les mammifères. En plus des variations spécifiques aux espèces dans le nombre de gènes, nous avons également mis en évidence une variation chez des orthologues dans le nombre de motifs de doigt de zinc et la présence de domaines effecteurs, ces derniers étant souvent perdus par dégénération. En conclusion, sur la base principale de ces résultats et de l'étude de la structure exon-intron des gènes C2H2-ZNF, nous proposons un nouveau modèle pour l'évolution de leurs sous-familles selon lequel les sous-familles les plus anciennes seraient dans l'ordre SCAN > SCAN-KRAB > KRAB.

Abstract

The C2H2/Kruppel zinc finger genes (C2H2-ZNF) encode the largest class of transcription factors in humans. These genes constitute one of the largest gene families in mammals and are often found in clusters. Using an extensive similarity search on the human genome to identify all C2H2-ZNF genes, we assembled a comprehensive repertoire of 718 human C2H2-ZNF genes. The genes were grouped into subfamilies based on the N-terminal effector domains they were associated with. We found that the KRAB-domain encoding subfamily constitutes ~ 45% of the total C2H2-ZNF genes and hence is the largest subfamily of zinc finger genes. In addition to this, we also identified 81 C2H2-ZNF clusters which constitute ~ 70% of the total genes. Almost 90% of the C2H2-ZNF belonging to the KRAB and SCAN subfamilies were found in clusters. We then assembled a comprehensive repertoire of all the human C2H2-ZNF clusters and their syntenic counterparts in chimpanzee, mouse, rat and dog genomes. A systematic analysis of all the syntenic clusters revealed a variation in the numbers of clusters and the genes within clusters among primates, rodents and canines indicating differential patterns of evolution in mammals. Evolutionary analysis of few selected C2H2-ZNF syntenic clusters in the five mammals studied suggested that not only differential duplication, but also gene loss has led to different repertoires in mammalian genomes. In addition to lineage- and species-specific variation in the number of genes, we also find a variation among orthologs in the number of zinc finger motifs and in the presence of the effector domains, the later being often lost by sequence degeneration. Finally, based on the above results and on the analysis of the exon-intron structure of the various C2H2-ZNF genes, we propose a model for the evolution of

their subfamilies suggesting that the more ancient subfamilies are in sequential order SCAN > SCAN-KRAB > KRAB.

Keywords: C2H2/Kruppel, zinc finger, gene family, tandem repeats, gene duplication, gene loss, evolution.

List of abbreviations

DNA: Deoxyribonucleic Acid

RNA: Ribonucleic Acid

BTB: Broad-Complex, Tramtrack and Bric-a-bric

POZ: Pox virus and Zinc finger

KRAB: Kruppel Associated Box

SCAN: SRE-ZBP, CTfin51, AW-1 and Number18 cDNA

KRI motif: KRAB Interior motif

Ig: Immunoglobulin

ZNF45: Zinc finger 45 (protein or gene)

ZNF91: Zinc finger 91 (protein or gene)

BLAST: Basic Local Alignment Search Tool

MUSCLE: Multiple Sequence Comparison by Log-Expectation

OR: Olfactory Receptor

V_H and V_L domains: Heavy & Light chains of the Variable domain of Immunoglobulin molecule

KRAB C2H2-ZNF: C2H2-Zinc finger proteins associated with a KRAB domain

SCAN C2H2-ZNF: C2H2-Zinc finger proteins associated with a SCAN domain

BTB C2H2-ZNF: C2H2-Zinc finger proteins associated with a BTB domain

KAP-1: KRAB associated protein 1

TIF1β: Transcription Intermediary Factor 1β

List of definitions

Homology: This is a concept that signifies common ancestry.

Orthologs: Genes in different species, which are similar to each other and originated from a common ancestor, regardless of their functions through a speciation event.

Paralogs: Genes that are derived from a duplication event, in the same species or different species. They may or may not have the same function.

Gene duplication: Duplication of a region of DNA that contains a gene; it may occur as an error in homologous recombination, a retrotransposition event, or duplication of an entire chromosome.

Phylogenetic tree: This is also called an evolutionary tree, and shows the evolutionary interrelationships among various species or other entities that are believed to have a common ancestor.

Synteny: This describes a common order of genes, especially between related species.

Acknowledgements

Questions and Answers are what life at the university seems to be about. While trying to answer the questions about zinc fingers during my thesis, I also seem to have learnt a lot about myself. These three years at UdeM have been a wonderful learning experience both academically and personally.

First and foremost, I would like to thank my thesis supervisor, Muriel Aubry for accepting and offering me the chance to be her student and work on zinc fingers and for all the guidance and encouragement. For teaching me that what you learn during the whole process of research is as important as the end result. For supporting me when I was struggling with my courses in French. For all the days and nights of constant guidance she gave me for the thesis and for all the weekends at North Hatley. For giving me the opportunity to go to the SMBE 07, which by far has been the most exciting experience of my life. Dr.Muriel, thank you very much for everything. This experience has made me the confident person I am today.

Gertraud Burger, my co-supervisor for her valuable guidance and suggestions. For giving me the opportunity to interact with everyone from the Bioinformatics group and supporting me to let me continue in the Masters program.

Franz.B.Lang, Nicolas Lartillot, Herve Philippe, Henner Brinkmann and Amy Hauth for all the helpful guidance, discussions and constructive comments. Allan Sun for the assistance with the hardware and software problems I had.

My labmates, Patricia, Delphine, Xavier, Imene, Phuong and Hadrian for all the help, for being so nice and always making me feel welcome in the lab.

I would like to thank my friends Uma, Reena and Ekta for supporting me during the difficult times I had. Karthik for being my computer guru. My girls Lakshmi, Gayatri, Sujata, Shivani and Ramaa for taking care of me and putting up with me during the difficult times of my thesis. Siva for helping me out at the university every time I had a problem. Nagu who always let me take my frustrations and bad moods on him and for always being there to talk. Preethi and Kavitha for just being my friends.

Last but not the least; this entire experience would be at most an unfulfilled dream were it not for my loving family. I would like to express my gratitude to my parents, Dr.Nagender Swamy and Vijaya Lakshmi for supporting my dreams and aspirations, for letting me take my own decisions, make mistakes, learn and grow. My sister Vamsee Priya, my brother Charan for always being there for me and always taking care of me no matter what and my brother-in-law Sanjay.

Table of contents

List of Figures

INTRODUCTION and DISCUSSION

ARTICLE

List of Supplementary Figures

ARTICLE

List of Tables

INTRODUCTION

List of Supplementary Tables

ARTICLE

Chapter 1. INTRODUCTION

1.1 Transcription Factors

A very important problem in biology is trying to understand the mechanisms by which particular genes are expressed in a temporal or a tissue-specific manner. The process through which a DNA sequence is copied by an RNA polymerase enzymatically to produce complementary RNA is called Transcription.

The transcription process in prokaryotes and eukaryotes differs in the fact that an RNA polymerase alone can initiate transcription in prokaryotes. In contrast, eukaryotes have a much more complex transcriptional regulatory mechanism. In addition to the RNA polymerase, eukaryotic genes need an initial assembly of transcription factors at the promoter (Pabo and Sauer 1992).

Transcription factors are proteins involved in the regulation of gene expression by binding to the promoter elements upstream of genes. They are composed mainly of two functional regions 1) a DNA-binding domain and 2) an Effector domain.

The DNA-binding domain consists of amino acids that recognize specific DNA bases generally near the start of transcription. Based on its structure, the DNA-binding domain is classified into different types as detailed in Table I.

1. Zinc finger
2. Helix-turn-helix
3. Leucine zipper domain
4. Winged helix
5. ETS domain

6. Helix-loop-helix
7. Immunoglobulin fold

In addition to a DNA-binding domain, transcription factors also contain an effector domain. This domain often interacts with proteins to either inhibit or activate transcription. Transcription factors can thus act as transcriptional activators or repressors that control gene expression by acting directly on the RNA-polymerase-containing complex bound at proximity of the transcription initiation sites and/or on proteins involved in the assembly of chromatin, the complex of DNA and proteins that make up chromosomes (Roberts 2000). Transcription factors bring about these changes either by themselves or indirectly by recruiting co-factors that are called co-repressors or co-activators (Roberts 2000) depending on their effect on transcription. Co-repressors or co-activators do not bind DNA directly, but are recruited to the gene by the effector domain of transcription factors.

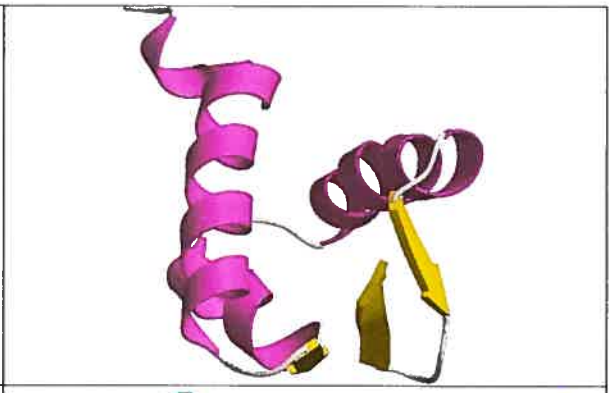
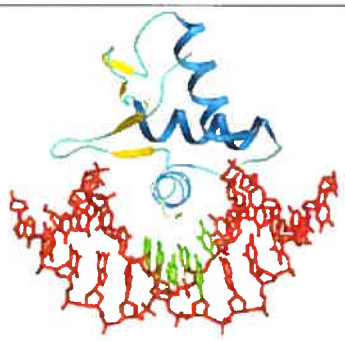
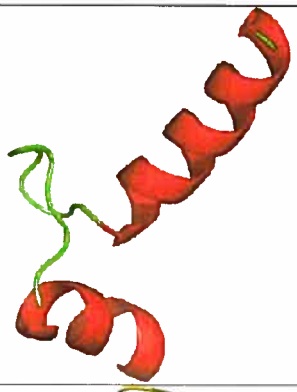
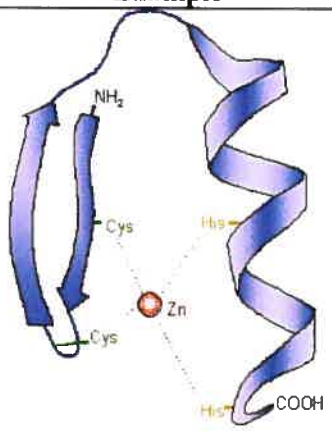
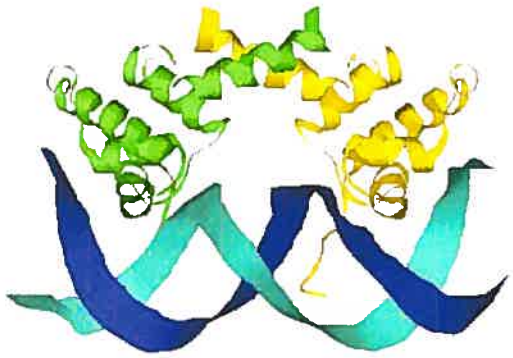
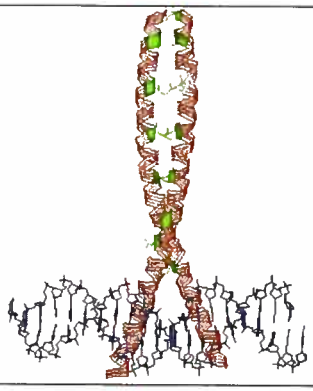
Winged helix: This motif has ~ 110 amino acids. Each domain has four alpha-helices and two beta-sheet strands. **Example:** Alpha helices are in purple and beta-strands are in yellow.	
ETS domain: This domain is 85-90 amino acids long. It was discovered in the ETS oncogene. Three alpha-helices and a 4-strand beta sheet fold into a domain. The third helix is the recognition helix. **Example:** The Elk1-E74DNA complex, where Elk-1 is a member of a large group of eukaryotic transcription factors with ETS domain. Alpha helices are in blue and beta-strands in yellow	
Helix-Loop-Helix: This motif has two alpha helices connected by a loop. Generally transcription factors with this loop are dimeric. A smaller helix allows dimerization while the other larger helix facilitates DNA binding. **Example:** Two alpha helices (in Red) connected by a loop (in Green) to form a domain.	
Immunoglobulin fold: This is also called an all β protein fold, which has a 2-layer sandwich of 7 antiparallel β-strands arranged in two β-sheets. **Example:** Human Tenascin with its immunoglobulin fold, fibronectin type III, coloured from Blue (N-terminus) to red (C-terminus).	

Table 1: Different types of DNA binding domains

DNA-binding domain	Example
Zinc finger A zinc finger has two antiparallel β strands and an α helix. Two cysteines and two histidines interact with a zinc ion to form a finger like structure. **Example:** The two cysteines on the beta-strand in green can be seen interacting with the two histidines in orange on the alpha-helix. The interacting zinc ion is shown in red in the center.	
Helix-Turn-Helix: This is a major structural unit capable of binding DNA. It has two alpha-helices which are joined by a short stretch of amino acids (turn). **Example:** Helix-turn-helix (green and yellow) of bacteriophage lambda, which binds to DNA (blue and cyan).	
Leucine zipper domain: It consists of a short alpha helix with a leucine residue at every seventh position. **Example:** The Ap-1 dimer formed by Fos and Jun homologous proteins. The leucine zipper motif has two α-helices which look like a zipper with the leucine residues (in Green) lining the zipper.	

1.2 The C2H2 zinc finger gene family

Of the many large families encoding transcription factors that have been identified, zinc finger genes of the C2H2 type constitute the largest one (Schuh, Aicher et al. 1986; Bellefroid, Lecocq et al. 1989). The C2H2 motif encoded in these genes typically includes two cysteines and two histidines coordinating a zinc ion. This motif was first identified in the TFIIIA of Xenopus *leavis* and later in the Krüppel drosophila segmentation gene (Miller, McLachlan et al. 1985; Schuh, Aicher et al. 1986). Thus, the C2H2 zinc finger genes are often referred to as TFIIIA/Kruppel type of zinc finger genes.

Known to constitute one of the ten largest gene families **[Pfam database]**, these zinc finger genes are found not only in eukaryotes but also in prokaryotes. Members of C2H2 zinc finger family have now been identified in all kingdoms of life i.e. eubacteria, archaebacteria, protists, fungi, animals and plants (Bouhouche, Syvanen et al. 2000; Moreira and Rodriguez-Valera 2000). Throughout evolution, there has been a massive expansion in the numbers of the C2H2 zinc finger genes (Lander, Linton et al. 2001; Venter, Adams et al. 2001). Noticeably, human beings are predicted to have more than 700 zinc finger genes often found in a clustered organization (Bellefroid, Lecocq et al. 1989; Looman, Abrink et al. 2002).

While most of the C2H2 zinc finger genes characterized have been described as genes encoding transcription factors which bind to DNA, some are also known to encode RNA binding proteins that may thus participate in RNA metabolism or maturation (Theunissen, Rudt et al. 1992; Grondin, Bazinet et al. 1996).

1.2.1 Structure of the proteins encoded by C2H2 ZNF genes

The C2H2 zinc finger transcription factors generally consist of two essential regions 1) the C2H2 zinc finger region containing in most instances several zinc finger motifs organized in tandem and 2) the N-terminal regulatory domain

1.2.2 The tandemly organized C2H2 zinc finger motif

The C2H2 zinc finger proteins are composed of zinc finger motifs which form the zinc finger region of the protein. Each motif is a highly conserved sequence of 28 amino acids (**C**X_{2-4}**C**X_3FX_5LX_2**H**X_{3-4}**H**TGEKPYX, where X is any amino acid). Each motif is separated from the following one by a highly conserved linker region (TGEKPYX, where X is any amino acid) (Miller, McLachlan et al. 1985; Wolfe, Nekludova et al. 2000; Looman, Abrink et al. 2002). The basic conserved C2H2 zinc finger structural unit includes two cysteines and two histidines which interact with a zinc ion and are essential for the proper folding of the motif into a finger like structure (See Figure 1) (Looman, Abrink et al. 2002). C2H2 zinc finger proteins are composed of one or more tandemly organized zinc finger motifs. The number of zinc finger motifs in the protein varies from one to more than 30 in a few cases (Ruiz i Altaba, Perry-O'Keefe et al. 1987).

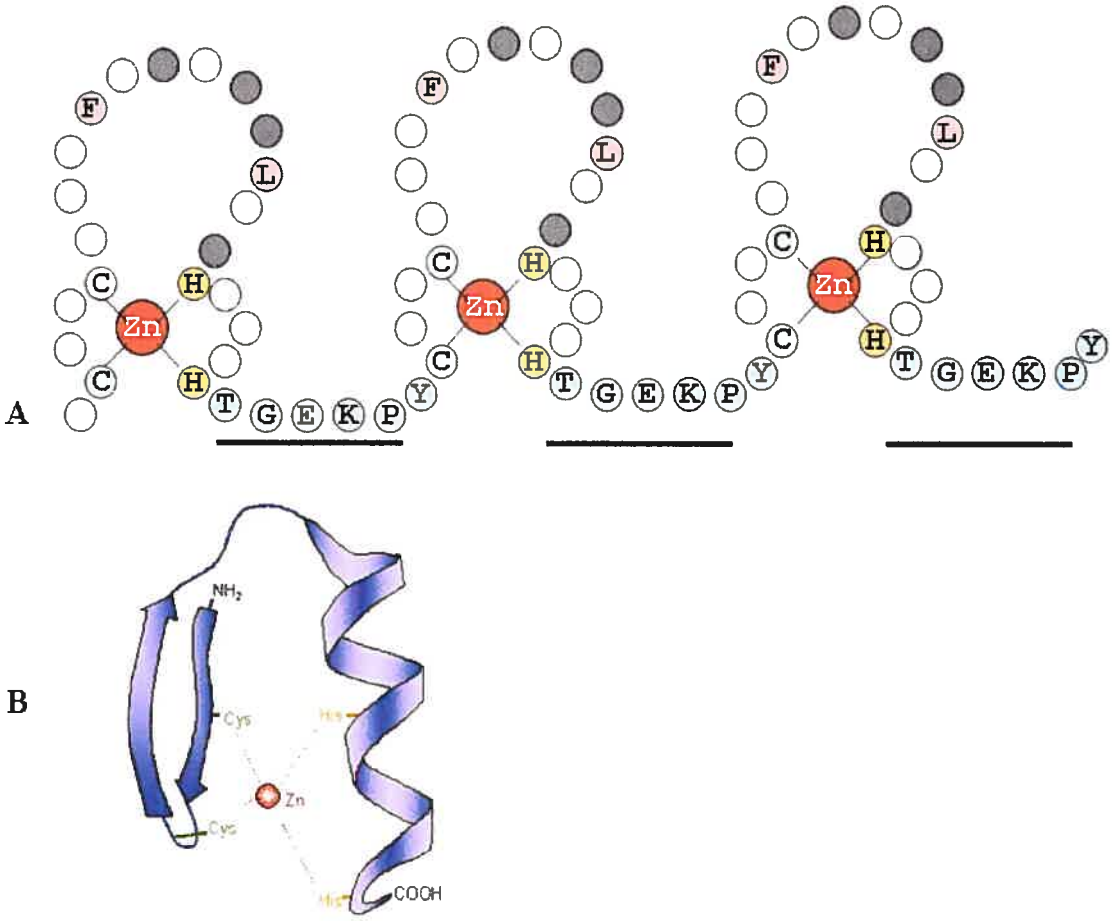


Figure 1: The basic structural unit of a C2H2 zinc finger protein.

(A) The C2H2 zinc finger motif is present in tandem in the protein. Three zinc finger motifs are connected by a conserved linker region (TGEKPY). The two cysteines and two histidines which interact with a zinc ion including the other conserved residues are shown with their single letter codes. The residues involved in DNA binding are shown in grey.

(B) The three-dimensional structure of a zinc finger binding domain. Two anti-parallel β-strands and one α-helix interact with a zinc ion as shown in the figure.

Nuclear magnetic resonance spectroscopy (NMR) was used to determine the three-dimensional structure of the C2H2 zinc finger motif (Lee, Gippert et al. 1989; Omichinski, Clore et al. 1990). Two beta strands and one alpha helix form an independently folded domain with a compact globular structure (See Figure 1). The zinc ion, that is tetrahedrally coordinated between two invariant pairs of cysteines and histidines, connects the β-sheet and the α-helix. Four amino acids on the surface of the α-helix in the zinc finger motif make base specific contacts with three to four bases in the major groove of the DNA helix (Frankel, Berg et al. 1987; Parraga, Horvath et al. 1988; Omichinski, Clore et al. 1992; Krishna, Majumdar et al. 2003). Although the zinc finger domain has been described as nucleic acid binding domain, not all the zinc finger motifs are involved in DNA or RNA binding. For example, in ZBRK1 zinc finger protein, only the first few fingers are involved DNA binding and all the others in protein-protein interactions (Zheng, Pan et al. 2000).

1.2.3 The N-terminal regulatory domain of C2H2 zinc finger proteins

In addition to the zinc finger region, C2H2 zinc finger proteins are also associated with an N-terminal regulatory domain (Figure 2), which regulates subcellular localization and the gene expression by acting as either a repressor or an activator by itself or by associating with other factors (Collins, Stone et al. 2001).

The regulatory domains associated with C2H2 Zinc finger proteins are

i. BTB/POZ domain

ii. KRAB domain

iii. SCAN domain

i. The BTB domain

The BTB domain (**B**road-Complex, **T**ramtrack and **B**ric-a-bric) also known as the POZ domain (Pox virus and Zinc finger) is a 120 amino acid conserved domain found to be associated with both DNA and actin-binding proteins. The BTB domain is involved in protein-protein interactions (Collins, Stone et al. 2001). As a part of DNA binding proteins, the BTB/POZ domain is known to be a dimerization domain which, in a few cases also recruits co-repressors (such as N-CoR, SIN3A or SMRT) and acts a repression domain. When found in association with C2H2 zinc finger transcription factors, the BTB domain is generally located N-terminal to the zinc finger region. Thus, by mediating oligomerization and in some instances interaction with co-factors, the BTB domain can lead to chromatin remodeling and change in gene expression (Melnick, Carlile et al. 2002).

ii. The Krüppel-Associated Box (KRAB) domain

Another well known example of an N-terminal regulatory domain associated with C2H2 zinc finger proteins is the Kruppel-Associated Box or the KRAB domain (Bellefroid, Poncelet et al. 1991; Rosati, Marino et al. 1991). KRAB domains are almost

always associated with C2H2 zinc finger proteins. An exception to this scenario is the SSX family of proteins. These proteins are associated with a "SSX KRAB domain" which is distantly related to the KRAB domain from C2H2 zinc finger proteins (~ 49% similar) but are not associated with zinc fingers (Collins, Stone et al. 2001; Urrutia 2003).

Unlike the C2H2 zinc finger proteins which are present in organisms ranging from bacteria to humans, the KRAB domain as seen in C2H2 zinc finger proteins is present only in vertebrates, more specifically in tetrapods (Looman, Abrink et al. 2002). However, a recent study identified a sea urchin homolog to the mammalian Meistez protein which includes a tandem array of C2H2 zinc finger motifs, a SET domain and a sequence with some similarity to the "SSX-KRAB domain" (Birtle and Ponting 2006). This suggests the presence of the KRAB domain in the common ancestor of echinoderms and vertebrates. A further study of these proteins in fungi and plants identified a 26 amino acid motif called the KRI motif which was found to be similar to the alpha-helical regions of KRAB and present in all eukaryotes. This indicated that the KRI motif was present in the last common ancestor of animals, plants and fungi and is the progenitor of the KRAB domain.

The KRAB domain is most abundant in mammals (Lander, Linton et al. 2001; Venter, Adams et al. 2001; Waterston, Lindblad-Toh et al. 2002). For example, about one-third of the mouse C2H2-ZNF are associated with KRAB (Benn, Antoine et al. 1991; Waterston, Lindblad-Toh et al. 2002). The KRAB domain is mostly associated with more than 5 C2H2 zinc finger motifs in a protein, justifying the name "Multifingered protein" (Bellefroid, Poncelet et al. 1991). Many genes encoding the KRAB containing proteins are found in a clustered organization as opposed to the ones found as singletons (Bellefroid,

Marine et al. 1993; Shannon, Kim et al. 1998; Chung, Schafer et al. 2002; Rousseau-Merck, Koczan et al. 2002; Hamilton, Huntley et al. 2003).

The KRAB domain is ~75 amino acids long and is divided into two boxes, Box A (~38 amino acids) and Box B (~32 amino acids) (Looman, Abrink et al. 2002; Urrutia 2003). A variant of the B box, called b box also exists. Some C2H2 zinc finger proteins have another box following the A box, called the C box (21 amino acids). Each of these boxes is encoded by different exons and separated by introns of varying lengths (Looman, Hellman et al. 2004). The KRAB domain functions as a potent repressor of transcription (Margolin, Friedman et al. 1994). The KRAB A box plays an important role in repression by binding to co-repressors, while the KRAB B box doesn't have transcriptional activity but is known to enhance the repression activity of the A box (Witzgall, O'Leary et al. 1994). The process of transcription repression is mediated by KAP-1, also called transcription intermediary factor 1β (TIF1β) which is a co-repressor that interacts with KRAB A (Friedman, Fredericks et al. 1996; Germain-Desprez, Bazinet et al. 2003). The KRAB domain of C2H2 zinc finger proteins recruits the KAP1 co-repressor to DNA, which results in the formation of a heterochromatin like complex and leads to gene silencing (Pengue, Calabro et al. 1994; Kim, Chen et al. 1996; Moosmann, Georgiev et al. 1996; Pengue and Lania 1996).

iii. The SCAN domain

The SCAN domain like the KRAB domain is another vertebrate specific domain only associated with C2H2 zinc finger proteins (Williams, Khachigian et al. 1995; Looman,

Abrink et al. 2002). The SCAN domain was estimated to be associated with 10% of the total C2H2-ZNF present in the human genome (Collins, Stone et al. 2001; Edelstein and Collins 2005). Also known as the LeR domain because of its leucine rich primary structure, the SCAN domain is named after the four proteins it was initially identified (**S**RE-ZBP, **C**Tfin51, **A**W-1 and **N**umber18 cDNA) (Urrutia 2003). In addition to being associated with the C2H2 zinc finger proteins, the SCAN domain containing proteins are sometimes associated with a KRAB domain having the organization SCAN-KRAB-$(C2H2)_x$ or in very few cases KRAB-SCAN-KRAB-$(C2H2)_x$ (Edelstein and Collins 2005; Huntley, Baggott et al. 2006).

Structural studies on the SCAN domain indicate that it has ~ 84 amino acids and is found to have three to five α-helices which are delineated by one or more proline residues. Proline residues are also present before and after the SCAN domain (Stone, Maki et al. 2002). The SCAN domain is a homo and hetero-dimerization domain mediating protein-protein interactions by self association and formation of heterodimers between SCAN family members (Sander, Haas et al. 2000; Schumacher, Wang et al. 2000). The importance of the dimerization for the transcriptional activity of SCAN-C2H2 zinc finger protein has not been clearly established.

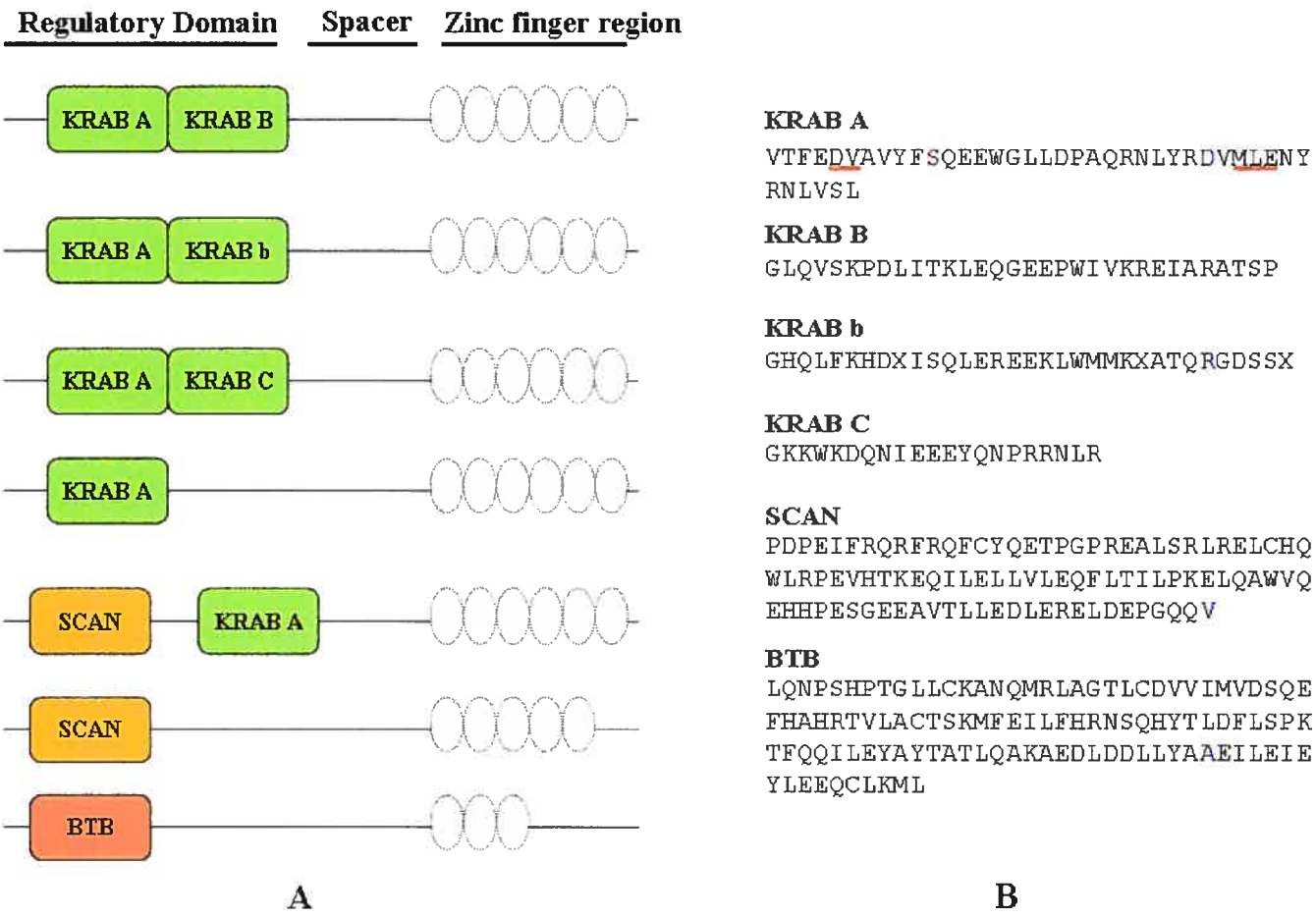


Figure 2: The Regulatory domains associated with C2H2 Zinc finger proteins.

(A) The different combinations of domains associated with zinc finger proteins are shown. Zinc finger proteins generally have three main regions: The Regulatory domain, the Spacer and the Zinc finger region. **(B)** The consensus sequence of the domains KRAB (A, B, b and C boxes), SCAN and BTB. The residues essential for binding KAP1 and thus for repression are shown in KRAB A underlined in red.

1.3 Gene families and Gene duplication

1.3.1 Gene Families

A gene family corresponds to a set of genes that are grouped based on their shared homology, biological or biochemical activity, sequence motifs or similarities in structure. Because they consist of a large number of genes, gene families are the most informative systems to study evolutionary dynamics of genes. Nuclear genomes have many multigene families and their studies provide clues to the evolutionary forces that have shaped these genomes (Ohta 2000; Thornton and DeSalle 2000). Mammalian genomes in particular have large numbers of genes organized in gene families (Demuth, Bie et al. 2006). Some gene families have uniform copy numbers of genes in all species (Thornton and DeSalle 2000), while there are gene families like the Immunoglobulin gene family, the Olfactory receptor gene family and the C2H2 zinc finger gene family which have a large variation in the number of genes across different species . The variation in the gene numbers of these families and diversity in function, suggests that gene duplication and/or gene loss have played an important role in shaping different mammalian genomes.

i. The Olfactory Receptor gene family

Olfactory receptor (OR) genes form the largest known multigene family in mammalian genomes (Glusman, Bahar et al. 2000) and code for membrane receptors that are responsible for olfaction, the sense of smell. OR genes are present in various vertebrates ranging from lampreys to humans. The OR proteins belong to the G-protein coupled

receptor family which have seven transmembrane domains. OR genes are divided into 2 classes based on their protein sequence similarity (Glusman, Bahar et al. 2000; Fuchs, Glusman et al. 2001). Of the two classes, Class I genes first identified in fish but also found in mammals are specialized in water-soluble odorants and the Class II genes specialized for airborne odorants are specific to tetrapods.

The number of the OR genes is quite varied in different genomes. Rodents have nearly twice as many as the number present in human, chimpanzee or dog (Niimura and Nei 2005). The Human genome has more than half of the ~900 OR genes as pseudogenes. In contrast, the mouse genome has ~1300 OR genes of which only one-fourth are pseudogenes. Studies on the human, chimpanzee and mouse OR gene repertoires indicate that there are species-specific expansion and pseudogenization signifying different selection pressures in humans, chimpanzees and mouse owing to their different sensory requirements (Sharon, Glusman et al. 1999; Glusman, Yanai et al. 2001; Lapidot, Pilpel et al. 2001; Gilad, Man et al. 2005; Niimura and Nei 2005). Evolutionary analysis of the human, mouse and chimpanzee datasets indicate the presence of clustered organization which is generally well conserved in these genomes. Analyses of the clusters indicate that there are tandem arrays of the OR genes which appear to have arisen by tandem duplication and several chromosomal rearrangements. The difference in the numbers of OR genes in human and mouse has been attributed to gene duplication and loss events (Sharon, Glusman et al. 1999; Niimura and Nei 2005).

ii. The Immunoglobulin gene family

The immunoglobulin gene family represents an example where its two subfamilies, the immunoglobulin heavy variable region sub-family and immunoglobulin light chain variable region subfamily, have co-evolved by varying in gene number and extent of diversity in different species (Sitnikova and Su 1998). An immunoglobulin molecule is a tetramer with two identical heavy chains and two identical light chains which form a Y-shaped structure. Each of these chains has a variable (V) and constant (C) domain. The V_H and V_L domains have the complementarity determining regions, called the CDRs which form the sites of interaction with antigens. Analyses of these two sub-families of genes from various species of amniotes identified that these gene families have diversified throughout the course of evolution (Sitnikova and Su 1998). Different coordinated loss and duplication events have led to different species-specific gene repertoires.

iii. The C2H2 zinc finger gene family

In addition to the above mentioned gene families, the C2H2 Zinc finger gene family is another example of a large multigene family with varying number of genes in different species. Over the course of evolution, this gene family has expanded drastically in mammalian genomes (e.g. ~ 400 in mouse and ~ 700 in human) (Venter, Adams et al. 2001; Waterston, Lindblad-Toh et al. 2002). Several studies involving these genes in the human genome have indicated that tandem duplication events are responsible for the

clustered organization of this family (Shannon, Kim et al. 1998; Elemento and Gascuel 2002; Elemento, Gascuel et al. 2002; Tang, Waterman et al. 2002; Bertrand and Gascuel 2005; Huntley, Baggott et al. 2006). A few instances of evolutionary studies of these genes, within the human genome and among a few mammalian genomes document cases of species-specific duplication (Dehal, Predki et al. 2001; Shannon, Hamilton et al. 2003; Huntley, Baggott et al. 2006).

All these examples of gene families suggest variation in number among different species involving different duplication and loss events. The gene family size could vary based on the functional relevance of the gene family in the organism. These examples also indicate the importance of studying the gene families to give clues on the evolutionary mechanisms which led to different sizes of gene families.

1.3.2 Gene Duplication and Gene Loss: Two important evolutionary mechanisms guiding the evolution of gene families in mammals

Considering the extremely large numbers of genes constituting gene families (Demuth, Bie et al. 2006), it is interesting to study their organization and the evolutionary mechanisms that created them. A study integrating the information from spatial organization of the genes with the phylogenetic relationships between the genes combined with evolutionary information of the species would help provide clues about the evolution of the gene families.

In the context of using phylogenetic studies to analyze the evolutionary relationships between genes in gene families, one significant term that features in all studies is "Homology". Homology forms the central and basic concept of comparative genomics but is also a term that is often misrepresented and misinterpreted. The term homology was introduced by Richard Owen in 1848, where he defined homology as "the same organ under every variety of form and function"(Francis Darwin 1903) . The importance of structure and function is emphasized more in this definition. In an attempt to give an evolutionary explanation to homologous structures, Darwin defined homology as "A structure is similar among related organisms because those organisms have all descended from a common ancestor that had an equivalent trait" (Darwin 1837) (Figure 3).

When put in the context of molecular sequence comparison, in today's times, homology[1] refers in an abstract way to a relationship which implies a possible common ancestry and should be differentiated from identity[2] or similarity[3] of sequences. However, to be substantiated, homology must be confirmed by appropriate phylogenetic studies. It is important to note that homology does not say anything about functional similarity (Thornton and DeSalle 2000))(Fitch 2000).

131

[1] Homology: A hypothesis that signifies common ancestry between sequences (nucleotide or amino acid) which is primarily based on sequence similarity.

[2] Identity: The extent to which two (nucleotide or amino acid) sequences are invariant.

[3] Similarity: The extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation

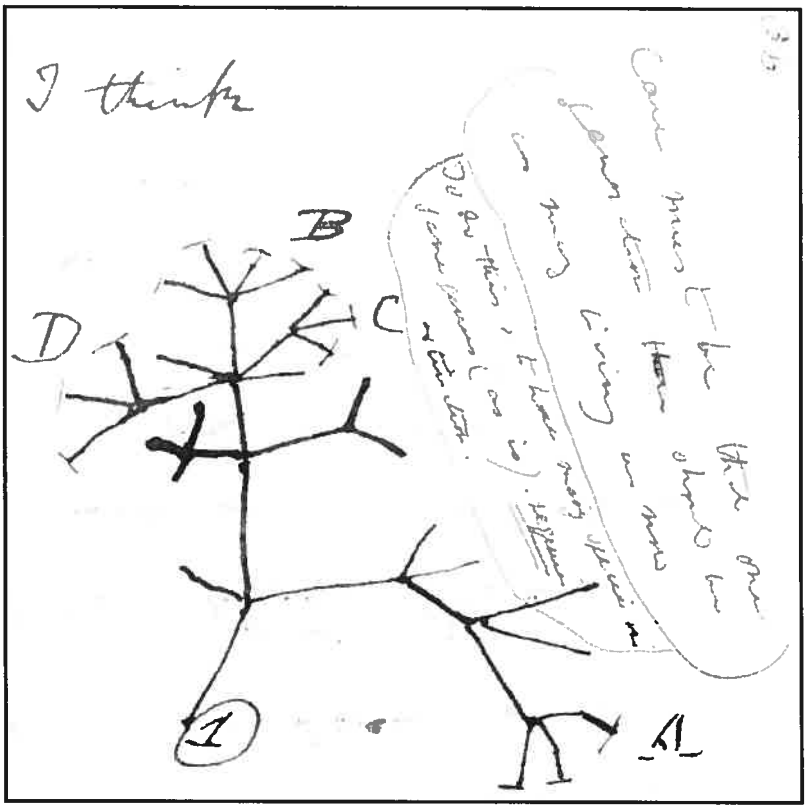


Figure 3: Darwin's evolutionary tree.

The figure is Charles Darwin's first ever sketch of an evolutionary tree from his book titled "First Notebook on Transmutation of Species (1837)".

There are three major types of homology in a phylogenetic context which are orthology, paralogy and xenology. *Orthology* as described by Fitch in 1970 is the relationship between two genes in two different species which originated from a common ancestor. Two homologous sequences are considered to be "orthologous" if a speciation event separates them. In contrast, *Paralogy* signifies the relationship between two genes which have been formed by a gene duplication event. *Xenology*, another type of homology relationship describes the relationship between two genes which have been transferred between two species by horizontal gene transfer.

Studying the homologous relationships of genes within and between various genomes and differentiating between orthologs and paralogs is a central aspect of comparative genomics. *Figure 4* shows a very simple explanation of the difference between orthologs and paralogs. The genes A1, B1 and B2 have evolved from an ancestral gene by speciation followed by a duplication event in species B. Gene A1 from species A is an ortholog of gene B1 and gene B2 in species B illustrating that one gene in a particular species may have more than one ortholog in the other. Gene B1 and gene B2 in species B, which were formed by gene duplication, are paralogs to each other.

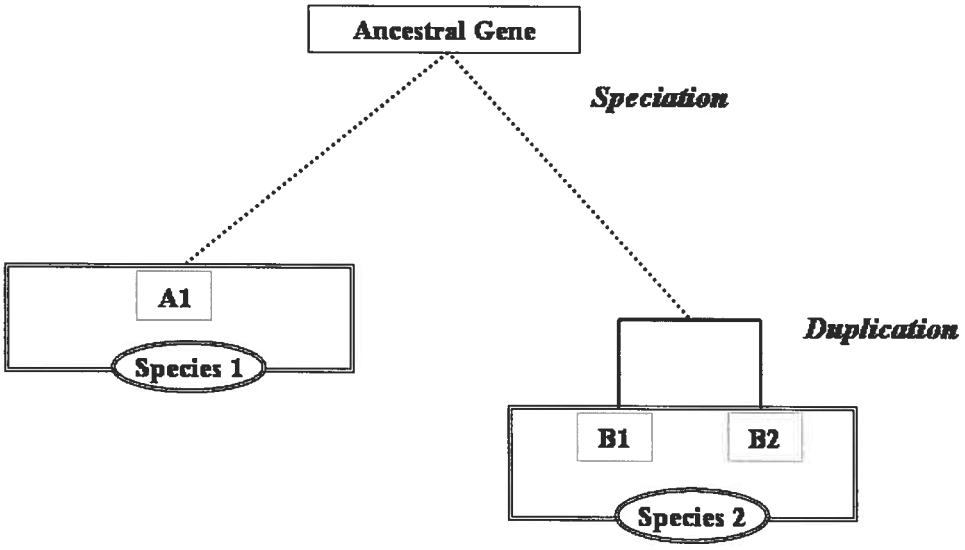


Figure 4: Schematic representation of speciation and duplication

Genes A1, B1 and B2 are formed from an ancestral gene by a speciation and duplication event. Gene A1 from species A has two orthologs in species B, genes B1 and B2. B1 and B2 are paralogs.

Figure 5 depicts different evolutionary scenarios that one encounters while studying the evolution of genes in gene families are depicted to explain the relationships between genes in terms of orthologs, paralogs and gene loss. An ancestral gene undergoes duplication in species 0 to give the genes, A, B and C. This is followed by a speciation event with genes A1, B1 and C1 in Species 1 and genes A2 and B2 in Species 2. The gene A1 is an ortholog of A2 and B1 is an ortholog of B2. The gene C1 does not have a corresponding ortholog, as the Species 2 lost the gene after speciation. The genes A1, B1 and C1 are paralogs within species 1 and, A2 and B2 are paralogs within species 2. Furthermore, as explicitly pointed out recently by Fitch (Fitch 2000) and as often ignored, gene A1 (species 1) is also a paralog of gene B2 (species 2), gene B1 (species 1) is a paralog of gene A2 (species 2) and gene C1 is paralog of gene A2 and B2 (species 2). From these explanations, it is clear that orthologs are homologous genes residing in different species, while paralogs may not only refer to the homology relationship between genes from the same species but also from different species. It is essential to understand that both orthologs and paralogs are free to diverge and do not necessarily always have the same function (Thornton and DeSalle 2000).

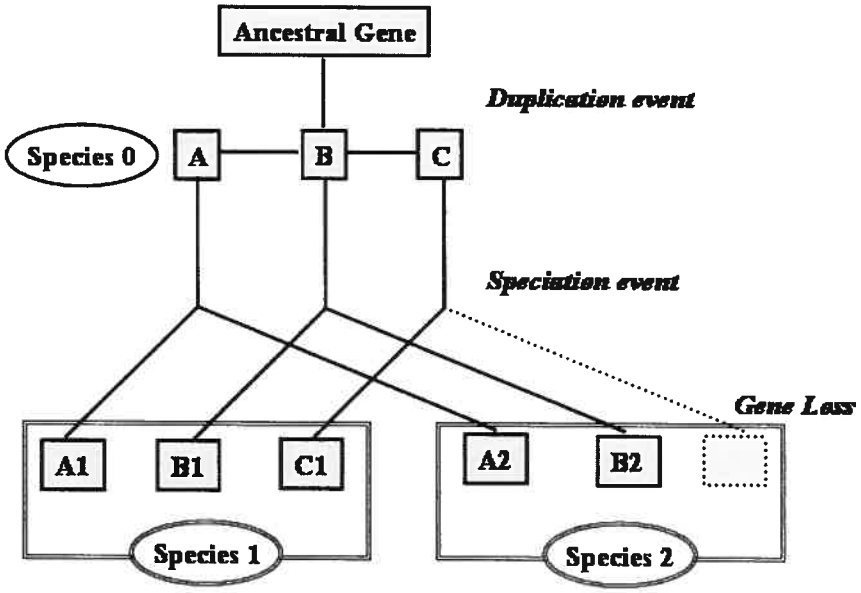


Figure 5: Schematic representation of different evolutionary processes shaping the gene families in different species.

Gene duplication, speciation and loss lead to the formation of genes A1, B1, C1 in species 1 and A2 and B2 in species 2.

1.4 Inferring gene duplication and gene loss

That two genes are homologous is a hypothesis that needs to be studied and analyzed to be able to derive the relationships between the genes to be either orthologs or paralogs. Studying and analyzing the relationships between gene families i.e. evaluation of orthology or paralogy requires a well formulated approach. In order to be able to postulate theories on how related genes evolved from an ancestral gene i.e. by gene duplication, gene loss or by speciation, one needs to assess homology relationships using a well founded phylogeny.

The first step in assessing homology is a sequence alignment of the molecular sequences be it nucleotides or amino acids. This gives a preliminary measure of possible homology which can then be assessed using a phylogeny. A well supported phylogeny gives the evolutionary relationships between the genes in relation to one another. Comparison of a gene phylogeny between genes with the taxonomic relationships between species, allows gene duplication and loss events to be assessed and roughly dated. As an example, *Figure 6* shows the different scenarios of gene duplication and gene loss. In *Figure 6A*, it can be seen that a duplication event prior to the speciation event resulted in species 2, 3 and 4 created the paralogous gene groups A, B and eventually C. In a hypothetical situation, suppose the genes 2A, 3C and 4B are missing from the gene tree. Assuming that the studied sequences were derived from completely sequenced genomes, the missing genes could either be due to a loss of genes or their possible pseudogenization in the respective species. That the gene duplication occurred prior to speciation can still be

resolved by superimposing the gene tree with the species tree. Reconciliation between the species and the gene tree will help resolve the absence of genes 2A, 3C and 4B as can be seen in *Figure 6B*. This kind of study can hence be used to infer the evolution of genes belonging to gene families within and among species in a phylogenetic context.

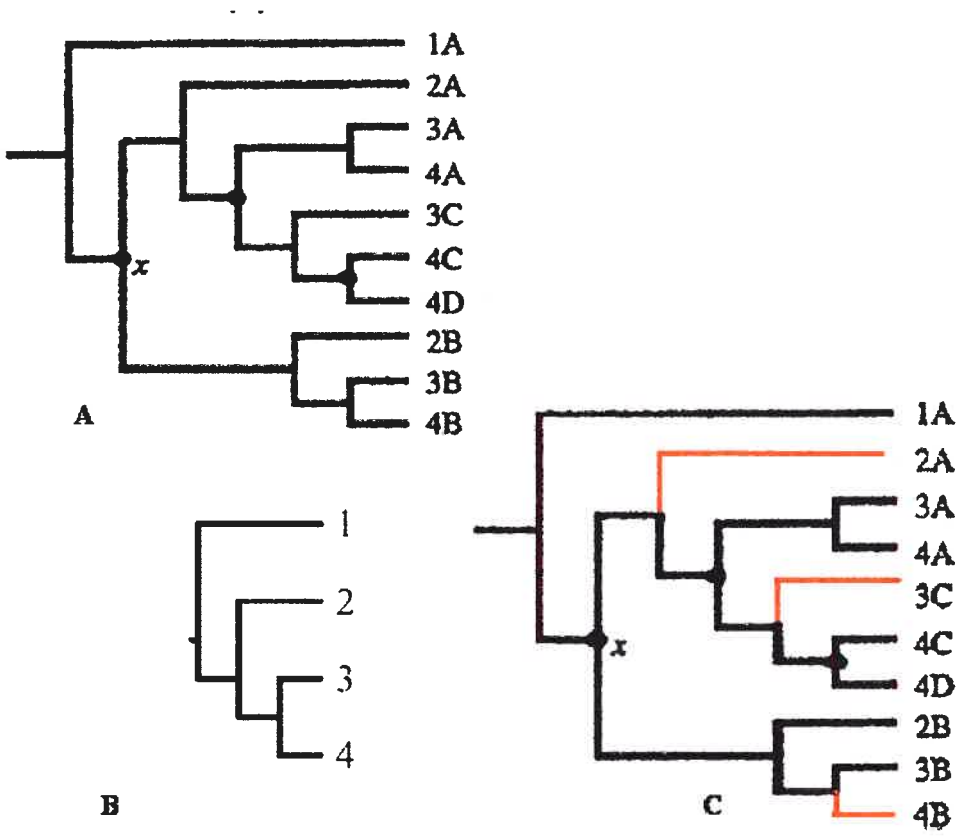


Figure 6: Inferring gene duplication and loss events from a gene tree in comparison with the species tree.

(A) A gene tree showing the phylogeny between genes belonging to species 1, 2, 3 and 4. Genes are represented as A, B, C and D. A gene duplication represented as *x* occurred prior to the speciation event leading to Species 2, 3 and 4.

(B) A species tree showing the relation between species 1, 2, 3 and 4.

(C) A hypothetical situation where the genes 2A, 3C and 4B are missing from the tree as shown in red. Reconciliation of the phylogenetic tree from (A) with the species tree from (B) helps identify the not only the duplication event but also the missing genes to be able to infer loss. Adapted from (Thornton and DeSalle 2000)

1.5 Previous Studies addressing zinc finger gene evolution

About 2000 of the ~ 30,000 genes in the human genome code for transcription factors (Venter, Adams et al. 2001). C2H2-ZNF are the most common of all the eukaryotic transcription factors present in the human genome (encoded by ~ 700 genes). Owing to these facts, the C2H2 zinc finger gene family has been considered to be an evolutionary playground for genes to develop and differentiate and hence is an interesting family to study (Looman, Abrink et al. 2002).

The studies pertaining to C2H2-ZNF have mostly been restricted to those associated with a KRAB domain and more specifically to the human genome. A recent study identified 423 KRAB C2H2-ZNF loci organized into 65 clusters on the human genome (Huntley, Baggott et al. 2006). Evolutionary studies involving these KRAB C2H2-ZNF genes indicated that the evolutionary relatedness within and among clusters was not only associated with physical proximity evolving through tandem duplications but also through distributed duplication and post-duplication rearrangement events, which have led to the dramatic increase in the gene numbers of this family in humans (Hamilton, Huntley et al. 2003; Hamilton, Huntley et al. 2006; Huntley, Baggott et al. 2006). Though present in clusters, the KRAB C2H2-ZNF are not co-regulated and they show different patterns of expression (Huntley, Baggott et al. 2006).

A study of one KRAB C2H2-ZNF gene cluster on human chromosome 19, suggested an evolutionary model showing the presence of certain beta-satellite repeat structures symmetrically ordered with the zinc finger genes in the cluster which have coevolved with the cluster accommodating the expansion of the genes within this cluster (Eichler, Hoffman et al. 1998).

A statistical analysis using phylogenetic models on four human C2H2-ZNF clusters on chromosome 19 indicated that positive selection is the driving force involved in the diversification of the KRAB C2H2 zinc finger genes (Schmidt and Durrett 2004).

Not much is known about the evolutionary histories of these genes in different mammalian genomes and very few studies have been carried out to comparatively analyze their evolution. A preliminary report on species-specific expansion of these genes, resulted from one study on a C2H2-ZNF cluster on human chromosome 19 and its syntenically homologous cluster on mouse chromosome 7 (Shannon, Hamilton et al. 2003) . A study on the evolution of members of the primate-specific ZNF91 KRAB subfamily, which are mainly found in a chromosome 19 cluster, revealed that this gene subfamily evolved before the split of humans and apes. But after the split, these genes have continued to evolve differentially be it through tandem duplications or segmental duplications, leading to species-specific genes (Dehal, Predki et al. 2001; Hamilton, Huntley et al. 2006).

Inspite of several studies dealing with these genes, there has never been a comprehensive study on the C2H2-ZNF genes and their evolution or their functions. In

order to systematically define and analyze the extent of species-specific duplication and the role of gene loss in the evolution of these genes, it is important to conduct a comprehensive study of these gene clusters in mammalian genomes to obtain clues on their evolution and their possible implications on functions specific to each species.

1.6 Hypothesis and Objective

Previous studies on zinc finger genes have provided evidence that zinc finger genes have undergone a huge expansion in vertebrate genomes, with a specific increase in humans. Studies have shown that these genes have been subjected to expansion through tandem duplication and also of the existence of species-specific duplication events (Shannon, Kim et al. 1998; Shannon, Hamilton et al. 2003; Hamilton, Huntley et al. 2006; Huntley, Baggott et al. 2006). A contribution of gene loss in the evolution of C2H2 zinc finger genes has been suggested but never tested rigorously.

The main objective of this thesis is to systematically determine to what extent zinc finger genes are submitted to species-specific expansion and to assess the potential contribution of gene loss in the evolution of this gene family in mammals. To this end, we have:

1. Assembled a curated database of all C2H2-ZNF genes in the human genome and identify all the C2H2-ZNF clusters in the human genome

2. Searched for syntenically homologous clusters in other completely sequenced mammalian genomes, namely chimpanzee, mouse, rat and dog genomes.
3. Performed a phylogenetic analysis of C2H2-ZNF genes from the syntenically homologous clusters.
4. Performed a reconciliation of both phylogenetic analyses and physical maps of the clusters with the species tree accounting for the evolutionary history of the species in order to infer gene loss and gain.

These studies should allow us to determine the nature of evolutionary events that shaped this large gene family in mammals. In particular, this study will help us to better infer orthology in the various mammals and better understand the evolution and relationships between the different C2H2-ZNF subfamilies.

Chapter 2. ARTICLE

Evolution of C2H2-zinc finger genes in mammals: Species-specific duplication and loss at the level of clusters, genes and their functional domains.

Hamsa Dhwani Tadepally, Gertraud Burger and Muriel Aubry*

Department of Biochemistry, Université de Montreal, C.P.6128, Succ.Centre-Ville, Montreal, QC, H3C 3J7, Canada

To whom correspondence and reprints should be addressed:
Muriel Aubry, Ph.D.
Department of Biochemistry
Université de Montréal
C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7
Canada

Key words: C2H2/Kruppel, zinc finger, gene family, tandem repeats, gene duplication, gene loss, evolution.

ABSTRACT

C2H2 zinc finger genes (C2H2-ZNF) constitute the largest class of transcription factors in humans and one of the largest gene families in mammals. Often arranged in clusters in the genome, these genes are thought to have undergone a massive expansion in vertebrates by a process involving tandem duplication. However, this view is based on limited datasets restricted to single chromosome or a specific subfamily of C2H2-ZNF genes. Here, we present the first comprehensive study of the dynamic evolution of the C2H2-ZNF family in mammals. We assembled the complete repertoire of human C2H2-ZNF genes (718 in total), about 70 % of which are organized into 81 clusters across all chromosomes. Based on an analysis of their N-terminal effector domains, we identified SET- and HOMEO domain-encoding C2H2-ZNF genes as members of two new C2H2-ZNF subfamilies. We searched for the syntenic counterparts of human clusters in other mammals for which complete gene data are available: chimpanzee, mouse, rat and dog. Cross-species comparisons show a large variation in the numbers of C2H2-ZNF genes within homologous mammalian clusters suggesting differential patterns of evolution. Phylogenetic analysis of selected C2H2-ZNF clusters reveals that differences in C2H2-ZNF gene repertoires across mammals not only originate from differential gene duplication but also gene loss. Further, we find variations among orthologs in the number of zinc finger motifs and association of the effector domains, the later often undergoing sequence degeneration. Based on these results and an analysis of the exon-intron organization of genes from the large SCAN and KRAB domains-containing subfamilies, we propose a new model for the evolution of these subfamilies.

This manuscript includes two supplementary Figures and four supplementary Tables

INTRODUCTION

The human genome sequence uncovered a large number of gene families often arranged in a clustered organization (Ohta 2000; Thornton and DeSalle 2000; Venter, Adams et al. 2001). C2H2 zinc finger (C2H2-ZNF) genes make up ~ 2 % of all the human genes and represent the second largest gene family in humans after the odorant receptor family (Lander, Linton et al. 2001) (Schuh, Aicher et al. 1986; Bellefroid, Lecocq et al. 1989; Messina, Glasscock et al. 2004). The first identified members of the C2H2-ZNF family are *Xenopus* TFIIIA and Drosophila Kruppel and thus genes of this family are often called zinc finger genes of the TFIIIA or Kruppel type (Miller, McLachlan et al. 1985; Schuh, Aicher et al. 1986).

Most of the characterized C2H2-ZNF genes code for transcription factors which bind DNA through their zinc finger region; others bind RNA and their exact function is yet unknown (Theunissen, Rudt et al. 1992; Grondin, Bazinet et al. 1996). The zinc finger region is composed of a basic structural unit of 28 amino acids ($CX_{2-4}CX_3FX_5LX_2HX_{3-4}$HTGEKPYX, where X is any amino acid), called the zinc finger motif, that is often repeated in tandem. The two cysteines and two histidines in this motif interact with a zinc ion, stabilizing the proper folding of this motif (Klug and Rhodes 1987; Lee, Gippert et al. 1989; Rhodes and Klug 1993). C2H2-ZNF proteins often contain an effector domain always located N-terminal to the zinc finger region, such as the KRAB (Kruppel-Associated-Box), SCAN (SRE-ZBP, CTfin51, AW-1 and Number18 cDNA) and BTB (Broad-Complex, Tramtrack and Bric-a-bric) domains. The first two domains are

vertebrate-specific (Bellefroid, Poncelet et al. 1991; Rosati, Marino et al. 1991; Collins, Stone et al. 2001), while BTB is also present in insects. The KRAB domain includes the box KRAB A (~38 amino acids) involved in transcriptional repression and often a second box, usually KRAB B (~32 amino acids) or in few cases KRAB b or KRAB C (~21 amino acids) box (Witzgall, O'Leary et al. 1994; Looman, Abrink et al. 2002; Urrutia 2003; Looman, Hellman et al. 2004). The KRAB A box and the second KRAB B, b or C box are encoded by separate exons, which are alternatively spliced. The SCAN, also called the leucine-rich (LeR) domain (~ 84 amino acids) (Stone, Maki et al. 2002) mediates protein-protein interactions through dimerization (Sander, Haas et al. 2000; Schumacher, Wang et al. 2000). The BTB domain (~ 120 amino acids) is a dimerization domain that also acts as a repression domain in some cases (Melnick, Carlile et al. 2002). In contrast to the SCAN and KRAB domains which are only present in C2H2-ZNF proteins, the BTB domain is also found as a part of actin-binding proteins (Collins, Stone et al. 2001). C2H2-ZNF proteins are grouped into different subfamilies based on the type of N-terminal effector domain present.

Initial studies on the C2H2-ZNF gene family focused on human chromosome 19, which is particularly enriched in clusters of these genes (Bellefroid, Marine et al. 1993; Eichler, Hoffman et al. 1998). More recent studies dealt more specifically with the KRAB subfamily (Mark, Abrink et al. 1999; Looman, Abrink et al. 2002; Shannon, Hamilton et al. 2003; Huntley, Baggott et al. 2006). The current view is that C2H2-ZNF genes have undergone a massive expansion during vertebrate evolution by a process involving tandem duplication (Dehal, Predki et al. 2001; Looman, Abrink et al. 2002; Hamilton, Huntley et

al. 2003; Shannon, Hamilton et al. 2003; Hamilton, Huntley et al. 2006; Huntley, Baggott et al. 2006). Yet, this view may be biased because it is extrapolated from small subsets of C2H2-ZNF genes.

In this report, we reconstructed a global picture of the evolution of the C2H2-ZNF gene repertoires during mammalian speciation, based on a comprehensive catalogue of all human C2H2-ZNF genes and their syntenic counterparts present in clusters in other mammals. Our study clearly demonstrates that this gene family expanded and contracted not only in human but across mammals and in a lineage-specific fashion. In addition, we discovered evolutionary change of individual C2H2-ZNF orthologs involving both differential duplication of zinc finger motifs and loss of N-terminal effector domains. Speciation of mammals is characterized by divergent evolutionary trends at the level of individual C2H2-ZNF genes as well as the entire family. This led us to propose a model for the evolution of SCAN, SCAN-KRAB and KRAB subfamilies and points to the importance of comparing complete repertoires rather than C2H2-ZNF genes from specific subfamilies for gaining insights into the possible orthologous relationships between genes from various genomes.

METHODS

Collection of human C2H2 zinc finger genes

We conducted an extensive similarity search to identify the complete repertoire of C2H2-ZNF genes in the human genome (assembly NCBI 36). First, we identified all the genes annotated as C2H2 and/or Kruppel zinc finger genes by performing an initial text term search via Entrez (www.ncbi.nlm.nih.gov). Second, we used PROSITE (http://www.expasy.com) to identify all the proteins which had a zinc finger motif of the C2H2 type as well as the N-terminal effector domain, if present.

From these searches, the genomic coordinates, chromosome number, position on the chromosome, number of fingers and identified domains were collected for each of the gene and protein sequences (initial dataset). A TBLASN (e-value cutoff 1e-3) (Gertz, Yu et al. 2006) search was done against the genome using each of the gene sequences from the initial dataset as a query. The blast hits were used to generate the final dataset of all the identified C2H2-ZNF genes (Supplementary Table S1).

Identification of C2H2-ZNF gene clusters in the human genome

We analyzed the relative positions of C2H2-ZNF genes in the human genome in order to identify the C2H2-ZNF clusters. A distribution of the distances between neighboring C2H2-ZNF genes in the human genome is presented in Supplementary Figure S1. Two consecutive C2H2-ZNF genes are said to belong to a cluster if the distance between them is $\leq$500 kb regardless of the presence of other genes within the cluster , a

threshold classically used in gene family studies (Niimura and Nei 2003). Clusters were determined for each human chromosome.

Identification of mammalian C2H2-ZNF clusters syntenically homologous to human clusters

We searched for clusters homologous to the human C2H2-ZNF clusters (i.e. syntenically homologous clusters) in other mammals for which complete genome sequences are available. The assemblies used for *Pan troglodytes)*, *Mus musculus*, *Rattus norvegicus* and *Canis familiaris* were chimpanzee Pan Tro- 2.1, mouse NCBI m36, rat RGSC 3.4 and dog Can Fam 2.0. We used the linkage maps of Ensembl (http://www.ensembl.org); assignment of syntenic clusters is based on the genes flanking each human cluster and which were mapped in all the species. Four flanking genes at each extremity were mapped in most instances. Then, we conducted TBLASTN analysis of the syntenic regions comprised between the flanking genes, using as queries the amino acid sequence of the zinc finger region from all the known human C2H2-ZNF genes from the corresponding region. A hit with e-value $\leq$ 1e-4 confirmed the respective homologous clusters in the five mammalian genomes. A comprehensive catalogue of the human C2H2-ZNF clusters and their syntenic counterparts in other mammals is compiled in Supplementary Table S4.

Phylogenetic analysis

Phylogenetic analysis was conducted using the amino acid sequences of the zinc finger region (identified using PROSITE) of C2H2-ZNF genes from selected human clusters and their syntenically homologous clusters in chimpanzee, mouse, rat and dog. A multiple sequence alignment of the zinc finger regions of the C2H2-ZNF genes was generated using the program MUSCLE (Edgar 2004). The alignments were edited to remove gaps using the program GBLOCKS (Castresana 2000). Maximum Likelihood (ML) and Bayesian Inference (BI) methods were used to infer the phylogenetic trees and estimate the clade support. For ML analysis, the program RAxML (RAxML-VI-HPC Version 2.2.1) (Stamatakis, Ludwig et al. 2005) employing the WAG model of amino acid substitution was used to reconstruct the best tree. Bootstrapping of 100 datasets was implemented. The posterior probabilities were determined by a Bayesian MCMC method implemented in the program Mr.Bayes v.3.1 (Huelsenbeck and Ronquist 2001) to test the robustness of the topology of the tree inferred through ML. One million generations were run and the trees were sampled after every 10 generations.

To determine appropriate outgroups for our analysis, we searched the nr database to look for close homologs in non-mammals using TBLASTN (e-value cut off 1e-4). In addition to the Xfin sequence from *Xenopus laevis*, we obtained a set of zinc finger genes from Chicken *(Gallus gallus,* Assembly WASHUC2) specifically selected for each human C2H2-ZNF cluster based on an extensive similarity search. To select the chicken outgroup, a TBLASTN (e-value cutoff 1e-4) search was done against the chicken genome using each of the human C2H2-ZNF sequences derived from the selected cluster of interest as a query.

The top 10 hits for each query sequence were all analysed using a CD-HIT analysis (Identity threshold = 100%, 95% and 90%) (Li, Jaroszewski et al. 2001) to produce a final set of non-redundant representative chicken sequences, all used as a part of the outgroups.

Sequence analysis to confirm the loss of domains

In the case where loss of a domain was suspected, we conducted an extensive sequence analysis to rule out the possibility that these domains would have been missed either due to a frame-shift or inadequate exon-intron splicing of the gene and thus inappropriate amino acid translation, preventing recognition by PROSITE (http://www.expasy.com). Firstly, for each particular C2H2-ZNF genes where loss of an N-terminal domain was suggested, we systematically collected the nucleotide sequence of the region ranging from the stop of translation of the previous gene to the start of translation of the next gene. We conducted a TBLASTN search of this region using the amino acid sequence of the domain of interest (present in the corresponding orthologs and the consensus of the domains selected from randomly selected sequences) as a query to confirm the absence of the domain in the C2H2-ZNF gene of interest. Secondly, we obtained the exon-intron structure of these genes using the Ensembl Genome Browser (http://www.ensembl.org). In order to search for exonic or intronic sequences which may exhibit significant identity with the nucleotide sequence of the domain of interest. For this purpose we conducted a BLAST analysis of the individual exon and intron sequences with the nucleotide sequence of the various domains that are present in the corresponding orthologs.

Flowchart of the study

Figure 1 summarizes the flowchart of our analysis procedure of C2H2-ZNF genes and clusters in mammals.

RESULTS

Compilation of a comprehensive catalogue of human C2H2-ZNF genes

Previous studies reported the existence of at least 564 C2H2-ZNF genes in the human genome and suggested that this family may include approximately 700-800 genes (Bellefroid, Lecocq et al. 1989; Bellefroid, Poncelet et al. 1991). As a first step to study the evolution of C2H2-ZNF genes, we established a comprehensive catalogue of the C2H2-ZNF genes in the human genome. By conducting an extensive similarity search (see Methods), we identified 718 C2H2- ZNF genes (compiled in Supplementary Table S1). Of the 718 genes, 66 are annotated as pseudogenes in GenBank. For all genes, we determined their exact position on the chromosomes, their orientation, the number of finger motifs and the effector domains.

These genes are distributed across all chromosomes of the human genome (Supplementary Table S2). As reported earlier, chromosome 19 has the highest number (Venter, Adams et al. 2001) and density of C2H2-ZNF genes, including 40 % (289) of the 718 human C2H2-ZNF genes, whereas this chromosome corresponds to only 2.1 % of the human genome. More than half (58%) of the C2H2-ZNF genes encode conserved N-terminal domains, the KRAB, SCAN and BTB domains (Figure 2A), typically involved in transcriptional regulation (Kim, Chen et al. 1996; Collins, Stone et al. 2001) and form different C2H2-ZNF subfamilies. Further, we discovered two additional domains typical of transcription regulators, the SET and HOMEO domains that are also encoded by C2H2-ZNF genes. While the KRAB subfamily represents almost half of the C2H2-ZNF genes (45%), SET and HOMEO C2H2-ZNF genes together with members of all the other

subfamilies account for only a small percentage (~12%) of the C2H2-ZNF genes (Figure 2A).

Clustered organization of human C2H2-ZNF genes

It was reported earlier that chromosome 19 is particularly rich in tandemly duplicated C2H2-ZNF genes and that KRAB C2H2-ZNF genes, are clustered on several other chromosomes (Dehal, Predki et al. 2001; Rousseau-Merck, Koczan et al. 2002). In order to trace the duplication history of the entire C2H2-ZNF repertoire, we studied the distribution of these genes across the whole human genome. Two consecutive C2H2-ZNF genes were considered to belong to a cluster if the distance between them is $\leq$500 Kb, regardless of the presence of other genes or pseudogenes within the cluster (see Methods). Using this definition, we identified 81 human C2H2-ZNF clusters accounting for 72 % of the total number of C2H2-ZNF genes (518 of the 718) (Supplementary Tables S2 and S3). The remaining genes are dispersed as singletons. Among these clusters, 31 % include exclusively tandemly organized C2H2-ZNF genes with no other intervening genes (Figure 2B, Supplementary Table S3). The number of C2H2-ZNF genes per cluster ranges from 2 to 76 with an average of 6. As illustrated in the Figure 2B, about 75 % of the total number of C2H2-ZNF clusters has between two to six genes. Consistent with previous reports, chromosome 19 not only has the largest number of C2H2-ZNF clusters (Supplementary Table S2) but also hosts the largest clusters (>12 genes) (see Figure 2B and Supplementary Table S3).

We find that the large majority of KRAB (89 %) and SCAN (90 %) types of C2H2-ZNF genes are arranged in clusters (Figure 2A and Supplementary Table S2). This contrasts with the BTB subfamily of C2H2-ZNF genes or those lacking regulatory domains which occur more often as singletons in the genome. An analysis of the composition of individual clusters revealed that two-third of the clusters contains a mixture of various C2H2-ZNF subfamilies ('mixed clusters', Supplementary Table S3). The few clusters made up of a single C2H2-ZNF gene subfamily ('pure clusters') are of small size (< 4 genes).

Identification and comparison of syntenic C2H2-ZNF clusters across mammals

With the ultimate goal to study the evolution of zinc finger genes, we identified and compiled clusters in completely sequenced mammalian genomes (i.e. chimpanzee, mouse, rat and dog) that are syntenically homologous to those of human. Syntenically homologous clusters were identified by the genes flanking each cluster. Then, all the C2H2-ZNF genes found within the delimited syntenic regions were identified using a TBLASTN search (see Methods). The 81 human C2H2-ZNF clusters and their syntenic counterparts in other mammals are listed in Supplementary Table S4, which also includes information on the orientation of the genes in the clusters, their associated domains, the number of zinc finger motifs and the flanking genes.

Primates (*Homo sapiens* and *Pan troglodytes*) stood out for their large number of both C2H2-ZNF clusters and genes within them, as compared to rodents (*Mus musculus*

and *Rattus norvegicus*) and *Canis familiaris* (Figure 3A). The most parsimonious explanation is that a large expansion of C2H2-ZNF genes occurred in primates, and more particularly in human (518 genes in human *versus* 397 in chimpanzee) after divergence from rodents and canines. Rat has slightly less C2H2-ZNF genes than dog (7%), but 25% less than mouse. Considering the evolutionary relationship of the species (Figure 3A), these data suggest that not only species-specific duplication events, as reported earlier (Dehal, Predki et al. 2001; Hamilton, Huntley et al. 2003; Shannon, Hamilton et al. 2003), but also loss of family members (suggested here in rodents) may have occurred during the evolution of mammals. Differential species-specific expansion was reported previously for a subset of genes from the human ZNF45 subfamily on chromosome 19 compared with its mouse counterpart (Shannon, Hamilton et al. 2003). Furthermore, expansion of the human KRAB C2H2-ZNF subfamily was also shown earlier based on draft versions of the genomes of chimpanzee, mouse and dog (Huntley, Baggott et al. 2006). However, evidence of C2H2-ZNF gene or cluster loss could not be definitively obtained in these studies as it required detailed analysis of more than two completely sequenced genomes.

Comparing individual syntenic clusters in the mammalian genomes

To distinguish whether differences in the number of C2H2-ZNF clusters are due to species-specific gene gain or loss, we systematically compared individual syntenic clusters in the five mammalian genomes studied. The results of this analysis point to a differential evolutionary history in mammals. About 60 % of the human clusters (49) have syntenically homologous counterparts in all the species studied indicating that these C2H2-ZNF clusters

predate the divergence of dog, rodents and primates (Supplementary Table S4 and Supplementary Figure S2). In addition, we found (i) primate specific clusters (14 including 2 human specific clusters), (ii) clusters, present in primates and dog, that were lost in rodents (8 clusters including 3 present in mouse but absent in rat) and (iii) clusters present in primates and rodents but absent in dog (10 clusters) (examples in Figure 3B). Essentially all the primate clusters have larger number of genes than rodent or dog clusters which reflects a global primate-specific expansion of C2H2-ZNF (Supplementary Figure S2). Further, in 40% of all primate clusters, those from human contain more C2H2-ZNF genes than those from chimpanzee. This indicates that most of the evolutionary changes (duplication and/or loss) occurred late in the primate branch. A similar pattern was seen in rodents, where almost all mouse C2H2-ZNF clusters exhibit more genes than their syntenic rat clusters. While these results illustrate that the C2H2-ZNF gene family is rapidly and independently evolving within different lineages, insights into the role of gene duplication and loss in the history of this gene family required rigorous phylogenetic analysis.

Phylogeny of C2H2-ZNF clusters in mammalian genomes

For addressing the relative contribution of gene duplication and loss in the evolution of C2H2-ZNF genes in mammals, we focused our study on selected large human C2H2-ZNF clusters and their syntenic counterparts in four other mammals. We expected that larger clusters would be more informative and possibly more representative of the whole genomes. Because of the clarity of evolutionary scenarios observed in the tree, we present here a detailed phylogenetic analysis of the second largest human C2H2-ZNF cluster (43

genes) located on chromosome 19q13.4, that we named cluster *19.12,* and of its syntenic clusters (Supplementary Table S3 and S4) in other species. For phylogenetic analysis, we used the predicted amino acid sequences of the zinc finger regions. Genes annotated as pseudogenes in Genbank or genes containing less than three zinc finger motifs were not considered in the phylogenetic analysis (noise is expected to be too high if sequences of only 56 amino acids corresponding to 2 fingers motifs or less were included). Our total data set of C2H2-ZNF sequences from the human cluster *19.12* and their syntenic homologs in chimpanzee, mouse, rat and dog consists of 101 protein sequences, including the outgroup sequences from *Xenopus* and Chicken. We constructed a phylogenetic tree using Maximum Likelihood and Bayesian methods. We subdivided the tree into three groups (Figure 5) based on the kind of evolutionary scenarios observed i.e. one-to-one and one-to-many orthologous relationships between genes as well as gene loss as defined in Figure 4. The number of C2H2-ZNF sequences from each species is highlighted for each group. Two of these groups are monophyletic with significant ($\geq$95%) support in both the Maximum Likelihood and Bayesian analysis (Group I and III).

A detailed analysis of the tree revealed four clades that underwent species-specific expansion, and two clades, with gene loss in some species. For example, a dog-specific expansion is seen in the monophyletic Group I, which includes three clustering genes from human (hZNF331), chimpanzee (pZNF331) and dog (cZNF331) which in turn grouped within a larger clade containing nine additional C2H2-ZNF genes from dog. In addition, this clade indicates a loss in rodents, due to the absence of mouse or rat genes. Group I alone illustrates how both species-specific duplication in dog and loss in rodents can

account for the higher number of genes seen in dog C2H2-ZNF clusters as compared to rodents.

Group II shows more pronounced expansion in human as seen in several clusters. For example, one of the primate-specific clades includes 17 human genes and 7 chimpanzee C2H2-ZNF genes (Figure 5). Of the 17 human genes present in the clade, only 6 genes show a one-to-one orthologous pairing with chimpanzee genes. Another well supported clade includes a single human gene (hZNF677) clustered with two dog genes (LOC484331 AND LOC476394). In this clade, the absence of a chimpanzee or rodent counterparts to these three genes suggests a loss in these species. For chimpanzee, however, loss by pseudogenization is possibly involved (see physical maps described below); note that the percentage of C2H2-ZNF genes annotated as pseudogenes was higher in chimpanzee that in human C2H2-ZNF clusters (62, Supplementary Table S4).

In group III, the relationship of the four rodent genes with the dog and primate genes could not be resolved (bootstrap values < 95 %). However, a rodent-specific clade revealed a mouse-specific duplication exhibiting a higher number of C2H2-ZNF genes in mouse than rat, as seen in several other cases in our study.

Superimposition of the phylogenetic trees with the physical maps of clusters

Comparison of gene trees, species tree and physical map information of cluster 19.12 genes and its syntenic homologs provide better insights into the processes underlying the evolution of the C2H2-ZNF clusters. The phylogenetic tree obtained for cluster 19.12 (Figure 5) suggests a simultaneous differential expansion and loss of C2H2-ZNF genes

throughout evolution. In perfect agreement with the phylogenetic tree, genes of the monophyletic groups I and III were found to be physically clustered together on the chromosomes across mammals (Figure 6). Evidence for a tandem duplication event is provided by the comparison of the relationship within C2H2-ZNF genes of Group I on the tree with their spatial relationships in the physical maps that showed that the sequences of the dog clade form a tandem array on the chromosome (Figure 5). In addition to tandem duplication of individual genes within this group, e.g. cLOC484324 and cLOC484323 (Figure 5) which are next to each other on the chromosome and exhibit the same orientation (Figure 6), we also discovered tandem duplication of multiple genes. For instance, three genes (LOC482273, LOC611599, LOC480782/ orientation -, +, +) appear as a tandem repeat of three other genes (LOC611583, LOC484328, LOC484326/ orientation -, +, +) in this group (Figures 5 and 6).

The group II mainly contains primate-specific C2H2-ZNF genes that cluster on the phylogenetic tree in two well supported clades (≥ 97 % bootstrap) and a sub-group of weaker support (93% bootstrap). Almost all these genes also cluster physically together on the chromosome. Human orthology assignments for ten of the twelve chimpanzee genes from group II (underlined in Figure 6) were corroborated by two lines of evidence i.e. from the phylogeny, which was supported by the topology on the chromosome. Furthermore, genes from 7 out of 10 of the C2H2-ZNF ortholog pairs from this primate-specific cluster exhibit the same number of zinc finger motifs and the same type of N-terminal motif.

Species-specific variation in the number of finger motifs and the presence of N-terminal conserved domains

When analysing the C2H2-ZNF genes from the 81 human clusters and their syntenic homologs in mammals, we noticed that the average number of zinc finger motifs varied depending on the C2H2-ZNF gene subfamilies. Noticeably in all the mammalian species studied, genes with KRAB and SCAN-KRAB motifs have a higher number of zinc finger motifs than those from the other subfamilies (Figure 7A). For example, member of the KRAB subfamily have an average of 10 to 17 zinc finger motifs, while members of the BTB subfamily have only 2 to 3 (Figure 7A). We also noted species-specific variation in the number of zinc finger motifs within mammalian C2H2-ZNF genes. In particular, dog tends to have a much higher number of zinc finger motifs in most C2H2-ZNF gene sub-families (Figure 7). Strikingly, LOC484264, a dog KRAB C2H2-ZNF gene exhibits 70 zinc finger motifs which is to our knowledge the highest number of zinc finger motifs to be reported for a zinc finger gene. Study of cluster 19.12 (Figure 5) illustrates more specifically the trend of dog genes to exhibit more zinc finger motifs; the dog LOC484338 gene (group III), for example, has six times more zinc finger motifs than its human ortholog. Furthermore, the dog gene LOC484326 has nearly twice as many motifs as its closest paralog LOC480782 (group I) (Figure 5). This indicates a quite recent and drastic expansion of zinc finger motifs within dog C2H2-ZNF genes, after the separation of dog from rodents and primates. In several cases, the C2H2-ZNF mammalian orthologs revealed differences in their numbers of finger motifs even within primate or within rodent lineages (Figure 7B and Supplementary Table 4).

In addition to the difference in the number of finger motifs in C2H2-ZNF orthologs and paralogs, we also found a variation in the presence of the N-terminal effector domains. As an example, orthologs of the C2H2-ZNF genes in the human cluster 6.2 show a variation in the presence of the KRAB or SCAN domains (Figure 7B), suggesting frequent and multiple losses and/or gains of KRAB and SCAN domains during evolution. To reconstruct these events, we analyzed in detail the exon-intron structure and sequences of these genes (See Methods). Serendipitously, this analysis led us to the observation that a large majority of the C2H2-ZNF containing a SCAN-KRAB or SCAN domain had each a typical exon-intron organization (Figure 7C). For example, both human genes, ZNF192 (SCAN-KRAB) and ZNF187 (SCAN) and their respective orthologs in other species (Figure 6B) share the predominant exon-intron organization most typical of SCAN-KRAB and SCAN C2H2-ZNF, respectively. While the dog LOC488318 has only a SCAN domain, its corresponding orthologs in human, mouse and rat have a SCAN-KRAB. When the nucleotide sequence of the exon which would have been predicted to encode a KRAB domain in dog (third exon after the SCAN) was compared with those of human, mouse and rat, the dog sequence exhibits a high conservation at the nucleotide level (>82 %) but no significant similarity at the amino acid level. This indicates that the loss of the KRAB domain in dog was due to sequence degeneration. Similarly, while the chimpanzee ZNF187 and its rat ortholog encode a SCAN domain, a degenerate SCAN domain was identified in the corresponding exon of their human and mouse orthologs. For the human SCAN-KRAB ZNF307 gene, we noticed that it exhibits an exon-intron organization typical of SCAN-KRAB C2H2-ZNF (Figure 7C) whereas its orthologs in the chimpanzee, mouse

and rat encode solely a SCAN domain and present an exon-intron structure more typical of SCAN C2H2-ZNF. However, it was found that, in chimpanzee, a sequence similar to the KRAB sequence (99% at the nucleotide level) was embedded in the intron preceding the exon encoding the zinc finger domain. No KRAB related sequence could be detected in the rodent orthologs even with a detailed analysis of their sequences. Thus, either the KRAB sequence was gain in the primate lineage or lost in the rodent lineage. For reasons explained in the discussion, we believe that loss, rather than gain, is a more likely hypothesis.

DISCUSSION

Comparative studies in genome research focused on the extensive similarities existing between the human genome and the genomes from various other model organisms which provide valuable insights into biological function and aetiology of human diseases. However, differences existing among genomes have received less attention inspite of the importance they may have in the physiological, morphological and behavioural distinctive traits observed among species. A few studies on various gene families, such as the odorant receptor family, pointed out to some differences existing between genes of closely related species (Sitnikova and Su 1998; Lapidot, Pilpel et al. 2001; Niimura and Nei 2003; Gilad, Man et al. 2005; Niimura and Nei 2005). Our study of the C2H2-ZNF gene family reveals that there is an extensive variation of the C2H2-ZNF gene content and organization in the genomes from various mammals as well as in the domain composition of orthologous genes among species. It also provides the first clear demonstration of the contribution of gene loss in the C2H2-ZNF family during evolution which occurs at the level of clusters, genes and their functional domains. We provide the first genome scale confirmation of the rapid evolution of C2H2-ZNF gene clusters that occurs independently within related species which also supports conclusions drawn from smaller-scale studies on individual genes, clusters and C2H2-ZNF subfamilies (Dehal, Predki et al. 2001; Shannon, Hamilton et al. 2003; Hamilton, Huntley et al. 2006; Huntley, Baggott et al. 2006).

Substantial variation in the C2H2-ZNF gene family size and clustering across mammals

We report here the first complete catalogue of all human C2H2-ZNF gene clusters and their syntenic homologs in chimpanzee, mouse rat and dog. This catalogue reveals that in human, a large proportion of the genes from the C2H2-ZNF family (>70%) are organized in clusters. Comparative studies of the five mammalian genomes indicated that the total number of genes found in clusters varied considerably from 172 in rat to 518 in human (number of genes found in clusters in human > chimpanzee > mouse > dog > rat). Significantly, human and mouse have a larger number of clustered C2H2-ZNF genes (>30%) as compared to chimpanzee and rat, respectively, indicating that independent evolutionary events occurred after the divergence of the two primates (within the last ~ 6-10 million years) and two rodents (within ~ 30-46 million years). We distinguish two kinds of events: first, a variation in the number of C2H2-ZNF genes in syntenically homologous clusters and second, the existence of lineage- and species-specific clusters in primates, rodents and canines. This can be accounted for by independent evolution of C2H2-ZNF genes in these closely related species. Previous studies focusing on KRAB C2H2-ZNF from chromosome 19 had identified and analyzed a primate-specific cluster (Mohrenweiser 1998) including members of the primate-specific ZNF91 subfamily of C2H2-ZNF (Hamilton, Huntley et al. 2006). Other studies on the KRAB C2H2-ZNF subfamily identified a differential expansion between a human KRAB C2H2-ZNF cluster and its syntenic counterpart in mouse (Shannon, Hamilton et al. 2003) and more recently other species-specific expansions based on draft of various mammalian genomes (Huntley,

Baggott et al. 2006). We illustrate and confirm at a larger scale the existence of an on-going process of genome dynamics with several lineage- and species-specific rearrangements and continuous repertoire expansion taking place independently in all evolutionary branches, particularly in primates. This finding was only possible through the analysis of a complete catalogue of all the subfamilies of C2H2-ZNF clusters and their syntenic counterparts in mammals.

Gene duplication and loss: Two counteracting forces in the evolution of C2H2-ZNF genes

An overview of 81 human C2H2-ZNF clusters identified here revealed that a third of them are pure clusters (with 2 to 24 C2H2-ZNF genes), i.e. they are not interspersed with other genes. Earlier observations of pure C2H2-ZNF gene clusters have led to the hypothesis that C2H2-ZNF genes in primates have expanded massively by tandem duplication (Bellefroid, Marine et al. 1993; Eichler, Hoffman et al. 1998; Elemento and Gascuel 2002; Schmidt and Durrett 2004; Bertrand and Gascuel 2005; Huntley, Baggott et al. 2006). We revisited this question based on our catalogue of human C2H2-ZNF clusters and their syntenic counterparts in chimpanzee, mouse, rat and dog. Here, we confirmed gene duplication and loss based on a reconciliation of both physical maps and the surperimposition of gene trees onto the known species tree (Page and Charleston 1997). Our results clearly show that both gene gain and gene loss events have occurred multiple times and independently in all the mammals studied. Combined with physical map data, our phylogenetic studies indicate that the expansion of C2H2-ZNF genes evidenced during the

evolution of the five species studied results from the combined action of single-gene duplication and multiple gene duplication (for instances, duplication of all or part of the genes within a cluster). These duplication events were however counteracted by the loss of individual genes or clusters as exemplified in several cases where related genes or clusters found in primates and canine were absent in both or in any of the two rodents studied. This study represents the first clear demonstration of the involvement of gene and cluster loss in the evolution of C2H2-ZNF genes and suggests that during mammalian evolution the duplication events outnumbered the loss events. Our results provide convincing support that the C2H2-ZNF gene family evolved according to the "Birth and Death" model as proposed by Nei and colleagues (Nei, Gu et al. 1997; Nei 2000). According to this model, new genes are created by duplication including tandem duplication and block gene duplication (birth). While certain copies remain relatively unchanged in the genome for a long time, others diverge functionally by acquiring a new function. Some get deleted from the genome or become pseudogenes following deleterious mutations (death through elimination or inactivation). In the case of C2H2-ZNF genes pseudogenization seems to be limited, as suggested by expression studies and statistical analysis showing positive selection based on the analysis of specific clusters (Schmidt and Durrett 2004; Huntley, Baggott et al. 2006). This makes the C2H2-ZNF family different from the other gene families such as the olfactory receptor gene family (Glusman, Yanai et al. 2001; Niimura and Nei 2003) . Noticeably, gene loss by pseudogenization was prominent for the olfactory receptors with humans accumulating a higher number of olfactory receptor pseudogenes as compared to other primates and mouse (Sitnikova and Su 1998; Lapidot,

Pilpel et al. 2001; Niimura and Nei 2005). These variations in the numbers of pseudogenes and functional genes have been associated with the differential chemosensory dependence in these species (Sharon, Glusman et al. 1999; Quignon, Kirkness et al. 2003). In comparison, besides the fact that they are known to function as regulator of transcription, the functions of only a few C2H2-ZNF proteins are known (Krebs, Larkins et al. 2003). Further studies of C2H2-ZNF genes in mammals could shed light on the functional consequences of different repertoires of these genes in different species. Until now, the clustered organization of these genes has made knock-out studies in animal models inefficient, possibly due to redundancy. However, based on a better knowledge of the organization/content of C2H2-ZNF genes in the various genomes, large chromosomal deletions of pure C2H2-ZNF clusters or other types of gene disruption or targeting approaches could provide insights into the functions of these genes in different animal models.

Evolution of C2H2-ZNF genes through gain and loss of finger motifs and N-terminal effector domains

Evidence of the variation in the numbers of zinc finger motifs among orthologs was previously reported for a subset of human chromosome 19 C2H2-ZNF genes and their mouse homologs (Looman, Abrink et al. 2002). It was shown that this variation is due to both differential duplication of finger motifs and loss due to degeneration. In our study, such variation in the number of zinc finger motifs among orthologs was observed recurrently among all mammals. Since the zinc finger motifs appears as a flexible motif

with the ability to bind DNA, RNA and/or proteins, changes in the zinc finger motif sequences and number within C2H2-ZNF genes could differentially alter binding specificities and thus protein function. Both changes in the number of C2H2-ZNF genes and in the number of finger motifs encoded by orthologous genes may be determinant in species-specific related function.

The rapid evolution of the C2H2-ZNF genes observed in the mammalian lineage was not limited to the variation in the number of genes and zinc finger motifs. Variation in the presence of N-terminal effector domains, such as SCAN or KRAB, was observed in orthologs and could be accounted for by either gain or loss of these motifs in the various species. Loss by sequence degeneration of both the SCAN and the KRAB sequences was confirmed in several cases in our study. In some cases, neither loss nor gain could be resolved. A puzzling question remains however if one assumes that gain of KRAB and SCAN sequences can occur recurrently within C2H2-ZNF genes. It is indeed difficult to explain that these effector domains are always found in association with and N-terminal to the zinc finger motifs of C2H2-ZNF proteins and that the SCAN domain is always positioned N-terminal to the KRAB domain of SCAN-KRAB C2H2-ZNF proteins. Interestingly, by analyzing the exon-intron structure of C2H2-ZNF genes from the clusters, we found that most SCAN C2H2-ZNF and SCAN-KRAB C2H2-ZNF genes have each a typical exon-intron structure (Figure 7C and Figure 8A). This suggests that the acquisition of a SCAN and KRAB sequences by C2H2-ZNF genes corresponds most likely to singular events. This led us to propose the model described in Figure 8. Considering that the SCAN domain is found in all vertebrates and is more ancient than the KRAB domain only found

in tetrapods, we suggest that first a SCAN-C2H2 ZNF gene was formed in an ancestor of vertebrates through the gain of SCAN sequence and that later, after the emergence of the tetrapods, the gain of a KRAB sequence by a SCAN C2H2-ZNF gene gave rise to a SCAN-KRAB C2H2-ZNF gene (Figure 8B). These two gain events possibly occurred through an exon-shuffling mechanism. Diversification of the C2H2-ZNF repertoires from each subfamily then occurred dynamically through on-going gene duplications and loss by deletion or degeneration of the SCAN and/or KRAB sequences. As implied by this model, the birth of the SCAN-KRAB C2H2-ZNF subfamily occurred earlier than that of the KRAB C2H2-ZNF subfamily. This was consistent with previous data showing that SCAN-KRAB-ZNF genes do not group together on one evolutionary clade but intermix with KRAB-ZNF genes in phylogenetic trees of the KRAB sequence (Looman, Abrink et al. 2002; Huntley, Baggott et al. 2006) (Huntley, Baggott et al. 2006). On the whole, our model is in agreement with the fact that C2H2-ZNF orthologs often belong to different C2H2-ZNF subfamilies and that we observed intermingling of C2H2-ZNF genes from the SCAN, KRAB and SCAN-KRAB subfamilies in many C2H2-ZNF clusters. Our study clearly indicated that the evolution of C2H2-ZNF subfamilies is tightly linked and stresses that the assignment of proper orthology requires comprehensive analysis of all C2H2-ZNF genes rather than the individual analysis of specific C2H2-ZNF subfamilies. It also points to the importance of loss/contraction and secondary simplification whose role in the dynamics of evolution is often underestimated. The underlying mechanisms in the expansion of C2H2-ZNF genes and the functional consequences of the important changes (gain and loss) occurring in their repertoires of various mammals are unclear. These

variations, for example, may be at the advantage of complex organisms by providing more subtle and species-specific control in gene expression for morphogenesis or cognitive functions.

ACKNOWLEDGEMENTS

The authors thank Franz B. Lang for suggestions and help in phylogenetic analysis and Nicolas Lartillot (Universite de Montreal) for constructive comments on tree building and datasets. We also acknowledge Herve Philippe, Henner Brinkmann and Amy Hauth for critical discussions and Allan Sun for assistance in hardware and software installations. This work was supported by a grant from Natural Sciences and Engineering Research Council of Canada (to M.A). M.A is a Chercheur National from the Fonds de la recherché Sante du Quebec (FRSQ) and G.B is an associate from CIFAR (Canadian Institute for Advanced Research).

REFERENCES

Bellefroid, E. J., P. J. Lecocq, et al. (1989). "The human genome contains hundreds of genes coding for finger proteins of the Kruppel type." DNA **8**(6): 377-87.

Bellefroid, E. J., J. C. Marine, et al. (1993). "Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells." Embo J **12**(4): 1363-74.

Bellefroid, E. J., D. A. Poncelet, et al. (1991). "The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins." Proc Natl Acad Sci U S A **88**(9): 3608-12.

Bertrand, D. and O. Gascuel (2005). "Topological rearrangements and local search method for tandem duplication trees." IEEE/ACM Trans Comput Biol Bioinform **2**(1): 15-28.

Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." Mol Biol Evol **17**(4): 540-52.

Collins, T., J. R. Stone, et al. (2001). "All in the family: the BTB/POZ, KRAB, and SCAN domains." Mol Cell Biol **21**(11): 3609-15.

Dehal, P., P. Predki, et al. (2001). "Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution." Science **293**(5527): 104-11.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-7.

Eichler, E. E., S. M. Hoffman, et al. (1998). "Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12." Genome Res **8**(8): 791-808.

Elemento, O. and O. Gascuel (2002). "An efficient and accurate distance based algorithm to reconstruct tandem duplication trees." Bioinformatics **18 Suppl 2**: S92-9.

Gertz, E. M., Y. K. Yu, et al. (2006). "Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST." BMC Biol **4**: 41.

Gilad, Y., O. Man, et al. (2005). "A comparison of the human and chimpanzee olfactory receptor gene repertoires." Genome Res **15**(2): 224-30.

Glusman, G., I. Yanai, et al. (2001). "The complete human olfactory subgenome." Genome Res **11**(5): 685-702.

Grondin, B., M. Bazinet, et al. (1996). "The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix." J Biol Chem **271**(26): 15458-67.

Hamilton, A. T., S. Huntley, et al. (2003). "Lineage-specific expansion of KRAB zinc-finger transcription factor genes: implications for the evolution of vertebrate regulatory networks." Cold Spring Harb Symp Quant Biol **68**: 131-40.

Hamilton, A. T., S. Huntley, et al. (2006). "Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes." Genome Res **16**(5): 584-94.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-5.

Huntley, S., D. M. Baggott, et al. (2006). "A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors." Genome Res **16**(5): 669-77.

Kim, S. S., Y. M. Chen, et al. (1996). "A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins." Proc Natl Acad Sci U S A **93**(26): 15299-304.

Klug, A. and D. Rhodes (1987). "Zinc fingers: a novel protein fold for nucleic acid recognition." Cold Spring Harb Symp Quant Biol **52**: 473-82.

Krebs, C. J., L. K. Larkins, et al. (2003). "Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression." Genes Dev **17**(21): 2664-74.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lapidot, M., Y. Pilpel, et al. (2001). "Mouse-human orthology relationships in an olfactory receptor gene cluster." Genomics **71**(3): 296-306.

Lee, M. S., G. P. Gippert, et al. (1989). "Three-dimensional solution structure of a single zinc finger DNA-binding domain." Science **245**(4918): 635-7.

Li, W., L. Jaroszewski, et al. (2001). "Clustering of highly homologous sequences to reduce the size of large protein databases." Bioinformatics **17**(3): 282-3.

Looman, C., M. Abrink, et al. (2002). "KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution." Mol Biol Evol **19**(12): 2118-30.

Looman, C., L. Hellman, et al. (2004). "A novel Kruppel-Associated Box identified in a panel of mammalian zinc finger proteins." Mamm Genome **15**(1): 35-40.

Mark, C., M. Abrink, et al. (1999). "Comparative analysis of KRAB zinc finger proteins in rodents and man: evidence for several evolutionarily distinct subfamilies of KRAB zinc finger genes." DNA Cell Biol **18**(5): 381-96.

Melnick, A., G. Carlile, et al. (2002). "Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors." Mol Cell Biol **22**(6): 1804-18.

Messina, D. N., J. Glasscock, et al. (2004). "An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression." Genome Res **14**(10B): 2041-7.

Miller, J., A. D. McLachlan, et al. (1985). "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes." Embo J **4**(6): 1609-14.

Nei, M., and Kumar S. (2000). Molecular Evolution and Phylogenetics, New York: Oxford University Press.

Nei, M., X. Gu, et al. (1997). "Evolution by the birth-and-death process in multigene families of the vertebrate immune system." Proc Natl Acad Sci U S A **94**(15): 7799-806.

Niimura, Y. and M. Nei (2003). "Evolution of olfactory receptor genes in the human genome." Proc Natl Acad Sci U S A **100**(21): 12235-40.

Niimura, Y. and M. Nei (2005). "Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice." Gene **346**: 13-21.

Ohta, T. (2000). "Evolution of gene families." Gene **259**(1-2): 45-52.

Page, R. D. and M. A. Charleston (1997). "From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem." Mol Phylogenet Evol **7**(2): 231-40.

Quignon, P., E. Kirkness, et al. (2003). "Comparison of the canine and human olfactory receptor gene repertoires." Genome Biol **4**(12): R80.

Rhodes, D. and A. Klug (1993). "Zinc fingers." Sci Am **268**(2): 56-9, 62-5.

Rosati, M., M. Marino, et al. (1991). "Members of the zinc finger protein gene family sharing a conserved N-terminal module." Nucleic Acids Res **19**(20): 5661-7.

Rousseau-Merck, M. F., D. Koczan, et al. (2002). "The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in EnsEMBL." Cytogenet Genome Res **98**(2-3): 147-53.

Sander, T. L., A. L. Haas, et al. (2000). "Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero- and homoprotein associations." J Biol Chem **275**(17): 12857-67.

Schmidt, D. and R. Durrett (2004). "Adaptive evolution drives the diversification of zinc-finger binding domains." Mol Biol Evol **21**(12): 2326-39.

Schuh, R., W. Aicher, et al. (1986). "A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene." Cell **47**(6): 1025-32.

Schumacher, C., H. Wang, et al. (2000). "The SCAN domain mediates selective oligomerization." J Biol Chem **275**(22): 17173-9.

Shannon, M., A. T. Hamilton, et al. (2003). "Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters." Genome Res **13**(6A): 1097-110.

Sharon, D., G. Glusman, et al. (1999). "Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes." Genomics **61**(1): 24-36.

Sitnikova, T. and C. Su (1998). "Coevolution of immunoglobulin heavy- and light-chain variable-region gene families." Mol Biol Evol **15**(6): 617-25.

Stamatakis, A., T. Ludwig, et al. (2005). "RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees." Bioinformatics **21**(4): 456-63.

Stone, J. R., J. L. Maki, et al. (2002). "The SCAN domain of ZNF174 is a dimer." J Biol Chem **277**(7): 5448-52.

Theunissen, O., F. Rudt, et al. (1992). "RNA and DNA binding zinc fingers in Xenopus TFIIIA." Cell **71**(4): 679-90.

Thornton, J. W. and R. DeSalle (2000). "Gene family evolution and homology: genomics meets phylogenetics." Annu Rev Genomics Hum Genet **1**: 41-73.

Urrutia, R. (2003). "KRAB-containing zinc-finger repressor proteins." Genome Biol **4**(10): 231.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Witzgall, R., E. O'Leary, et al. (1994). "The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression." Proc Natl Acad Sci U S A **91**(10): 4514-8.

Figure 1: Flowchart of the analysis procedure of C2H2-ZNF genes and clusters

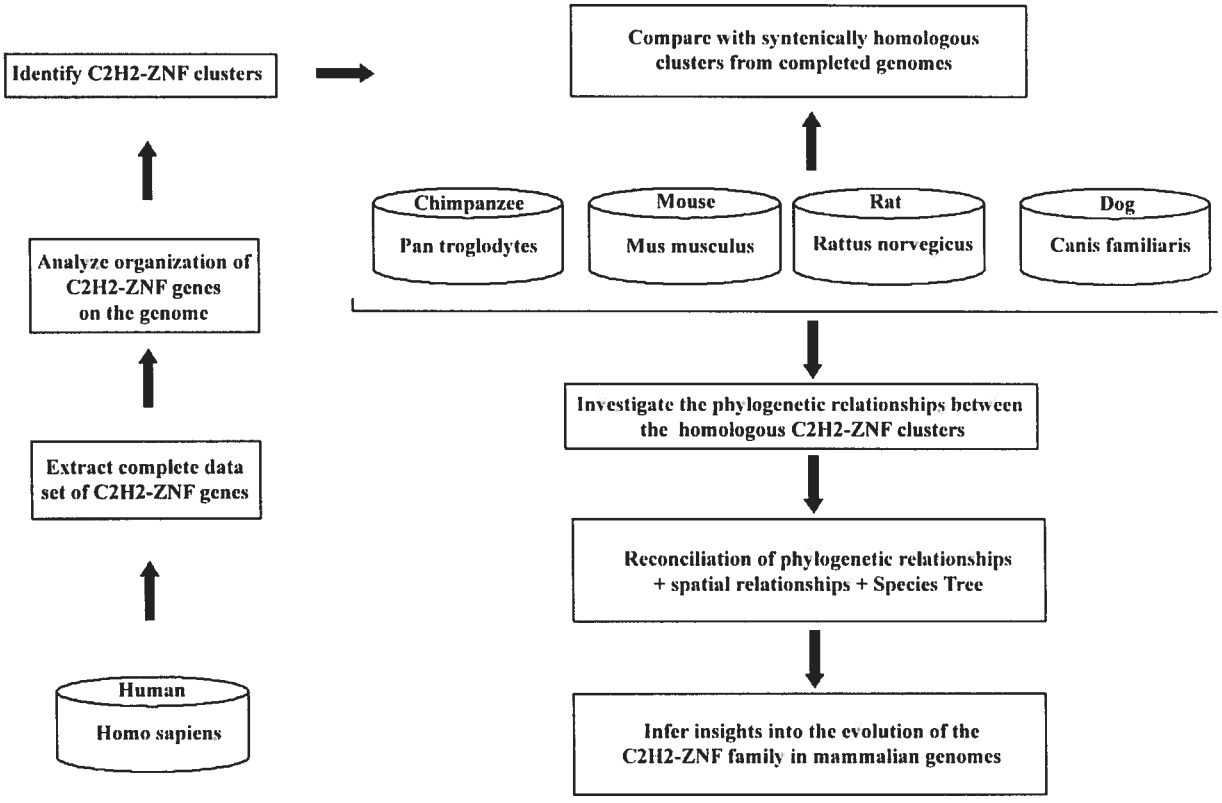


Figure 2: Distribution of all the singletons and clustered genes from the various human C2H2-ZNF sub-families and gene composition of the C2H2-ZNF clusters

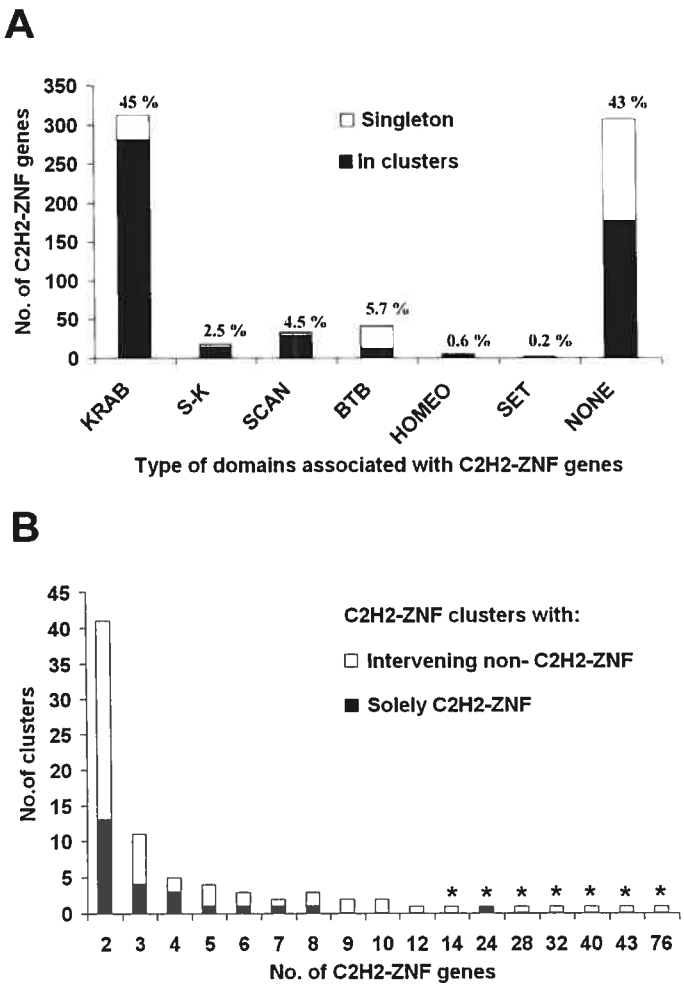


Figure 2

Distribution of all the singletons and clustered genes from the various human C2H2-ZNF sub-families and gene composition of the C2H2-ZNF clusters.

A) The number of genes belonging to the various C2H2-ZNF subfamilies are shown as well as the proportion of genes found as singletons or as part of clusters. C2H2-ZNF genes associated with KRAB and SCAN domains are more often found to be clustered. S-K= C2H2-ZNF containing both a SCAN and a KRAB domain. NONE= C2H2-ZNF without any conserved domain associated. The percentage distribution is mentioned on top of each bar for each sub-family.

B) The number of C2H2-ZNF clusters is shown with respect to the number of genes present in each cluster. The proportion of clusters composed of solely C2H2-ZNF without any intervening gene or with intervening genes other than C2H2 ZNF (Non-C2H2-ZNF) is also represented. A star (*) identifies large clusters present on chromosome 19.

Figure 3: Differential expansion and loss of C2H2-ZNF clusters in five mammalian genomes

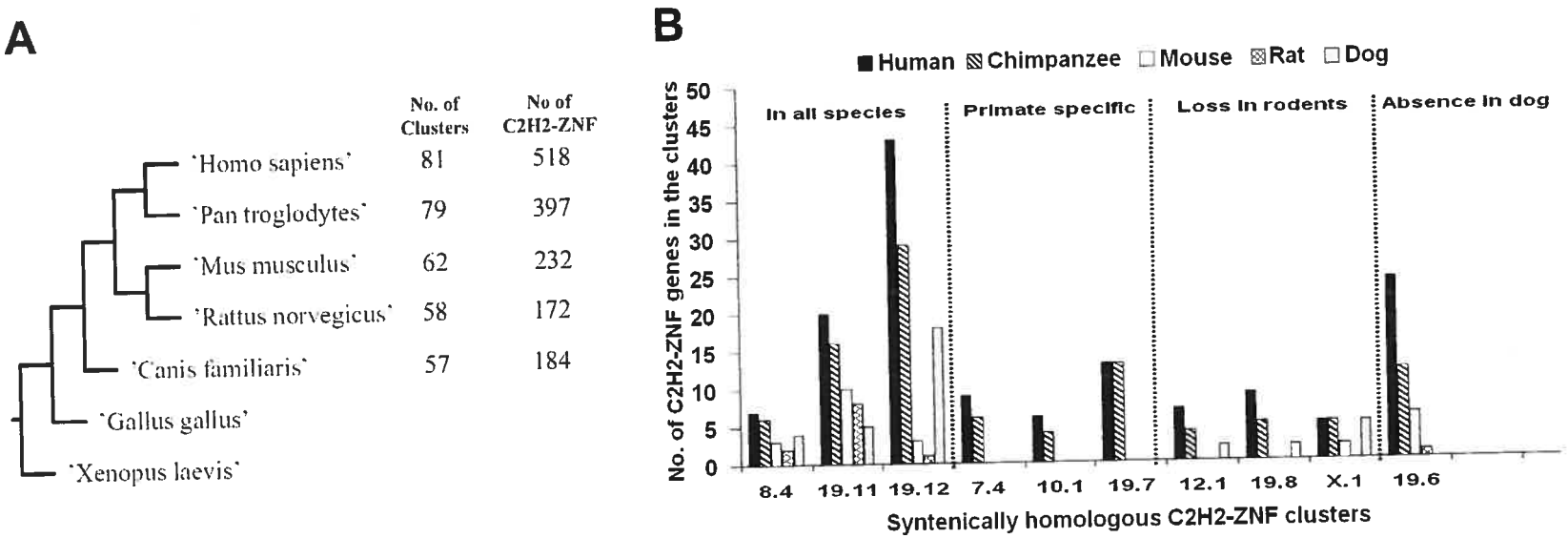


Figure 3

Differential expansion and loss of C2H2-ZNF clusters in five mammalian genomes.

A) Evolution of the C2H2-ZNF repertoires in primates, rodents and dog. The number of C2H2-ZNF clusters and the total number of C2H2-ZNF found in these clusters are mentioned on the species tree. Since *Xenopus laevis* and *Gallus gallus* C2H2-ZNF are used as an outgroup in phylogenetic studies, these species are also positioned on the tree.

The figure indicates the primate-specific increase in the number of C2H2-ZNF as compared to rodents and dog.

B) A graphical representation of different scenarios seen in the evolution of human C2H2-ZNF clusters and its syntenically homologous C2H2-ZNF clusters in chimpanzee, mouse, rat and dog. The human clusters selected and named on the graph as well as their syntenic counterparts were 1) present in all species, 2) primate-specific, 3) lost in rodents or 4) absent in dog. For each human C2H2-ZNF cluster named on the graph, the first number indicates the chromosome number and the second is the number attributed to that cluster on the chromosome. Supplementary Figure S2 provides a more comprehensive graphical representation including the 40 human clusters that contain at least 3 C2H2-ZNF and their syntenic counterparts in the four other mammals.

Figure 4: Evolutionary scenarios in the phylogenetic tree

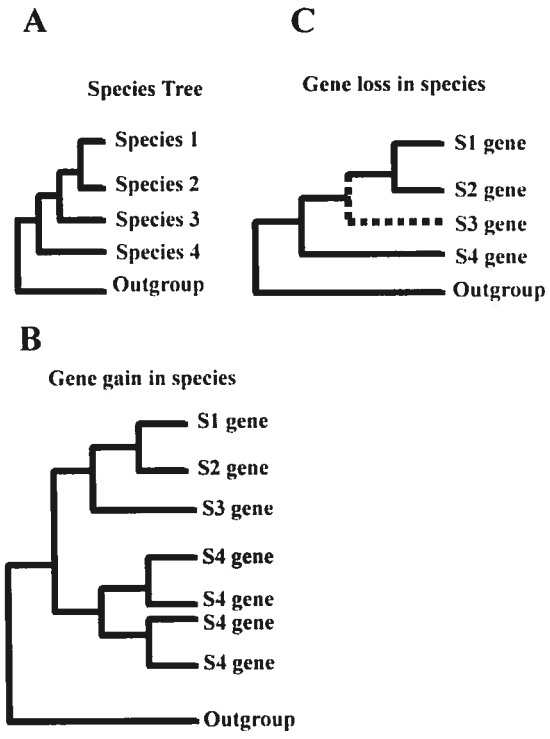


Figure 4

Evolutionary scenarios in the phylogenetic tree

The different kinds of evolutionary scenarios seen in the phylogenetic tree are shown. **A)** Species tree showing the evolutionary relationship between the species, 1, 2, 3 and 4. **B)** A species-specific gain of genes appears as a clade including a single homolog from one species and multiple homologs from the other. Phylogeny between genes S1 gene, S2 gene, S3 gene and S4 gene from species 1, 2, 3 and 4 respectively. Gene gain in species 4 is observed. **C)** Species-specific gene loss appears as the absence of a corresponding ortholog for one species on the tree and is deduced from the evolutionary relationships of the species considered with the other species. Loss of the corresponding gene (S3 gene) in species 3.

Figure 5: Phylogenetic analysis of C2H2-ZNF genes in cluster 19.12 of human and its syntenic counterparts in other mammals

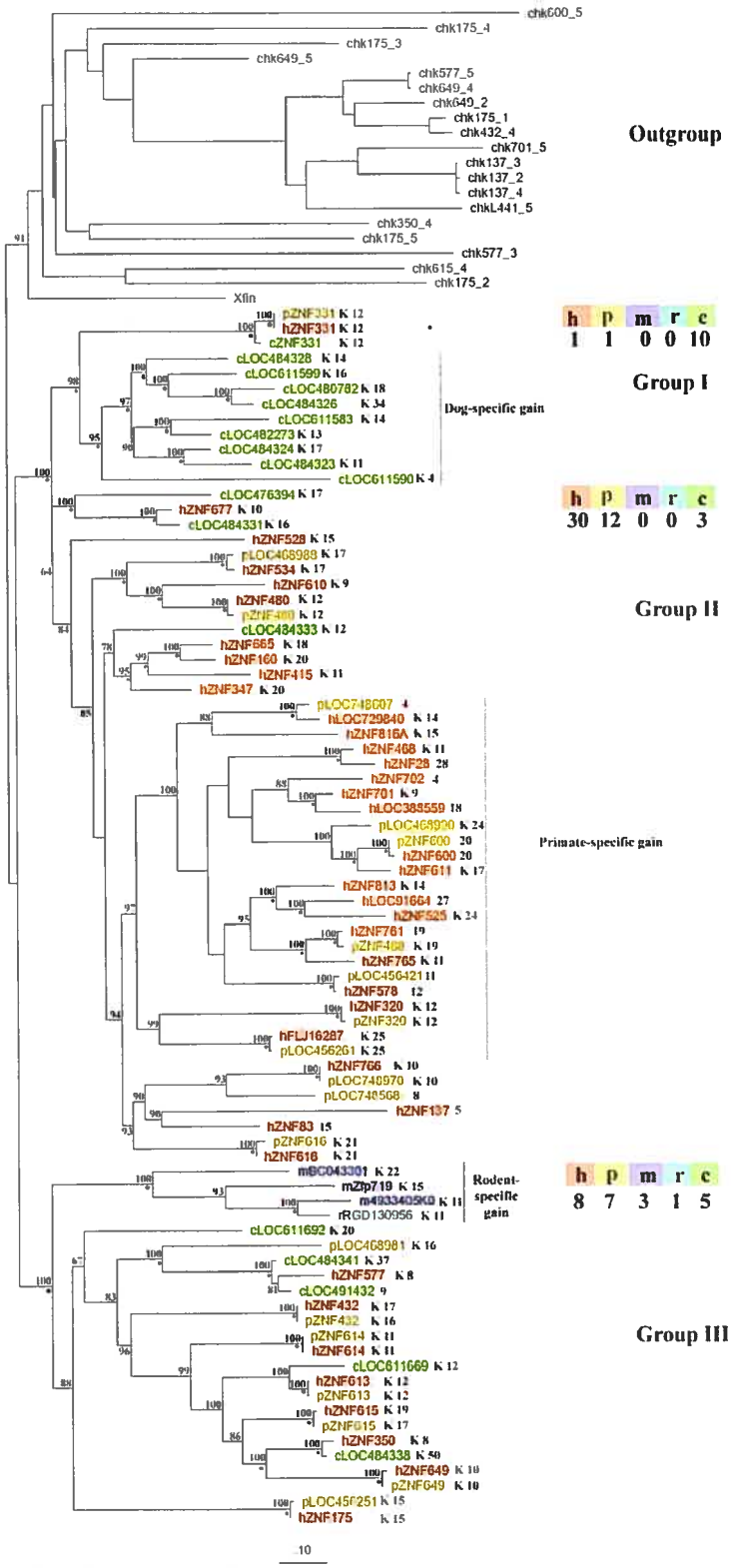


Figure 5

Phylogenetic analysis of C2H2-ZNF genes in cluster 19.12 of human and its syntenic counterparts in other mammals.

A phylogenetic tree was built using the amino acid sequences corresponding to the zinc finger regions of the various human C2H2-ZNF from cluster 19.12 and their syntenic counterparts in chimpanzee, mouse, rat and dog. The tree was generated using a maximum likelihood method (RaxML) and verified using a bayesian method (Mr.Bayes). 346 sites from 101 sequences (including the 20 outgroup sequences from chicken and *Xenopus*) were used in the analysis. The tree is divided into three major Groups (I-III). A tabulation of the number of genes present in each group is indicated for each species (h: human, p: chimpanzee, m: mouse, r: rat, c: dog. The bootstraps values are indicated for each node on the tree. A small black circle is also represented at each node in cases where the posterior probability value is equal to 1.00. This cluster contains only C2H2-ZNF genes that are either from the KRAB subfamily or that do not encode any conserved N-terminal domain. Next to the name of each C2H2-ZNF gene, the presence of an N-terminal KRAB domain is indicated by a K and number of zinc finger motifs is mentioned. A clear evidence of differential expansion is seen in primates and dog. Loss of C2H2-ZNF in the rodent lineage is also observed.

Figure 6: Physical maps showing the organization of the human C2H2-ZNF from cluster 19.12 localized on 19q13.4 and its syntenically homologous counterparts in other mammals

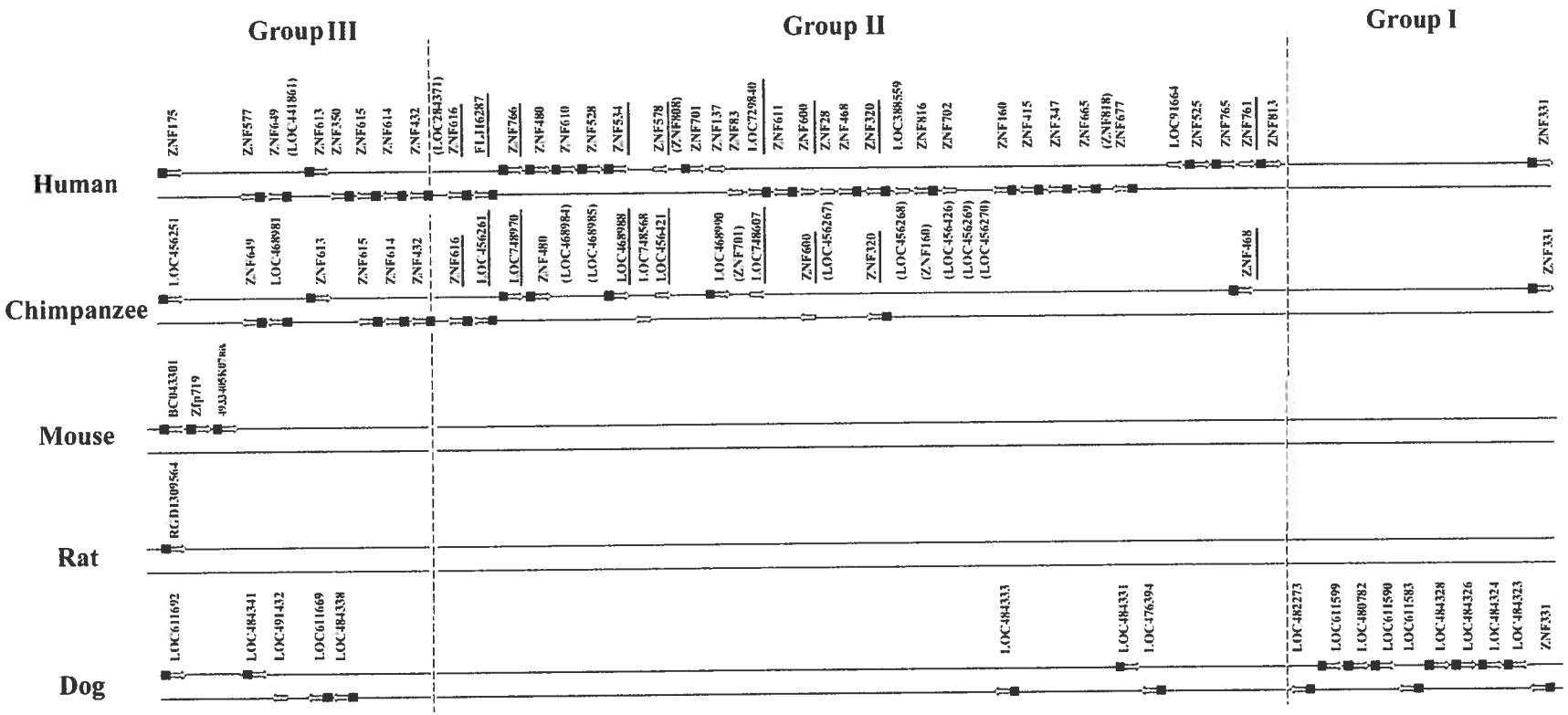


Figure 6

Physical maps showing the organization of the human C2H2-ZNF from cluster 19.12 localized on 19q13.4 and its syntenically homologous counterparts in other mammals. For the large C2H2-ZNF cluster 19.12 and its syntenically homologous counterparts in chimpanzee, mouse, rat and dog, each C2H2-ZNF genes is represented by an open arrow which indicates its orientation on the chromosome strands; this excludes the pseudogenes whose name appears in parenthesis. For these clusters which contain only C2H2-ZNF that are from the KRAB subfamily or that do not encode any conserved N-terminal domain, the presence of a conserved N-terminal KRAB domain is indicated by as square positioned in front of the open arrow representing the gene. Genes identified as orthologs, based on the phylogenetic tree and physical maps, are underlined and are aligned vertically on their respective chromosomes. Dotted lines separate the genes belonging to Group I, Group II and Group III defined in the phylogenetic tree (Figure 5). The two species specific groups from dog and primates are seen in Group I and Group II, respectively.

Figure 7: Variation in the numbers of zinc finger motifs in mammals and in the presence of conserved N-terminal domains in orthologs

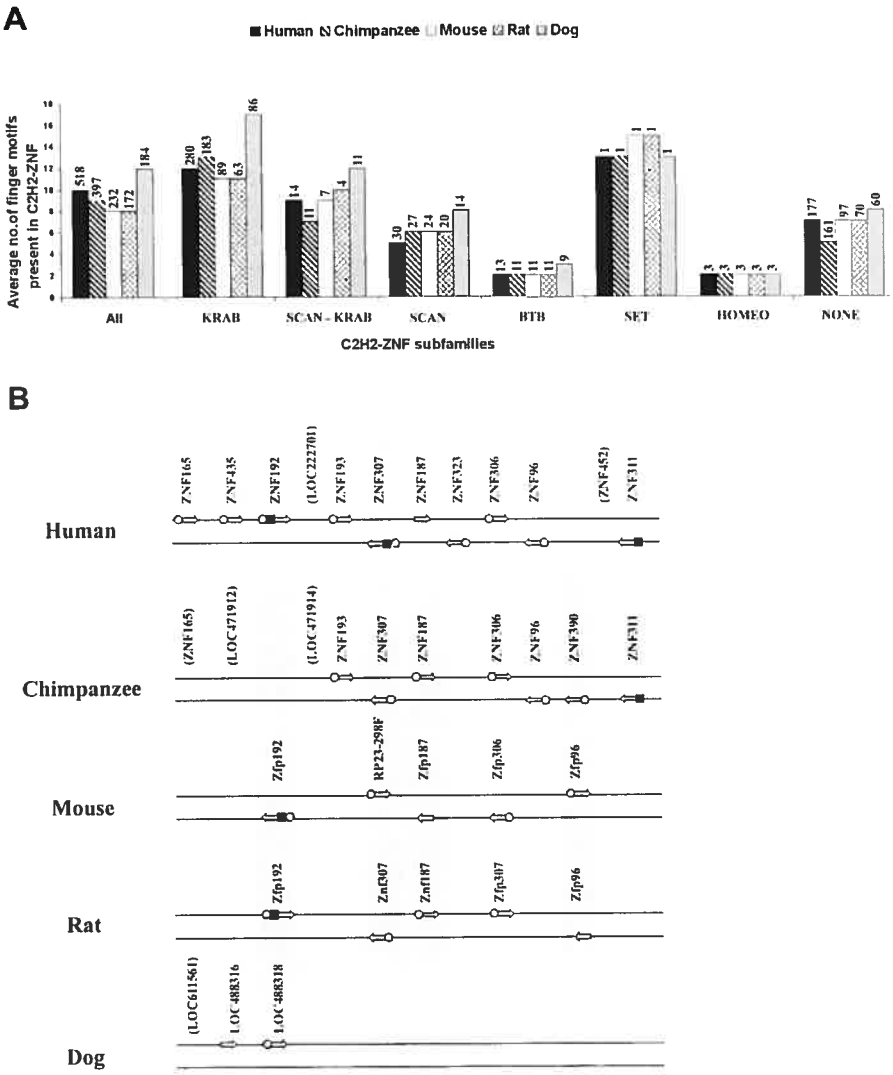


Figure 7

Variation in the numbers of zinc finger motifs in mammals and in the presence of conserved N-terminal domains in orthologs.

A) The average number of zinc finger motifs was calculated for all the C2H2-ZNF from the 81 human clusters identified and their corresponding syntenically homologous clusters in the other mammals; for each species, the average number for the total C2H2-ZNF (All) and for members of the various C2H2-ZNF sub-families (KRAB, SCAN, SCAN-KRAB, BTB, HOMEO, SET and, NONE = no conserved domain associated) is presented. For each category, the number of genes in each species is listed above the bars in the following order (human, chimpanzee, mouse, rat and dog).

B) For the human C2H2-ZNF cluster 6.2 (chromosome 6p22.1) and its syntenically homologous counterparts in chimpanzee, mouse, rat and dog, each C2H2-ZNF genes is represented by an open arrow which indicates its orientation on the chromosome strands; this excludes the pseudogenes whose name appears in parenthesis. For these clusters which contain C2H2-ZNF that are from the KRAB or SCAN subfamily or that do not encode any conserved N-terminal domain, the presence of a conserved N-terminal is indicated by as square for a KRAB domain or an open circle for a SCAN domain both being positioned in front of the open arrow representing the gene. Genes identified as orthologs, based on the phylogenetic tree and physical maps, are aligned vertically on their respective chromosomes. Cases where domain shuffling was observed among orthologs from the different mammals are marked by a grey box.

C) Exon-Intron organization of most human C2H2-ZNF from the SCAN-KRAB and SCAN subfamilies. 80% of the human SCAN-KRAB C2H2-ZNF (11/14) and 55% of the SCAN C2H2-ZNF (16/29) found in clusters have the presented exon-intron structures shown. The exons encoding the SCAN, KRAB (A box) and ZNF are indicated.

Figure 8: Model for the evolution of the SCAN, SCAN-KRAB and KRAB C2H2-ZNF subfamilies

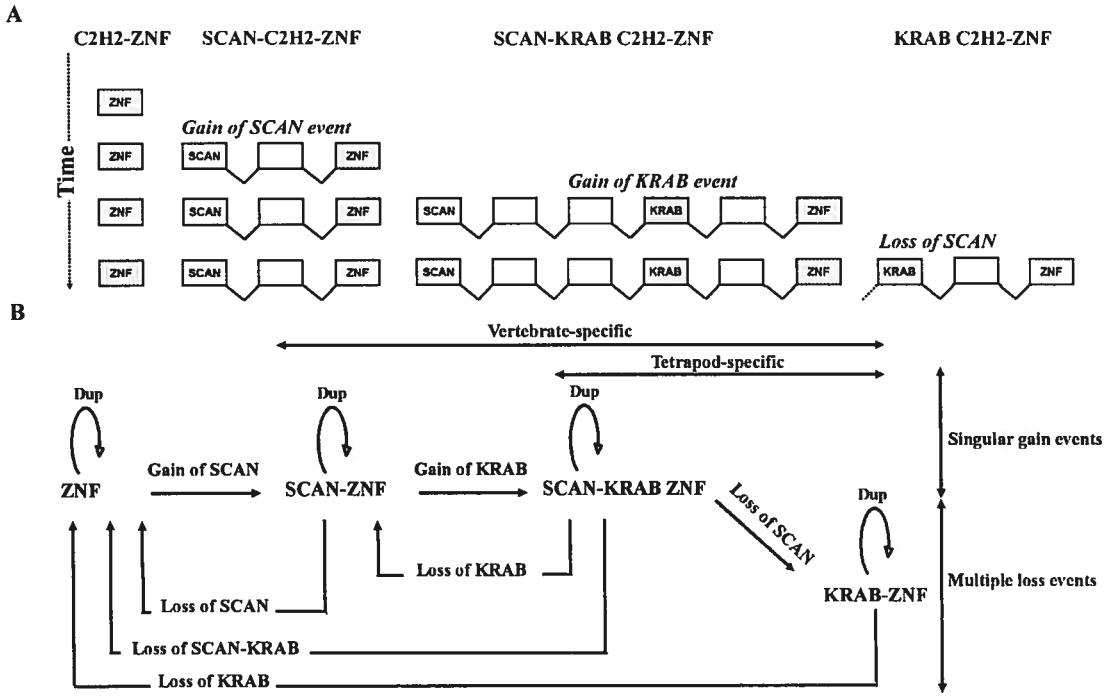


Figure 8

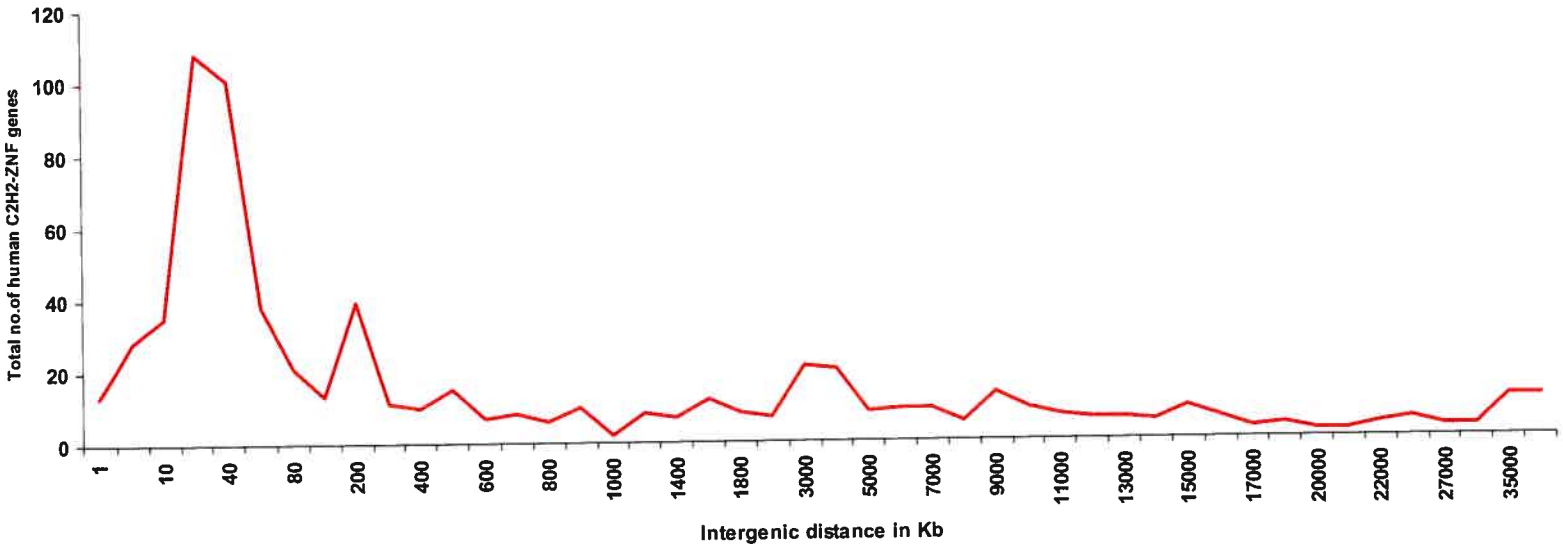
Model for the evolution of the SCAN, SCAN-KRAB and KRAB C2H2-ZNF subfamilies

A) Sequential events of exon shuffling leading to the birth of SCAN-C2H2-ZNF and SCAN-KRAB C2H2-ZNF subfamilies. Most of the SCAN-C2H2-ZNF and the SCAN-KRAB C2H2 ZNF have the exon-intron structure shown (boxes represent exons). Birth of new families may have occurred by an exon shuffling mechanism leading presumably first to the acquisition of a SCAN domain by a C2H2-ZNF and later of a KRAB domain by a SCAN-C2H2-ZNF. Most SCAN-KRAB C2H2-ZNF have a single exon placed in between the exon encoding the KRAB A box (identified as KRAB) and the exon encoding the zinc finger domain (ZNF). This exon encodes in most instances the so-called KRAB B, b, or C boxes.

B) Dynamic evolution of C2H2-ZNF after birth of the SCAN and SCAN-KRAB subfamilies through gene duplication and recurrent loss of effector domains. A first SCAN C2H2-ZNF appeared in an ancestor of vertebrates following the gain of a SCAN domain by a C2H2-ZNF (in grey box); duplication then led to the establishment of the SCAN C2H2-ZNF subfamily. The gain of a KRAB domain at the emergence of tetrapods by a SCAN C2H2-ZNF gave rise to a SCAN-KRAB C2H2-ZNF (in grey box). This was followed by duplication and establishment of the SCAN-KRAB subfamily. Loss of SCAN domain by deletion or degeneration from some SCAN-KRAB C2H2-ZNF genes followed in many instances by duplication led to the expansion of the KRAB C2H2-ZNF. Duplication and loss of SCAN or KRAB domains by deletion or degeneration from SCAN, SCAN-KRAB

and KRAB C2H2-ZNF subfamilies are seen as a recurrent theme shaping the repertoires of the C2H2-ZNF subfamilies.

Supplementary Figure 1: Distribution of intergenic distances between 718 C2H2-ZNF in the human genome



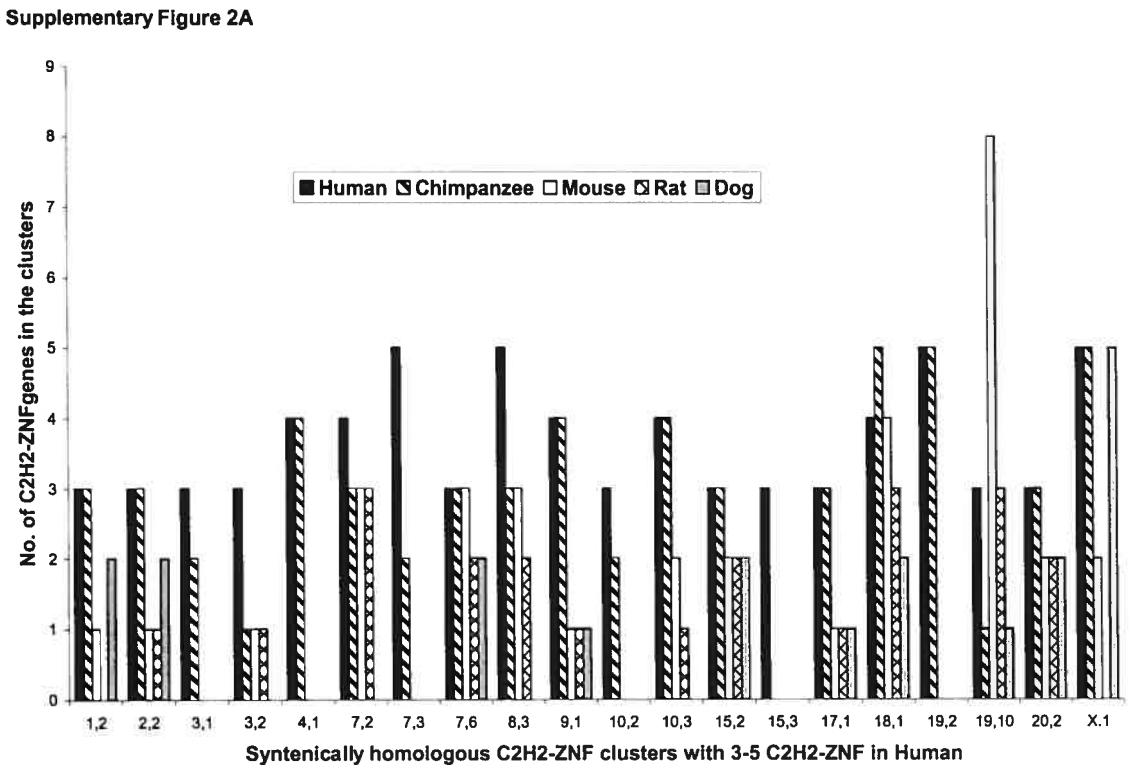
Supplementary Figure 1: Distribution of intergenic distances between 718 C2H2-ZNF in the human genome.

Supplementary Figure 1

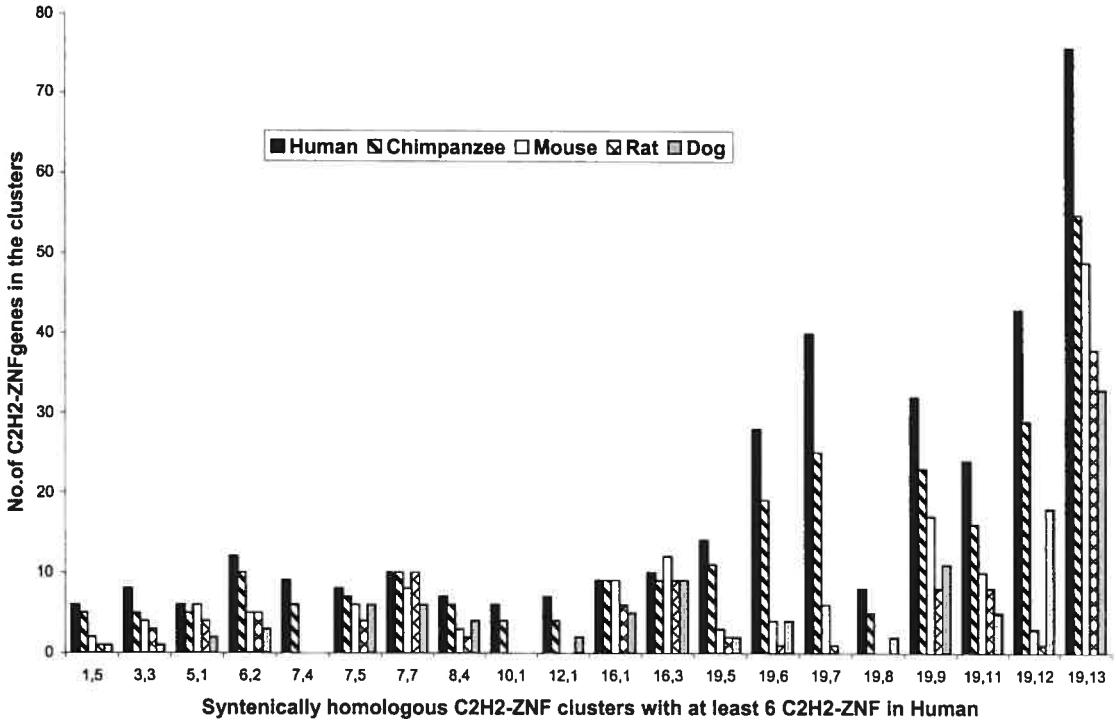
Distribution of intergenic distances between 718 C2H2-ZNF in the human genome.

The intergenic distances between the consecutive C2H2-ZNF on each chromosome was calculated for each C2H2-ZNF of the human genome. For the 718 C2H2-ZNF, the number of C2H2-ZNF found within the range of intergenic distances indicated on the x axis is plotted on the y axis. For example, there are 108 C2H2-ZNF within 10 to 20 Kb from a consecutive C2H2-ZNF.

Supplementary Figure 2: Comparison of the number of C2H2-ZNF genes in the 40 human clusters containing at least 3 C2H2-ZNF and their syntenic counterparts in four other mammals



Supplementary Figure 2B



Supplementary Figure 2

Comparison of the number of C2H2-ZNF genes in the 40 human clusters containing at least 3 C2H2-ZNF and their syntenic counterparts in four other mammals.

For each human C2H2-ZNF cluster named of the graph, the first number indicates the chromosome number and the second is the number attributed to that cluster on the chromosome. C2H2-ZNF clusters with six or more (A) and three to five (B) genes in human and their syntenic counterparts Chimpanzee, Mouse, Rat and Dog. This figure provides evidence of C2H2-ZNF differential species-specific expansion and gene loss in rodents.

Supplementary Table S1

Comprehensive catalogue of the 718 C2H2-ZNF genes in the human genome.

Chr[1]	Position[2]	Cluster[3]	Pseudo[4]	Name[5]	Description[6]	Domain[7]	F[8]	O[9]	L[10]	Start[11]	Stop[12]
1	1p36.3	-		HKR3	GLI-Kruppel family member HKR3	BTB	11	+	688	6562698	6571926
1	1p36.21	-		PRDM2	Retinoblastoma protein interacting ZNF	SET	7	+	1718	13903937	14024162
1	1p36.2-p36.1	-		ZBTB17	Zinc finger and BTB domain containing 17	BTB	13	-	803	16140951	16175101
1	1p36	-		ZNF436	Zinc finger protein 436	KRAB	12	-	470	23559055	23567466
1	1p36.11	1.1a		ZNF593	Zinc finger protein 593	-	1	+	116	26361096	26369951
1	1p36.11	1.1b		ZNF683	Zinc finger protein 683	-	4	-	509	26560712	26571853
1	1p35.1	-		ZBTB8	Zinc finger and BTB domain containing 8	BTB	2	+	512	32777359	32844129
1	1p34.3	-		ZNF31	Zinc finger protein 31(KOX 29)	SCAN	10	+	977	33710846	33734582
1	1p34.2	1.2a		ZNF643	Zinc finger protein 643	KRAB	9	+	432	40688366	40701939
1	1p34.2	1.2b		ZNF642	Zinc finger protein 642	KRAB	9	+	505	40715889	40734602
1	1p34.2	1.2c		ZNF684	Zinc finger protein 684	KRAB	8	+	378	40769820	40786425
1	1p34.2	-		ZNF691	Zinc finger protein 691	-	7	+	284	43084867	43090735
1	1p34.1	1.3a		ZNF393	Zinc finger protein 393	-	3	+	389	44357109	44373399
1	1p34.1	1.3b	Ψ	LOC128208	similar to dJ675G8.1 (novel zinc finger protein)	-	-	-	-	45272586	45686366
1	1p32.3	-		GLIS1	GLIS family zinc finger 1	-	5	-	620	53744494	53972465
1	1p22.2	-		ZNF644	Zinc finger protein 644	-	3	-	1327	91153443	91260259
1	1p22	-		GFI1	Zinc finger protein Gfi-1	-	6	-	422	92712909	92725021
1	1q22	-		ZBTB7B	Zinc finger protein and BTB domain 7B	BTB	4	+	539	153253548	153256078
1	1q25.1	-	Ψ	ZBTB37	Zinc finger and BTB domain containing 37	BTB	-	+	361	172104155	172109404
1	1q25.3	-		ZNF648	Zinc finger protein 648	-	10	-	568	180290328	180297470
1	1q31.1	-	Ψ	LOC441918	similar to Zinc finger protein 132	-	-	+	-	183273244	183280109
1	1q31.2	-	Ψ	LOC391146	similar to zinc finger protein 101	-	-	-	-	189694321	189695464
1	1q31.3	-		ZBTB41	Zinc finger and BTB domain containing 41	BTB	14	-	909	195389437	195436295
1	1q32.1	-		ZNF281	Zinc finger protein 281	-	4	-	895	198642043	198645789
1	1q42.13	1.4a		ZNF678	Zinc finger protein 678	-	15	+	525	225817867	225910754
1	1q42.13	1.4b		Gm127	Similar to zinc finger protein ZFP	KRAB	7	-	714	225951873	225961023
1	1q43	-		LOC441927	similar to zinc finger protein 532	-	2	-	179	240540626	240546819
1	1q44-qter	-		ZNF238	Zinc finger protein 238	BTB	4	+	531	242281208	242287401
1	1q44	1.5a	Ψ	ZNF695	Zinc finger protein 695	KRAB	-	-	133	245175487	245237946
1	1q44	1.5b		ZNF670	Zinc finger protein 670	KRAB	9	-	389	245266710	245308692
1	1q44	1.5c		ZNF669	Zinc finger protein 669	KRAB	9	-	464	245329916	245334251
1	1q44	1.5d		ZNF124	Zinc finger protein 124 (HZF-16)	KRAB	7	-	289	245385826	245401941

1	1q44	1.5e		LOC729806	similar to Zinc finger protein 492	KRAB	10	-	741	245419499	245430449
1	1q44	1.5f		ZNF496	Zinc finger protein 496	SCAN-KRAB	5	-	249	245530245	245561668
1	1q44	1.6a		ZNF672	Zinc finger protein 672	-	13	+	452	247099153	247110337
1	1q44	1.6b		ZNF692	Zinc finger protein 692	-	5	-	519	247110828	247119894
2	2p25	-		KLF11	Kruppel-like factor 11	-	3	+	512	10101133	10112414
2	2p23.3	2.1a		ZNF513	Zinc finger protein 513	-	7	-	541	27453606	27457097
2	2p23.3	2.1b		ZNF512	Zinc finger protein 512	-	2	+	567	27659397	27699467
2	2p16.1	-		BCL11A	B-cell CLL/lymphoma 11A(zinc finger protein)	-	3	-	779	60531806	60634137
2	2p13.2-p13.1	-	Ψ	ZNF638	Zinc finger protein 638	-	-	+	-	71412397	71515697
2	2p13	-		EGR4	Early growth response 4	-	3	-	486	73371570	73374118
2	2q11.1	2.2a		ZNF514	Zinc finger protein 514	KRAB	7	-	400	95177127	95188990
2	2q11.2	2.2b		ZNF2	Zinc finger protein 2(A1-5)	KRAB	9	+	425	95194910	95213792
2	2q11.1	2.2c		LOC344065	Similar to zinc finger protein 135	-	10	+	524	95236998	95245078
2	2q13	2.3a	Ψ	LOC343938	Similar to zinc finger protein532	-	-	+	-	110109590	110125737
2	2q13	2.3b	Ψ	LOC442041	Similar to zinc finger protein 532	-	-	-	-	110552807	110569031
2	2q14	-		GLI2	GLI-Kruppel family member GLI2	-	4	+	1258	121266327	121466321
2	2q21.2	-		LOC442049	Similar to zinc finger protein 285	KRAB	8	+	653	144834358	144972846
2	2q31.2-q31.3	-		ZNF533	Zinc finger protein 533	-	4	-	471	180014954	180434312
2	2q32	-		KLF7	Kruppel-like factor 7	-	3	-	302	207653774	207738859
2	2q34-q35	-		ZNF142	Zinc finger protein 142	-	18	-	1524	219210883	219232599
2	2q34	-		ZNFN1A2	Zinc finger protein , subfamily 1A, 2(Helios)	-	4	-	526	213579589	213723303
3	3p24.3	-		LOC389099	Similar to zinc finger protein 533	-	1	-	145	22150080	22186054
3	3p22.1	3.1a		ZNF619	Zinc finger protein 619	-	10	+	371	40493641	40504881
3	3p22.1	3.1b		ZNF620	Zinc finger protein 620	KRAB	8	+	422	40522534	40534042
3	3p22.1	3.1c		ZNF621	Zinc finger protein 621	KRAB	7	+	439	40541508	40556047
3	3p22.1	3.2a		ZNF651	Zinc finger protein 651	-	8	+	371	42675878	42684076
3	3p22.1	3.2b		ZNF662	Zinc finger protein 662	KRAB	8	+	426	42922406	42934136
3	3p22.1	3.2c	Ψ	LOC339903	Similar to zinc finger protein 621	KRAB	-	+	128	42953061	42959288
3	3p21.32	3.3a		ZNF445	Zinc finger protein 445	SCAN-KRAB	14	-	1031	44457410	44494166
3	3p21.32	3.3b		LOC285346	Similar to zinc finger protein ZFP1	KRAB	12	-	544	44512201	44516816
3	3p22.3-p21.1	3.3c		ZNF167	Zinc finger protein 167	SCAN-KRAB	13	+	754	44571717	44599979
3	3p21.32	3.3d		ZNF660	Zinc finger protein 660	-	10	+	331	44601460	44612561
3	3p21	3.3e		ZNF197	Zinc finger protein 197	SCAN-KRAB	22	+	1029	44641515	44664967
3	3p22-p21	3.3f		ZNF35	Zinc finger protein 35	-	11	+	519	44665259	44677280
3	3p21.31	3.3g		ZNF502	Zinc finger protein 502	-	14	+	544	44729142	44740327

3	3p21.31	3.3h		ZNF501	Zinc finger protein 501	-	9	+	262	44746128	44753579
3	3p21	-		ZNF589	Zinc finger protein 589	KRAB	4	+	421	48257649	48287484
3	3p14.2	-		ZNF312	Zinc finger protein 312	-	6	-	459	62330399	62334061
3	3p12.3	-		ZNF717	Zinc finger protein 717	KRAB	19	-	907	75841490	75917386
3	3p11.1	-		ZNF654	Zinc finger protein 654	-	1	+	299	88271165	88273912
3	3q12.3	-		ZBTB11	Zinc finger and BTB domain containing 11	BTB	12	-	1053	102850978	102878607
3	3p12-qter	3.4a		ZNF80	Zinc finger protein 80	-	7	-	273	115436168	115439115
3	3q13.2	3.4b		ZBTB20	Zinc finger and BTB domain containing 20	BTB	5	-	668	115540207	116348817
3	3q21	-		ZNF148	Zinc finger protein 148	-	4	-	794	126433609	126576781
3	3q13-q21	-		KLF15	Kruppel-like factor 15	-	3	-	416	127544168	127558926
3	3q24	3.5a		ZIC4	Zic family member 4	-	4	-	334	148586527	148607097
3	3q24	3.5b		ZIC1	Zic family member 1	-	4	+	447	148609871	148617196
3	3q26.32	3.6a		ZNF639	Zinc finger protein 639	-	5	+	485	180524245	180536014
3	3q26.3-q27	3.6b		WIG1	P53 target zinc finger protein	-	2	-	288	180224196	180272278
3	3q27	-		BCL6	Zinc finger protein 51	BTB	6	-	706	188921859	188946169
3	3q29	-	Ψ	LOC285388	Similar to zinc finger protein 161	-	-	-	-	194355947	194357460
4	4p16.3	4.1a		ZNF595	Zinc finger protein 595	KRAB	18	+	648	43227	78099
4	4p16.3	4.1b		MGC26356	Similar to zinc finger protein 595	-	6	+	274	196418	239748
4	4p16.3	4.1c	Ψ	LOC654254	Similar to zinc finger protein 595	-	-	-	-	252840	301265
4	4p16.3	4.1d		ZNF141	Zinc finger protein 141	KRAB	11	+	474	321617	359047
4	4p16.3	-		ZNF509	Zinc finger protein 509	BTB	7	+	765	4342879	4374410
4	4p16.1	-		LOC441007	Similar to zinc finger protein 596	KRAB	-	-	178	9048008	9087407
4	4p14-p15.1	-		KLF3	Kruppel-like factor 3	-	3	+	345	38342212	38376796
4	4q12	-		REST	RE1-silencing transcription factor	-	6	+	1097	57468799	57493097
4	4q31.1-q31.2	-		ZNF330	Zinc finger protein 330	-	-	+	320	142361547	142375301
4	4q35.2	-		Zfp42	Zinc finger protein 42	-	3	+	310	189153919	189163193
5	5p15.33	-		ZFP62	Zinc finger protein 62 homolog	-	16	-	498	12667	25741
5	5p15.1	-		ZNF622	Zinc finger protein 622	-	-	-	477	16504628	16518894
5	5p13.3	-		ZFR	Zinc finger RNA binding protein	-	-	-	1074	32390213	32480601
5	5p12	-	Ψ	LOC442134	similar to zinc finger protein 35	-	-	-	-	42460563	42504182
5	5p11-p12	-		ZNF131	Zinc finger protein 131	-	4	+	510	43157399	43211593
5	5q13.2	-		ZNF366	Zinc finger protein 366	-	11	-	744	71774990	71839005
5	5q35.3	-		EGR1	Zinc finger protein 225	-	3	+	543	137829080	137832903
5	5q33.1	-		ZNF300	Zinc finger protein 300	KRAB	12	-	604	150254157	150264584
5	5q35.2	-		ZNF346	Zinc finger protein 346	-	4	+	294	176382303	176426364

5	5q35.3	5.1a		ZNF354A	Zinc finger protein 354A	KRAB	13	-	605	178071128	178090309
5	5q35.3	5.1b		ZNF354B	Zinc finger protein 354B	KRAB	13	+	612	178219607	178244021
5	5q35.3	5.1c		ZFP2	Zinc finger protein 2 homolog	-	13	+	461	178255522	178292816
5	5q35.3	5.1d		ZNF454	Zinc finger protein 454	KRAB	12	+	522	178300830	178326040
5	5q35.3	5.1e		DKFZp686E2433	similar to hypothetical protein 9630041N07	KRAB	13	+	685	178383412	178394655
5	5q35.2	5.1f		ZNF354C	Zinc finger protein 354C	KRAB	11	+	554	178420213	178440301
6	6p22.1	-		ZNF322A	Zinc finger protein 322A	-	11	-	402	26744497	26767741
6	6p21.3	-	Ψ	ZNF204	Zinc finger protein 204	-	-	-	-	27433581	27447283
6	6p22.1	6.1a		LOC346157	Similar to dJ153G14.3 (novel C2H2 type ZNF)	-	9	+	407	27474095	27477205
6	6p21.3	6.1b		ZNF184	Zinc finger protein 184	KRAB	19	-	751	27526506	27548858
6	6p21.3	6.2a		ZNF165	Zinc finger protein 165	SCAN	6	+	485	28154551	28165320
6	6p22.1	6.2b		ZNF435	Zinc finger protein 435	SCAN	4	+	348	28200366	28205836
6	6p21.3	6.2c		ZNF192	Zinc finger protein 192	SCAN-KRAB	9	+	578	28217695	28233215
6	6p22.1	6.2d	Ψ	LOC222701	Similar to zinc finger protein 192	-	-	+	269	28237530	28245348
6	6p21.3	6.2e		ZNF193	Zinc finger protein 193	SCAN	5	+	394	28301049	28309239
6	6p21.33-p21.31	6.2f		ZNF307	Zinc finger protein 307	SCAN-KRAB	7	-	545	28320469	28327981
6	6p21.31	6.2g		ZNF187	Zinc finger protein 187	-	8	+	325	28342872	28353960
6	6p21.31	6.2h		ZNF323	Zinc finger protein 323	SCAN	6	-	406	28400493	28429951
6	6p22.1	6.2i		ZNF306	Zinc finger protein 306	SCAN	7	+	538	28425738	28442503
6	6p22.2-p21.3	6.2j		ZNF305	Zinc finger protein 305	SCAN	11	-	604	28454973	28475487
6	6p22.1	6.2k	Ψ	ZNF452	Zinc finger protein 452	SCAN	-	-	1325	28647386	28663091
6	6p22.1	6.2l		ZNF311	Zinc finger protein 311	KRAB	14	-	574	29070573	29081016
6	6p22.1	-		ZFP57	Zinc finger protein 57 homolog	KRAB	7	-	508	29748239	29756866
6	6p21.33	-		ZBTB12	Zinc finger and BTB domain containing 12	BTB	3	-	459	31975373	31977748
6	6p21.3	6.3a		ZNF297	Zinc finger protein 297	BTB	2	-	634	33390173	33393472
6	6p21.32	6.3b		ZBTB9	Zinc finger and BTB domain containing 9	BTB	1	+	473	33530334	33533299
6	6p21.3-p21.2	-		ZNF76	Zinc finger protein 76	-	7	+	570	35335488	35371738
6	6pter-p12.1	-		ZNF318	Zinc finger protein 318	-	-	-	2099	43411786	43445159
6	6p12.1	-		ZNF451	Zinc finger protein 451	-	5	+	1013	57019470	57143057
6	6q15	-		ZNF292	Zinc finger protein 292	-	11	+	2895	87921986	88030633
6	6q21	-	Ψ	LOC442240	similar to zinc finger protein 259	-	-	-	-	109213585	109214949
6	6q21	-		ZBTB24	Zinc finger and BTB domain containing 24 (ZNF450)	BTB	8	-	697	109890412	109911133
6	6q25	-		PLAG1	Zinc finger protein PLAGL1	-	7	-	463	144303132	144371236
6	6q25.1	-		ZBTB2	Zinc finger and BTB domain containing 2	BTB	3	-	514	151777365	151804791
7	7p22.1	-		LOC441192	Similar to zinc finger protein 16	-	2	-	385	4934111	4957467

7	7p22.1	7.1a		LOC441193	Similar to zinc finger protein 469	-	-	-	303	5426140	5428044
7	7p22.1	7.1b		ZNF815	Zinc finger protein 815	KRAB	3	+	517	5829273	5853913
7	7p22.1	7.2a		DKFZp434J1015	hypothetical protein DKFZp434J1015	-	5	+	563	6428035	6437162
7	7p22.1	7.2b		DKFZp547K054	hypothetical protein DKFZp547K054	KRAB	15	+	1393	6435920	6467482
7	7p22.1	7.2c	Ψ	LOC442283	Similar to zinc finger protein 11b	-	-	+	-	6700537	6713023
7	7p22.1	7.2d		ZNF325	Zinc finger protein 325 (ZNF12)	KRAB	15	-	697	6696844	6713094
7	7p13	-		GLI3	GLI-kruppel family member GLI3	-	5	-	1580	41970196	42241712
7	7p13-p11.1	-		ZNFN1A1	Zinc finger protein, subfamily 1A, 1	-	4	+	519	50411724	50438053
7	7p11.2	-		ZNF713	Zinc finger protein 713	KRAB	6	+	430	55754540	55782642
7	7p11.2	7.3a		LOC442311	Similar to zinc finger protein 43	-	12	+	392	56308323	56309501
7	7p11.2	7.3b	Ψ	LOC222032	Similar to zinc finger protein 208	-	-	-	-	56661026	56708371
7	7p11.2	7.3c		ZNF479	Zinc finger protein 479	KRAB	10	-	878	57191263	57211513
7	7p11.2-p11.1	7.3d	Ψ	LOC340223	Similar to zinc finger protein 479	-	-	-	-	57320482	57343922
7	7p11.1	7.3e		ZNF716	Zinc finger protein 716	KRAB	12	+	495	57501998	57533597
7	7q11.21	7.4a		ZNF679	Zinc finger protein 679	KRAB	9	+	411	63346841	63364744
7	7q11.21	7.4b		LOC728927	Similar to zinc finger protein 92	KRAB	9	+	438	63429463	63446960
7	7q11.21	7.4c		ZNF680	Zinc finger protein 680	KRAB	12	-	530	63617697	63660923
7	7q11.2	7.4d		ZFD25	Zinc finger protein (ZFD25) (ZNF588)	-	24	+	783	63763946	63808839
7	7q11.21-q11.23	7.4e		ZNF138	Zinc finger protein 138	-	6	+	262	63892241	63931140
7	7q11.21	7.4f		ZNF273	Zinc finger protein 273	KRAB	13	+	504	64001101	64028773
7	7q11.2	7.4g		ZNF117	Zinc finger protein 117	KRAB	9	-	383	64075795	64088849
7	7q11.21	7.4h		H-plK	Kruppel-related zinc finger protein	-	13	-	483	64089025	64090719
7	7q11.21	7.4i		ZNF92	Zinc finger protein 92	KRAB	14	+	586	64476203	64503433
7	7q22.1	7.5a		ZNF789	Zinc finger protein 789	KRAB	8	+	425	98908451	98923153
7	7q22.1	7.5b		ZNF394	Zinc finger protein 394	SCAN-KRAB	7	-	561	98928790	98935813
7	7q22	7.5c		ZFP95	Zinc finger protein 95 homolog (mouse)	SCAN-KRAB	13	+	839	98940232	98969380
7	7q22.1	7.5d		ZNF655	Zinc finger protein 655	-	6	+	491	98993981	99012012
7	7q22.1	7.5e		ZNF498	Zinc finger protein 498	-	7	+	544	99052507	99067976
7	7q21.3-q22.1	7.5f		ZKSCAN1	Zinc finger protein 36	KRAB	6	+	563	99451155	99473339
7	7q22.1	7.5g		ZNF38	Zinc finger protein 38	KRAB	8	+	473	99485353	99500599
7	7q22.1	7.5h		ZNF3	Zinc finger protein 3	KRAB	8	-	410	99517266	99499406
7	7q22.1	7.6a		LOC643641	Hypothetical protein LOC643641	KRAB-KRAB	-	+	1163	99500406	99500106
7	7q22.1	7.6b		LOC649746	Hypothetical protein LOC649746	KRAB	3	+	345	99800106	99801106
7	7q22.1	7.6c		LOC567641	Hypothetical protein LOC567641	KRAB	4	+	267	100001106	100003106
7	7q31.1	-		ZNF277	Zinc finger protein (C2H2 type) 277	-	2	+	438	111633879	111770320
7	7q31.32	-		FEZF1	Zinc finger protein FEZ	-	5	-	471	121729287	121731835
7	7q36.1	7.7a		ZNF786	Zinc finger protein 786	KRAB	15	-	782	148397513	148418720

7	7q36.1	7.7b		ZNF425	Zinc finger protein 425	KRAB	19	-	752	148430809	148454311
7	7q36.1	7.7c		ZNF398	Zinc finger protein 398	KRAB	8	+	642	148454441	148511052
7	7q36.1	7.7d		ZNF282	Zinc finger protein 282	KRAB	5	+	671	148553199	148554267
7	7q36.1	7.7e		ZNF212	Zinc finger protein 212	KRAB	4	+	495	148567707	148583630
7	7q36.1	7.7f		ZNF783	Zinc finger protein 783	KRAB	4	+	546	148590195	148625325
7	7q36.1	7.7g		ZNF777	Zinc finger protein 777	KRAB	9	-	831	148759394	148783306
7	7q36.1	7.7h		ZNF746	Zinc finger protein 746	KRAB	4	-	644	148800818	148825727
7	7q36.1	7.7i		ZNF767	Zinc finger protein 767	-	12	-	595	148875178	148952751
7	7q36.1	7.7j		ZNF467	Zinc finger protein 467	-	12	-	595	149092385	149101228
8	8p23.3	-		ZNF596	Zinc finger protein 596	KRAB	11	+	504	172382	187339
8	8p23.1	-		LOC283202	Similar to zinc finger protein 75	KRAB	-	-	178	7188906	7226546
8	8p23.1	-		ZNF705B	Zinc finger protein 705B	KRAB	-	+	300	7821316	7856224
8	8p23.1	8.1a	Ψ	ZNF705C	Zinc finger protein 705C	-	-	+	-	12237465	12263633
8	8p23.1	8.1b		LOC441341	Similar to zinc finger protein 10	-	-	+	301	12214635	12261425
8	8p21.1	-		ZNF395	Zinc finger protein 395	-	1	-	513	28259021	28299896
8	8q11.1	-	Ψ	LOC392215	Similar to zinc finger protein 92	-	-	+	-	47815662	47816648
8	8q21.11	-		ZFHX4	Zinc finger homeodomain 4	HOMEO-4	7	+	3571	77778835	77940711
8	8q13-q21.1	-		ZBTB10	Zinc finger and BTB domain containing 10	BTB	2	+	847	81561003	81595322
8	8q22.2	-		LOC442392	Similar to zinc finger protein 317	-	3	+	293	94576479	94728501
8	8q22.2	-		KLF10	Kruppel-like factor 10	-	3	-	480	103730188	103737128
8	8q23	-		ZFPM2	zinc finger protein, multitype 2	-	2	+	1151	106400323	106885943
8	8q24.12	-		TRPS1	Zinc finger transcription factor TRPS1	-	1	-	1281	116489900	116750402
8	8q24.12	-	Ψ	LOC392264	Similar to zinc finger protein 532	-	-	-	-	120561696	120612050
8	8q24.13	8.2a		ZHX2	Zinc fingers and homeoboxes 2	HOMEO-3	1	+	837	123863082	124055936
8	8q24.13	8.2b		ZHX1	Zinc fingers and homeoboxes 1	HOMEO-4	1	-	873	124329877	124355728
8	8q24.13	-		ZNF572	Zinc finger protein 572	-	12	+	529	126054733	126060809
8	8q24.22	-		ZNF406	Zinc finger protein 406	-	13	-	1243	135559213	135794463
8	8q24.3	8.3a		ZFP41	Zinc finger protein 41 homolog	-	4	+	198	144400484	144416250
8	8q24.3	8.3b		GLI4	GLI-Kruppel family member GLI4	-	7	+	376	144420982	144430476
8	8q24.3	8.3c		ZNF696	Zinc finger protein 696	-	9	+	374	144444971	144451539
8	8q24.3	8.3d		ZNF623	Zinc finger protein 623	-	13	+	536	144802973	144809731
8	8q24.3	8.3e		ZNF707	Zinc finger protein 707	KRAB	7	+	369	144838650	144849515
8	8q24.3	8.4a		ZNF251	Zinc finger protein 251	-	7	-	293	145948140	145952607
8	8q24.3	8.4b		ZNF34	Zinc finger protein 34	KRAB	12	-	549	145969309	145981498
8	8q24.3	8.4c		ZNF517	Zinc finger protein 517	KRAB	10	+	492	145995065	146006265
8	8q24	8.4d		ZNF7	Zinc finger protein 7	KRAB	14	+	686	146023807	146039409

8	8q24.3	8.4e		ZNF647	Zinc finger protein 647 (ZNF250)	KRAB	13	-	560	146077337	146097632
8	8q24.3	8.4f		ZNF16	Zinc finger protein 16	-	17	-	682	146126548	146147078
8	8q24.3	8.4g		LOC642914	Similar to zinc finger protein 135	KRAB	10	-	335	146172603	146196311
9	9p24.2	-		GLIS3	GLIS family zinc finger 3	-	6	-	775	3817676	4142183
9	9p13.1	-	Ψ	LOC653501	Similar to zinc finger protein 658	-	-	-	-	39433814	39354471
9	9p13.2	-		ZBTB5	Zinc finger and BTB domain containing 5	BTB	2	-	677	37428111	37455396
9	9p12	-		ZNF658	Zinc finger protein 658	KRAB	21	-	1059	40761412	40782063
9	9p12	-		ZNF658B	Zinc finger protein 658B	-	21	-	819	41578833	41582207
9	9q22.31	-		ZNF484	Zinc finger protein 484	KRAB	18	-	816	94648172	94680111
9	9q22.32	-		ZNF169	Zinc finger protein 169	KRAB	13	+	603	96080861	96103545
9	9q22.32	9.1a		ZNF367	Zinc finger protein 367	-	2	-	350	98190057	98220490
9	9q22.33	9.1b		ZNF510	Zinc finger protein 510	KRAB	10	-	683	98557968	98580149
9	9q22.33	9.1c		ZNF782	Zinc finger protein 782	KRAB	14	-	699	98619094	98656210
9	9q22.33	9.1d		ZNF322B	Zinc finger protein 322B	-	11	-	402	98999358	99001731
9	9q22-q31	-		ZNF189	Zinc finger protein 189	KRAB	16	+	626	103200984	103212763
9	9q31.2	9.2a		ZNF462	Zinc finger protein 462	-	9	+	2506	108665199	108813628
9	9q31	9.2b		KLF4	Kruppel-like factor 4	-	3	-	470	109286956	109291576
9	9q31.3	-		ZNF483	Zinc finger protein 483	SCAN-KRAB	-	+	256	113327260	113379945
9	9q32	9.3a		LOC169834	hypothetical protein LOC169834	-	13	-	530	114799221	114814293
9	9q32	9.3b		ZFP37	Zinc finger protein 37 homolog (mouse)	KRAB	12	-	630	114843995	114858817
9	9q32	-		ZNF618	Zinc finger protein 618	-	4	+	861	115678380	115852293
9	9q33.2	9.4a		ZNF482	Zinc finger protein 482	BTB	4	-	424	124710150	124715430
9	9q33.2	9.4b		ZBTB26	Zinc finger and BTB domain containing 26	BTB	4	-	441	124720199	124733600
9	9q33-q34	-		ZNF297B	Zinc finger protein 297B	BTB	3	+	467	128607182	128637318
9	9q34	-		ZNF79	Zinc finger protein 79	KRAB	11	+	498	128662244	128687979
9	9	-		ZBTB34	Zinc finger and BTB domain containing 34	BTB	3	+	532	129226482	129247471
10	10p15	-		KLF6	Kruppel- like factor 6	-	3	-	283	3808188	3817455
10	10p11.2	10.1a		ZNF248	Zinc finger protein 248	KRAB	8	-	579	38157905	38186492
10	10p11	10.1b	Ψ	BA775A3.1	KRAB box zinc finger pseudogene	-	-	+	-	38212553	38213031
10	10p11	10.1c	Ψ	BA393J16.4	Zinc finger pseudogene	-	-	+	-	38223526	38225869
10	10p11.21	10.1d		ZNF25	Zinc finger protein 25	KRAB	12	-	456	38278801	38300744
10	10p11.2	10.1e		ZNF33A	Zinc finger protein 33a	KRAB	16	+	810	38341276	38388310
10	10p11.2	10.1f		ZNF37A	Zinc finger protein 37a	KRAB	12	+	561	38423281	38452286
10	10q11.21	10.2a		LOC401642	Similar to zinc finger protein 91	-	16	-	597	42151534	42172713
10	10q11.21	10.2b		ZNF37B	Zinc finger protein 37b	KRAB	8	-	525	42367418	42368286

10	10q11.2	10.2c		ZNF11B	Zinc finger protein 11b	KRAB	16	-	778	42404561	42453998
10	10q11.21	10.3a		ZNF487	Zinc finger protein 487	KRAB	3	+	421	43252288	43298636
10	10q11.22-q11.23	10.3b		ZNF239	Zinc finger protein 239	-	9	-	458	43371801	43383913
10	10q11.21	10.3c		ZNF485	Zinc finger protein 485	KRAB	11	+	402	43421881	43433358
10	10q22-q25	10.3d		ZNF32	Zinc finger protein 32	-	7	-	273	43459313	43464332
10	10q11	-		ZNF22	Zinc finger protein 22	-	5	+	224	44815928	44820780
10	10q11.22	-		ZNF488	Zinc finger protein 488	-	2	+	340	47975095	47993872
10	10q21.2	-		ZNF365	Zinc finger protein 365	-	-	+	462	63803957	64101777
10	10q22.2	-		ZNF503	Zinc finger protein 503	-	1	-	646	76827915	76831431
10	10q22.3	-	Ψ	LOC399783	Similar to zinc finger protein 532	-	-	+	-	79149361	79164538
10	10q24.1	-		ZNF518	Zinc finger protein 518	-	4	+	1483	97879494	97912480
10	10q26	-		ZNFN1A5	zinc finger protein, subfamily 1A, 5	-	4	-	420	124741955	124758311
10	10q26.3	-		ZNF511	zinc finger protein 511	-	3	+	252	134972413	134976656
11	11p15.5	-		ZNF195	zinc finger protein 195	KRAB	10	-	557	3336572	3356891
11	11p15.4	11.1a		ZNF215	zinc finger protein 215	SCAN-KRAB	4	+	517	6904230	6935854
11	11p15.4	11.1b		ZNF214	zinc finger protein 214	KRAB	11	-	606	6977125	6998117
11	11p15.4	-		ZNF143	zinc finger protein 143 (clone pHZ-1)	-	7	+	626	9439089	9506188
11	11p14.3	-		LOC341002	similar to dJ568F9.1 (zinc finger protein 133	KRAB	15	+	653	23767378	23768482
11	11p13	-	Ψ	ZNF	Kruppel like zinc finger protein	-	-		-	32365900	32413653
11	11p11.2	-		ZNF408	zinc finger protein 408	-	10	+	720	46678944	46684037
11	11q12	11.2a		ZFP91	zinc finger protein 91 homolog (mouse)	-	4	+	570	58103225	58145091
11	11q12.2	11.2b		ZFP91-CNTF	zinc finger protein 91 homolog (mouse), CNF	-	-	+	529	58103225	58149778
11	11q12.3	-		ZBTB3	zinc finger and BTB domain containing 3	BTB	2	-	574	62275011	62278190
11	11q13.4	-		LOC440053	similar to zinc finger protein 596	KRAB	-	-	178	71196287	71232113
11		-		ZNF75C	Zinc finger protein 75C	-	3		90	77774116	77774385
11	11q23.1	-		ZBTB16	zinc finger and BTB domain containing 16	BTB	9	+	673	113435659	113626608
11	11q23.3	-		ZNF202	zinc finger protein 202	SCAN-KRAB	8	-	648	123100207	123117573
11	11q24.3	-		ZBTB15	BTB(POZ) domain containing 15	BTB	2	+	539	129601789	129689717
12	12p12	-		ZNF384	zinc finger protein 384	-	6	-	516	6645904	6668930
12	12p13.31	-		ZNF705A	Zinc finger protein 705A	KRAB	5	+	300	8216417	8223909
12	12q13	-	Ψ	ZNF75B	zinc finger protein 75b	-	-	-	-	42687843	42690385
12	12q13.11	-		ZNF641	Zinc finger protein 641	KRAB	5	-	438	47022179	47030844
12	12q13.13	-		ZNF385	zinc finger protein 385	-	3	-	366	53049187	53064748
12	12q13	-		ZNFN1A4	zinc finger protein, subfamily 1A, 4	-	4	+	544	54704791	54717887
12	12q13.2-q13.3	-		GLI	Glioma-associated oncogene homolog	-	5	+	1106	56140201	56152312

12	12q24.31	-		ZNF664	Zinc finger protein 664	-	9	+	261	123023623	123065922
12	12q24.33	12.1a		ZNF605	zinc finger protein 605	KRAB	17	-	641	132008120	132042941
12	12q24.33	12.1b		ZNF26	zinc finger protein 26 (KOX 20)	KRAB	13	+	533	132073129	132099226
12	12q24.33	12.1c		ZNF84	zinc finger protein 84 (HPF2)	KRAB	19	+	738	132124239	132146261
12	12q24.32-q24.33	12.1d		ZNF140	Zinc finger protein 140	KRAB	10	+	457	132167110	132194333
12	12q24.33	12.1e		LOC440122	similar to KRAB zinc finger protein 6D	KRAB	6	-	224	132206943	132208683
12	12q24.33	12.1f		ZNF10	zinc finger protein 10 (KOX 1)	KRAB	11	+	573	132217287	132246124
12	12q24.33	12.1g		ZNF268	zinc finger protein 268	-	24	+	947	132268068	132289534
13	13q22.1	13.1a		KLF5	Kruppel like factor 5	-	3	+	457	72531143	72549677
13	13q22	13.1b		KLF12	Kruppel like factor 12	-	3	-	402	73158150	73606043
13	13q32.3	13.2a		ZIC5	Zic family member 5 (odd-paired homolog, Drosophi	-	4	-	639	99415452	99422179
13	13q32	13.2b		ZIC2	Zic family member 2 (odd-paired homolog, Drosophi	-	4	+	532	99432320	99437020
13	13q32.1	-		DZIP1	zinc-finger protein DZIP1	-	1	-	780	99600736	99601385
14	14q11.1	-		LOC441666	similar to ZNF43 protein	KRAB	17	-	645	18241488	18260616
14	14q11	-		ZNF219	zinc finger protein 219	-	6	-	722	20628076	20636639
14	14q11.2	14.1a		ZFHX2	Zinc finger homeobox 2	HOMEO-3	4	-	1517	23059904	23065839
14	14q11.2	14.1b		ZNF409	zinc finger protein 409	-	1	-	862	23066981	23090692
14	14q22.1	-	Ψ	ZNF405P	zinc finger protein 405 pseudogene	-	-	-	-	50240463	50241980
14	14q23-q24	14.2a		ZBTB25	Zinc finger and BTB domain containing 25	BTB	2	-	435	64023308	64040307
14	14q23.3	14.2b		ZBTB1	zinc finger and BTB domain containing 1	BTB	2	+	644	64041174	64070161
14	14q24.3	-		ZNF410	zinc finger protein 410	-	5	+	478	73423339	73468556
14	14q32.2	-		BCL11B	B-cell CLL/ zinc finger protein	-	6	-	894	98705377	98807575
14	14q11.1-q12	-		SALL2	Sal-like 2	-	7		1007	21059071	21075177
15	15q12	-		KLF13	Kruppel-like factor 13	-	3	+	288	29406375	29457394
15	15q15.3	-		ZNF690	Zinc finger protein 690	SCAN	6	-	851	41440563	41449400
15	15q22.31	-		ZNF609	Zinc finger protein 609	-	1	+	1411	62578672	62765320
15	15q24	-		ZNF291	Zinc finger protein 291	-	-	-	1399	74427592	74963247
15	15q24	15.1a		BTBD1	BTB (POZ) domain containing 1	BTB	-	-	482	81476179	81527110
15	15q25.2	15.1b		BNC1	Zinc finger protein basonuclein	-	3	-	994	81715659	81744472
15	15q25.2	15.2a		ZFP29	Zinc finger protein 29 (ZSCAN2)	SCAN	14	+	614	82945253	82967951
15	15q25-q26	15.2b		SCAND2	SCAN domain containing protein 2	SCAN	-	+	306	82975695	82985528
15	15q25.3	15.2c		ZNF592	Zinc finger protein 592	-	4	+	1267	83092584	83147530
15	15q26.1	15.3a		LOC390636	Similar to zinc finger protein 495 (pseudo)	-	2	+	133	88280506	88281204
15	15q26.1	15.3b		ZNF710	Zinc finger protein 710	-	11	+	664	88345756	88425029

15	15q26.1	15.3c	ZNF774	Zinc finger protein 774	-	12	+	483	88696546	88705719	
16	16p13.3	-	ZNF598	zinc finger protein 598	-	-	-	904	1987769	1999764	
16	16p13.3	16.1a	ZNF206	zinc finger protein 206	SCAN	14	-	725	3078896	3082862	
16	16p13.3	16.1b	ZNF205	zinc finger protein 205	KRAB	8	+	554	3102607	3110519	
16	16p13.3	16.1c	ZNF213	zinc finger protein 213	SCAN-KRAB	5	+	459	3125140	3132806	
16	16p13.3	16.1d	ZNF200	zinc finger protein 200	-	5	-	395	3212343	3225410	
16	16p13.3	16.1e	ZNF263	zinc finger protein 263	SCAN-KRAB	9	+	683	3273488	3281461	
16	16p13.11	16.1f	ZNF75A	Zinc finger protein 75a	KRAB	5	+	296	3295485	3308575	
16	16p13.3	16.1g	ZNF434	Zinc finger protein 434	-	6	-	485	3372086	3391026	
16	16p13.3	16.1h	ZNF174	zinc finger protein 174	SCAN	3	+	407	3391245	3399365	
16	16p13.3	16.1i	ZNF597	zinc finger protein 597	-	7	-	424	3426111	3433491	
16	16p13.3	16.2a	GLIS2	GLIS family zinc finger 2	-	4	+	524	4322226	4327803	
16	16p13.3	16.2b	ZNF500	zinc finger protein 500	SCAN	5	-	480	4740816	4757167	
16	16p12.1	-	ZNF694	Zinc finger protein 694	SCAN-KRAB	6	-	967	25154823	25176343	
16	16p11.2	16.3a	ZNF553	zinc finger protein 553	-	12	+	618	30314558	30318216	
16	16p11.2	16.3b	ZNF768	Zinc finger protein 768	-	10	-	520	30442826	30445411	
16	16p11.2	16.3c	ZNF747	Zinc finger protein 747	KRAB	-	-	191	30450280	30453695	
16	16p11.2	16.3d	ZNF764	Zinc finger protein 764	KRAB	7	-	408	30472586	30477085	
16	16p11.2	16.3e	ZNF688	Zinc finger protein 688	KRAB	2	-	276	30488508	30491229	
16	16p11.2	16.3f	ZNF785	Zinc finger protein 785	KRAB	7	-	405	30499495	30504511	
16	16p11.2	16.3g	ZNF689	Zinc finger protein HIT-39 (ZNF689)	KRAB	11	-	500	30522187	30529183	
16	16p11.2	16.3h	ZNF629	zinc finger protein 629	-	19	-	1056	30697271	30706024	
16	16p11.2	16.3i	ZNF668	Zinc finger protein 668	-	16	-	619	30979672	30993005	
16	16p11.2	16.3j	ZNF646	Zinc finger protein 646	-	29	+	1832	30993269	31002334	
16	16p11.2	16.4a	LOC342426	similar to zinc finger protein 267	-	13	-	580	31520510	31522350	
16	16p11.2	16.4b	ZNF267	zinc finger protein 267	KRAB	14	+	743	31632096	31680365	
16	16q12	-	ZNF423	zinc finger protein 423	-	23	-	1284	48082022	48418419	
16	16q13	-	ZNF319	zinc finger protein 319	-	15	-	582	56586074	56591263	
16	16q22	16.5a	ZNF23	Zinc finger protein 23 (KOX 16)	KRAB	17	-	643	70039000	70053618	
16	16q22	16.5b	ZNF19	Zinc finger protein 19 (KOX 12)	KRAB	10	-	458	70065563	70080742	
16	16q22.3	-	ZFP1	zinc finger protein 1 homolog (mouse)	-	8	+	352	73739926	73763486	
16	16q24	16.6a	ZNF469	zinc finger protein 469	-	3	+	3446	87021380	87034666	
16	16q24.2	16.6b	ZFPM1	zinc finger protein, multitype 1	-	2	+	1004	87047226	87128890	
16	16q24.3	-	ZFP276	ZNF276 homolog (mouse)	-	4	+	539	88314934	88333811	
17	17p13.2	17.1a	ZFP3	zinc finger protein 3 homolog (mouse)	-	13	+	502	4922478	4940393	

17	17p13-p12	17.1b	ZNF232	zinc finger protein 232	SCAN	5	-	444	4949755	4967121
17	17p13	17.1c	ZNF594	zinc finger protein 594	-	22	-	1120	5023554	5028416
17	17p13.1	-	ZBTB4	zinc finger and BTB domain containing 4	BTB	6	-	1013	7303421	7328241
17	17p11.2	-	ZNF18	zinc finger protein 18 (KOX 11)	SCAN-KRAB	5	-	549	11821487	11841414
17	17p11.2	-	ZNF286	zinc finger protein 286	KRAB	10	+	521	15544054	15561963
17	17p11.2	17.2a	ZNF287	zinc finger protein 287	SCAN-KRAB	14	-	754	16395426	16413189
17	17p11.2	17.2b	ZNF624	Zinc finger protein 624	KRAB	21	-	865	16464776	16497883
17	17q11.2	-	ZNF207	zinc finger protein 207	-	2	+	478	27414039	27721583
17	17q12	-	ZNF403	zinc finger protein 403	-	-	+	697	31974917	32020391
17	17q21	-	ZNFN1A3	zinc finger protein, subfamily 1A, 3	-	1	-	509	35174724	35273967
17	17q21.32	-	ZNF652	Zinc finger protein 652	-	9	-	606	44727485	447948834
17	17q22	-	ZNF161	zinc finger protein 161	-	5	-	521	53403909	53420614
18	18pter-p11.2	-	ZFP161	zinc finger protein 161 homolog (mouse)	BTB	5	-	449	5279379	5283313
18	18p11.21	-	ZNF519	zinc finger protein 519	KRAB	10	-	540	14094724	14122429
18	18q11.2	-	LOC441816	similar to zinc finger protein 586	-	-	-	122	20341044	20341412
18	18q11.2	-	ZNF521	zinc finger protein 521	-	24	-	1311	20895889	21186114
18	18q12.2	18.1a	ZNF397	zinc finger protein 397	SCAN	-	+	275	31075017	31092357
18	18q12	18.1b	ZNF271	zinc finger protein 271	-	5	+	423	31124298	31142072
18	18q12	18.1c	ZNF24	zinc finger protein 24 (KOX 17)	SCAN	4	-	368	31169957	31178405
18	18q12	18.1d	ZNF396	zinc finger protein 396	SCAN	2	-	333	31200659	31211299
18	18q21.1	-	APM-1	BTB/POZ-zinc finger protein-like	BTB	4	-	766	43807731	43821492
18	18q21.32	-	ZNF532	zinc finger protein 532	-	8	+	1301	54681041	54804689
18	18q23	-	ZNF407	Zinc finger protein 407	-	7	+	1001	70474282	70762386
18	18q23	18.2a	ZNF516	zinc finger protein 516	-	7	-	2852	72198625	72296128
18	18q22-q23	18.2b	ZNF236	zinc finger protein 236	-	25	+	1558	72665104	72811671
18	18q23	-	SALL3	Sal-like 3	-	8	+	1300	74841263	74859182
19	19p13.3	19.1aa	KLF16	Kruppel-like factor 16	-	3	-	252	1803399	1814496
19	19p13.3	19.1ab	BTBD2	BTB domain containing 2	BTB	-	-	525	1936447	1966702
19	19p13.3	19.2aa	ZNF554	Zinc finger protein 554	KRAB	7	+	487	2770917	2786469
19	19p13.3	19.2ab	ZNF555	Zinc finger protein 555	KRAB	15	+	628	2792482	2805036
19	19p13.3	19.2ac	ZNF556	Zinc finger protein 556	KRAB	9	+	456	2818333	2829501
19	19p13.3	19.2ad	ZNF57	Zinc finger protein 57	KRAB	13	+	555	2851964	2869474
19	19p13.3	19.2ae	ZNF77	Zinc finger protein 77 (pT1)	KRAB	12	-	545	2884217	2895930
19	19p13.3	19.3aa	KIAA1086	Similar to zinc finger protein	-	3	-	939	3755010	3820026
19	19p13.3	19.3ab	ZBTB7A	Zinc finger and BTB domain containing 7	BTB	4	-	584	3996217	4017816

19	19p13.2	19.4aa		ZNF557	Zinc finger protein 557	KRAB	10	+	430	7020721	7034589
19	19p13.3-p13.2	19.4ab		ZNF358	Zinc finger protein 358	-	9	+	481	7487075	7491911
19	19p13.2	19.5aa		ZNF414	Zinc finger protein 414	-	1	-	312	8485032	8482224
19	19p13.2	19.5ab		ZNF558	Zinc finger protein 558	KRAB	9	-	402	8781382	8794565
19	19p13	19.5ac		ZNF317	Zinc finger protein 317	KRAB	13	+	595	9112073	9135084
19	19p13.2	19.5ad		ZNF699	Zinc finger protein 699	KRAB	16	-	642	9265957	9281384
19	19p13.2	19.5ae		ZNF559	Zinc finger protein 559	KRAB	11	+	538	9295928	9315521
19	19p13.2	19.5af		ZNF177	Zinc finger protein 177	KRAB	7	+	321	9334696	9353866
19	19p13.2	19.5ag		ZNF266	Zinc finger protein 266	KRAB	14	-	549	9384272	9407234
19	19p13.2	19.5ah		ZNF560	Zinc finger protein 560	KRAB-KRAB	14	-	790	9438003	9470279
19	19p13.2	19.5ai		ZNF426	Zinc finger protein 426	KRAB	12	-	554	9499683	9510303
19	19p13.2	19.5aj		ZNF121	Zinc finger protein 121	-	10	-	390	9537292	9556209
19	19p13.2	19.5ak		ZNF561	Zinc finger protein 561	-	10	-	417	9580131	9592899
19	19p13.2	19.5al		ZNF562	Zinc finger protein 562	-	9	-	354	9620341	9632550
19	19p13.2	19.5am	Ψ	LOC729648	Similar to zinc finger protein 561	-	-	-	-	9661814	9667794
19	19p13.2	19.5an		LOC162993	Hypothetical protein LOC162993	KRAB	12	-	533	9729151	9740410
19	19p13.2	19.6aa		ZNF653	Zinc finger protein 653	-	4	-	615	11455246	11477654
19	19p13.2	19.6ab		ZNF627	Zinc finger protein 627	KRAB	11	+	461	11569327	11590974
19	19p13.2	19.6ac		LOC401898	Similar to hypothetical protein FLJ38281	-	6	+	187	11611591	11624258
19	19p13.2	19.6ad		HSZFP36	Zinc finger protein ZFP-36	KRAB	16	-	610	11693080	11710731
19	19p13.2	19.6ae		ZNF441	Zinc finger protein 441	-	19	+	626	11738907	11754301
19	19p13.2	19.6af		ZNF491	Zinc finger protein 491	-	13	+	437	11770400	11780306
19	19p13.2	19.6ag		ZNF440	Zinc finger protein 440	KRAB	12	+	595	11801554	11806031
19	19p13.2	19.6ah		ZNF439	Zinc finger protein 439	KRAB	11	+	499	11837844	11841306
19	19p13.2	19.6ai		ZNF69	Zinc finger protein 69	KRAB	-	+	149	11859670	11886144
19	19p13.2	19.6aj		ZNF700	Zinc finger protein 700	KRAB	21	+	742	11896900	11922578
19	19p13.2	19.6ak		ZNF440L	Zinc finger protein 440 like	KRAB	8	+	397	11936869	11952214
19	19p13.2	19.6al		ZNF433	Zinc finger protein 433	KRAB	19	-	673	11986573	11990116
19	19p13.2	19.6am		LOC729747	Similar to zinc finger protein 709	KRAB	15	-	578	12015620	12028127
19	19p13.2	19.6an		FLJ14959	Hypothetical protein FLJ14959	KRAB	8	+	666	12036528	12049631
19	19q13.43	19.6ao		ZNF788	Hypothetical protein LOC388507	-	16	+	615	12064078	12086499
19	19p13.3-p13.2	19.6ap		ZNF20	Zinc finger protein 20 (KOX 13)	KRAB	13	-	536	12103803	12112116
19	19p13.2	19.6aq		ZNF625	Zinc finger protein 625	-	8	-	306	12116705	12128529
19	19p13.2-p13.12	19.6ar		ZNF136	Zinc finger protein 136	KRAB	14	+	540	12134919	12161064
19	19p13.2	19.6as		ZNF44	Zinc finger protein 44 (KOX 7)	KRAB	16	-	637	12219007	12266637
19	19p13.2	19.6at		ZNF563	Zinc finger protein 563	KRAB	8	-	476	12289291	12305502
19	19p13.2	19.6au		ZNF442	Zinc finger protein 442	KRAB	14	-	627	12321185	12337447

19	19p13.2	19.6av		ZNF799	Zinc finger protein 799	KRAB-KRAB	26	-	1130	12351256	12373052	
19	19p13.2	19.6aw		ZNF443	Zinc finger protein 443	KRAB	19	-	671	12401520	12412824	
19	19p13.2	19.6ax		ZNF709	Zinc finger protein 709	KRAB	19	-	641	12435743	12456632	
19	19p13.2	19.6ay		ZNF564	Zinc finger protein 564	KRAB	16	-	553	12497540	12523317	
19	19p13.2	19.6az		ZNF490	Zinc finger protein 490	KRAB	13	-	529	12547645	12582595	
19	19p13.2-p13.13	19.6ba		ZNF791	Zinc finger protein 791	KRAB	17	+	576	12582753	12601676	
19	19p13.13-p13.12	19.6bb		KLF1	Kruppel-like factor 1	-	3	-	362	12859017	12856239	
19	19p13	-		ZNF333	Zinc finger protein 333	KRAB-KRAB	10	+	665	14661870	14692772	
19	19p13.1	-		WIZ	Widely-inter spaced zinc finger motifs	-	7	-	1752	15392824	15421762	
19	19p13.13-p13.11	-		KLF2	Kruppel-like factor 2	-	3	+	355	16296651	16299337	
19	19p13.11	19.7aa		ZNF101	Zinc finger protein 101	KRAB	10	+	436	19640663	19652138	
19	19p13.3-p13.2	19.7ab		ZNF140	Zinc finger protein 14 (KOX 6)	KRAB	19	-	642	19682282	19704921	
19	19p13.11	19.7ac		ZNF506	Zinc finger protein 506	KRAB	8	-	444	19757648	19793560	
19	19p12	19.7ad		LOC730008	Similar to zinc finger protein 506	-	9	+	281	19805375	19807882	
19	19p12	19.7ae		ZNF253	Zinc finger protein 253	KRAB	3	+	275	19837714	19865293	
19	19p12	19.7af		ZNF505	Zinc finger protein 505	KRAB	17	+	620	19872787	19906766	
19	19p12	19.7ag		ZNF682	Zinc finger protein 682	KRAB	11	-	498	19996186	19977814	
19	19p12	19.7ah		ZNF90	Zinc finger protein 90	KRAB	15	+	657	20049817	20091937	
19	19p12	19.7ai		ZNF486	Zinc finger protein 486	KRAB	10	+	463	20139023	20172670	
19	19p12	19.7aj		FLJ44894	Similar to zinc finger protein 91	-	3	-	177	20367763	20399602	
19	19p12	19.7ak		LOC163233	Similar to zinc finger protein 93	KRAB	14	-	693	20518768	20540439	
19	19p12	19.7al		ZNF626	Zinc finger protein 626	KRAB	-	-	92	20619349	20636246	
19	19p12	19.7am		ZNF66	Zinc finger protein 66	-	8	+	335	20765278	20782245	
19	19p13.1-p12	19.7an		ZNF85	Zinc finger protein 85	KRAB	15	+	595	20897934	20925343	
19	19p12	19.7ao		ZNF430	Zinc finger protein 430	KRAB	12	+	570	20995337	21033493	
19	19p12	19.7ap		ZNF714	Hypothetical protein LOC148206	-	12	+	400	21056849	21099736	
19	19p12	19.7aq		ZNF431	Zinc finger protein 431	KRAB	12	+	576	21116728	21160645	
19	19p12	19.7ar		ZNF708	Zinc finger protein 708	KRAB	15	-	563	21265803	21304052	
19	19p12	19.7as		ZNF493	Zinc finger protein 493	KRAB	-	+	103	21371798	21383441	
19	19p13.1	19.7at		ZNF429	Zinc finger protein 429	KRAB	17	+	674	21480303	21512565	
19	19p12	19.7au		ZNF100	Zinc finger protein 100	KRAB	12	-	542	21697408	21742163	
19	19p13.1-p12	19.7av		ZNF43	Zinc finger protein 43 (HTF6)	KRAB	22	-	809	21779592	21810810	
19	19p12	19.7aw		ZNF208	Zinc finger protein 208	KRAB	34	-	1167	21945325	21963552	
19	19q13	19.7ax		ZNF257	Zinc finger protein 257	KRAB	12	+	535	22027189	22064570	
19	19p12	19.7ay		ZNF676	Zinc finger protein 676	KRAB	15	-	588	22153733	22155823	
19	19p12	19.7az		LOC148198	Similar to zinc finger protein 492	KRAB	13	-	572	22363750	22396874	
19	19p12	19.7az	Ψ	LOC441843	Similar to zinc finger protein 429	-	-	+	-	22401608	22447337	

19	19p12	19.7az		ZNF492	Zinc finger protein 492	KRAB	28	+	1132	22608966	22642312
19	19p12	19.7az		ZNF99	Zinc finger protein 99	KRAB	30	-	1036	22730847	22744624
19	19p12	19.7az		LOC646864	Similar to zinc finger protein 430	KRAB	7	+	651	22781774	22833075
19	19p12	19.7az	Ψ	LOC388523	Similar to Zinc finger protein 208	-	-	-	-	22949451	22977818
19	19p12	19.7az	Ψ	ZNF724P	Similar to zinc finger protein 43	-	-	-	-	23196873	23258859
19	19p13.1-p12	19.7az		ZNF91	Zinc finger protein 91 (HPF7, HTF10)	KRAB	35	-	1191	23333876	23370089
19	19p12	19.7az	Ψ	ZNF725	Zinc finger protein 725	-	-	-	-	23466157	23490947
19	19p12	19.7az		ZNF675	Zinc finger protein 675 (TIZ)	KRAB	14	-	568	23627812	23661782
19	19p12	19.7az		ZNF681	Zinc finger protein 681	-	16	-	576	23718000	23733479
19	19p12	19.7az		LOC646895	Similar to zinc finger protein 539	-	6	+	193	23806558	23807139
19	19p12	19.7az		LOC730084	Similar to zinc finger protein 539	-	6	+	171	23807338	23807853
19	19p12	19.7az	Ψ	LOC730087	Similar to zinc finger protein 91	-	-	+	-	23889628	23910116
19	19p12	19.7az		ZNF254	Zinc finger protein 254	KRAB	4	+	353	24061816	24103022
19	19q12	-		ZNF536	Zinc finger protein 536	-	8	+	1300	35555168	35740805
19	19q12	-		ZNF537	Zinc finger protein 537	HOMEO	2	-	1081	36457693	36462015
19	19q13.11	-		ZNF507	Zinc finger protein 507	-	5	+	953	37528393	37570413
19	19q13.11	19.8aa		LOC441847	Similar to zinc finger protein 239	-	9	-	345	39817307	39818566
19	19q13.11	19.8ab		ZNF302	Zinc finger protein 302	KRAB	7	+	478	39860439	39869137
19	19q13.11	19.8ac		ZNF181	Zinc finger protein 181 (HHZ181)	KRAB	11	+	507	39916933	39925613
19	19q13.11	19.8ad		ZNF599	Zinc finger protein 599	KRAB	14	-	588	39940819	39955960
19	19q13.11	19.8ae		LOC643825	Similar to zinc finger protein 396	SCAN	-	+	417	39985078	40008745
19	19q13.11	19.8af		LOC441848	Similar to zinc finger protein 113	-	1	-	587	40049101	40049676
19	19q13.11	19.8ag		ZNF30	Zinc finger protein 30 (KOX 28)	-	18	+	542	40109724	40127912
19	19q13.11	19.8ah		ZNF92	Zinc finger protein 92	-	13	-	553	40139098	40143044
19	19q13.1	19.9aa		TZFP	Testis zinc finger protein	BTB	2	+	487	40895670	40899780
19	19q13.12	19.9ab		ZNF565	Zinc finger protein 565	KRAB	12	-	499	41364889	41384792
19	19q13.1	19.9ac		ZNF146	Zinc finger protein 146	-	10	+	292	41411488	41421506
19	19q13.12	19.9ad		ZFP14	Zinc finger protein 14-like	KRAB	13	-	533	41550715	41519002
19	19q13.12	19.9ae		ZNF545	Zinc finger protein 545	KRAB	13	-	532	41574701	41601390
19	19q13.12	19.9af		ZNF566	Zinc finger protein 566	KRAB	7	-	418	41630415	41659358
19	19q13.12	19.9ag		ZFP260	Zinc finger protein 260	-	13	-	412	41693770	41711012
19	19q13.13	19.9ah		ZNF529	Zinc finger protein 529	-	11	-	458	41727130	41756030
19	19q13.12	19.9ai		ZNF382	Zinc finger protein 382	KRAB	10	+	550	41788061	41811339
19	19q13.12	19.9aj		GIOT-1	Gonadotropin inducible TRF	KRAB	12	-	563	41820123	41849579
19	19q13.12	19.9ak		ZNF567	Zinc finger protein 567	KRAB	15	+	616	41872142	41904066
19	19q13.12	19.9al		LOC342892	Hypothetical protein LOC342892	KRAB	32	-	1090	41930509	41955571
19	19q13.12	19.9a m		MGC62100	Hypothetical protein LOC388536	KRAB	13	-	636	41996710	42021121

19	19q13.12	19.9an	ZNF345	Zinc finger protein 345	-	15 + 488	42033103	42062311	
19	19q13.12	19.9ao	ZNF568	Zinc finger protein 568	KRAB	12 + 417	42173736	42180341	
19	19q13.12	19.9ap	LOC653284	Zinc finger protein LOC653284	-	12 + 439	42156286	42181436	
19	19q13.12	19.9aq	ZNF420	Zinc finger protein 420	KRAB	19 + 688	42261222	42312502	
19	19q13.12	19.9ar	ZNF585A	Zinc finger protein 585A	KRAB	23 - 714	42332844	42355455	
19	19q13.12	19.9as	ZNF585B	Zinc finger protein 585B	KRAB	23 - 769	42367562	42393293	
19	19q13.12	19.9at	ZNF383	Zinc finger protein 383	KRAB	11 + 475	42409206	42426414	
19	19q13.12	19.9au	HKR1	GLI-kruppel family member HKR1	KRAB	13 + 659	42517420	42547197	
19	19q13.1	19.9av	ZNF527	Zinc finger protein 527	KRAB	12 + 609	42553899	42575806	
19	19q13.12	19.9aw	ZNF569	Zinc finger protein 569	KRAB	18 - 686	42593902	42650179	
19	19q13.12	19.9ax	ZNF570	Zinc finger protein 570	KRAB	11 + 536	42651822	42668082	
19	19q13.12	19.9ay	LOC390927	Similar to zinc finger protein 569	KRAB	5 + 374	42689538	42720601	
19	19q13.12	19.9az	ZNF540	Zinc finger protein 540	KRAB	17 + 660	42734148	42796836	
19	19q13.12	19.9ba	ZNF571	Zinc finger protein 571	KRAB	17 - 609	42746995	42777513	
19	19q13.13	19.9bb	ZFP30	Similar to Hypothetical zinc finger protein	KRAB	13 - 519	42815229	42838153	
19	19q13.12	19.9bc	FJL37549	Hypothetical protein FLJ37549	-	3 - 327	42850491	42874992	
19	19q13.1	19.9bd	ZNF607	Zinc finger protein 607	KRAB	20 - 696	42879116	42902531	
19	19q13.12	19.9be	ZNF573	Zinc finger protein 573	-	19 - 607	42921028	42956231	
19	19q13.12-q13.13	19.9bf	LOC401915	Similar to hypothetical protein FLJ32191	-	8 + 721	42974701	43002122	
19	19q13.2	19.10aa	ZNF546	Zinc finger protein 546	KRAB	22 + 836	45194869	45215355	
19	19q13.2	19.10ab	ZNF780B	Hypothetical protein LOC163131	KRAB	23 - 833	45232104	45248079	
19	19q13.2	19.10ac	LOC284323	Hypothetical protein LOC284323	KRAB	17 - 607	45272263	45280891	
19	19q13.2	19.11aa	ZNF574	Zinc finger protein 574	-	19 + 896	47266358	47277544	
19	19q13.2	19.11ab	ZNF526	Zinc finger protein 526	-	11 + 670	47416332	47424193	
19	19q13.31	19.11ac	ZNF575	Zinc finger protein 575	-	6 + 245	48729166	48732130	
19	19q13.31	19.11ad	ZNF576	Zinc finger protein 576	-	3 + 170	48792384	48795996	
19	19q13.31	19.11ae	ZNF283	Zinc finger protein 283	KRAB	15 + 852	49016379	49044884	
19	19q13.31	19.11af	ZNF404	Zinc finger protein 404	KRAB	15 - 552	49068497	49079956	
19	19q13.2	19.11ag	ZNF45	Zinc finger protein 45	KRAB	18 - 682	49108621	49121397	
19	19q13.2	19.11ah	ZNF221	Zinc finger protein 221	KRAB	15 + 617	49147237	49163701	
19	19q13.2-q13.32	19.11ai	ZNF155	Zinc finger protein 155	KRAB	12 + 538	49180195	49194317	
19	19q13.31	19.11aj	ZNF230	Zinc finger protein 230	KRAB	10 + 474	49198949	49209912	
19	19q13.2	19.11ak	ZNF222	Zinc finger protein 222	KRAB	10 + 451	49221356	49229100	
19	19q13.2	19.11al	ZNF223	Zinc finger protein 223	KRAB	10 + 482	49248004	49263982	
19	19q13.31	19.11a m	ZNF284	Zinc finger protein 284	KRAB	14 + 370	49268137	49283463	
19	19q13.2	19.11an	ZNF224	Zinc finger protein 224	KRAB	18 + 707	49290337	49304317	
19	19q13.2	19.11ao	ZNF225	Zinc finger protein 225	KRAB	18 + 706	49309381	49328824	

19	19q13.31	19.11ap		ZNF234	Zinc finger protein 234	KRAB	19	+	700	49337618	49354127	
19	19q13.2	19.11aq		ZNF226	Zinc finger protein 226	KRAB	19	+	803	49361089	49373678	
19	19q13.32	19.11ar		ZNF227	Zinc finger protein 227	KRAB	19	+	799	49408531	49433260	
19	19q13.31	19.11as		ZNF233	Zinc finger protein 233	KRAB	8	+	670	49455916	49471308	
19	19q13.2	19.11at		ZNF235	Zinc finger protein 235	KRAB	16	-	738	49482440	49501010	
19	19q13.2	19.11au		ZNF228	Zinc finger protein 228	KRAB	17	-	907	49522546	49552666	
19	19q13.32	19.11av		ZNF285	Zinc finger protein 285	KRAB	11	-	590	49581648	49597605	
19	19q13.31	19.11aw	Ψ	LOC147711	Similar to zinc finger protein 285	-	-	+	-	49654547	49669406	
19	19q13.2	19.11ax		ZNF180	Zinc finger protein 180 (HHZ168)	KRAB	12	-	692	49671701	49696395	
19	19q13.2	-		ZNF342	Zinc finger protein 342	-	6	-	475	50266599	50271528	
19	19q13.32	-		ZNF541	Zinc finger protein 541	-	2	-	792	52715759	52739966	
19	19q13.32	-		ZNF114	Zinc finger protein 114	KRAB	4	+	417	53466466	53482675	
19	19q13.33	-		ZNF473	Zinc finger protein 473	KRAB	20	+	871	55221024	55243845	
19	19q13.4	19.12aa		ZNF175	Zinc finger protein 175	KRAB	15	+	711	56766343	56784803	
19	19q13.41	19.12ab		ZNF577	Zinc finger protein 577	KRAB	8	-	478	57066365	57083009	
19	19q13.41	19.12ac		ZNF649	Zinc finger protein 649	KRAB	10	-	505	57100059	57084301	
19	19q13.41	19.12ad	Ψ	LOC441861	Similar to zinc finger protein 84	-	-	-	-	57109726	57112852	
19	19q13.41	19.12ae		ZNF613	Zinc finger protein 613	KRAB	12	+	581	57122500	57140817	
19	19q13.41	19.12af		ZNF350	Zinc finger protein 350	KRAB	8	-	532	57159406	57181880	
19	19q13.41	19.12ag		ZNF615	Zinc finger protein 615	KRAB	19	-	731	57186400	57203270	
19	19q13.41	19.12ah		ZNF614	Zinc finger protein 614	KRAB	11	-	585	57208391	57223429	
19	19q13.41	19.12ai		ZNF432	Zinc finger protein 432	KRAB	17	-	652	57228490	57243885	
19	19q13.41	19.12aj		LOC284371	Hypothetical protein LOC284371	-	-	-	924	57259531	57290830	
19	19q13.41	19.12ak		ZNF616	Zinc finger protein 616	KRAB	21	-	781	57308967	57335003	
19	19q13.41	19.12al		FLJ16287	Similar to zinc finger protein 616	KRAB	25	-	936	57349937	57366556	
19	19q13.41	19.12a m		ZNF766	Zinc finger protein 766	KRAB	10	+	468	57464636	57487788	
19	19q13.41	19.12an		ZNF480	Zinc finger protein 480	KRAB	12	+	516	57492263	57520987	
19	19q13.41	19.12ao		ZNF610	Zinc finger protein 610	KRAB	9	+	462	57540494	57561923	
19	19q13	19.12ap		ZNF528	Zinc finger protein 528	KRAB	15	+	628	57592933	57613469	
19	19q13.41	19.12aq		ZNF534	Zinc finger protein 534	KRAB	17	+	674	57624252	57634511	
19	19q13.41	19.12ar		ZNF578	Zinc finger protein 578	-	12	+	365	57706025	57708794	
19	19q13.41	19.12as		ZNF808	Similar to zinc finger protein 600	-	2	+	903	57738331	57750693	
19	19q13.41	19.12at		ZNF701	Zinc finger protein 701	KRAB	9	+	465	57765340	57779861	
19	19q13.4	19.12au		ZNF137	Zinc finger protein 137	-	5	+	207	57791719	57795214	
19	19q13.3	19.12av		ZNF83	Zinc finger protein 83 (HPF1)	-	15	-	516	57807443	57833450	
19	19q13.41	19.12aw		LOC729840	Similar to zinc finger protein 160	KRAB	14	-	626	57847611	57885574	
19	19q13.41	19.12ax		ZNF611	Zinc finger protein 611	KRAB	17	-	705	57899284	57924947	

19	19q13.41	19.12ay	ZNF600	Zinc finger protein 600	-	20 - 722	57959279	57981828	
19	19q	19.12az	ZNF28	Zinc finger protein 28 (KOX 24)	-	18 - 665	57992474	57994974	
19	19q13.4	19.12ba	ZNF468	Zinc finger protein 468	KRAB	11 - 469	58033684	58052682	
19	19q13.41	19.12bb	ZNF320	Zinc finger protein like	KRAB	12 - 509	58075392	58076918	
19	19q13.41	19.12bc	LOC388559	Similar to zinc finger protein 600	-	18 - 682	58101230	58112874	
19	19q13.41	19.12bd	ZNF816	Zinc finger protein 816	KRAB	15 - 651	58144446	58157926	
19	19q13.41	19.12be	ZNF702	Zinc finger protein 702	-	4 129	58163317	58188596	
19	19q13.41	19.12bf	ZNF160	Zinc finger protein 160	KRAB	20 - 818	58260992	58298488	
19	19q13.42	19.12bg	ZNF415	Zinc finger protein 415	KRAB	11 - 555	58302945	58327957	
19	19q13.42	19.12bh	ZNF347	Zinc finger protein 347	KRAB	20 - 839	58334460	58348858	
19	19q13.42	19.12bi	ZNF665	Zinc finger protein 665	KRAB	18 - 678	58388415	58357489	
19	19q13.42	19.12bj	ZNF818	Zinc finger protein 818	-	2 + 136	58408040	58411160	
19	19q13.42	19.12bk	ZNF677	Zinc finger protein 677	KRAB	10 - 584	58449923	58430450	
19	19q13.42	19.12bl	LOC91664	Hypothetical protein LOC91664	-	27 + 920	58507227	58549611	
19	19q13.42	19.12bm	ZNF525	Zinc finger protein 525	KRAB	24 + 425	58574446	58579145	
19	19q13.41	19.12bn	ZNF765	Zinc finger protein 765	KRAB	11 + 523	58590209	58607074	
19	19q13.42	19.12bo	ZNF761	Zinc finger protein 761	-	19 + 692	58644962	58653326	
19	19q13.42	19.12bp	ZNF813	Zinc finger protein 813	KRAB	14 + 408	58662809	58689358	
19	19q13.3-q13.4	19.12bq	ZNF331	Zinc finger protein 331	KRAB	12 + 463	58716183	58773147	
19	19q13.42	19.13aa	ZNF628	Zinc finger protein 628	-	17 + 1048	60679511	60687662	
19	19q13.42	19.13ab	KLP1	K562 leucine zipper like protein	-	- + 206	60688406	60690747	
19	19q13.42	19.13ac	ZNF579	Zinc finger protein 579	-	8 - 562	60780711	60783999	
19	19q13.42	19.13ad	FIZ1	FLT3-interacting zinc finger 1	-	11 - 496	60794558	60802705	
19	19q13.42	19.13ae	ZNF524	Zinc finger protein 524	-	4 + 264	60803542	60806316	
19	19q13.42	19.13af	ZNF784	Zinc finger protein 784	-	5 - 323	60823919	60827753	
19	19q13.42	19.13ag	ZNF580	Zinc finger protein 580	-	3 + 172	60844204	60846648	
19	19q13.42	19.13ah	ZNF581	Zinc finger protein 581	-	4 + 197	60846798	60848801	
19	19q13.42	19.13ai	ZNF787	Zinc finger protein 787	-	7 - 382	61290544	61324461	
19	19q13.43	19.13aj	ZNF444	zinc finger protein 444	SCAN	4 + 327	61344368	61364074	
19	19q13.43	19.13ak	LOC342933	Similar to zinc finger protein 495	SCAN	5 - 447	61392870	61396157	
19	19q13.43	19.13al	LOC649137	Similar to zinc finger protein and SCAN 5	SCAN	5 + 358	61409239	61412381	
19	19q13.43	19.13a m	ZSCAN5	Zinc finger SCAN domain 5	SCAN	5 - 496	61424491	61431471	
19	19q13.43	19.13an	LOC646698	Similar to zinc finger and SCAN 5	SCAN	3 + 324	61447224	61450678	
19	19q13.43	19.13ao	ZNF542	zinc finger protein 542	-	8 + 241	61571320	61582709	
19	19q13.43	19.13ap	ZNF582	zinc finger protein 582	KRAB	10 - 517	61586460	61596701	
19	19q13.43	19.13aq	ZNF583	zinc finger protein 583	KRAB	12 + 569	61607530	61628212	
19	19q13.43	19.13ar	ZNF667	Zinc finger protein 667	KRAB	15 - 610	61643012	61680555	

19	19q13.43	19.13as		ZNF471	zinc finger protein 471	KRAB	15	+	626	61711024	61732082	
19	19q13.43	19.13at		ZFP28	Zinc finger protein 28 mouse homolog	KRAB-KRAB	15	+	868	61742129	61759982	
19	19q13.43	19.13au		ZNF470	Zinc finger protein 470	KRAB	17	+	717	61768362	61781931	
19	19q13.4	19.13av		ZNF71	zinc finger protein 71 (Cos26)	-	13	+	489	61798504	61827362	
19	19q13.43	19.13aw		BC37295_3	Hypothetical protein BC37295_3	-	14	-	559	61866765	61876058	
19	19q13.4	19.13ax		ZIM2	Zinc finger, imprinted 2	KRAB	5	-	527	61977742	62043887	
19	19q13.4	19.13ay		PEG3	Paternally expressed 3	SCAN	12	-	1588	62015615	62043876	
19	19q13.4	19.13az		ZIM3	Zinc finger, imprinted 3	KRAB	11	-	472	62337276	62348382	
19	19q13.4	19.13ba		ZNF264	zinc finger protein 264	KRAB	13	+	627	62394681	62422351	
19	19q13.43	19.13bb		ZNF805	Zinc finger protein 805	KRAB	11	+	377	62456615	62465479	
19	19q13.4	19.13bc		ZNF272	Zinc finger protein 272	KRAB	11	+	562	62483745	62496618	
19	19q13.43	19.13bd		ZNF543	Zinc finger protein 543	KRAB	13	+	600	62523689	62533956	
19	19q13.4	19.13be		ZNF304	Zinc finger protein 304	KRAB	15	+	659	62554487	62563078	
19	19q13.43	19.13bf		ZNF574	Zinc finger protein 547	KRAB	9	+	402	62566691	62582739	
19	19q13.43	19.13bg		ZNF548	Zinc finger protein 548	KRAB	11	+	533	62593030	62604598	
19	19q13.4	19.13bh		ZNF175	Zinc finger protein 17 (KOX 10)	KRAB	17	+	662	62614359	62624983	
19	19q13.4	19.13bi		ZNF749	Zinc finger protein 749	-	17	+	691	62646590	62648665	
19	19q13.43	19.13bj		ZNF772	Zinc finger protein 772	KRAB	10	-	489	62672766	62680750	
19	19q13.43	19.13bk		ZNF419	Zinc finger protein 419	KRAB	11	+	510	62690945	62697860	
19	19q13.43	19.13bl		ZNF773	Zinc finger protein 773	KRAB	9	+	442	62703121	62711338	
19	19q13.43	19.13bm		ZNF549	Zinc finger protein 549	KRAB	15	+	627	62730505	62743943	
19	19q13.43	19.13bn		ZNF550	Zinc finger protein 550	KRAB	8	-	381	62759537	62750155	
19	19q13.4	19.13bo		ZNF416	Zinc finger protein 416	KRAB	12	-	594	62782055	62774746	
19	19q13.43	19.13bp		ZIK1	Zinc finger protein ZIK1	KRAB	9	+	487	62787440	62795570	
19	19q13.43	19.13bq		ZNF530	Zinc finger protein 530	KRAB	13	+	599	62803065	62811444	
19	19q13.4	19.13br		ZNF134	Zinc finger protein 134	-	10	+	427	62817440	62826536	
19	19q13.4	19.13bs		ZNF211	Zinc finger protein 211	KRAB	12	+	564	62836396	62845948	
19	19q13.43	19.13bt		ZSCAN4	Zinc finger and SCAN domain 4	SCAN	4	+	433	62872115	62882317	
19	19q13.43	19.13bu		ZNF551	Zinc finger protein 551	KRAB	16	+	654	62885217	62892991	
19	19q13.4	19.13bv		ZNF154	Zinc finger protein 154	KRAB	10	-	449	62904779	62912391	
19	19q13.43	19.13bw		ZNF671	Zinc finger protein 671	KRAB	10	-	534	62922931	62930795	
19	19q13.43	19.13bx		ZNF776	Zinc finger protein 776	-	10	+	476	62954023	62961337	
19	19q13.43	19.13by		ZNF586	Zinc finger protein 586	KRAB	10	+	402	62972850	62983757	
19	19q13.43	19.13bz		ZNF552	Zinc finger protein 552	KRAB	8	-	407	63010264	63018093	
19	19q13.43	19.13ca		ZNF587	Zinc finger protein 587	KRAB	13	+	575	63053081	63071765	
19	19q13.43	19.13cb	Ψ	ZNF814	Zinc finger protein 814	-	-	-	-	63075442	63092226	
19	19q13.43	19.13cc		ZNF417	Zinc finger protein 417	KRAB	13	-	575	63110026	63119756	

19	19q13.43	19.13cd	ZNF418	Zinc finger protein 418	KRAB	16 - 676	63125064	63138552	
19	19q13.43	19.13ce	ZNF256	Zinc finger protein 256	KRAB	15 - 627	63144013	63150889	
19	19q13.4	19.13cf	ZNF606	Zinc finger protein 606	KRAB	16 - 792	63180252	63206526	
19	19q13.43	19.13cg	ZSCAN1	Zinc finger and SCAN domain containing 1	SCAN	3 + 408	63237246	63257811	
19	19q13.4	19.13ch	ZNF135	Zinc finger protein 135	KRAB	16 + 658	63262424	63272588	
19	19q13.43	191.13ci	ZNF447	Zinc finger protein 447	SCAN	2 - 510	63287018	63301389	
19	19q13.43	19.13cj	ZNF329	Zinc finger protein 329	-	12 - 541	63329507	63353960	
19	19qter	19.13ck	ZNF274	Zinc finger protein 274	KSK	5 + 653	63386208	63416739	
19	19q13.43	19.13cl	ZNF544	Zinc finger protein 544	KRAB	13 + 715	63432646	63467285	
19	19q13.43	19.13cm	ZNF8	Zinc finger protein 8	KRAB	7 + 575	63482130	63499066	
19	19q13.43	19.13cn	HKR2	GLI-Kruppel family member HKR2	SCAN	8 + 491	63530197	63545510	
19	19q13.43	19.13co	ZNF497	Zinc finger protein 497	-	14 - 498	63557537	63565932	
19	19q13.43	19.13cp	LOC116412	Hypothetical protein BC012365	-	8 - 531	63570549	63584224	
19	19q13.43	19.13cq	ZNF584	Zinc finger protein 584	KRAB	8 + 421	63611875	63621506	
19	19q13.4	19.13cr	ZNF132	Zinc finger protein 132	KRAB	18 - 706	63635994	63643401	
19	19q13.43	19.13cs	ZNF324B	Zinc finger protein 324B	KRAB	9 + 544	63654783	63661011	
19	19q13.43	19.13ct	ZNF324	Zinc finger protein 324	KRAB	9 + 553	63670275	63676577	
19	19q13.43	19.13cu	ZNF446	Zinc finger protein 446	SCAN	3 + 450	63679607	63684409	
19	19q13.43	19.13cv	ZNF499	Zinc finger protein 499	BTB	4 - 511	63716709	63722733	
19	19q13.2-q13.4	19.13cw	ZNF42	Zinc finger protein 42	SCAN	13 - 734	63765096	63776754	
19	19q13.2	19.13cx	ZNF93	Zinc finger protein 93	KRAB	17			
20	20p13	-	SCRT2	Sratch homolog 2, zinc finger protein	-	5 - 307	590240	604823	
20	20p13	-	ZNF343	Zinc finger protein 343	KRAB	12 - 599	2410463	2437778	
20	20pter-q11.23	20.1a	ZNF339	Zinc finger protein 339	-	4 - 275	17952796	17986521	
20	20p11.23-p11.22	20.1b	ZNF133	Zinc finger protein 133	KRAB	15 + 653	18217157	18245640	
20	20p12.3-p11.21	-	ZNF336	Zinc finger protein 336	BTB	10 + 711	23293021	23301683	
20	20p11.21	-	ZNF337	Zinc finger protein 337	KRAB	20 - 751	25602851	25625469	
20	20q11.21	-	PLAGL2	Zinc finger protein PLAGL2	-	6 - 496	30243968	30259192	
20	20q11.22	-	ZNF341	Zinc finger protein 341	-	12 + 847	31783469	31843736	
20	20q11.1-q11.23	-	SCAND1	SCAN domain containing protein 1	SCAN	- - 179	34004960	34005842	
20	20q11.21-q13.12	20.2a	ZNF335	Zinc finger protein 335	-	13 - 1342	44010699	44034240	
20	20q13.12	20.2b	ZNF663	Zinc finger protein 663	-	1 - 106	44476274	44521322	
20	20q13.12	20.2c	ZNF334	Zinc finger protein 334	KRAB	14 - 680	44563114	44575601	
20	20q13.1-q13.2	-	SNAI1	Zinc finger protein snail homolog	-	4 + 264	48032934	48038830	
20	20q13.13-q13.2	20.3a	SALL4	Sal-like 4	-	7 - 1053	49833988	49852421	
20	20q13.2	20.3b	ZFP64	Zinc finger protein 64 homolog	-	13 - 645	50133957	50241931	

20	20q13.2	-		ZNF217	Zinc finger protein 217	-	7	-	1048	51617017	51633043	
20	20q13.31	-		CTCFL	Zinc finger protein CTCF-T	-	11	-	663	55505630	55533560	
20	20q13.33	-		BTBD4	BTB (POZ) domain containing 4	BTB	2	-	589	61846322	61907300	
21	21q21.1	-	Ψ	ZNF299P	zinc finger protein 299 pseudogene	-	-	-	-	23383998	23385465	
21	21q22.3	21.1a		PRDM15	Zinc finger protein 298	SET	13	-	1507	42091454	42172660	
21	21q22.3	21.1b		ZNF295	Zinc finger protein 295 (ZBTB21)	BTB	6	-	1066	42280009	42303519	
22	22p	-		ZNF73	Zinc finger protein 73	-	8	+	326	450078	452598	
22	22q11.1	-	Ψ	LOC343927	similar to zinc finger protein 91	-	-	+	-	15007571	15054750	
22	22q11.1	-	Ψ	LOC391288	Similar to zinc finger protein 532	-	-	-	-	15713058	15761696	
22	22q11.21	-		ZNF74	Zinc finger protein 74	KRAB	12	+	572	19078478	19091970	
22	22q11.21	-		HIC2	ZBTB30	BTB	5	+	615	20101693	20135748	
22	22q11.2	22.1a		SUHW2	Zinc finger protein 279	-	1	-	543	21168767	21193505	
22	22q11.22	22.1b		SUHW1	Zinc finger protein 280	-	1	-	542	21198060	21204613	
22	22q11.23	-		ZNF70	Zinc finger protein 70	-	11	-	446	22413772	22423279	
22	22q12.2	-		ZNF278	Zinc finger protein 278	BTB	7	-	687	30051790	30072249	
22	22q11.2	-		ZNF69	Zinc finger protein 69	KRAB	-	+	323	30751319	30765974	
X	Xp21.3	-		ZFX	Zinc finger protein X-linked	-	13	+	805	24077824	24142549	
X	Xp11.3	-		ZNF674	Zinc finger protein 673	KRAB	11	-	581	46243490	46289820	
X	Xp11.2	X.1a		ZNF157	Zinc finger protein 157	KRAB	12	+	506	47114926	47158338	
X	Xp11.23	X.1b		ZNF41	Zinc finger protein 41	KRAB	18	-	779	47191347	47227289	
X	Xp11.23	X.1c		ZNF81	Zinc finger protein 81	KRAB	13	+	661	47581245	47666550	
X	Xp22.11-p11.23	X.1d		ZNF21	Zinc finger protein 21	KRAB	15	-	639	47719194	47748321	
X	Xp11.1-11.3	X.1e		ZNF630	Zinc finger protein 630	KRAB	13	-	657	47802547	47815739	
X	Xp11.23	-	Ψ	LOC139163	Similar to Sal-like protein 1	-	-	-	-	49316398	49325222	
X	Xp11.21	-		KLF8	Kruppel-like factor 8	-	3	+	359	56275632	56328255	
X	Xp11.21	X.2a		ZXDB	Zinc finger, X-linked, duplicated B	-	9	+	803	57634994	57640635	
X	Xp11.1	X.2b		ZXDA	Zinc finger, X-linked, duplicated A	-	9	-	799	57950922	57953792	
X	Xq13.2	-	Ψ	LOC260337	Zinc finger protein Np97 pseudogene	-	-	+	-	73242135	73245219	
X	Xq21.1-q21.2	-		ZNF6	Zinc finger protein 6	-	11	+	761	84385694	84415024	
X	Xq23	-		ZBTB33	Zinc finger and BTB domain containing 33	BTB	3	+	672	119268635	119276279	
X	Xq26.3	X.3a		ZNF75	Zinc finger protein 75	SCAN-KRAB	5	-	510	134247385	134305623	
X	Xq26.3	X.3b		ZNF449	Zinc finger protein 449	SCAN	7	+	518	134306387	134325004	
X	Xq26.2	-		ZIC3	Zinc finger protein of the cerebellum 3	-	4	+	467	136476012	136481925	
X	Xq28	X.4a		ZNF275	Zinc finger protein 275	-	11	+	376	152262060	152270249	

X	Xq28	X.4b		LOC139735	Similar to zinc finger protein 92	KRAB	8	+	495	152336765	152340280
Y	Yp11.3	-		ZFY	zinc finger protein, Y-linked	-	12	+	801	2863546	2909891
Y	Yq11.223	-	Ψ	ZNF381P	zinc finger protein 381, Y-linked pseudogene	-	-	+	-	23605570	23606703
Y	Yq11.223	Y.1a	Ψ	LOC392603	similar to Zinc finger protein 43 (Zinc protein HTF6)	-	-	+	-	25235272	25240390
Y	Yq11.23	Y.1b	Ψ	LOC442486	similar to zinc finger protein 91	-	-	-	-	25540738	25545902

For each C2H2-ZNF in the dataset

[1] *Chromosome number*

[2] *Position on the chromosome*

[3] *The cluster number to which the gene belongs '-' for a gene found as a singleton rather than a cluster. If the gene belongs to a cluster, the cluster number is indicated : The first number indicates the chromosome number, The second number indicates the number of the cluster on the chromosome. For example, a cluster number 1.1 indicates Chromosome1.Cluster1*

[4] *Status as a pseudogene as reported in Genbank. ' Ψ' - Identified as a pseudogene*

[5 6] *The name of the C2H2-ZNF and its description*

[7] *The domain associated with the C2H2-ZNF : KRAB, SCAN, SCAN-KRAB, BTB, HOMEO, SET an without an encoded conserved N-terminal domain (Φ)*

[8] *The number of zinc finger motifs present*

[9] *The orientation*

[10] *The amino acid sequence length*

[11 12] *The start and stop of translation*

Supplementary Table S2

Comprehensive summary of the organization of all C2H2-ZNF found as singletons or in clusters on each human chromosome and classified with respect to the various C2H2-ZNF sub-families.

				C2H2-ZNF sub-families													
	Total		No. C2H2-ZNF in	KRAB		SCAN-KRAB		SCAN		BTB		SET		HOMEO		[1] ø	
Chr	No. C2H2-ZNF	No. Clusters	Clusters	[1] S	[1] C	S	C	S	C	S	C	S	C	S	C	S	C
1	36	6	17	1	9		1	1		7		1				9	7
2	17	3	7	1	2											9	5
3	30	6	20	2	5		3			2	1					6	11
4	10	1	4	1	2					1						4	2
5	15	1	6	1	5											8	1
6	28	3	16	1	2		2		7	3	2					8	3
7	47	7	41	1	27		2									5	12
8	30	4	16	3	6					1				1	2	10	8
9	23	4	10	5	3	1				3	2					4	5
10	22	3	13		8											9	5
11	15	2	4	3	1	1	1			3						4	2
12	15	1	7	2	6											6	1
13	5	2	4													1	4
14	10	2	4	1							2				1	5	1
15	12	3	8					1	2		1					3	5
16	33	6	27		10	1	2		3							5	12
17	13	2	5	1	1	1	1		1	1						5	2
18	14	2	6	1					3	2						5	3
19	289	13	279	3	185		1		13		4			1		6	76
20	18	3	7	2	2			1		2						6	5
21	3	1	2								1		1			1	
22	10	1	2	2						2						4	2
X	19	4	11	1	6		1		1	1						6	3
Y	4	1	2													2	2
Total	**718**	**81**	**518**	**32**	**280**	**4**	**14**	**3**	**30**	**28**	**13**	**1**	**1**	**2**	**3**	**130**	**177**

[1] *No. of C2H2-ZNF without an encoded conserved N-terminal domain (ø), found as singletons (S) or in C2H2-ZNF clusters (C)*

Supplementary Table S3

Gene organization of the 81 human C2H2-ZNF clusters.

Cluster[1]	Position	No. Of C2H2-ZNF	Clusters with Solely C2H2-ZNF	Order of the genes from the different C2H2-ZNF subfamilies[2]	Cluster composition[4]
1.1	1p36.11	2	No	ø/ø	Pure
1.2	1p34.2	3	No	K/K/K	Pure
1.3	1p34.1	2	No	ø/ø	Pure
1.4	1q42.13	2	No	ø/K	Mixed
1.5	1q44	6	No	K/K/K/K/K/S-K	Mixed
1.6	1q44	2	Yes	ø/ø	Pure
2.1	2p23.3	2	No	ø/ø	Pure
2.2	2q11.1	3	Yes	K/K/ø	Mixed
2.3	2q13	2	No	ø/ø	Pure
3.1	3p22.1	3	Yes	ø/K/K	Mixed
3.2	3p22.1	3	No	ø/K/K	Mixed
3.3	3p21.32	8	Yes	S-K/K/S-K/ø/S-K/ø/ø/ø	Mixed
3.4	3q13.2	2	No	ø/B	Mixed
3.5	3q24	2	Yes	ø/ø	Pure
3.6	3q26.32	2	No	ø/ø	Pure
4.1	4p16.3	4	Yes	K/ø/ø/K	Mixed
5.1	5q35.3	6	No	K/K/ø/K/K/K	Mixed
6.1	6p22.1	2	No	ø/K	Mixed
6.2	6p22.1	12	No	S/S/S-K/ø/S/S-K/ø/S/S/S/S/K	Mixed
6.3	6p21.3	2	No	B/B	Pure
7.1	7p22.1	2	No	ø/K	Mixed
7.2	7p22.1	4	No	ø/K/ø/K	Mixed
7.3	7p11.2	5	No	ø/ø/K/ø/K	Mixed
7.4	7q11.21	9	No	K/K/K/ø/ø/K/K/ø/K	Mixed
7.5	7q22.1	8	No	K/S-K/S-K/ø/ø/K/K/K	Mixed
7.6	7q22.1	3	Yes	K/K/K	Pure
7.7	7q36.1	10	No	K/K/K/K/K/K/K/ø/K-K	Mixed
8.1	8p23.1	2	No	ø/ø	Pure
8.2	8q24.13	2	No	H/H	Pure
8.3	8q24.3	5	No	ø/ø/ø/ø/K	Mixed
8.4	8q24.3	7	No	ø/K/K/K/K/ø/K	Mixed
9.1	9q22.32	4	No	ø/K/K/ø	Mixed
9.2	9q31.2	2	No	ø/ø	Pure
9.3	9q32	2	Yes	ø/K	Mixed
9.4	9q33.2	2	Yes	B/B	Pure
10.1	10p11.21	6	Yes	K/ø/ø/K/K/K	Mixed
10.2	10q11.21	3	No	ø/K/K	Mixed
10.3	10q11.21	4	Yes	K/ø/K/ø	Mixed
11,1	11p15.4	2	Yes	S-K/K	Mixed
11.2	11q12.2	2	Yes	ø/ø	Pure
12.1	12q24.33	7	Yes	K/K/K/K/K/K/ø	Mixed
13.1	13q22.1	2	No	ø/ø	Pure
13.2	13q32.3	2	Yes	ø/ø	Pure
14.1	14q11.2	2	Yes	H/ø	Mixed
14.2	14q23.3	2	Yes	B/B	Pure
15.1	15q24	2	No	B/ø	Mixed
15.2	15q25.3	3	No	S/S/ø	Mixed
15.3	15q26.1	3	No	ø/ø/ø	Pure
16.1	16p13.3	9	No	S/K/S-K/ø/S-K/K/ø/S/ø	Mixed

	16.2	16p13.3	2	No	ø/S	Mixed
	16.3	16p11.2	10	No	[3] ø/ø/5K/ø/ø/ø	Mixed
	16.4	16p11.2	2	No	ø/K	Mixed
	16.5	16q22	2	Yes	K/K	Pure
	16.6	16q24.2	2	Yes	ø/ø	Pure
	17.1	17p13.2	3	No	ø/S/ø	Mixed
	17.2	17p11.2	2	No	S-K/K	Mixed
	18.1	18q12	4	Yes	S/ø/S/S	Mixed
	18.2	18q23	2	No	ø/ø	Pure
	19.1	19q13.2	3	Yes	K/K/K	Pure
	19.1	19p13.3	2	No	ø/B	Mixed
	19.11	19q13.31	24	Yes	[3] 19K/5ø	Mixed
	19.12	19q13.41	43	No	[3] 30K/13ø	Mixed
	19.13	19q13.43	76	No	[3] 43K/1S-K/12S/1B/19ø	Mixed
	19.2	19p13.3	5	Yes	K/K/K/K/K	Pure
	19,3	19p13.3	2	No	ø/B	Mixed
	19.4	19p13.2	2	No	K/ø	Mixed
	19.5	19p13.2	14	No	[3] 9K/5ø	Mixed
	19.6	19p13.2	28	No	[3] 21K/7ø	Mixed
	19.7	19p13.11	40	No	[3] 28K/12ø	Mixed
	19.8	19q13.11	8	No	[3] 3K/S/3ø	Mixed
	19.9	19q13.12	32	No	[3] 23K/1B/8ø	Mixed
	20.1	20p11.23	2	No	K/ø	Mixed
	20.2	20q13.12	3	No	ø/ø/K	Mixed
	20.3	20q13.2	2	Yes	ø/ø	Pure
	21.1	21q22.3	2	No	Se/B	Mixed
	22.1	22q11.22	2	No	ø/ø	Pure
	X.1	Xp11.23	5	No	K/K/K/K/K	Pure
	X.2	Xp11.1	2	No	ø/ø	Pure
	X.3	Xq26.3	2	Yes	S-K/S	Mixed
	X.4	Xq28	2	No	ø/K	Mixed
	Y.1	Yq11.23	2	No	ø/ø	Pure
Total	**81 Clusters**		**518 C2H2-ZNF**	**Yes= 25 ; No = 56**		**Pure= 29 ; Mixed= 52**

[1] *For the cluster name, the first number correspond to the chromosome on which the cluster is found and the second to the number attributed to the cluster.*

[2] *Sequential order of the genes from the different C2H2-ZNF subfamilies such as KRAB (K), SCAN (S), SCAN-KRAB (S-K), SET (Se), HOMEO (H) and without an encoded conserved N-terminal domain (ø)*

[3] *For the very large clusters, the number of C2H2-ZNF from each subfamilies is specified (eg: 23 K means that 23 consecutive genes from the KRAB-C2H2-ZNF subfamily are found in the cluster).*

[4] *'Pure' = The cluster is composed of C2H2-ZNF from a single subfamily; 'Mixed' = different subfamilies of C2H2-ZNF are present in the cluster.*

Note: 'Pure' clusters with solely tandemly repeated C2H2-ZNF are in grey

Supplementary Table 4

Comprehensive catalog of the C2H2-ZNF genes from the 81 human clusters and their syntenic counterparts from other mammalian genomes (Chimpanzee, Mouse, Rat and Dog)

Human Cluster 1.1				*Chimpanzee*				*Mouse*				*Rat*				*Dog*			
h chr1				*p chr1*				*m chr4*				*r chr5*				*c chr2*			
	D	F	O		D	F	O		D	F	O		D	F	O		D	F	O
FG TRIM63, PDIK1L				*FG TRIM63, PDIK1L*				*FG Trim63, Pdik1l*				*FG Trim63, Pdik1l*				*FG TRIM63, PDIK1L*			
GRAP1				*GRAP1*				*Grap1*				*Grap1*				*GRAP1*			
ZNF593	-	1	+	**ZNF593**	-	1	+	**Zfp593**	-	1	-	**Zfp593_pred**	-	1	-	**LOC487360**	-	1	+
ZNF683	-	4	-	**ZNF683**	-	4	-									**LOC487354**	-	4	-
FG LIN28, DHDD3				*FG LIN28, DHDD3*				*FG Lin28, Dhdds*				*FG Lin28, Dhdds*				*FG LIN28, DHDD3*			
HMGN2				*HMGN2*				*Hmgn2*				*Hmgn2*				*HMGN2*			

Human Cluster 1.2				*Chimpanzee*				*Mouse*				*Rat*				*Dog*			
h chr1				*p chr1*				*m chr4*				*r chr5*				*c chr15*			
FG ZMPSTE24,COL9A2	D	F	O	*FG ZMPSTE24, LOC456791*	D	F	O	*FG Zmpste24,Col9a2*	D	F	O	*FG Zmpste24,Col9a2*	D	F	O	*FG: LOC607622,*	D	F	O
LOC388621,SMAP1L				*LOC748691, SMAP1L*				*Smap1l*				*Smap1l*				*LOC607609 LOC482458*			
ZNF643	K	9	+	**ZNF643**	-	9	+	**Zfp69**	K	9	-					**LOC482457**	K	17	-
ZNF642	K	9	+	**ZNF642**	K	9	+									**LOC482454**	K	8	-
ZNF684	K	8	+	**LOC456797**	K	8	+												
FG LOC728633, RIMS3				*FG LOC748746, RIMS3*				*FG Rims3,Nfyc*				*FG Rims3,Nfyc*				*FG:LOC607486*			
NFYC,KCNQ4				*LOC456799, KCNQ4*				*Kcnq4*				*Kcnq4*				*LOC475312, LOC482451*			

Human Cluster 1.3				*Chimpanzee*				*Mouse*				*Rat*				*Dog*			
h chr1				*p chr1*				*m chr4*				*m chr5*				*c chr15*			
FG ATP6VOB, BAGALT	D	F	O	*FG ATP6VOB, BAGALT*	D	F	O	*FG Atp6vob,Bagalt2*	D	F	O	*FG Atp6vob,Bagalt2*	D	F	O	*FG: ATP6VOB, BAGALT*	D	F	O
2,CCD24,SLC6A9				*2,CCD24,SLC6A9*				*Slc6a9*				*Slc6a9*				*2,CCD24,SLC6A9*			
KLF17	-	3	+	**KLF17**	-	3	+	**Klf17**	-	3	-	**Zfp393_pred**	-	3	-	**KLF17**	-	3	+
LOC128208	-	5	-																
FG DMAP1, PRNPIP				*FG DMAP1, PRNPIP*				*FG Dmap1, Pmpip*				*FG Dmap1, Pmpip*				*FG DMAP1, PRNPIP*			
TMEM53				*TMEM53*				*Tmem53*				*Tmem53*				*TMEM53*			

Human Cluster 1.4				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr11				r chr10				c chr14			
FG CDC42BPA	D	F	O	FG CDC42BPA	D	F	O	FG Cdc42bpa	D	F	O	FG Cdc42bpa	D	F	O	FG CDC42BPA	D	F	O
CTD-29DI				CTD-29DI				Ctd-29di				Ctd-29di				CTD-29DI			
ZNF678	-	15	+	LOC469695	-	15	+												
gm127	K	3	-																
FG JMJD4, MPN2				FG JMJD4, MPN2				FG Jmjd4				FG Jmjd4				FG JMJD4, MPN2			
WNT9A				WNT9A				Wnt9a				Wnt9a				WNT9A			

Human Cluster 1.5				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr8				r chr10				c chr11			
FG TFB2M,SCCPDH	D	F	O	FG TFB2M,SCCPDH	D	F	O	FG Tfb2m, Sccpdh	D	F	O	FG Tfb2m, Sccpdh	D	F	O	FG LOC480102	D	F	O
LOC149134,AHCTF1				LOC149134,AHCTF1				Gm1305, Ahctf1				Gm1305, Ahctf1				LOC480105			
ZNF695	K	-	-	ZNF670	K	9	-	BC050078	K	12	-	Zfp496	S	5	-	LOC490575	S	5	-
ZNF670	K	9	-	LOC457877	K	9	-	Zfp496	SK	5	-								
ZNF669	K	9	-	ZNF124	K	7	-												
ZNF124	K	7	-	LOC457880	K	10	-												
LOC729806	K	10	-	ZNF496	SK	5	-												
ZNF496	SK	5	-																
FG CIAS1, OR2B11				FG CIAS1, OR2B11				FG Cias1, Olfr222				FG Cias1, Olfr222				FG CIAS1, OR2B11			
OR2W5				OR2W5				LOC668157				LOC668157				OR2W5			

Human Cluster 1.6				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr11				r chr10				c chr16			
FG OR2T227, OR5BU1	D	F	O	FG OR2T227, OR5BU1	D	F	O	FG Or2t227	D	F	O	FG Or2t227	D	F	O	FG OR2T227, OR5BU1	D	F	O
SH3BP5L				SH3BP5L				Sh3bp5l				Sh3bp5l				SH3BP5L			
ZNF672	-	13	+	ZNF672	-	13	+	Zfp672	-	13	-	Zfp672	-	13	-	LOC482699	-	12	+
ZNF692	-	5	-	ZNF692	-	5	-	Zfp692	-	5	+	Zfp692	-	5	+	LOC482698	-	3	-
FG PGBD2				FG PGBD2				FG Pgbd2				FG Pgbd2				FG PGBD2			

Human Cluster 2.1				Chimpanzee				Mouse				Rat				Dog			
h chr2				p chr2				m chr5				r chr6				c chr17			
FG MPV17, GTF3C2	D	F	O	FG MPV17, GTF3C2	D	F	O	FG Gtf3c2, Eif2b4	D	F	O	FG Gtf3c2, Eif2b4	D	F	O	FG MPV17, GTF3C2	D	F	O
EIF2B4, SNX17				EIF2B4, SNX17				Snx17				Snx17				EIF2B4, SNX17			
ZNF513	-	7	-	ZNF513	-	7	-	Zfp513	-	7	-	Zfp513	-	7	+	LOC483012	-	7	-
ZNF512	-	2	+	ZNF512	-	2	+	Zfp512	-	2	+	RGD156152	-	2	-	LOC608296	-	1	+
FG CCDC121, XAB1				FG CCDC121, XAB1				FG Xab1, Supttl				FG Xab1, Supttl				FG CCDC121, XAB1			
SUPTTL				SUPTTL												SUPTTL			

Human Cluster 2.2				Chimpanzee				Mouse				Rat				Dog			
h chr2				p chr2				m chr2				r chr3				c chr17			
FG:TEKT4, LOC442048	D	F	O	FG:TEKT4, MAL	D	F	O	FG: Tekt4, Mal	D	F	O	FG: Tekt4, Mal	D	F	O	FG: LOC483056	D	F	O
MAL, MRPS5				MRPS5				Mrps5				Mrps5				MAL,LOC475746			
ZNF514	K	7	-	LOC470431	K	7	-	Zfp661	K	9	-	Zfp661	K	9	-	LOC611670	K	7	+
ZNF2	K	9	+	ZNF2	K	9	+									LOC483055	K	28	-
LOC344065	-	10	+	LOC459400	-	8	+												
FG: PROM2, KCNIP3				FG: PROM2, KCNIP3				FG Prom2,Kcnip3				FG Prom2,Kcnip3				FG: LOC611664			
FAHD2A				FAHD2A				Fahd2a				Fahd2a				LOC609135,LOC425745			

Human Cluster 2.3				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1															
FG:ADRA2B, ASTL	D	F	O	FG:ADRA2B, ASTL	D	F	O	FG:ADRA2B, ASTL	D	F	O	FG:ADRA2B, ASTL	D	F	O	FG:ADRA2B, ASTL	D	F	O
DUSP2				DUSP2				DUSP2				DUSP2				DUSP2			
LOC343938	-	-	-																
LOC442041	-	-	-																
FG:NCAPH				FG:NCAPH				FG:NCAPH				FG:NCAPH				FG:NCAPH			
LINCR				LINCR				LINCR				LINCR				LINCR			

Human Cluster 3.1				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr9				r chr8				c chr23			
FG:MYRIP,EIF1B	D	F	O	FG: EIF1B,ENTPD3	D	F	O	FG:MYRIP,EIF1B	D	F	O	FG:MYRIP,EIF1B	D	F	O	FG:MYRIP,EIF1B	D	F	O
ENTPD3, RPL14				LOC735759,LOC735578				Entpd3, Rpl14				Entpd3, Rpl14				ENTPD3, RPL14			
ZNF619	-	10	+	LOC470797	2K	26	+												
ZNF620	K	8	+	LOC470799	-	-	-												
ZNF621	K	7	+																
FG: LOC645807, LOC651625				FG: MRPS31P1,				FG: LOC645807, LOC651625				FG: LOC645807, LOC651625				FG: LOC645807,			
MRPS31P1, CTNN31				CTNN31				MRPS31P1, CTNN31				MRPS31P1, CTNN31				MRPS31P1, CTNN31			

Human Cluster 3.2				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr9				r chr8				c chr23			
FG:VIPR1, SEC22C	D	F	O	FG:VIPR1, SEC22C	D	F	O	FG: Sec22c, Ss18l2	D	F	O	FG: Sec22c, Ss18l2	D	F	O	FG:VIPR1, SEC22C	D	F	O
SS18L2, NKTR				SS18L2, NKTR				Nktr				Nktr				SS18L2, NKTR			
ZNF651	-	8	+	ZNF662	K	8	+	Zfp651	B	8	+	RGD1562434	B	8	+				
ZNF662	K	8	+																
LOC339903	K	-	+																
FG:SNRK, TMEM16K				FG:SNRK, TMEM16K				FG: Snrk, Tmem16k				FG: Snrk, Tmem16k				FG:SNRK, TMEM16K			
ABHD5				ABHD5				Abhd5				Abhd5				ABHD5			

Human Cluster 3.3				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr9				r chr8				c chr23			
FG SNRK TMEM16K	D	F	O	FG SNRK TMEM16K	D	F	O	FG Snrk Tmem16K	D	F	O	FG Snrk Tmem16K	D	F	O	FG SNRK TMEM16K	D	F	O
ABHD5 FLJ36157				ABHD5 FLJ36157				Abhd5 LOC666218				Abhd5 LOC666218				ABHD5			
ZNF445	SK	14	-	ZNF445	SK	4	-	Zfp445	SK	12	-	RGD1559144	SK	12	-	ZNF445	SK	14	-
LOC285346	K	12	-	ZNF167	SK	23	+	Zfp167	S	11	+	Znf167	S	22	+				
ZNF167	SK	13	+	ZNF35	-	11	+	Zfp660	-	10	+	Zfp105	-	11	+				
ZNF660	-	10	+	ZNF502	-	14	+	Zfp105	-	11	+								
ZNF197	SK	22	+	LOC470807	-	9	+												
ZNF35	-	11	+																
ZNF502	-	14	+																
ZNF501	-	9	+																
FG KIAA1143 KIF15				FG KIAA1143 KIF15				FG 1110059G10RIK Kif15				FG 1110059G10RIK Kif15				FG TMEM42 KIF15			
TMEM42 TGM4				TMEM42 TGM4				Tmem42 Tgm4				Tmem42 Tgm4							

Human Cluster 3.4				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr16				r chr11				c chr17			
FG QTRTD1	D	F	O	FG QTRTD1	D	F	O	FG Qtrtd1	D	F	O	FG Qtrtd1	D	F	O	FG QTRTD1	D	F	O
DRD3				DRD3				Drd3				Drd3				DRD3			
ZNF80	-	7	-	LOC470886	-	7	-	Zbtb20	B	5	+	Zbtb20	B	5	-				
ZBTB20	B	5	-	ZBTB20	B	5	-												
FG GAP43				FG GAP43				FG Gap43				FG Gap43				FG GAP43			
LSAMP				LSAMP				Lsamp				Lsamp				LSAMP			

Human Cluster 3.5				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr9				r chr8				c chr23			
FG PLSCR2 PLSCR1	D	F	O	FG PLSCR2 PLSCR1				FG Plscr2				FG Plscr2				FG PLSCR2 PLSCR1			
PLSCR5				PLSCR5				Plscr5				Plscr5				PLSCR5			
ZIC4	-	4	-	LOC470956	-	4	-	Zic4	-	4	+	Zic4	-	4	+	LOC485704	-	4	-
ZIC1	-	4	+	LOC460759	-	4	+	Zic1	-	4	-	Zic1	-	4	-	LOC611554	-	4	+
FG RPL38P1 AGTR1				FG RPL38P1 AGTR1				FG Rpl38p1				FG Rpl38p1				FG RPL38P1 AGTR1			
CPB1				CPB1				Cpb1				Cpb1				CPB1			

Human Cluster 3.6				Chimpanzee				Mouse				Rat				Dog			
h chr3				p chr3				m chr3				r chr2				c chr34			
FG ZMAT3, PIK3CA	D	F	O	FG ZMAT3, PIK3CA	D	F	O	FG Zmat3, Pik3ca	D	F	O	FG Zmat3, Pik3ca	D	F	O	FG ZMAT3, PIK3CA	D	F	O
KCNMB3				KCNMB3												KCNMB3			
WIG1	-	2	-	WIG1	-	2	-	Wig1	-	2	-	Wig1	-	2	-	WIG1	-	2	-
ZNF639	-	5	+	ZNF639	-	5	+	Zfp639	-	5	+	Zfp639	-	5	+	ZNF639	-	5	+
FG MFN1,				FG MFN1,				FG Mfn1				FG Mfn1				FG MFN1,			
GNB4				GNB4				Gnb4				Gnb4				GNB4			

Human Cluster 4.1				Chimpanzee				Mouse				Rat				Dog			
h chr4				p chr4				m chr5				r chr1				c chr23			
FG LOC737253	D	F	O	FG LOC737253	D	F	O	FG LOC737253	D	F	O	FG LOC737253	D	F	O	FG LOC737253	D	F	O
ZNF595	K	18	+	ZNF595	K	18	+												
MGC26356	-	6	+	ZNF718	K	11	+												
LOC654254	-	-	-	LOC461038	-	-	-												
ZNF141	K	11	+	ZNF721	K	28	-												
FG PIGG				FG PIGG				FG Pigg				FG Pigg				FG PIGG			
LOC461041, ATP5I				LOC461041, ATP5I				Atp5I				Atp5I				LOC461041, ATP5I			

Human Cluster 5.1				Chimpanzee				Mouse				Rat				Dog			
h chr5				p chr5				m chr11				r chr10				c chr11			
FG COL23A1,MRPL50P3	D	F	O	FG COL25A1, CLK4	D	F	O	FG:Col23a1	D	F	O	FG:Col23a1	D	F	O	FG:LOC607556	D	F	O
CLK4								Clk4				Clk4				LOC474645			
ZNF354A	K	13	-	ZNF354A	K	13	-	Zfp354a	K	13	+	Zfp354a	K	13	+	ZNF594	K	35	-
ZNF354B	K	13	+	ZNF354B	K	13	+	Zfp354b	K	13	-	RGD1560080	KK	13	-	LOC474650	K	24	-
ZFP2	-	13	+	ZNF71	K	-	+	Zfp2	-	9	-	RGD156327	K	-	-				
ZNF454	K	12	+	ZFP2	K	11	+	Zfp454	K	12	-	Zfp354c	K	11	+				
DKFZp686E2433	K	13	+	ZNF454	-	12	+	9630041N07Rik	K	13	-								
ZNF354C	K	11	+					Zfp354c	K	11	-								
FG ADAMTS2				FG GRM6, ADAMTS2				FG Adamts2				FG Adamts2				FG LOC481453			
LOC391859								Rufy1				Rufy1				LOC482454			

Human Cluster 6.1				Chimpanzee				Mouse				Rat				Dog			
h chr6				p chr6				m chr13				r chr10				c chr35			
FG POM121167, FKSG83	D	F	O	FG POM121167, FKSG83	D	F	O	FG Pom12162	D	F	O	FG Pom12162	D	F	O	FG POM121167, FKSG83	D	F	O
RP1-153G14 3				RP1-153G14 3				Fksg83				Fksg83				RP1-153G14 3			
LOC346157	-	9	+	LOC472231	-	9	+	Zfp184	K	19	+	Zfp184	K	19	+	LOC478746	-	9	+
ZNF184	K	19	-	ZNF184	K	19	-												
FG HISTIH2A1				FG HISTIH2A1				FG HISTIH2A1				FG HISTIH2A1				FG HISTIH2A1			
HIST1H3H,HIST1H2AJ				HIST1H3H,HIST1H2AJ				HIST1H3H,HIST1H2AJ				HIST1H3H,HIST1H2AJ				HIST1H3H,HIST1H2AJ			

Human Cluster 6.2				Chimpanzee				Mouse				Rat				Dog			
h chr6				p chr6				m chr13				r chr17				c chr35			
FG OR2B7P,OR2B8P	D	F	O	FG LOC471910	D	F	O	FG Olfr1370,Olfr1369	D	F	O	FG Olfr1370,Olfr1369	D	F	O	FG LOC611561,	D	F	O
OR1F12P				LOC462510				Olfr42,Olfr1368				Olfr42,Olfr1368				LOC488312,Cor2y2			
ZNF165	S	6	+	ZNF165	-	-	-	Zfp96	S	8		Zfp192	SK	10	+	LOC611561	-	-	+
ZNF435	S	4	+	LOC471912	-	-	-	Zfp306	S	7		Znf307	SK	3	-	LOC488316	-	8	+
ZNF192	SK	9	+	LOC471914	-	-	-	Zfp187	-	7		Znf187	SK	7	+	LOC488318	S	15	+
LOC222701	-	-	+	ZNF193	S	5	+	RP23-298F22.2	S	7		Zfp307	S	7	+				
ZNF193	S	5	+	ZNF307	-	7	-	Zfp102	SK	9		Zfp96	-	8	-				
ZNF307	SK	7	-	ZNF306	S	7	+												
ZNF187	-	8	+	ZNF96	S	18	-												
ZNF323	S	6	-	ZNF390	S	5	-												
ZNF306	S	7	+	ZNF452	-	-	-												
ZNF305	S	11	-	ZNF311	K	14	-												
ZNF452	S	-	-																
ZNF311	K	14	-																
FG OR2W1,OR2P1P				FG LOC471926				FG Olfr1366,Olfr1365				FG Olfr1366,Olfr1365				FG LOC488327			
LOC646260				LOC471927				Olfr1364				Olfr1364				LOC488328			

Human Cluster 6.3				Chimpanzee				Mouse				Rat				Dog			
h chr6				p chr6				m chr17				r chr20				c chr12			
FG WDR46, PFDN6	D	F	O	FG WDR46, PFDN6	D	F	O	FG Wdr46, Pfdn6	D	F	O	FG Wdr46, Pfdn6	D	F	O	FG WDR46, PFDN6	D	F	O
RGL2, TAPBP				RGL2, TAPBP				Rgl2,Tapbp				Rgl2,Tapbp				RGL2, TAPBP			
ZNF297	B	2	-	ZNF297	B	2	-	Zbtb22	B	3	+	Zbtb22	B	3	+	LOC607900	B	2	-
ZBTB9	B	1	+	ZBTB9	B	1	+	Zbtb9	B	9	-	Zbtb9	B	9	-	LOC607940	B	1	+
FG BAK1, ITPR3				FG BAK1, ITPR3				FG Bak1, Itpr3				FG Bak1, Itpr3				FG BAK1, ITPR3			

Human Cluster 7.1				Chimpanzee				Mouse				Rat				Dog			
h chr7				p chr7				m chr5				r chr12				c chr23			
FG SLC29A4	D	F	O	FG SLC29A4	D	F	O	FG Slc29a4	D	F	O	FG Slc29a4	D	F	O	FG SLC29A4,	D	F	O
KIAA1856				KIAA1856				KIAA1856				KIAA1856				KIAA1856			
LOC441193	-	-		LOC441193	-	-													
ZNF815	K	3	+	ZNF185	K	3	+												
FG PMS2, JTV1				FG PMS2, JTV1				FG Pms2, Jtv1				FG Pms2, Jtv1				FG PMS2, JTV1			
EIF2AK1				EIF2AK1				Eif2ak1				Eif2ak1				EIF2AK1			

Human Cluster 7.2				Chimpanzee				Mouse				Rat				Dog			
h chr7				p chr7				m chr5				r chr12				c chr6			
FG RAC1,DAGLB	D	F	O	FG RAC1,DAGLB	D	F	O	FG Rac1, Dag1b	D	F	O	FG Rac1, Dag1b	D	F	O	FG RAC1,DAGLB	D	F	O
KDELK2, GRID21P				KDELK2, GRID21P				Gnd2ip				Gnd2ip				KDELK2, GRID21P			
DKFZp434J1015	-	5	+	LOC742263	K	15	+	Gm792	-	5	-	RGD1563095	-	5	+				
DKFZp547K054	K	15	+	LOC472281	K	15	+	Zfp316	K	15	-	Zfp316	-	15	+				
LOC442283	-	-	-	ZNF325	K	15	-	Zfp12	K	15	+	RGD1564396	K	15	-				
ZNF325	K	15	-																
FG OR7E39, OR7E136P				FG OR7E39, OR7E136P				FG Or7e39, Or7e136p				FG Or7e39, Or7e136p				FG OR7E39, OR7E136P			
OR7E59P				OR7E59P				Or7e59p				Or7e59p				OR7E59P			

Human Cluster 7.3				Chimpanzee				Mouse				Rat				Dog			
h chr7				p chr7															
FG COL23A1,MRPL50P3	D	F	O	FG COL23A1,MRPL50P3	D	F	O												
CLK4				CLK4															
LOC442311	-	12	+	ZNF479	K	10	-												
LOC222032	-	-	-	ZNF716	K	12	+												
ZNF479	K	10	-																
LOC340223	-	-	-																
ZNF716	K	12	+																
FG HIST1H2A1				FG HIST1H2A1															
HIST1H3H,HIST1H2AJ				HIST1H3H,HIST1H2AJ															

Human Cluster 7.4				Chimpanzee				Mouse	Rat	Dog
h chr7				p chr7						
FG: LOC285900, LOC402273	D	F	O	FG:LOC472386,LOC463343	D	F	O			
LOC442320				LOC472387						
ZNF679	K	9	+	ZNF679	K	9	+			
LOC728927	K	9	+	LOC472390	K	10	+			
ZNF680	K	12	-	ZNF680	K	12	+			
ZFD25	-	24	+	LOC738367	-	23	+			
ZNF138	-	6	+	ZNF273	K	13	+			
ZNF273	K	13	+	LOC463440	-	-	-			
ZNF117	K	9	-							
H-plK	-	13	-							
ZNF92	K	14	+							
FG: LOC644586, LOC653491				FG: LOC644586, LOC653491						
LOC402279				LOC402279						

Human Cluster 7.5				Chimpanzee				Mouse				Rat				Dog			
h chr7				p chr7				m chr5				r chr12				c chr6			
FG:PDAP1, BUD31	D	F	O	FG:PDAP1, BUD31	D	F	O	FG: Pdap1,Bud31	D	F	O	FG:MGC108785	D	F	O	FG:LOC489860,LOC489859	D	F	O
PTCD1,CPSF4,ATP5J2				PTCD1,CPSF4,ATP5J2				Ptcd1,Cpsf4				LOC690441				LOC479742,LOC608757			
ZNF789	K	8	+	LOC737149	K	-	+	Zfp99	S	7	-	Zfp99	S	7	+	LOC479741	SK	7	+
ZNF394	SK	7	-	LOC463573	SK	7	-	Zfp95	SK	12	+	Zfp95	S	12	-	LOC489856	SK	12	-
ZFP95	SK	13	+	ZFP95	SK	19	+	Zfp655	-	7	+	Zfp655	-	7	-	LOC489855	-	7	-
ZNF655	-	6	+	LOC463577	-	-	-	LOC666311	S	7	+	Znf498	S	7	-	LOC489853	S	7	-
ZNF498	-	7	+	ZKSCAN1	SK	6	+	LOC667050	-	2	-					LOC489849	SK	6	-
ZKSCAN1	K	6	+	LOC463584	S	7	+	LOC630654	-	2	-					LOC489848	S	7	-
ZNF38	K	8	+	ZNF3	K	8	-												
ZNF3	K	8	-																
FG: COPS6,MCM7				FG: COPS6,MCM7				FG:Cops6,Mcm7				FG:Cops6,Mcm7				FG: COPS6,MCM7			
AP4M1,TAF6				AP4M1,TAF6				Ap4m1,Taf6				Ap4m1,Taf6				AP4M1,TAF6			

Human Cluster 7.7				Chimpanzee				Mouse				Rat				Dog			
h chr7				p chr7				m chr6				r chr4				c chr16			
FG: EZH2,RNY5	D	F	O	FG: EZH2,RNY5	D	F	O	FG:Cul1,Ezh2	D	F	O	FG:Cul1,Ezh2	D	F	O	FG: EZH2,RNY5	D	F	O
PDIA4,COX6BP1				PDIA4,COX6BP1				Rn4,Pdia4				Rn4,Pdia4				PDIA4,COX6BP1			
ZNF786	K	15	-	LOC472581	K	15	-	Zfp786	K	8	-	RGD156212	2K	8		LOC482782	K	28	-
ZNF425	K	19	-	ZNF425	-	18	-	Zfp398	K	7	+	Zfp282	K	5		LOC482785	K	4	+
ZNF398	K	8	+	ZNF398	K	8	+	Zfp282	K	5	+	Zfp212	K	4		LOC610802	K	4	+
ZNF282	K	5	+	ZNF282	K	5	+	Zfp212	-	4	+	LOC681104	-	-		LOC482786	K	9	-
ZNF212	K	4	+	ZNF212	K	4	+	AI894139	-	6	+	RGD1304879	-	4		LOC610814	K	2	-
ZNF783	K	4	+	LOC463826	K	4	+	Zfp797	K	9	-	RGD156605	2K	9		LOC610822	-	-	+
ZNF777	K	9	-	LOC745595	-	-	-	Zfp746	K	4	-	RGD1306209	K	4		LOC482787	-	12	-
ZNF746	K	4	-	LOC463827	K	9	+	Zfp467	-	10	-	RGD130696	K	-					
ZNF767	-	12	-	LOC463829	-	-	-					LOC500110	-	12					
ZNF467	-	12	-	LOC463832	K	-	+					LOC68127	K	-					
FG: ATP6V0E2L,ACTR3B				FG: ATP6V0E2L,ACTR3B				FG: Atp6v0e2,Lrrc61				FG: Atp6v0e2,Lrrc61				FG: LOC482989			
LRRC61,RARRES2				LRRC61,RARRES2				Rarres2				Rarres2				LOC610852,LOC475532			

Human Cluster 8.1				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr11				r chr10				c chr16			
FG DEFB130	D	F	O	FG DEFB130	D	F	O	FG EG654465	D	F	O	FG Defb41	D	F	O	FG DEFB130	D	F	O
LOC389631	-	-	-																
LOC441341	-	-	-																
FG RPL38P1, AGTR1				FG RPL38P1, AGTR1				FG Agtr1				FG Agtr1				FG RPL38P1, AGTR1			
CPB1				CPB1				Cpb1				Cpb1				CPB1			

Human Cluster 8.2				Chimpanzee				Mouse				Rat				Dog			
h chr8				p chr8				m chr15				r chr7				c chr13			
FG SNTB1, HAS2	D	F	O	FG SNTB1, HAS2	D	F	O	FG Sntb1, Has2				FG Sntb1, Has2				FG SNTB1, HAS2	D	F	O
HASNT, MRPS36P3				HASNT, MRPS36P3				Mrps36p3				Mrps36p3				HASNT, MRPS36P3			
ZHX2	H	1	+	ZHX2	H	1	+	Zhx2	H	1	+	Zhx2	H	1	+	LOC482033	H	1	+
ZHX1	H	1	-	ZHX1	H	1	-	Zhx1	H	1	-	Zhx1	H	1	-	LOC475089	H	1	-
FG ATAD2				FG ATAD2				FG Atad2				FG Atad2				FG ATAD2			

Human Cluster 8.3				Chimpanzee				Mouse				Rat				Dog			
h chr8				p chr8				m chr15				r chr7				c chr13			
FG LY6E, HHCM	D	F	O	FG LY6E, HHCM	D	F	O	FG Ly6e, Hhcm	D	F	O	FG Ly6e, Hhcm	D	F	O	FG Ly6e, Hhcm	D	F	O
LY6H				LY6H				Ly6h				Ly6h				Ly6h			
ZFP41	-	4	+	GLI4	-	7	+	Zfp41	-	4	+	Zfp41	-	4	+				
GLI4	-	7	+	ZNF623	-	13	+	Zfp623	-	13	+	LOC550893	-	7	+				
ZNF696	-	9	+	ZNF707	K	6	+	Zfp707	K	7	+								
ZNF623	-	13	+																
ZNF707	K	7	+																
FG BREA2, MAPK15				FG BREA2, MAPK15				FG, Brea2, Mapk15				FG Brea2, Mapk15				FG Brea2, Mapk15			
FAM83H				FAM83H				Fam83h				Fam83h				Fam83h			

Human Cluster 8.4				Chimpanzee				Mouse				Rat				Dog			
h chr8				p chr8				m chr15				r chr7				c chr13			
FG:RECQL4,LRRC14, LRRC24,KIAA1688	D	F	O	FG:RECQL4,LRRC14, LRRC24,KIAA1688	D	F	O	FG.Recql4,Lrrc14 Lrrc24	D	F	O	FG.Recql4,Lrrc14 Lrrc24	D	F	O	FG:LOC482101 LOC482102 LOC482103	D	F	O
ZNF251	-	7	-	LOC742563	K	14	-	Zfp251	K	12	-	Znf251	K	12	-	LOC475129	K	15	-
ZNF34	K	12	-	ZNF34	K	12	-	Zfp7	K	16	+	Zfp647	K	13	-	LOC482106	K	10	-
ZNF517	K	10	+	ZNF7	K	14	+	Zfp647	K	13	-					LOC482107	K	12	+
ZNF7	K	14	+	ZNF250	K	13	-									ZNF347	K	20	-
ZNF647	K	13	-	ZNF16	-	17	-												
ZNF16	-	17	-	LOC747863	-	-	-												
LOC642914	K	10	-																
FG:TMED10P,C8Orf77 C8Orf33				FG: LOC464478 LOC464479				FG: 1110038F14Rik Tmed10P				FG: 1110038F14Rik Tmed10P				FG:LOC10016,LOC482110 LOC482111			

Human Cluster 9.1				Chimpanzee				Mouse				Rat				Dog			
h chr9				p chr9				m chr13				r chr17				c chr1			
FG:HSD17B3, SLC35D2	D	F	O	FG:HSD17B3, SLC35D2	D	F	O	FG: Hsd17b3 Slc35d2	D	F	O	FG: Hsd17b3 Slc35d2	D	F	O	FG:HSD17B3, SLC35D2	D	F	O
ZNF367	-	2	-	ZNF367	-	2	-	Zfp367	-	2	+	Zfp367	-	2	+	LOC476295	-	2	+
ZNF510	K	10	-	ZNF510	K	10	-												
ZNF782	K	14	-	LOC742564	K	13	-												
ZNF322B	-	11	-	ZNF322B	-	11	-												
FG:KIAA1529, TDRD7 TMOD1, NCBP1				FG:KIAA1529, TDRD7 TMOD1, NCBP1				FG: Tdrd7 Tmod1, Ncbp1				FG: Tdrd7 Tmod1, Ncbp1				FG: Tdrd7 Tmod1, Ncbp1			

Human Cluster 9.2				Chimpanzee				Mouse				Rat				Dog			
h chr9				p chr9				m chr4				r chr5				c chr15			
FG: FCMD, TAC2 TMEM38B	D	F	O	FG: FCMD, TAC2 TMEM38B	D	F	O	FG: Fcmd, Tac2 Tmem38b	D	F	O	FG: Fcmd, Tac2 Tmem38b	D	F	O	FG: FCMD, TAC2 TMEM38B	D	F	O
ZNF462	-	9	+	ZNF462	-	9	+	Zfp462	-	9	+	Zfp462	-	9	+	LOC481655	-	9	+
KLF4	-	3	-	LOC464640	-	3	-	Klf4	-	3	-	Klf4	-	3	-	LOC481657	-	2	-
FG: ACTL7B, ACTL7A IKBKAP				FG: ACTL7B, ACTL7A IKBKAP				FG: Actl7b Ikbkap				FG: Actl7b Ikbkap				FG: ACTL7B, ACTL7A IKBKAP			

Human Cluster 9.3				Chimpanzee				Mouse				Rat				Dog			
h chr9				p chr9				m chr4				r chr5				c chr11			
FG SNX30, TSCOT	D	F	O	FG SNX30, TSCOT	D	F	O	FG Snx30, Tscot	D	F	O	FG Snx30, Tscot	D	F	O	FG SNX30, TSCOT	D	F	O
LOC169834	-	13	-																
ZFP37	K	12	-	ZFP37	K	12	-	Zfp37	K	12	-	Zfp37	K	12	-				
FG SLC31A2, FKBP15 SLC31A1, CDC26				FG SLC31A2, FKBP15 SLC31A1, CDC26				FG Slc31a2,Fkbp15 Cdc26				FG Slc31a2,Fkbp15 Cdc26				FG SLC31A2, FKBP15 SLC31A1, CDC26			

Human Cluster 9.4				Chimpanzee				Mouse				Rat				Dog			
h chr9				p chr9				m chr2				r chr3				c chr9			
FG PDCL, RC3H2	D	F	O	FG PDCL, RC3H2	D	F	O	FG Pdcl, Rc3h2	D	F	O	FG Pdcl, Rc3h2	D	F	O	FG PDCL, RC3H2	D	F	O
ZNF482	B	4	-	ZNF482	B	4	-	Zfp482	B	4	-	Zfp482	B	4	-	ZNF482	B	4	-
ZBTB26	B	4	-	ZBTB26	B	4	-	Zbtb26	B	4	-	Zbtb26	B	4	-	ZBTB26	B	4	-
FG RABGAP1, GPR21				FG RABGAP1, GPR21				FG Rabgap1, Gpr21				FG Rabgap1, Gpr21				FG RABGAP1, GPR21			

Human Cluster 10.1				Chimpanzee				Mouse	Rat	Dog
h chr10				p chr10						
FG LOC646352,ANKRD30A LOC219752	D	F	O	FG LOC646352,ANKRD30A LOC219752	D	F	O			
ZNF248	K	8	-	ZNF248	K	8	-			
BA775A3.1	-	-	+	LOC466277	K	12	-			
BA393J16.4	-	-	+	ZNF33A	K	16	+			
ZNF25	K	12	-	ZNF37A	K	12	+			
ZNF33A	K	16	+							
ZNF37A	K	12	+							
FG LOC646419 LOC646423,LOC646426				FG LOC646419 LOC646423,LOC646426						

Human Cluster 10.2				Chimpanzee				Mouse	Rat	Dog
h chr10				p chr10						
FG CCNYL2	D	F	O	FG CCNYL2	D	F	O			
MGC16291				MGC16291						
LOC401642	-	16	-	LOC450411	-	-	-			
ZNF37B	K	8	-	LOC740109	-	8	-			
ZNF11B	K	16	-							
FG DUXAP3				FG DUXAP3						
BMS1 RET				BMS1, RET						

Human Cluster 10.3				Chimpanzee				Mouse			Rat			Dog			
h chr10				p chr10				m chr6			r chr4			c chr28			
FG:GALNACT-2, RASGEF1A	D	F	O	FG:GALNACT-2, RASGEF1A	D	F	O	FG Rasgef1a, Fxyd4	D	F	O	FG Rasgef1a, Fxyd4	D	F	O	FG:GALNACT-2, RASGEF1A	D	F	O
FXYD4,HNRPF				FXYD4,HNRPF				Hnrpf				Hnrpf				FXYD4,HNRPF			
ZNF487	K	3	+	LOC745336	K	-	+	Zfp239	-	7	+	Zfp637	-	7	+				
ZNF239	-	9	-	ZNF239	-	9	-	Zfp637	-	7	+								
ZNF485	K	11	+	LOC745499	K	11	+												
ZNF32	-	7	-	ZNF32	-	7	-												
FG HNRPA3P1				FG HNRPA3P1				FG Hnrpa3p1				FG Hnrpa3p1				FG: HNRPA3P1			
CXCL12				CXCL12				Cxcl12				Cxcl12				CXCL12			

Human Cluster 11.1				Chimpanzee				Mouse			Rat			Dog			
h chr11				p chr11				m chr7			r chr1			c chr21			
FG OR10A2,OR2D2	D	F	O	FG OR10A2,OR2D2	D	F	O	FG Or10a2	D	F	O	FG Or10a2	D	F	O	FG OR10A2,OR2D2	D	F	O
OR2D3, OR1DA4				OR2D3, OR1DA4				Or2d2				Or2d2				OR2D3, OR1DA4			
ZNF215	SK	4	+	ZNF215	SK	4	+									LOC485362	-	4	+
ZNF214	K	11	-	ZNF214	K	11	-									LOC485363	K	11	-
FG NLKP14, HNRNPG-T				FG NLKP14, HNRNPG-T				FG Nlkp14				FG Nlkp14				FG NLKP14, HNRNPG-T			
SYT9				SYT9				Syt9				Syt9				SYT9			

Human Cluster 11.2				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr11				r chr10				c chr16			
FG OR5B12, OR5B21	D	F	O	FG OR5B12, OR5B21	D	F	O	FG Or5b12	D	F	O	FG Or5b12	D	F	O	FG OR5B12, OR5B21	D	F	O
LPXN				LPXN				Lpxn				Lpxn				LPXN			
ZFP91	-	4	+	ZFP91	-	4	+	Zfp91	-	4	+	Zfp91	-	4	+	LOC475962	-	4	+
ZFP91-CNTF	-	-	+																
FG GLYAT				FG GLYAT				FG Glyat				FG Glyat				FG GLYAT			
GLYATL2				GLYATL2				Glyatl2				Glyatl2				GLYATL2			

Human Cluster 12.1				Chimpanzee				Mouse				Rat				Dog			
h chr12				p chr12				m chr5				r chr12				c chr5			
FG CHFR	D	F	O	FG LOC452391	D	F	O	FG Chfr	D	F	O	FG Chfr	D	F	O	FG CHFR	D	F	O
ZNF605	K	17	-	ZNF26	K	13	+									LOC486220	K	17	-
ZNF26	K	13	+	ZNF84	K	19	+									LOC486219	K	13	-
ZNF84	K	19	+	ZNF140	K	21	+												
ZNF140	K	10	+	ZNF268	K	24	+												
LOC440122	K	6	-																
ZNF10	K	11	+																
ZNF268	-	24	+																
FG End Of Chromsome				FG End Of Chromsome				FG End Of Chromsome				FG End Of Chromsome				FG End Of Chromsome			

Human Cluster 13.1				Chimpanzee				Mouse				Rat				Dog			
h chr13				p chr13				m chr14				r chr15				c chr22			
FG DACH1	D	F	O	FG DACH1	D	F	O	FG Dach1	D	F	O	FG Dach1	D	F	O	FG DACH1	D	F	O
DIS3				DIS3				Dis3				Dis3				DIS3			
KLF5	-	3	+	KLF5	-	3	+	Klf5	-	3	+	Klf5	-	3	+	KLF5	-	3	+
KLF12	-	3	-	KLF12	-	3	-	Klf12	-	3	-	Klf12	-	3	-	KLF12	-	3	-
FG GCG-172, BIMD6				FG GCG-172, BIMD6				FG Gcg-172, Bimd6				FG Gcg-172, Bimd6				FG GCG-172, BIMD6			
TBC1b4				TBC1b4				Tbc1b4				Tbc1b4				TBC1b4			

Human Cluster 13.2				Chimpanzee				Mouse				Rat				Dog			
h chr13				p chr13				m chr14				r chr15				c chr22			
FG TM9SF2, CLYBL	D	F	O	FG TM9SF2, CLYBL	D	F	O	FG Tm9sf2, Clybl	D	F	O	FG Tm9sf2, Clybl	D	F	O	FG TM9SF2, CLYBL	D	F	O
ZIC5	-	4	-	ZIC5	-	4	-	Zic5	-	4	-	Zic5	-	4	-	ZIC5	-	4	-
ZIC2	-	4	+	ZIC2	-	4	+	Zic2	-	4	+	Zic2	-	4	+	ZIC2	-	4	+
FG PCCA, RPS26L				FG PCCA, RPS26L				FG Pcca, Rps26l				FG Pcca, Rps26l				FG PCCA, RPS26L			
TMTC4				TMTC4				Tmtc4				Tmtc4				TMTC4			

Human Cluster 14.1				Chimpanzee				Mouse				Rat				Dog			
h chr14				p chr14				m chr14				r chr15				c chr8			
FG MYH6, MYH7,NGDN	D	F	O	FG MYH6, MYH7,NGDN	D	F	O	FG Myh6, Myh7 Ngdn	D	F	O	FG Myh6, Myh7 Ngdn	D	F	O	FG MYH6, MYH7,NGDN	D	F	O
ZFHX2	H	4	-	ZFHX2	H	4	-	Zfhx2	H	4	-	Zfhx2	H	4	-	LOC490613	H	4	-
ZNF409	-	1	-	ZNF409	-	1	-												
FG THTPA, AP1G2				FG THTPA, AP1G2				FG Thtpa, Ap1g2				FG Thtpa, Ap1g2				FG THTPA, AP1G2			
JPH4				JPH4				Jph4				Jph4				JPH4			

Human Cluster 14.2				Chimpanzee				Mouse				Rat				Dog			
h chr14				p chr14				m chr112				r chr6				c chr8			
FG ESR2, MTHFD1	D	F	O	FG ESR2, MTHFD1	D	F	O	FG Esr2, Mthfd1	D	F	O	FG Esr2, Mthfd1	D	F	O	FG ESR2, MTHFD1	D	F	O
AKAP5				AKAP5				Akap5				Akap5				AKAP5			
ZBTB25	B	2	-	ZBTB25	B	2	-	Zbtb25	B	2	-	Zbtb25	B	2	-	LOC490730	B	2	-
ZBTB1	B	2	+	ZBTB1	B	2	+	Zbtb1	B	2	+	Zbtb1	B	2	+	LOC490731	B	2	+
FG HSPA2,NUP50P1				FG HSPA2,NUP50P1				FG Hspa2,Nup50p1				FG Hspa2,Nup50p1				FG HSPA2,NUP50P1			
PLEKHG3, SPTB				PLEKHG3, SPTB				Sptb				Sptb				PLEKHG3, SPTB			

Human Cluster 15.1				Chimpanzee				Mouse				Rat				Dog			
h chr1				p chr1				m chr11				r chr10				c chr16			
FG WHDC1, HOMER2	D	F	O	FG WHDC1, HOMER2	D	F	O	FG Whdc1, Homer2	D	F	O	FG Whdc1, Homer2	D	F	O	FG WHDC1, HOMER2	D	F	O
FAM103A1				FAM103A1				Fam103A1				Fam103A1				FAM103A1			
BTBD1	B	-	-	BTBD1	B	-	-	Btbd1	B	-	-	Btbd1	B	-	-				
BNC1	-	3	-	BNC1	-	3	-	Bnc1	-	3	-	Bnc1	-	3	-				
FG SH3GL3, ADAMTSL3				FG SH3GL3, ADAMTSL3				FG Sh3gl3, Adamtsl3				FG Sh3gl3, Adamtsl3				FG SH3GL3, ADAMTSL3			

Human Cluster 15.2				Chimpanzee				Mouse				Rat				Dog			
h chr15				h chr15				m chr7				r chr1				c chr3			
FG LOC440302,LOC388163	D	F	O	FG LOC440302,LOC388163	D	F	O	FG LOC440302,LOC388163	D	F	O	FG LOC440302,LOC388163	D	F	O	FG LOC440302,LOC388163	D	F	O
FLJ40113, E2Q2				FLJ40113, E2Q2				FLJ40113, E2Q2				FLJ40113, E2Q2				FLJ40113, E2Q2			
ZFP29	S	14	+	ZSCAN2	S	14	+	Zscan2	S	14	+	LOC686118	S	-	+	LOC488749	S	21	+
SCAND2	S	-	+	SCAND2	-	-	-	Zfp592	-	4	+	Zfp592	-	4	+	LOC488751	-	4	+
ZNF592	-	4	+	LOC453608	-	-	-												
FG ALPK3, SLC28A1, PDE8A				FG ALPK3, SLC28A1, PDE8A				FG ALPK3, SLC28A1, PDE8A				FG ALPK3, SLC28A1, PDE8A				FG ALPK3, SLC28A1, PDE8A			

Human Cluster 15.3				Chimpanzee				Mouse				Rat				Dog			
h chr15				p chr15				m chr7				r chr1				c chr3			
FG.MESP2, ANPEP	D	F	O	FG.MESP2, ANPEP	D	F	O	FG.MESP2, ANPEP	D	F	O	FG.MESP2, ANPEP	D	F	O	FG.MESP2, ANPEP	D	F	O
AP3S2				AP3S2				AP3S2				AP3S2				AP3S2			
LOC390636	-	2	+																
ZNF710	-	11	+																
ZNF774	-	12	+																
FG.IQGAP1, CRTC3				FG.IQGAP1, CRTC3				FG.IQGAP1, CRTC3				FG.IQGAP1, CRTC3				FG.IQGAP1, CRTC3			
BLM, FURIN				BLM, FURIN				BLM, FURIN				BLM, FURIN				BLM, FURIN			

Human Cluster 16.1				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr17				r chr10				c chr6			
FG.THOC6,MMP25	D	F	O	FG.THOC6,MMP25	D	F	O	FG.Thoc6, Mmp25	D	F	O	FG.Thoc6, Mmp25	D	F	O	FG.LOC490047	D	F	O
MMPL1,IL32				MMPL1,IL32												LOC490046			
ZNF206	S	14	-	LOC747677	S	14	-	Zfp206	S	14	+	Zfp206	S	14	+	LOC490045	S	15	+
ZNF205	K	8	+	ZNF205	K	8	+	Zfp13	K	8	-	Zfp13	-	8	-	LOC479870	K	13	-
ZNF213	SK	5	+	ZNF213	SK	5	+	Zfp213	SK	5	-	Znf213	S	5	-	LOC490040	SK	9	+
ZNF200	-	5	-	ZNF200	-	5	-	Zfp40	K	19	-	Zfp263	S	9	-	LOC490042	S	5	-
ZNF263	SK	9	+	ZNF263	SK	9	+	LOC545191	-	5	-	Znf174	S	-	-	LOC490035	S	6	-
ZNF75A	K	5	+	LOC453861	SK	5	+	LOC433078	K	-	+	Zfp597	-	7	+				
ZNF434	S	6	-	ZNF434	S	6	-	1300003B13Rik	K	11	+								
ZNF174	S	3	+	ZNF174	S	3	+	Zfp758	K	13	+								
ZNF597	-	7	-	LOC467889	-	7	-	633041620	K	10	+								
FG.FLJ14154,LOC646174				FG.FLJ14154,LOC646174				FG Cluap1				FG Cluap1				FG LOC609456			
LOC390671,CLUAP1,NOD3				LOC390671,CLUAP1,NOD3				Nod3				Nod3				LOC479868,LOC490033			

Human Cluster 16.2				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr16				r chr10				c chr20			
FG ADCY9, SRL	D	F	O	FG ADCY9, SRL	D	F	O	FG Adcy9, Srl	D	F	O	FG Adcy9, Srl	D	F	O	FG ADCY9, SRL	D	F	O
TFAP4				TFAP4				Tfap4				Tfap4				TFAP4			
GLIS2	-	4	+	GLIS2	-	4	+	Glis2				Glis2_pred				LOC490028			
ZNF500	S	5	-	ZNF500	S	5	-												
FG SEPT12,ROGDI				FG SEPT12,ROGDI				FG Sept12				FG Sept12				FG SEPT12,ROGDI			
								Rogdi				Rogdi							

Human Cluster 16.3				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr7				r chr1				c chr6			
FG CD2BP2,TBC1D10B	D	F	O	FG CD2BP2,TBC1D10B	D	F	O	FG Cd2bp2,Tbc1d10b	D	F	O	FG Cd2bp2,Tbc1d10b	D	F	O	FG CD2BP2,TBC1D10B	D	F	O
MYLPF,SEPT1				MYLPF,SEPT1				Mylpf,Sept1				Mylpf,Sept1				MYLPF,SEPTI			
ZNF553	-	12	+	LOC740214	-	12	+	Zfp553	-	12	+	RGD1561639				LOC489901	-	12	-
ZNF768	-	10	-	LOC454044	-	8	+	Zfp771	-	8	+	RGD1305903				LOC489906	K	17	+
ZNF747	K	-	-	LOC454367	K	14	-	Zfp768	-	10	-	LOC691885				LOC489908	K	2	+
ZNF764	K	7	-	LOC454046	K	7	-	6430604K15Rik	K	9	-	LOC691887				LOC607501	K	-	+
ZNF688	K	2	-	LOC467950	2K	9	-	9130019022Rik	K	13	-	Zfp688				LOC489910	K	11	+
ZNF785	K	7	-	LOC467951	-	-	-	E430018J23Rik	-	9	-	Hit39				LOC489911	K	9	+
ZNF689	K	11	-	ZNF629	-	19	-	Zfp764	K	9	-	Zfp629				LOC489914	-	19	+
ZNF629	-	19	-	LOC467958	-	15	-	Zfp688	-		-	Zfp668				LOC489920	-	16	+
ZNF668	-	16	-	ZNF646	-	29	+	Zfp689	K	2	-					LOC489922	-	28	-
ZNF646	-	29	+					Zfp629	-	19	-								
								Zfp668	-	16	-								
								6820420M01	-	27	+								
FG VKORC1,LOC647097				FG LOC467959,BCKDK				FG VkorC1,Bckdk				FG VkorC1,Bckdk							
BCKDK,MYST1,PRSS8				PRSS8				Myst1,Prss8				Myst1,Prss8							

Human Cluster 16.4				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr5				r chr1							
FG PPP2CBP	D	F	O	FG PPP2CBP	D	F	O	FG Ppp2cbp	D	F	O	FG Ppp2cbp	D	F	O	FG PPP2CBP	D	F	O
VN1R3				VN1R3				Vn1r3				Vn1r3				VN1R3			
LOC342426	-	12	-	LOC743274	-	12	-	Zfp267	K	14	+	Zfp267	K	14	+				
ZNF267	K	14	+	ZNF267	K	14	+												
FG LOC647126,				FG LOC647126,				FG LOC647126,				FG LOC647126,				FG LOC647126,			
LOC388248				LOC388248				LOC388248				LOC388248				LOC388248			

Human Cluster 16.5				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr8				r chr19				c chr5			
FG ABBA1, VAC14	D	F	O	FG ABBA1, VAC14	D	F	O	FG Abba1, Vac14	D	F	O	FG Abba1, Vac14	D	F	O	FG ABBA1, VAC14	D	F	O
HYDIN, CALB2				HYDIN, CALB2				Hydin, Calb2				Hydin, Calb2				HYDIN, CALB2			
ZNF23	K	17	-	LOC468017	K	17	-	Zfp612	K	16	+	Zfp612_pred	K	16	+	LOC489720	K	17	-
ZNF19	K	10	-	ZNF19	K	10	-									LOC489721	K	10	-
FG CHST4,TAT				FG CHST4,TAT				FG Chst4				FG Chst4				FG CHST4,TAT			
MARVELD3				MARVELD3				Tat, Marveld3				Tat, Marveld3				MARVELD3			

Human Cluster 16.6				Chimpanzee				Mouse				Rat				Dog			
h chr16				p chr16				m chr8				r chr19				c chr5			
FG SLC7A5,CA5A	D	F	O	FG SLC7A5,CA5A	D	F	O	FG Slc7a5, Ca5a	D	F	O	FG Slc7a5, Ca5a	D	F	O	FG CA5A	D	F	O
BANP				BANP				Banp				Banp				BANP			
ZNF649	-	3	+	ZNF649	-	3	+	Gm22	-	4	+	LOC691499	-	4	+	LOC489666	-	4	-
ZFPM1	-	2	+	ZFPM1	-	2	+	ZfpM1	-	2	+	LOC691504	-	2	+				
FG NHN1, IL17C				FG NHN1, IL17C				FG Nhn1, Il17c				FG Nhn1, Il17c				FG NHN1, IL17C			
CYBA				CYBA				Cyba				Cyba				CYBA			

Human Cluster 17.1				Chimpanzee				Mouse				Rat				Dog			
h chr17				p chr17				m chr11				r chr10				c chr5			
FG KIF1C, GPR1728	D	F	O	FG KIF1C, GPR1728	D	F	O	FG Kif1c	D	F	O	FG Kif1c	D	F	O	FG KIF1C, GPR1728	D	F	O
								Gpr1728				Gpr1728							
ZFP3	-	3	+	LOC468455	-	13	+	Zfp3	-	13	+	RGD1565881	-	13	+	LOC489452	-	13	-
ZNF232	S	3	-	LOC455260	S	5	-												
ZNF594	-	22	-	ZNF594	-	32	-												
FG UNQ5783,RABEP1				FG UNQ5783,RABEP1				FG Unq5783				FG Unq5783				FG UNQ5783,RABEP1			
NUP88				NUP88				Nup88				Nup88				NUP88			

Human Cluster 17.2				Chimpanzee				Mouse				Rat				Dog			
h chr17				p chr17				m chr11				r chr10				c chr5			
U8B,TRPV2	D	F	O	U8B,TRPV2	D	F	O	FG U8b	D	F	O	FG U8b	D	F	O	U8B,TRPV2	D	F	O
SNORD49B				SNORD49B				Trpv2				Trpv2				SNORD49B			
ZNF287	SK	14	-	ZNF287	SK	14	-	Zfp287				Zfp287				ZNF287			
ZNF624	K	21	-	ZNF624	K	21	-												
FG RNASEH1P2				FG RNASEH1P2				FG Rnaseh1p2				FG Rnaseh1p2				FG RNASEH1P2			

Human Cluster 18.1				Chimpanzee				Mouse				Rat				Dog			
h chr18				p chr18				m chr18				r chr18				c chr7			
FG NOL4,DTNA	D	F	O	FG NOL4 DTNA	D	F	O	FG Nol4, Dtna	D	F	O	FG Nol4, Dtna	D	F	O	FG LOC490491	D	F	O
MAPRE2				MAPRE2				Mapre2				Mapre2				LOC490488 LOC490487			
ZNF397	S	-	+	LOC455370	S	9	+	Zfp397	S	9	+	Znf397	S	-	-	LOC490486	S	7	+
ZNF271	-	5	+	LOC455372	S	7	-	C230097124Rik	S	-	-	Zfp239	-	18	+	LOC480162	S	4	+
ZNF24	S	4	-	LOC468520	-	20	+	Zfp35	-	18	+	Zfp191	S	4	-				
ZNF396	S	2	-	ZNF24	S	4	-	Zfp191	S	4	-								
				ZNF396	S	2	-												
FG GALNT1, LOC648532				FG GALNT1, LOC648532				FG Galnt1,				FG Galnt1,				FG LOC480161			
P15RS				P15RS				P15rs				P15rs				LOC480160			

Human Cluster 18.2				Chimpanzee				Mouse				Rat				Dog			
h chr18				p chr18				m chr18				m chr18				c chr1			
FG TSHZ1, LOC284274	D	F	O	FG TSHZ1, LOC284274	D	F	O	FG Tshz1	D	F	O	FG Tshz1	D	F	O	FG TSHZ1, LOC284274	D	F	O
LOC728662				LOC728662												LOC728662			
ZNF516	-	7	-	ZNF516	-	7	-	Zfp516	-	7	-	RGD1306817	-	7	-	LOC483930	-	7	-
ZNF236	-	25	+	ZNF236	-	25	+	Zfp236	-	29	+	Zfp236	-	30	+	LOC483929	-	30	+
FG MBP, GALR1				FG MBP, GALR1				FG Mbp, Galr1				FG Mbp, Galr1				FG MBP, GALR1			

Human Cluster 19.1				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr				m chr10				r chr7				c chr16			
FG ATPD8B3, REXO1	D	F	O	FG ATPD8B3, REXO1	D	F	O	FG Atpd8b3, Rexo1	D	F	O	FG Atpd8b3, Rexo1	D	F	O				
KLF16	-	3	-	KLF16	-	3	-	Klf16	-	3	-	LOC690820	-	3	+				
BTBD2	B	-	-	BTBD2	B	-	-	Btbd2	B	-	-	RGD1566094	B	-	+				
FG MKNK2, MOBKL2A				FG MKNK2, MOBKL2A				FG Mknk2, Mobkl2a				FG Mknk2, Mobkl2a							

Human Cluster 19.2				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr10				r chr7				c chr20			
FG SLC39A3,SGTA	D	F	O	FG SLC39A3,SGTA	D	F	O	FG Sgta	D	F	O	FG Sgta	D	F	O	FG SLC39A3,SGTA	D	F	O
THOP1				THOP1				Thop1				Thop1				THOP1			
ZNF554	K	7	+	ZNF554	-	7	+												
ZNF555	K	15	+	ZNF555	K	15	+												
ZNF556	K	9	+	ZNF556	K	9	+												
ZNF57	K	13	+	ZNF57	K	13	+												
ZNF77	K	12	-	ZNF77	K	12	-												
FG TLE6				FG TLE6				FG Tle6				FG Tle6				FG TLE6			
TLE2				TLE2				Tle2				Tle2				TLE2			

Human Cluster 19.3				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr11				r chr10				c chr20			
FG_MBD3L2	D	F	O	FG_MBD3L2	D	F	O	FG_Mbd3l2	D	F	O	FG_Mbd3l2	D	F	O	FG_MBD3L2	D	F	O
ZNF557	K	10	+	ZNF557	K	10	+									ZNF557	K	10	+
ZNF358	-	9	+	LOC455655	-	9	+									ZNF358	-	9	+
FG:MCOLN1,PNPLA6				FG:MCOLN1,PNPLA6				FG_Mcoln1				FG_Mcoln1				FG:MCOLN1,PNPLA6			
KIAA153,PCP2				KIAA153,PCP2				Pcp2				Pcp2				KIAA153,PCP2			

Human Cluster 19.4				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr17				r chr9				c chr20			
FG_VAV1,EMR1	D	F	O	FG_VAV1,EMR1	D	F	O	FG_Vav1	D	F	O	FG_Vav1	D	F	O	FG_VAV1,EMR1	D	F	O
EMR4				EMR4				Emr4				Emr4				EMR4			
ZNF557	K	10	+	ZNF557	K	10	+									ZNF557	K	10	+
ZNF358	-	9	+	ZNF358	-	9	+												
FG_MCOLN1, PNPLA6				FG_MCOLN1, PNPLA6				FG_Mcoln1				FG_Mcoln1				FG_MCOLN1, PNPLA6			
KIAA1543, XAB2				KIAA1543, XAB2				Xab2				Xab2				KIAA1543, XAB2			

Human Cluster 19.5				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr17				r chr8				c chr20			
FG_MARCH2,HNRPM	D	F	O	FG_MARCH2,HNRPM	D	F	O	FG_Hnrpm	D	F	O	FG_Hnrpm	D	F	O	FG_MARCH2,HNRPM	D	F	O
PRAM1				PRAM1				Pram1				Pram1				PRAM1			
ZNF414	-	1	-	LOC743829	-	-	-	Zfp414	-	2	+	Zfp101	K	15	-	LOC611636	-	1	-
ZNF558	K	9	-	LOC468704	K	9	-	Zfp101	K	15	-	Zfp81	K	13	-	ZNF558	K	9	-
ZNF317	K	13	+	ZNF317	K	13	+	Zfp81	K	13	-								
ZNF699	K	16	-	LOC744819	K	5	-												
ZNF559	K	11	+	ZNF177	K	7	+												
ZNF177	K	7	+	ZNF266	K	14	-												
ZNF266	K	14	-	ZNF560	KK	14	-												
ZNF560	KK	14	-	ZNF426	K	12	-												
ZNF426	K	12	-	ZNF121	-	10	-												
ZNF121	-	10	-	ZNF561	K	10	-												
ZNF561	-	10	-	LOC455686	K	11	-												
ZNF562	-	9	-																
LOC729648	-	-	-																
LOC162993	K	12	-																
FG_UBE2L4,FBXL1,UBL5				FG_UBE2L4,FBXL1,UBL5				FG_Ube2l4, Fbxl12				FG_Ube2l4, Fbxl12				FG_UBE2L4,FBXL1,UBL5			
PIN1,OLFM2				PIN1,OLFM2				Ubl5, Pin1				Ubl5, Pin1				PIN1,OLFM2			

Human Cluster 19.6				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr9				r chr8				c chr20			
FG RGL3,PRKCSH	D	F	O	FG RGL3,PRKCSH	D	F	O	FG Prkcsh	D	F	O	FG Prkcsh	D	F	O	FG LOC484941	D	F	O
ELAVL3				ELAVL3				Elavl3				Elavl3				LOC484940			
ZNF653	-	4	-	ZNF653	-	4	-	Zfp653	-	4	-	Zfp653	-	-	-	LOC484939	-	4	+
ZNF627	K	11	+	LOC468723	-	-	-	9530015io7Rik	K	13	+					LOC611075	K	15	+
LOC401898	-	6	+	LOC455735	-	-	-	Zfp809	K	7	+					LOC484934	K	16	+
HSZFP36	K	16	-	HSZFP36	-	16	-	BC050092	K	11	-					LOC484933	K	17	-
ZNF441	-	19	+	LOC468726	-	3	-	Zfp810	K	12	-								
ZNF491	-	13	+	LOC468727	-	-	-												
ZNF440	K	12	+	ZNF440	K	21	+												
ZNF439	K	11	+	LOC746850	-	3	+												
ZNF69	K	-	+	LOC468730	-	-	-												
ZNF700	K	21	+	LOC468731	K	19	-												
ZNF440L	K	8	+	ZNF208	K	39	+												
ZNF433	K	19	-	ZNF136	-	1	+												
LOC729747	K	15	-	ZNF44	K	15	-												
FLJ14959	K	8	+	ZNF443	K	18	-												
ZNF788	-	16	+	LOC468733	K	36	-												
ZNF20	K	13	-	ZNF564	K	15	-												
ZNF625	-	8	-	ZNF490	K	13	-												
ZNF136	K	14	+	LOC468735	-	16	+												
ZNF44	K	16	-	LOC455740	-	-	-												
ZNF563	K	8	-																
ZNF442	K	14	-																
ZNF799	KK	26	-																
ZNF443	K	19	-																
ZNF709	K	19	-																
ZNF564	K	16	-																
ZNF490	K	13	-																
ZNF791	K	17	+																
KLF1	-	3	-																
FG MAN2B1, MORG1				FG MAN2B1, MORG1				FG Man2b1, Morg1				FG Man2b1, Morg1				FG MAN2B1			
DHPS FBXW9				DHPS FBXW9				Dhps				Dhps				MORG1			

Human Cluster 19.7				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr8				r chr16				c chr20			
FG GMIP	D	F	O	FG GMIP	D	F	O	FG Atp13a1	D	F	O	FG Atp13a1	D	F	O				
ATP13A1				ATP13A1				Gmip				Gmip							
ZNF101	K	10	+	ZNF101	K	10	+	D330038006Rik	K	12	-	MGC72612	K	9	+				
ZNF140	K	19	-	ZNF208	K	36	-	9830167H18Rik	K	7	-								
ZNF506	K	8	-	ZNF93	K	15	+	1200003107RIK	K	12	-								
LOC730008	-	9	+	ZNF91	K	37	+	EG636741	K	10	+								
ZNF253	K	3	+	ZNF85	-	-	+	AI449175	K	9	-								
ZNF505	K	17	+	ZNF430	K	12	+	D10627	K	13	+								
ZNF682	K	11	-	LOC455891	-	-	+												
ZNF90	K	15	+	ZNF431	K	14	+												
ZNF486	K	10	+	ZNF85	K	14	-												
FLJ44894	-	3	-	ZNF91	-	-	+												
LOC163233	K	14	-	ZNF90	-	-	+												
ZNF626	K	-	-	ZNF429	-	-	+												
ZNF66	-	8	+	ZNF492	K	12	-												
ZNF85	K	15	+	ZNF100	K	12	-												
ZNF430	K	12	+	ZNF43	-	-	-												

ZNF714	-	12	+	ZNF208	K	36	-			
ZNF431	K	12	+	LOC468804	K	9	+			
ZNF708	K	15	-	ZNF43	K	16	-			
ZNF493	K	-	+	ZNF93	K	34	-			
ZNF429	K	17	+	LOC740583	-	-	-			
ZNF100	K	12	-	LOC468806	-	5	+			
ZNF43	K	22	-	ZNF675	K	14	-			
ZNF208	K	34	-	LOC468808	K	16	-			
ZNF257	K	12	+	LOC740901	-	5	+			
ZNF676	K	15	-	LOC455907	-	-	+			
LOC148198	K	13	-							
LOC441843	-	-	+							
ZNF492	K	28	+							
ZNF99	K	30	-							
LOC646864	K	7	+							
LOC388523	-	-	-							
ZNF724P	-	-	-							
ZNF91	K	35	-							
ZNF725	-	-	-							
ZNF675	K	14	-							
ZNF681	-	16	-							
LOC646895	-	6	+							
LOC730084	-	6	+							
LOC730087	-	-	+							
ZNF254	K	4	+							
FG CTP25I, UQCRFS1				*FG CTP25I, UQCRFS1*				*FG Ctp25i, Uqcrfs1*	*FG Ctp25i, Uqcrfs1*	
POP4, PLEKHF1				*POP4, PLEKHF1*				*Pop4*	*Pop4*	

Human Cluster 19.8				***Chimpanzee***				***Mouse***				***Rat***				***Dog***			
h chr19				***p chr19***				***m chr7***				***r chr1***				***c chr1***			
FG PDCD2L,UBA2	D	F	O	*FG PDCD2L,UBA2*	D	F	O	*FG Uba2*	D	F	O	*FG Uba2*	D	F	O	*FG LOC611942,LOC484594*	D	F	O
WTIP,LOC284402				*WTIP,LOC284402*				*Wtip*				*Wtip*				*LOC476490*			
LOC441847	-	9	-	LOC468824	-	-	-									LOC484590	K	28	-
ZNF302	K	7	+	ZNF302	-	-	-									LOC484586	K	13	+
ZNF181	K	11	+	ZNF181	K	11	+												
ZNF599	K	14	-	LOC468825	-	14	-												
LOC643825	S	-	+	ZNF30	K	16	+												
LOC441848	-	1	-	LOC468828	K	13	-												
ZNF30	-	18	+																
ZNF92	-	13	-																
FG GRAMD1A, SCN1B				*FG GRAMD1A, SCN1B*				*FG Gramd1a, Scn1b*				*FG Gramd1a, Scn1b*				*FG SCNN1B*			
HPN,FXYD3				*HPN,FXYD3*				*Hpn*				*Hpn*				*LOC484585*			

Human Cluster 19.9				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr7				r chr1				c chr1			
FG:COX6B1,UPK1A	D	F	O	FG:LOC455974,LOC455975	D	F	O	FG Cox6b1,Ukp1a,Ckap1	D	F	O	FG:Cox6b1,Ukp1a,Ckap1	D	F	O	FG:LOC612644,LOC484579	D	F	O
CKAP1,CAPNS1,COX7A1				LOC468844				Capns1,Cox7a1				Capns1,Cox7a1							
TZFP	B	2	+	LOC455976	K	12	-	Zbtb32	B	2	-	Znf382	K	10	+	LOC484577	B	2	-
ZNF565	K	12	-	LOC455977	K	11	-	Zfp146	-	10	-	zZfp260	-	13	+	LOC484564	-	3	-
ZNF146	-	10	+	LOC455978	K	23	+	5930415A09Rik	K	9	+	RGD1563239	K	33	-	LOC484561	K	27	+
ZFP14	K	13	-	LOC468846	K	23	+	Zfp260	-	12	+	RGD1560682	K	11	+	LOC484559	K	10	+
ZNF545	K	13	-	LOC455979	2K	34	-	Zfp566	K	7	-	Zfp569	K	7	+	LOC484548	K	12	+
ZNF566		7	-	LOC468848	K	10	+	Zfp82	K	13	-	Zfp74	K	18	-	LOC484547	-	29	-
ZFP260	-	13	-	LOC455983	K	13	+	Zfp14	K	13	-	LOC499120	K	39	+	LOC484545	K	35	-
ZNF529	-	11	-	ZNF208	K	39	+	Zfp568	2K	11	+	Zfp84	K	23	+	LOC484542	K	31	+
ZNF382	K	10	+	ZNF567	K	14	-	LOC625421	K	-	+					LOC484541	K	15	+
GIOT-1	K	12	-	ZNF461	K	12	+	Zfp74	-	1	-					LOC484540	K	22	-
ZNF567	K	15	+	ZNF382	K	10	-	Zfp383	-	11	+					LOC484538	K	33	-
LOC342892	K	32	-	ZNF529	K	11	+	Zfp27	K	22	-								
MGC62100	K	13	-	ZNF260		13	+	B230312I18Rik	-	19	+								
ZNF345	-	15	+	ZNF566	K	8	+	BC027344	K	10	-								
ZNF568	K	12	+	ZNF545	K	13	+	6330581L23Rik	K	13	+								
LOC653284	-	12	+	ZFP14	K	13	+	Zfp30	K	13	+								
ZNF420	K	19	+	HKR1	K	13	+	Zfp84	K	11	+								
ZNF585A	K	23	-	ZNF569	K	39	-												
ZNF585B	K	23	-	ZNF571	K	17	-												
ZNF383	K	11	+	ZFP30	K	13	-												
HKR1	K	13	+	LOC468857	K	15	-												
ZNF527	K	12	+	ZNF607	K	20	-												
ZNF569	K	18	-	LOC468859	K	9	-												
ZNF570	K	11	+																
LOC390927	K	5	+																
ZNF540	K	17	+																
ZNF571	K	17	-																
ZFP30	K	13	-																
FJL37549	-	3	-																
ZNF607	K	20	-																
ZNF573	-	19	-																
LOC401915	-	8	+																
FG NVD,SP11,SIPA1L3				FG:LOC468860,LOC468861				FG 4930432E11Rik				FG 4930432E11Rik				FG LOC484537			
DPF1,PPP1R14A				LOC468862				Dpf3,Ppp1r149				Dpf3,Ppp1r149				LOC612600,LOC484536			

Human Cluster 19.10				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr7				r chr1				c chr1			
FG_FBL_FCGBP	D	F	O	FG_FBL_FCGBP	D	F	O	FG_Fbl_Fcgbp	D	F	O	FG_Fbl_Fcgbp	D	F	O	FG_FBL_FCGBP	D	F	O
PSMC4				PSMC4				Psmc4				Psmc4				PSMC4			
ZNF546	K	22	+	ZNF546	K	22	+	C130069109Rik	K	18	-	LOC685610	K	15	-	ZNF546	K	22	+
ZNF780B	K	23	-					Zfp780b	K	22	-	LOC685578	K	9	+				
LOC284323	K	17	-					1700049G17Rik	-	22	-	LOC499110	K	8	+				
								4732475C15Rik	K	18	+								
								Zfp59	K	16	+								
								4933426I21Rik	K	18	+								
								Zfp60	K	19	+								
								C030039L03Rik	K	17	+								
FG:MAP3K10, TTC9B				FG:MAP3K10, TTC9B				FG_Map3k10, Ttc9b				FG_Map3k10, Ttc9b				FG MAP3K10, TTC9B			
CNTD2				CNTD2				Cntd2				Cntd2				CNTD2			

Human Cluster 19.11				Chimpanzee				Mouse				Rat				Dog			
h chr19				c chr19				m chr7											
FG:KCNN4	D	F	O	FG:KCNN4	D	F	O	FG_Kcnn4	D	F	O	FG_Kcnn4	D	F	O	FG:KCNN4	D	F	O
LYPD5				LYPD5				Lypd5				Lypd5				LYPD5			
ZNF283	K	15	+	LOC468907	K	15	+	Zfp180	-	12	+	Zfp180	-	12	+	LOC484460	K	19	-
ZNF404	K	15	-	ZNF45	-	-	-	Zfp112	K	12	+	Zfp112	-	13	+	LOC484461	K	31	-
ZNF45	K	18	-	ZNF221	K	15	+	Zfp235	K	15	+	RGD1560762	K	15	+	LOC484463	K	26	-
ZNF221	K	15		ZNF155	K	12	+	Zfp114	K	12	+	Zfp111	K	19	-	LOC612388	K	15	+
ZNF155	K	12		ZNF230	K	10	+	Zfp111	K	18	-	RGD1564214	K	3	+	LOC613011	K	16	-
ZNF230	K	10		ZNF222	K	10	+	Zfp109	K	17	-	LOC499094	K	8	-				
ZNF222	K	10		ZNF223	K	10	+	Zfp108	K	13	+	Zfp94	K	9	-				
ZNF223	K	10		ZNF224	K	18	+	Zfp93	K	13	+	LOC680802	-	13	-				
ZNF284	K	14		ZNF225	K	47	+	Zfp61	K	8	-								
ZNF224	K	18		ZNF227	K	19	+	Zfp94	K	11	-								
ZNF225	K	18		LOC468909	-	-	-												
ZNF234	K	19		ZNF228	K	30	-												
ZNF226	K	19		ZNF285	K	11	-												
ZNF227	K	19		LOC456109	-	-	-												
ZNF233	K	8		LOC468912	-	-	-												
ZNF235	K	16	-	ZNF180	K	12	-												
ZNF228	K	17	-																
LOC147711	-	-	-																
ZNF180	K	12	-																
FG_CEACAM20				FG_CEACAM20				FG_Ceacam20				FG_Ceacam20				FG_CEACAM20			

Human Cluster 19.12				Chimpanzee				Mouse				Rat				Dog			
h chr19				p chr19				m chr7				r chr1				c chr1			
FG SIGLEC12,SIGLECP11 SIGLEC6,SIGLECP12	D	F	O	FG.SIGLEC12,SIGLECP11 SIGLEC6,SIGLECP12	D	F	O	FG 2210412E05Rik LOC434181,Cd33	D	F	O	FG 2210412E05Rik LOC434181,Cd33	D	F	O	FG.LOC611716,LOC476396 LOC611696,LOC611700	D	F	O
ZNF175	K	15	+	LOC456251	K	15	+	BC043301	K	22	+	RGD1309564	K	11	+	LOC611692	K	20	+
ZNF577	K	8	-	ZNF649	K	10	-	Zfp719	K	15	+					LOC484341	K	37	+
ZNF649	K	10	-	LOC468981	K	16	-	4933405K07Rik	K	11	+					LOC491432	-	9	-
LOC441861	-	-	-	ZNF613	K	2	+									LOC611669	K	12	-
ZNF613	K	12	+	ZNF615	K	17	-									LOC484338	K	50	-
ZNF350	K	8	-	ZNF614	K	11	-									LOC484333	K	12	-
ZNF615	K	19	-	ZNF432	K	16	-									LOC484331	K	16	+
ZNF614	K	11	-	ZNF616	K	21	-									LOC476394	K	17	-
ZNF432	K	17	-	LOC456261	K	25	-									LOC482273	K	13	-
LOC284371	-	-	-	LOC748970	K	10	+									LOC611599	K	16	+
ZNF616	K	21	-	ZNF480	K	12	+									LOC480782	K	18	+
FLJ16287	K	25	-	LOC468984	-	-										LOC611590	K	4	+
ZNF766	K	10	+	LOC468985	-	-										LOC611583	K	14	-
ZNF480	K	12	+	LOC468988	K	17	+									LOC484328	K	14	+
ZNF610	K	9	+	LOC748568	-	8	-									LOC484326	K	34	+
ZNF528	K	15	+	LOC456421	-	11	+									LOC484324	K	17	+
ZNF534	K	17	+	LOC468990	K	24	+									LOC484323	K	11	+
ZNF578	-	12	+	ZNF701	-	-										ZNF331	K	12	-
ZNF808	-	2	+	LOC748607	-	4	+												
ZNF701	K	9	+	ZNF600	-	20	-												
ZNF137	-	5	+	LOC456267	-	-													
ZNF83	-	15	-	ZNF320	K	12	-												
LOC729840	K	14	-	LOC456268	-	-													
ZNF611	K	17	-	ZNF160	-	-													
ZNF600	-	20	-	LOC456426	-	-													
ZNF28	-	18	-	LOC456269	-	-													
ZNF468	K	11	-	LOC456270	-	-													
ZNF320	K	12	-	ZNF468	K	19	+												
LOC388559	-	18	-	ZNF331	K	12	+												
ZNF816	K	15	-																
ZNF702	-	5																	
ZNF160	K	20	-																
ZNF415	K	11	-																
ZNF347	K	20	-																
ZNF665	K	18	-																

ZNF818	-	2	+																				
ZNF677	K	10	-																				
LOC91664	-	27	+																				
ZNF525	K	24	+																				
ZNF765	K	11	+																				
ZNF761	-	19	+																				
ZNF813	K	14	+																				
ZNF331	K	12	+																				
FG LOC284379				*FG LOC284379*				*FG Dc2, Dprx*				*FG Dc2, Dprx*				*FG LOC284379*					
DC2, DPRX				*DC2, DPRX*												*DC2, DPRX*					

Human Cluster 19.13				***Chimpanzee***				***Mouse***				***Rat***				***Dog***			
h chr19				***p chr19***				***m chr7***				***r chr1***				***c chr1***			
FG RPL28,UBE2S	D	F	O	*FG RPL28,UBE2S*	D	F	O	*FG Rpl28, Ube2s*	D	F	O	*FG Rpl28, Ube2s*	D	F	O	*FG RPL28,UBE2S*	D	F	O
ISOC2				*ISOC2*				*Isoc2*				*Isoc2*				*ISOC2*			
ZNF628	-	17	+	ZNF628	-	8	+	Zfp628	-	16	+	Zfp628	-	16	-	LOC484294	-	16	-
KLP1	-	-	+	ZNF579	-	8	-	Zfp579	-	8	-	Znf579	-	8	+	LOC484293	-	8	+
ZNF579	-	8	-	LOC456307	-	11	-	Fiz1	-	11	-	RGD1306359	-	11	+	LOC484292	-	17	+
FIZ1	-	11	-	LOC469015	-	4	+	Zfp524	-	4	+	RGD1565485	-	4	-	LOC611259	-	3	-
ZNF524	-	4	+	LOC456308	-	20	+	6430526N21Rik	-	20	+	LOC691396	-	20	-	LOC611254	-	4	-
ZNF784	-	5	-	LOC737491	-	4	+	Zfp784	-	6	-	RGD1565545	-	5	+	LOC484286	-	7	+
ZNF580	-	3	+	ZNF444	S	4	+	Zfp580	-	3	+	Zfp580	-	3	-	LOC611205	S	4	-
ZNF581	-	4	+	LOC456317	S	5	-	Zfp787	-	7	-	RGD1310536	-	7	+	LOC484282	K	12	+
ZNF787	-	7	-	LOC738992	S	5	+	Zfp444	S	4	+	Zfp444	S	4	-	LOC484278	K	17	-
ZNF444	S	4	+	LOC739105	S	5	-	Zfp371	S	5	+	RGD1559143	S	5	-	LOC611144	2K	14	-
LOC342933	S	5	-	ZSCAN5	S	5	+	Zfp667	K	14	+	RGD1562044	K	5	+	LOC612850	K	24	-
LOC649137	S	5	+	LOC469022	K	12	+	Zfp583	K	12	-	LOC691269	K	15	-	LOC484275	-	23	+
ZSCAN5	S	5	-	LOC469023	K	10	-	Zfp78	K	11	+	RGD1562235	K	15	-	LOC484269	SK	6	+
LOC646698	S	3	+	LOC469024	K	12	+	Zfp28	K	14	+	LOC691168	-	-	-	LOC476371	S	25	+
ZNF542	-	8	+	ZNF667	K	15	-	Zim2	-	2	-	LOC691166	K	8	+	LOC476369	K	1	-
ZNF582	K	10	-	ZNF471	K	15	+	Zim1	K	11	-	LOC308320	K	8	+	LOC484264	K	70	-
ZNF583	K	12	+	ZFP28	2K	15	+	Peg3	-	12	-	RGD1563793	K	7	+	LOC484257	K	8	+
ZNF667	K	15	-	ZNF470	K	15	+	LOC637583	K	3	+	RGD1308782	K	7	-	LOC484255	S	4	-
ZNF471	K	15	+	ZNF71	K	13	+	5730403M16Rik	K	2	-	LOC691135	K	14	+	LOC491462	-	10	-
ZFP28	2K	15	+	LOC469028	-	14	-	2810409K11Rik	-	-	-	RGD1311663	K	14	+	LOC491461	K	15	-
ZNF470	K	17	+	PEG3	S	12	-	Zfp418	K	14	+	LOC690887	K	19	-	LOC610966	-	10	+
ZNF71	-	13	+	LOC740808	-	2	-	BCO23179	K	8	-	LOC690661	K	19	-	LOC610960	S	4	+
BC37295_3	-	14	-	ZIM3	K	11	-	LOC665507	-	-	-	LOC690543	K	-	+	LOC484250	K	32	+
ZIM2	K	5	-	LOC469033	-	-	-	LOC628297	K	-	-	RGD1560046	K	19	+	LOC484247	K	38	+
PEG3	S	12	-	ZNF264	-	-	-	LOC665609	-	-	-	LOC690514	-	-	-	LOC484242	K	55	+
ZIM3	K	11	-	LOC469036	-	-	-	Zik1	K	9	-	LOC690502	K	16	+	LOC610913	K	29	+
ZNF264	K	13	+	ZNF304	K	-	+	BQ559217	S	-	-	LOC361487	K	19	+	LOC484238	SK	17	+
ZNF805	K	11	+	ZNF547	K	9	+	Gm397	S	4	+	Zfp53	K	15	+	LOC484234	-	12	+
ZNF272	K	11	+	ZNF17	K	17	+	EG545912	S	4	-	Zfp54	K	30	+	LOC476367	(S	11	-
ZNF543	K	13	+	LOC469041	K	10	-	EG545913	S	4	-	prdm7	-	-	-	LOC484226	SK	27	+
ZNF304	K	15	+	ZNF419A	-	-	-	EG665848	S	4	-					LOC484224	K	11	-
ZNF574	K	9	+	ZNF549	K	15	+	EG665853	-	3	+					LOC484222	B	4	+
ZNF548	K	11	+	LOC456330	-	-	-	LOC665902	S	4	+					LOC484220	S	13	+

ZNF175	K	17	+	ZNF416	K	12	-	LOC665913	S	4	+
ZNF749	-	17	+	ZNF211	K	22	+	Zfp551	K	13	-
ZNF772	K	10	-	ZNF551	-	-	-	Zfp606	K	15	+
ZNF419	K	11	+	LOC741520	K	15	-	Zfp329	-	12	-
ZNF773	K	9	+	ZNF256	K	15	-	Zfp110	SKK	5	+
ZNF549	K	15	+	ZNF606	K	16	-	Zfp128	K	7	+
ZNF550	K	8	-	LOC469048	S	-	+	Zscan22	S	8	+
ZNF416	K	12	-	LOC456340	S	4	-	Zfp324	K	9	+
ZIK1	K	9	+	ZNF329	-	12	-	Zfp446	S	3	+
ZNF530	K	13	+	ZNF274	SA	5	+	Zbtb45	B	4	-
ZNF134	-	10	+	LOC456344	-	-	-	Mzf1	S	13	-
ZNF211	K	12	+	ZNF8	K	7	+				
ZSCAN4	S	4	+	LOC742219	S	8	+				
ZNF551	K	16	+	ZNF497	-	14	-				
ZNF154	K	10	-	LOC456346	-	8	-				
ZNF671	K	10	-	ZNF584	K	8	+				
ZNF776	-	10	+	ZNF132	-	17	-				
ZNF586	K	10	+	LOC742660	K	-	+				
ZNF552	K	8	-	ZNF324	-	17	+				
ZNF587	K	13	+	ZNF446	S	3	+				
ZNF814	-	-	-	LOC456352	-	-	-				
ZNF417	K	13	-	ZNF42	S	13	-				
ZNF418	K	16	-								
ZNF256	K	15	-								
ZNF606	K	16	-								
ZSCAN1	S	3	+								
ZNF135	K	16	+								
ZNF447	S	2	-								
ZNF329	-	12	-								
ZNF274	KSK	5	+								
ZNF544	K	13	+								
ZNF8	K	7	+								
HKR2	S	8	+								
ZNF497	-	14	-								
LOC116412	-	8	-								
ZNF584	K	8	+								
ZNF132	K	18	-								
ZNF324B	K	9	+								
ZNF324	K	9	+								
ZNF446	S	3	+								
ZNF499	B	4	-								
ZNF42	S	13	-								
ZNF93	K	17	-								
FG LOC653789				*FG LOC653789*				*FG LOC653789*			
								FG LOC653789			
											FG LOC653789

Human Cluster 20.1				Chimpanzee				Mouse				Rat				Dog			
h chr20				p chr20				m chr2				r chr3				c chr24			
FG SNX5	D	F	O	FG SNX5	D	F	O	FG Snx5	D	F	O	FG Snx5	D	F	O	FG SNX5	D	F	O
PTMAP3				PTMAP3				Ptmap3				Ptmap3				PTMAP3			
ZNF339	-	4	-	ZNF339	-	4	-	Zfp339	-	4	-	Zfp339	-	4	-	ZNF339	-	4	-
ZNF133	K	15	+	ZNF133	K	15	+	Zfp133	K	15	+	Zfp133	K	15	+				
FG POLR3F, RBBP9				FG POLR3F, RBBP9				FG Polr3f, Rbbp9				FG Polr3f, Rbbp9				FG POLR3F, RBBP9			

Human Cluster 20.2				Chimpanzee				Mouse				Rat				Dog			
h chr20				p chr20				m chr2				r chr3				c chr24			
FG NEVRL2, RTSA	D	F	O	FG NEVRL2, RTSA	D	F	O	FG Nevrl2, Rtsa	D	F	O	FG Nevrl2, Rtsa	D	F	O	FG NEVRL2, RTSA	D	F	O
PLTP				PLTP				Pltp				Pltp				PLTP			
ZNF335	-	13	-	ZNF335	-	13	-	Zfp335	-	13	-	Zfp335	-	13	-	LOC485904	-	13	-
ZNF663	-	1	-	LOC742834	-	1	-	Zfp334	K	14	-	Zfp334	K	14	-	ZNF334	K	14	-
ZNF334	K	14	-	ZNF334	K	14	-												
FG SLC13A3, TP53RK				FG SLC13A3, TP53RK				FG Tp53rk				FG Tp53rk				FG SLC13A3, TP53RK			
SLC2A10, EYA2				SLC2A10, EYA2				Slc2a10, Eya2				Slc2a10, Eya2				SLC2A10, EYA2			

Human Cluster 20.3				Chimpanzee				Mouse				Rat				Dog			
h chr20				p chr20				m chr2				r chr3				c chr24			
FG HNFATC2, ATP9A	D	F	O	FG HNFATC2, ATP9A	D	F	O	FG Hnfatc2, Atp9a	D	F	O	FG Hnfatc2, Atp9a	D	F	O	FG HNFATC2, ATP9A	D	F	O
SALL4	-	7	-	SALL4	-	7	-	Sall4	-	7	-					LOC485931	-	7	-
ZFP64	-	13	-	ZFP64	-	13	-	Zfp64	-	13	-	Zfp64	-	13	-	LOC485932	-	13	-
FG ERP28P, MRPS33P4				FG ERP28P, MRPS33P4				FG Erp28p,Mrps33p4				FG Erp28p,Mrps33p4				FG ERP28P, MRPS33P4			

Human Cluster 21.1				Chimpanzee				Mouse				Rat				Dog			
h chr21				p chr21				m chr16				r chr11				c chr31			
FG HMX1, TMPRSS2	D	F	O	FG HMX1, TMPRSS2	D	F	O	FG Hmx1, Tmprss2	D	F	O	FG Hmx1, Tmprss2	D	F	O	FG HMX1, TMPRSS2	D	F	O
RIPK4				RIPK4				Ripk4				Ripk4				RIPK4			
PRDM15	Se	14	-	PRDM15	Se	14	-	Prdm15	Se	14	-	Prdm15	Se	14	-	LOC610905	Se	14	-
ZNF295	B	6	-	ZNF295	B	6	-	Zfp295	B	6	-	Zfp295	B	6	-	LOC487775	B	6	-
FG UMODL1, ABCG1				FG UMODL1, ABCG1				FG Umodl1, Abcg1				FG Umodl1, Abcg1				FG UMODL1, ABCG1			
TFF3, TFF2				TFF3, TFF2				Tff2				Tff2				TFF3, TFF2			

Human Cluster 22.1				Chimpanzee				Mouse				Rat				Dog			
h chr22				p chr22				m chr10				r chr10				c chr16			
FG UMODL1, ABCG1	D	F	O	FG UMODL1, ABCG1	D	F	O	FG Umodl1, Tff3	D	F	O	FG Umodl1, Tff3	D	F	O	FG UMODL1, ABCG1	D	F	O
TFF3				TFF3												TFF3			
SUHW2	-	1	-	SUHW2	-	1	-	Suhw2	-	1	-								
SUHW1	-	1	-	SUHW1	-	1	-												
FG IGLV2-34, IGLV2-33				FG IGLV2-34, IGLV2-33				FG Pom121l1				FG Pom121l1				FG IGLV2-34, IGLV2-33			
POM121L1,BCRL4				POM121L1,BCRL4				Bcrl4				Bcrl4				POM121L1,BCRL4			

Human Cluster X.1				Chimpanzee				Mouse				Rat				Dog			
h chrX				p chrX				m chrX				r chrX				c chrX			
FG UBE1,PCTK1	D	F	O	FG UBE1,PCTK1	D	F	O	FG Ube1x, Pctk1	D	F	O	FG Ube1x, Pctk1	D	F	O	FG UBE1, PCTK1	D	F	O
USP11				USP11				Usp11				Usp11				USP11			
ZNF157	K	12	+	ZNF157	K	12	+	Zfp182								LOC612509	K	12	+
ZNF41	K	18	-	ZNF41	K	18	-	D930016N04Rik								LOC480899	K	18	-
ZNF81	K	13	+	ZNF81	K	13	+									LOC491863	K	13	+
ZNF182	K	15	-	ZNF182	-	15	-									LOC491864	K	14	-
ZNF630	K	13	-	LOC473594	-	-	-									LOC491865	K	13	-
FG SSX6,psiSSX2				FG.SSX6,psiSSX2				FG Ssxa1				FG Ssxa1				FG.SSX6,psiSSX2			
LOC653317				LOC653317				Ssxa2				Ssxa2				LOC653317			

Human Cluster X.2				Chimpanzee				Mouse				Rat				Dog			
h chrX				p chrX				m chr				r chr				c chr			
FG SPIN2A, FAAH2	D	F	O	FG SPIN2A, FAAH2	D	F	O	FG Spin2a	D	F	O	FG Spin2a	D	F	O	FG SPIN2A, FAAH2	D	F	O
								Faah2				Faah2							
ZXDB	-	9	+	ZXDB	-	9	+									LOC491912	-	9	+
ZXDA	-	9	-	ZXDA	-	9	-												
FG KRT8P17,				FG KRT8P17,				FG Krt8p17				FG Krt8p17				FG KRT8P17,			
LOC653588				LOC653588												LOC653588			

Human Cluster X.3				Chimpanzee				Mouse				Rat				Dog			
h chrX				p chrX				m chrX				r chrX				c chrX			
FG CXorf48,LOC728470	D	F	O	FG CXorf48,LOC728470	D	F	O	FG CXorf48	D	F	O	FG CXorf48	D	F	O	FG CXorf48,LOC728470	D	F	O
LOC650024				LOC650024												LOC650024			
ZNF75	SK	5	-	ZNF75	SK	5	-	Zfp449	S	7	+					LOC611850	SK	5	+
ZNF449	S	7	+	ZNF449	S	7	+									LOC492160	S	7	+
FG DDX26B				FG DDX26B				FG Ddx26b				FG Ddx26b				FG DDX26B			

Human Cluster X.4				Chimpanzee				Mouse				Rat				Dog			
h chrX				p chrX				m chrX				r chrX				c chrX			
FG ANMAGA,MAGEA1	D	F	O	FG ANMAGA,MAGEA1	D	F	O	FG Anmaga Magea1	D	F	O	FG Anmaga Magea1	D	F	O		D	F	O
ZNF275	-	11	+	ZNF275	-	11	+	Zfp275	-	11	+	Zfp275	-	11	+	LOC492233	-	19	+
LOC139735	K	8	+	LOC739968	-	8	+	Zfp92	K	9	+								
FG TREX2, UCHL5IP				FG TREX2, UCHL5IP				FG Trex2, Uchl5ip				FG Trex2, Uchl5ip				FG TREX2, UCHL5IP			

Human Cluster Y.1				Chimpanzee				Mouse	Rat	Dog
h chrY				p chrY						
FG HTFFY4B,BPY2B	D	F	O	FG HTFFY4B,BPY2B	D	F	O			
LOC392603	-	-	-		-	-	-			
LOC442486	-	-	-		-	-	-			
FG BPY2C, TTTY4C				FG BPY2C, TTTY4C						

Supplementary Table S4

Comprehensive catalogue of the C2H2-ZNF genes from the 81 human clusters and their syntenically homologous clusters from other mammalian genomes (Chimpanzee, Mouse, Rat and Dog).

For the 81 human (***h***) C2H2-ZNF clusters and their corresponding syntenically homologous clusters in chimpanzee (***p***), mouse (***m***), rat (***r***) and dog (***c***), this table provides the cluster number, the position on the chromosome, the flanking genes (*FG*), the names of the C2H2-ZNF from the cluster (in bold), the domain associated (D) (K = KRAB, S = SCAN, S-K= SCAN-KRAB, B = BTB, H = HOMEO, Se = SET and ,'Φ ' = no domain associated), the number of zinc finger motifs present (F), the orientation (O).

Chapter 3. DISCUSSION

Many studies in biology focus on the extensive similarities between the genomes of human and model organisms, to extract insights into the molecular mechanisms and aetiology of human diseases. Our investigation of the C2H2-ZNF gene family in mammals reveals that there is an extensive variation of the C2H2-ZNF gene content and genomic organization as well as the domain composition of orthologous genes among species. In addition, our study is the first to provide a clear demonstration of the important contribution of gene loss in the evolution of C2H2-ZNF family and to demonstrate the rapid evolution of C2H2-ZNF genes that occurs between related species, our observations at the genomic scale provide insights into C2H2-ZNF gene evolution that confirm conclusions drawn from smaller-scale studies on individual genes, clusters and C2H2-ZNF subfamilies.

The major contributions of our study are:

i. The extensive analysis of all the C2H2-ZNF genes in the human genome.
ii. A comprehensive and systematic analysis of all the human C2H2-ZNF clusters and the identification of their syntenically homologous counterparts in other mammalian genomes.
iii. The distinction of species-specific expansion and loss in C2H2-ZNF clusters and genes in all mammals.
iv. The identification of variation in the number of zinc finger motifs and the presence or absence of the conserved N-terminal domains associated with C2H2-ZNF mammalian orthologs.

v. The tracing back of different evolutionary patterns of the C2H2-ZNF gene family within primates and rodents.

vi. The establishment of a model reconstructing the history and evolution of the SCAN, SCAN-KRAB and KRAB subfamilies.

In brief, our study reveals that the multiple and independent duplications and losses of C2H2-ZNF genes and their effector domains within different lineages and species has shaped and diversified C2H2-ZNF repertoires in mammals.

3.1 The C2H2-ZNF genes in the human genome

Earlier studies of C2H2-ZNF genes focussed on human chromosome 19 (Eichler, Hoffman et al. 1998; Dehal, Predki et al. 2001; Looman, Abrink et al. 2002; Shannon, Hamilton et al. 2003) and KRAB C2H2-ZNF subfamily in human (Huntley, Baggott et al. 2006). In contrast, our study provides a comprehensive and systematic analysis of all the C2H2-ZNF genes (Supplementary Table S1) in the human genome. We identified and analyzed the organization of 718 C2H2-ZNF genes in the human genome and classified them into different subfamilies of C2H2-ZNF (KRAB-C2H2-ZNF, SCAN-C2H2-ZNF, BTB-C2H2-ZNF and those without a conserved N-terminal domain). We also discovered two new C2H2-ZNF subfamilies, the HOMEO and SET subfamily which have a limited number of members (5 and 2, respectively) possibly due to a more recent appearance or to a different rate of duplication and loss.

Consistent with previous reports (Rousseau-Merck, Koczan et al. 2002; Huntley, Baggott et al. 2006), we observed a massive clustering of the C2H2-ZNF on the human genome. More than 70% of the genes are organized into clusters on the human genome. However, in addition to the earlier reported clusters on human chromosome 19, (Venter, Adams et al. 2001; Huntley, Baggott et al. 2006), we also located a substantial amount of clusters (83%) on the other chromosomes of the human genome (Supplementary Table S2). Interestingly, the distribution of C2H2-ZNF genes is positively biased toward chromosome 19, harbouring 40% of all C2H2-ZNF genes in humans. Most of the human C2H2-ZNF genes are organized into clusters (500) with more than 60% of these clusters containing intermixed sets of genes from different subfamilies (Supplementary table S3).

The above observations were only possible through the study of all the C2H2-ZNF sub-families at the whole genome level.

3.2 Variation in the numbers of C2H2-ZNF genes in mammalian clusters

A systematic and comprehensive analysis of the human C2H2-ZNF clusters and its syntenic counterparts in the chimpanzee, mouse, rat and dog genomes, allowed us to gain insights into the evolution of all the C2H2-ZNF gene subfamilies in mammals (Supplementary Table S4). The criterion to identify homologous clusters in syntenic regions was based on the flanking genes identified for each human cluster. Interestingly, this analysis revealed a high variation in the number of C2H2-ZNF genes within

syntenically homologous clusters of mammals. Considering primates, humans have 518 C2H2-ZNF forming 81 clusters whereas chimpanzee has only 397 C2H2-ZNF organized into 79 clusters. This suggests that almost all the C2H2-ZNF clusters in human have a syntenic counterpart in chimpanzee. However, humans have 30 % more C2H2-ZNF genes within the identified clusters than chimpanzee implying that C2H2-ZNF genes are evolving differently within the primate lineage. A similar pattern was observed within the rodent lineage, where mouse and rat have 232 and 172 C2H2-ZNF organized into 62 and 58 clusters, respectively. A differential expansion of C2H2-ZNF genes, particularly striking in primates was evident in mammals (human>chimpanzee>mouse>dog>rat for the number of genes within clusters and human>chimpanzee>mouse>rat>dog for the number of clusters) (Figure 2 and Supplementary Table S4). A closer look at the individual syntenic clusters in mouse, rat and dog indicates many cases where dog has more number of genes than rodents and more specifically than rat (Supplementary Figure S2).

Our study indicates that C2H2-ZNF genes are indeed rapidly evolving genes as evident for example within the primate and rodents lineages. The differential expansion observed in the various species may be accounted both by differential duplication and/or loss. If the high numbers of C2H2-ZNF genes found in human as compared to chimpanzee suggest an expansion specific to human through tandem duplication, it seems that this difference is not solely due to duplications but also involves loss in chimpanzee as seen in many gene families (Fortna, Kim et al. 2004). In agreement with this, there are more pseudogenes in chimpanzee clusters (as annotated in Genbank) than in the corresponding human clusters. Thus, the variation in the numbers of C2H2-ZNF genes observed within

primates could be attributed to both gene duplication and loss due to deletion or pseudogenization, gain being more predominant in human and loss in chimpanzee. Interestingly, a variation in the number of C2H2-ZNF genes is also evident in rodents. However, in this case in almost all the clusters, mouse has either a higher or equal number of C2H2-ZNF as compared to rat. Altogether, our study suggests that in addition to lineage- or species-specific increase in the numbers of the C2H2-ZNF genes, loss of these genes has also played a very important role in the evolution of this gene family. The evaluation of the relative contribution of gene duplication and loss requires detailed phylogenetic studies.

3.3 Evolution of C2H2-ZNF genes in mammals through differential expansion and loss

Phylogenetic analyses of human C2H2-ZNF clusters with their syntenic counterparts from other mammals provided a better estimation of the relative contribution of gene duplication and loss in the analyzed clusters.

For example, the phylogenetic analysis of the C2H2-ZNF genes from human cluster 19.12 and its syntenically homologous clusters in mammals combined with the physical maps of these clusters gives us insights into the gene rearrangement mechanisms that could have taken place during evolution. Consistent with previous individual reports of lineage-specific expansion, more specifically of KRAB C2H2-ZNF genes (Dehal, Predki et al. 2001; Shannon, Hamilton et al. 2003; Huntley, Baggott et al. 2006), a primate lineage-specific expansion but also a dog specific expansion and a mouse duplication of C2H2-

ZNF genes were clearly identified . In all species, tandem duplication was found responsible for the species-specific increase in the number of C2H2-ZNF genes as confirmed by the fact that the genes that group together in the tree are almost always physically clustered together in the cluster on the chromosome. Furthermore, the orientations of the genes belonging to the same clade in the phylogenetic tree and their orthologs were almost always the same. In a few instances, however, the orientations of the genes belonging to the same phylogenetic clade were different and genes within syntenic clusters were inverted in a few instances, revealing a lot of possible gene rearrangements.

Clear evidence of gene loss was also obtained by analyzing cluster 19.12 and other clusters. Considering that rodents are evolutionarily more related to primates than dog, an absence in rodents of C2H2-ZNF clusters or genes, that are present in primates and dog, suggests a loss in rodents. Several examples of genes loss in rodents were obtained. The phylogenetic analysis also indicated loss of genes in chimpanzee by pseudogenization as suggested above.

Altogether our studies indicate a predominant role of gene gain by tandem duplication over gene loss for the evolution of C2H2-ZNF genes in mammals. It should however be pointed out that more definitive conclusions about the pseudogene status of the various C2H2-ZNF genes and on the role of these genes requires detailed functional investigation of the individual genes.

3.4 Evolution of the C2H2-ZNF genes through duplication or loss of zinc finger and N-terminal effector motifs

In accordance with a previous study on the average number of zinc finger motifs from a few plant (1), yeast(1.5), nematode (2.5), insect (3.5) and human (8) C2H2-ZNF (Venter, Adams et al. 2001), an in depth analysis of the zinc finger motifs associated with the C2H2-ZNF genes found in clusters in human and their syntenic genes in chimpanzee, mouse, rat and dog indicated that there is a significant variation in the number of zinc finger motifs associated with C2H2-ZNF genes in these mammalian species. Noticeably, the C2H2-ZNF genes from dog were found to encode a higher average of zinc finger motifs as compared to the other mammals studied. It is possible that an increase in the number of fingers within genes could confer advantageous additional functionality to the C2H2-ZNF genes through a diversification of the possible nucleic acid and protein interactions.

We also observed a variation in the presence of N-terminal effector motifs, such as SCAN or KRAB among orthologs, accounted by either gain or loss of these motifs. However, loss of N-terminal effector domains by sequence degeneration was confirmed in several cases in our study. A thorough analysis of the exon-intron structure of C2H2-ZNF genes indicated a typical conserved exon-intron organization for C2H2-ZNF genes associated with a SCAN-KRAB. Based on these observations, we propose a model of evolution of C2H2-ZNF sub-families involving independent gain events of a SCAN and KRAB domain each by an exon-shuffling mechanism and subsequent gene duplications and loss of effector motifs by deletion or degeneration of the SCAN and/or KRAB domain.

3.5 Birth and Death model of evolution

It was suggested from a study of a few chromosome 19 C2H2-ZNF clusters that C2H2-ZNF genes evolve by positive selection (Schmidt and Durrett 2004). Our results and analyses of the human C2H2-ZNF clusters and their syntenic counterparts in other mammals suggests a "Birth and Death" model of evolution similar to that proposed by Nei and coll. (Nei, Gu et al. 1997; Nei 2000) (See Figure 7B). According to this model, new genes are created by duplication including tandem duplication and block gene duplication (birth). While some of them might acquire a new function and thus diverge functionally, the others may remain relatively unchanged in the genome for a long time. Again others become pseudogenes following deleterious mutations or get deleted from the genome (death through inactivation or elimination). Though functional information is known for only a handful of C2H2-ZNF proteins, the variations in the numbers C2H2-ZNF genes that we found throughout the evolution of mammals and our phylogenetic analysis points to duplication and loss as a guiding force in the evolution of these genes.

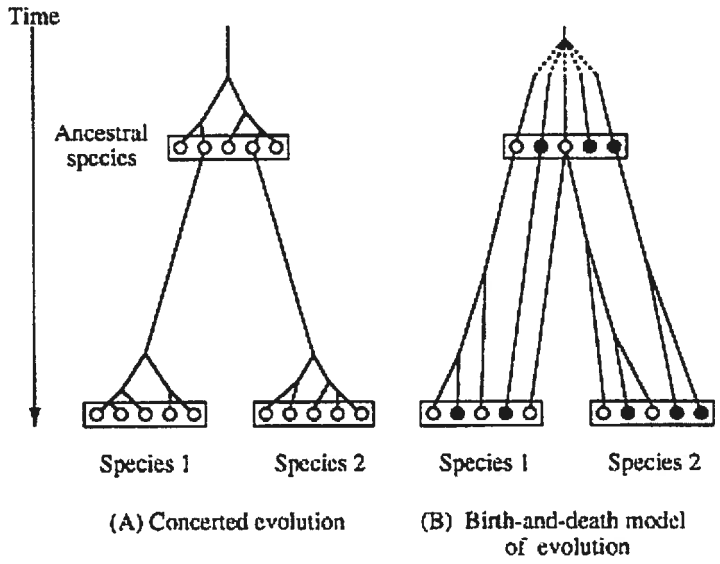


Figure 7: Birth-and-death model of evolution.

The figure shows the two models associated with the evolution of multigene families. Open circles represent functional genes and closed circles represent pseudogenes.

(A) In concerted evolution, related genes belonging to the ancestral species evolve in a concerted manner rather than independently in both Species 1 and Species 2.

(B) Birth and Death Model of evolution, where the genes evolve differently by duplication, few are maintained in the genome for longer, while the others are deleted or become pseudogenes.

3.6 C2H2-ZNF gene family: An analogy with the Olfactory Receptor gene family

The olfactory receptor genes constitute the largest mammalian gene family with more than 1000 members in human. However, 60% of these genes are pseudogenes. In contrast to this, the olfactory receptor gene family in mouse comprises of roughly the same number of genes as human, though the number of pseudogenes is only 20% (Glusman, Yanai et al. 2001; Niimura and Nei 2003; Niimura and Nei 2005). Comparative analyses of this gene family in human, mouse and non-human primates have revealed that differential expansion and loss have guided the evolution of this gene family (Sharon, Glusman et al. 1999; Lapidot, Pilpel et al. 2001). However, human counterparts have accumulated a lot of mutations, leading to the numerous pseudogenes in comparison to mouse or any other non-human primate. This is associated with the reduced chemosensory capacity in humans.

The olfactory receptor and the C2H2-ZNF gene families show similar patterns of evolution. Differential gene expansion and loss have played an important role in the evolution of both gene families in mammals. However, in contrast to the olfactory receptor genes, C2H2-ZNF genes apparently do not accumulate pseudogenes in humans irrespective of their large number. Studies on human C2H2-ZNF clusters have indicated that these genes are rapidly evolving through positive selection and may acquire new functions after duplication (Schmidt and Durrett 2004; Huntley, Baggott et al. 2006).

By making a correlation between the number of olfactory genes in human and mouse and their functions in the respective organisms, we do understand that a reduced chemosensory dependence in primates and non-human primates as compared to mouse can

be responsible for the large number of pseudogenes in human. Presently, the lack of large scale analysis of the expression profile and function of C2H2-ZNF genes preclude the establishment of such type of conclusions. The extremely high number of human C2H2-ZNF genes, the species-specific expansions and loss in mammals leading to differential evolution within primates and rodents, when put in perspective with functional information of these proteins could give us interesting insights into the evolution of this gene family.

3.7 A few concerns to the study

We must acknowledge here three possible sources of bias in our study.

First, errors in reporting the number of genes due to improper sequencing and annotation in the available databases could be a primary source of bias to our study. However, we did only consider genomes like human, chimpanzee, mouse, rat and dog which are >94% complete to minimize significantly this source of bias.

A second concern is that the loss of C2H2-ZNF clusters or genes that we see among the species could be due to the fact that the genes were dispersed onto different chromosomes due to translocation. Though we do accept this as a possibility, we conduct an extensive analysis to rule out this possibility for the clusters we studied in depth. For example, in group I of the phylogenetic tree (Figure 5; Article) of the human cluster 19.12 and its syntenic clusters in chimpanzee, mouse, rat and dog, we observe three orthologs (hZNF331, pZNF331 and cZNF331) from human, chimpanzee and dog. There is no rodent counterpart for these genes, which suggests a loss in rodents. To rule out the possibility that these genes were dispersed onto other regions of the genome by translocation, we conduct

an extensive homology TBLASTN search of the mouse and rat genomes using each of the three orthologs from human, chimpanzee and dog as a query. For the three queries, the top most blast hit was Zfp14 from mouse and LOC97124 from rat. We included these two sequences into the dataset used for the phylogenetic analysis (see Methods; Article) of cluster 19.12 in human and its syntenic counterparts in chimpanzee, mouse, rat and dog. The phylogenetic tree revealed that the two mouse and rat sequences group with the three orthologs from human, chimpanzee and dog in group I. However, a closer look at the sequence similarity between these sequences (< 60%) suggests that they cannot be orthologs and the grouping we see could possibly be because of the fact that they were the closest to the query sequences used

The third and final concern is that considering the extremely large numbers of C2H2-ZNF in the human genome, we cannot rule out the possibility of pseudogenes. Though we do not conduct an extensive search to look for possible pseudogenes, an analysis of the open reading frames of the C2H2-ZNF genes considered in this study with their translated sequences suggest that most of them are most likely not pseudogenes. A distribution curve (Figure 8) of the amino acid sequence length of the various C2H2-ZNF genes shows that almost all of the sequences have large open reading frames potentially translated and functional.

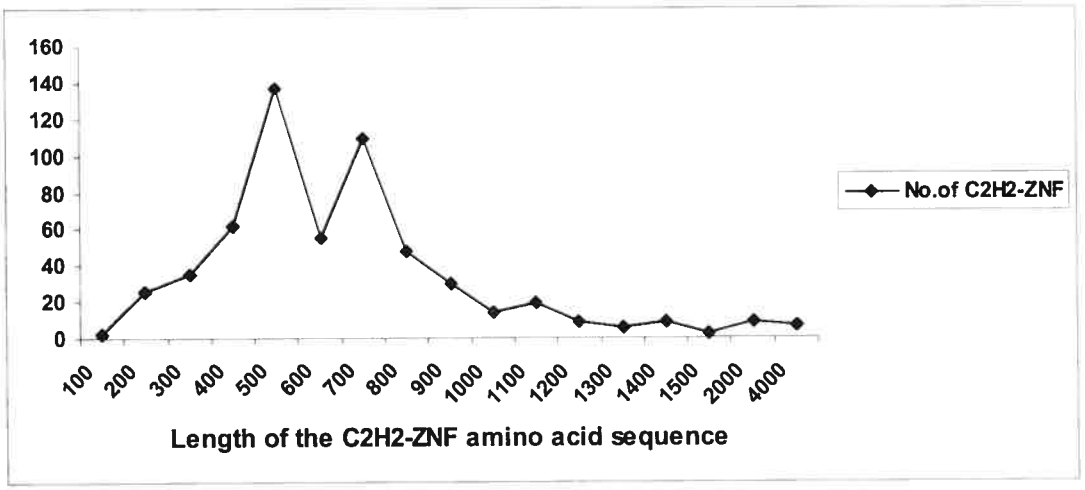


Figure 8: Plot of the amino acid sequence lengths of all the C2H2-ZNF in the human genome

3.8 Merits of the study

Our study provides a comprehensive insight into the evolution of C2H2-ZNF throughout several mammalian genomes. To summarize, the merits of our study are as follows.

- A good range of species, with completely sequenced genomes was considered to analyze the evolution of C2H2-ZNF genes in mammals.
- A stringent phylogenetic analysis of the syntenic clusters in human, chimpanzee, mouse, rat and dog was performed using both maximum likelihood (RAxML) and bayesian (Mr.Bayes) methods. Noticeably, unlike other studies on C2H2-ZNF which use Xfin as an outgroup (Looman, Abrink et al. 2002; Shannon, Hamilton et al. 2003; Huntley, Baggott et al. 2006), we conduct an extensive search to include chicken *(Gallus gallus)* homologs in addition to Xfin as an outgroup. The chicken sequences are considerably closer as an outgroup to the species studied (human, chimpanzee, mouse, rat and dog).

  The phylogenetic relationships we observed between C2H2-ZNF genes in the syntenic clusters from the different species was found consistent with the overall picture of the number of genes in the species and with the physical maps of the clusters.
- A model for the evolutionary relationship of SCAN, SCAN-KRAB and KRAB C2H2-ZNF subfamilies is proposed providing a possible explanation for previously unresolved questions in the field.

3.9 Perspectives

The following studies could be done as a future approach, to what is already known.

- The compilation of comprehensive catalogues of the C2H2-ZNF gene repertoires in chimpanzee, mouse, rat and dog. The detailed comparison of the organization and numbers of C2H2-ZNF in complete repertoires.
- The stringent phylogenetic analysis of these repertoires to gain insights into the various detailed mechanisms, which have taken place during the evolution of this gene family in mammals.
- The more detailed analysis of the physical mapping of genes within clusters to gain insight into the molecular mechanisms involved in the expansion of these genes. This could include the analysis of the possible repeated sequences that are bordering these C2H2-ZNF and that may be involved in this phenomenon, the analysis of the orientation, distances between genes and exon-intron organisation.
- The more comprehensive study of pseudogenes

Clearly, more detailed bio-informatics and functional studies are still required for a better understanding of the driving force behind the expansion of C2H2-ZNF genes in mammals.

REFERENCES

Bellefroid, E. J., P. J. Lecocq, et al. (1989). "The human genome contains hundreds of genes coding for finger proteins of the Kruppel type." DNA **8**(6): 377-87.

Bellefroid, E. J., J. C. Marine, et al. (1993). "Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells." Embo J **12**(4): 1363-74.

Bellefroid, E. J., D. A. Poncelet, et al. (1991). "The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins." Proc Natl Acad Sci U S A **88**(9): 3608-12.

Benn, A., M. Antoine, et al. (1991). "Primary structure and expression of a chicken cDNA encoding a protein with zinc-finger motifs." Gene **106**(2): 207-12.

Bertrand, D. and O. Gascuel (2005). "Topological rearrangements and local search method for tandem duplication trees." IEEE/ACM Trans Comput Biol Bioinform **2**(1): 15-28.

Birtle, Z. and C. P. Ponting (2006). "Meisetz and the birth of the KRAB motif." Bioinformatics **22**(23): 2841-5.

Bouhouche, N., M. Syvanen, et al. (2000). "The origin of prokaryotic C2H2 zinc finger regulators." Trends Microbiol **8**(2): 77-81.

Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." Mol Biol Evol **17**(4): 540-52.

Chung, H. R., U. Schafer, et al. (2002). "Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila." EMBO Rep **3**(12): 1158-62.

Collins, T., J. R. Stone, et al. (2001). "All in the family: the BTB/POZ, KRAB, and SCAN domains." Mol Cell Biol **21**(11): 3609-15.

Darwin, C. (1837). The First Notebook on Transmutation of Species.

Dehal, P., P. Predki, et al. (2001). "Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution." Science **293**(5527): 104-11.

Demuth, J. P., T. D. Bie, et al. (2006). "The evolution of Mammalian gene families." PLoS ONE **1**: e85.

Edelstein, L. C. and T. Collins (2005). "The SCAN domain family of zinc finger transcription factors." Gene **359**: 1-17.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-7.

Eichler, E. E., S. M. Hoffman, et al. (1998). "Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12." Genome Res **8**(8): 791-808.

Elemento, O. and O. Gascuel (2002). "An efficient and accurate distance based algorithm to reconstruct tandem duplication trees." Bioinformatics **18 Suppl 2**: S92-9.

Elemento, O., O. Gascuel, et al. (2002). "Reconstructing the duplication history of tandemly repeated genes." Mol Biol Evol **19**(3): 278-88.

Fitch, W. M. (2000). "Homology a personal view on some of the problems." Trends Genet **16**(5): 227-31.

Fortna, A., Y. Kim, et al. (2004). "Lineage-specific gene duplication and loss in human and great ape evolution." PLoS Biol **2**(7): E207.

Francis Darwin, A. C. S. (1903). More letters of Charles Darwin: A record of his work in a series of hitherto unpublished letters.

Frankel, A. D., J. M. Berg, et al. (1987). "Metal-dependent folding of a single zinc finger from transcription factor IIIA." Proc Natl Acad Sci U S A **84**(14): 4841-5.

Friedman, J. R., W. J. Fredericks, et al. (1996). "KAP-1, a novel corepressor for the highly conserved KRAB repression domain." Genes Dev **10**(16): 2067-78.

Fuchs, T., G. Glusman, et al. (2001). "The human olfactory subgenome: from sequence to structure and evolution." Hum Genet **108**(1): 1-13.

Germain-Desprez, D., M. Bazinet, et al. (2003). "Oligomerization of transcriptional intermediary factor 1 regulators and interaction with ZNF74 nuclear matrix protein revealed by bioluminescence resonance energy transfer in living cells." J Biol Chem **278**(25): 22367-73.

Gertz, E. M., Y. K. Yu, et al. (2006). "Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST." BMC Biol **4**: 41.

Gilad, Y., O. Man, et al. (2005). "A comparison of the human and chimpanzee olfactory receptor gene repertoires." Genome Res **15**(2): 224-30.

Glusman, G., A. Bahar, et al. (2000). "The olfactory receptor gene superfamily: data mining, classification, and nomenclature." Mamm Genome **11**(11): 1016-23.

Glusman, G., I. Yanai, et al. (2001). "The complete human olfactory subgenome." Genome Res **11**(5): 685-702.

Grondin, B., M. Bazinet, et al. (1996). "The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix." J Biol Chem **271**(26): 15458-67.

Hamilton, A. T., S. Huntley, et al. (2003). "Lineage-specific expansion of KRAB zinc-finger transcription factor genes: implications for the evolution of vertebrate regulatory networks." Cold Spring Harb Symp Quant Biol **68**: 131-40.

Hamilton, A. T., S. Huntley, et al. (2006). "Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes." Genome Res **16**(5): 584-94.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-5.

Huntley, S., D. M. Baggott, et al. (2006). "A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors." Genome Res **16**(5): 669-77.

Kim, S. S., Y. M. Chen, et al. (1996). "A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins." Proc Natl Acad Sci U S A **93**(26): 15299-304.

Klug, A. and D. Rhodes (1987). "Zinc fingers: a novel protein fold for nucleic acid recognition." Cold Spring Harb Symp Quant Biol **52**: 473-82.

Krebs, C. J., L. K. Larkins, et al. (2003). "Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression." Genes Dev **17**(21): 2664-74.

Krishna, S. S., I. Majumdar, et al. (2003). "Structural classification of zinc fingers: survey and summary." Nucleic Acids Res **31**(2): 532-50.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lapidot, M., Y. Pilpel, et al. (2001). "Mouse-human orthology relationships in an olfactory receptor gene cluster." Genomics **71**(3): 296-306.

Lee, M. S., G. P. Gippert, et al. (1989). "Three-dimensional solution structure of a single zinc finger DNA-binding domain." Science **245**(4918): 635-7.

Li, W., L. Jaroszewski, et al. (2001). "Clustering of highly homologous sequences to reduce the size of large protein databases." Bioinformatics **17**(3): 282-3.

Looman, C., M. Abrink, et al. (2002). "KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution." Mol Biol Evol **19**(12): 2118-30.

Looman, C., L. Hellman, et al. (2004). "A novel Kruppel-Associated Box identified in a panel of mammalian zinc finger proteins." Mamm Genome **15**(1): 35-40.

Margolin, J. F., J. R. Friedman, et al. (1994). "Kruppel-associated boxes are potent transcriptional repression domains." Proc Natl Acad Sci U S A **91**(10): 4509-13.

Mark, C., M. Abrink, et al. (1999). "Comparative analysis of KRAB zinc finger proteins in rodents and man: evidence for several evolutionarily distinct subfamilies of KRAB zinc finger genes." DNA Cell Biol **18**(5): 381-96.

Melnick, A., G. Carlile, et al. (2002). "Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors." Mol Cell Biol **22**(6): 1804-18.

Messina, D. N., J. Glasscock, et al. (2004). "An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression." Genome Res **14**(10B): 2041-7.

Miller, J., A. D. McLachlan, et al. (1985). "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes." Embo J **4**(6): 1609-14.

Moosmann, P., O. Georgiev, et al. (1996). "Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1." Nucleic Acids Res **24**(24): 4859-67.

Moreira, D. and F. Rodriguez-Valera (2000). "A mitochondrial origin for eukaryotic C2H2 zinc finger regulators?" Trends Microbiol **8**(10): 448-50.

Nei, M., and Kumar S. (2000). Molecular Evolution and Phylogenetics, New York: Oxford University Press.

Nei, M., X. Gu, et al. (1997). "Evolution by the birth-and-death process in multigene families of the vertebrate immune system." Proc Natl Acad Sci U S A **94**(15): 7799-806.

Niimura, Y. and M. Nei (2003). "Evolution of olfactory receptor genes in the human genome." Proc Natl Acad Sci U S A **100**(21): 12235-40.

Niimura, Y. and M. Nei (2005). "Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice." Gene **346**: 13-21.

Ohta, T. (2000). "Evolution of gene families." Gene **259**(1-2): 45-52.

Omichinski, J. G., G. M. Clore, et al. (1990). High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. Biochemistry. **29:** 9324-34.

Omichinski, J. G., G. M. Clore, et al. (1992). "High-resolution solution structure of the double Cys2His2 zinc finger from the human enhancer binding protein MBP-1." Biochemistry **31**(16): 3907-17.

Pabo, C. O. and R. T. Sauer (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem. **61:** 1053-95.

Page, R. D. and M. A. Charleston (1997). "From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem." Mol Phylogenet Evol **7**(2): 231-40.

Parraga, G., S. J. Horvath, et al. (1988). "Zinc-dependent structure of a single-finger domain of yeast ADR1." Science **241**(4872): 1489-92.

Pengue, G., V. Calabro, et al. (1994). "Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins." Nucleic Acids Res **22**(15): 2908-14.

Pengue, G. and L. Lania (1996). "Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements." Proc Natl Acad Sci U S A **93**(3): 1015-20.

Quignon, P., E. Kirkness, et al. (2003). "Comparison of the canine and human olfactory receptor gene repertoires." Genome Biol **4**(12): R80.

Rhodes, D. and A. Klug (1993). "Zinc fingers." Sci Am **268**(2): 56-9, 62-5.

Roberts, S. G. (2000). "Mechanisms of action of transcription activation and repression domains." Cell Mol Life Sci **57**(8-9): 1149-60.

Rosati, M., M. Marino, et al. (1991). "Members of the zinc finger protein gene family sharing a conserved N-terminal module." Nucleic Acids Res **19**(20): 5661-7.

Rousseau-Merck, M. F., D. Koczan, et al. (2002). "The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in EnsEMBL." Cytogenet Genome Res **98**(2-3): 147-53.

Ruiz i Altaba, A., H. Perry-O'Keefe, et al. (1987). "Xfin: an embryonic gene encoding a multifingered protein in Xenopus." Embo J **6**(10): 3065-70.

Sander, T. L., A. L. Haas, et al. (2000). "Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero- and homoprotein associations." J Biol Chem **275**(17): 12857-67.

Schmidt, D. and R. Durrett (2004). "Adaptive evolution drives the diversification of zinc-finger binding domains." Mol Biol Evol **21**(12): 2326-39.

Schuh, R., W. Aicher, et al. (1986). "A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene." Cell **47**(6): 1025-32.

Schumacher, C., H. Wang, et al. (2000). "The SCAN domain mediates selective oligomerization." J Biol Chem **275**(22): 17173-9.

Shannon, M., A. T. Hamilton, et al. (2003). "Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters." Genome Res **13**(6A): 1097-110.

Shannon, M., J. Kim, et al. (1998). "Tandem zinc-finger gene families in mammals: insights and unanswered questions." DNA Seq **8**(5): 303-15.

Sharon, D., G. Glusman, et al. (1999). "Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes." Genomics **61**(1): 24-36.

Sitnikova, T. and C. Su (1998). "Coevolution of immunoglobulin heavy- and light-chain variable-region gene families." Mol Biol Evol **15**(6): 617-25.

Stamatakis, A., T. Ludwig, et al. (2005). "RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees." Bioinformatics **21**(4): 456-63.

Stone, J. R., J. L. Maki, et al. (2002). "The SCAN domain of ZNF174 is a dimer." J Biol Chem **277**(7): 5448-52.

Tang, M., M. Waterman, et al. (2002). "Zinc finger gene clusters and tandem gene duplication." J Comput Biol **9**(2): 429-46.

Theunissen, O., F. Rudt, et al. (1992). "RNA and DNA binding zinc fingers in Xenopus TFIIIA." Cell **71**(4): 679-90.

Thornton, J. W. and R. DeSalle (2000). "Gene family evolution and homology: genomics meets phylogenetics." Annu Rev Genomics Hum Genet **1**: 41-73.

Urrutia, R. (2003). "KRAB-containing zinc-finger repressor proteins." Genome Biol **4**(10): 231.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Williams, A. J., L. M. Khachigian, et al. (1995). "Isolation and characterization of a novel zinc-finger protein with transcription repressor activity." J Biol Chem **270**(38): 22143-52.

Witzgall, R., E. O'Leary, et al. (1994). "The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression." Proc Natl Acad Sci U S A **91**(10): 4514-8.

Wolfe, S. A., L. Nekludova, et al. (2000). "DNA recognition by Cys2His2 zinc finger proteins." Annu Rev Biophys Biomol Struct **29**: 183-212.

Zheng, L., H. Pan, et al. (2000). "Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1." Mol Cell **6**(4): 757-68.