

# The Evolution and Functional Diversification of the Deubiquitinating Enzyme Superfamily

Caitlyn Vlasschaert<sup>1,2,3</sup>, David Cook<sup>2,4</sup>, Xuhua Xia<sup>3,5</sup>, and Douglas A. Gray<sup>1,2,\*</sup>

<sup>1</sup>Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ontario, Canada

<sup>2</sup>The Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

<sup>3</sup>Department of Biology, University of Ottawa, Ontario, Canada

<sup>4</sup>Department of Cellular and Molecular Medicine, University of Ottawa, Ontario, Canada

<sup>5</sup>Ottawa Institute of Systems Biology, Ottawa, Ontario, Canada

\*Corresponding author: E-mail: dgray@ohri.ca.

Accepted: February 6, 2017

## Abstract

Ubiquitin and ubiquitin-like molecules are attached to and removed from cellular proteins in a dynamic and highly regulated manner. Deubiquitinating enzymes are critical to this process, and the genetic catalogue of deubiquitinating enzymes expanded greatly over the course of evolution. Extensive functional redundancy has been noted among the 93 members of the human deubiquitinating enzyme (DUB) superfamily. This is especially true of genes that were generated by duplication (termed paralogs) as they often retain considerable sequence similarity. Because complete redundancy in systems should be eliminated by selective pressure, we theorized that many overlapping DUBs must have significant and unique spatiotemporal roles that can be evaluated in an evolutionary context. We have determined the evolutionary history of the entire class of deubiquitinating enzymes, including the sequence and means of duplication for all paralogous pairs. To establish their uniqueness, we have investigated cell-type specificity in developmental and adult contexts, and have investigated the coemergence of substrates from the same duplication events. Our analysis has revealed examples of DUB gene subfunctionalization, neofunctionalization, and nonfunctionalization.

**Key words:** deubiquitinating enzymes, molecular evolution, innate immunity, DNA repair.

## Introduction

Ubiquitin is a molecular tag, a small protein whose attachment to other proteins can influence their fate. The first role established for ubiquitin conjugation was in the regulation of protein stability, but attachment of ubiquitin was later found to influence the subcellular localization of substrates and their proclivity to interact with signaling complexes (reviewed by Varshavsky 2012). For some functions ubiquitin remains covalently attached as a monomer, but the full multiplicity of functions derives from the potential of ubiquitin to be assembled into chains of various topologies, generating what has been called the “ubiquitin code” (Komander and Rape 2012). The constellation of potential substrates and the varying conditions in which conjugation to any given substrate may or may not be desirable necessitates a specific and responsive enzymatic machinery for ubiquitin attachment and removal. The initial attachment of ubiquitin and assembly of chains is orchestrated by an understandably large repertoire of ubiquitin conjugases

and ligases (acting in concert), and is opposed by a somewhat smaller repertoire of deubiquitinating enzymes (DUBs). The opposing activities of ligases and DUBs can be modulated by posttranslational modifications, allowing rapid adjustments to be made through signaling input. In mammalian cells there are several hundred ubiquitin ligases, and in the order of a hundred DUBs (Hutchins et al. 2013). It is a safe assumption that the metazoan ubiquitin system in all its daunting complexity had its evolutionary origins in a simpler system; the unicellular eukaryote *Saccharomyces cerevisiae* reportedly has 68 ubiquitin ligases and 24 DUBs (Hutchins et al. 2013). The thermophilic archaea, for whom a plausible role as an endosymbiotic host has been proposed in the genesis of eukaryotes (Martin et al. 2015), appear to have ancestral versions of several eukaryotic systems (Koonin 2015) including a minimalist ubiquitin “toolkit” (Grau-Bové et al. 2015). As an extreme example the genome of *Candidatus Caldiarchaeum subterraneum* contains an operon incorporating a single

© The Author(s) 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

ligase and DUB in tandem with a ubiquitin-like gene (Nunoura et al. 2011).

The focus of the current work is the evolutionary origin of the DUBs. It is likely that gene duplication has provided raw material for the expansion of this part of the toolkit, but the extent to which gene duplication has contributed to the metazoan DUB repertoire has not been previously evaluated, nor have the types of duplication events and the subsequent specialization of duplicated genes been comprehensively explored. The two rounds (2R) of whole genome duplication (WGD) that are purported to have occurred early in vertebrate evolution (Dehal and Boore 2005) would have generated a surfeit of duplicated genes (designated “ohnologs” (Wolfe 2000) in honor of Ohno, who proposed the 2R-WGD hypothesis). By its very nature WGD globally preserves molecular stoichiometry, whereas duplication only of chromosomal regions (segmental duplication) disrupts the stoichiometry of unlinked genes with potentially deleterious consequences. DUBs can have many interacting partners (Sowa et al. 2009), and as network hubs could be very sensitive to dosage alterations. Stoichiometry can be restored by silencing a duplicated gene, and silencing is indeed the fate of most gene duplicates over a timeframe of a few million years (Lynch and Conery 2000). The sequence of duplicated genes is initially identical, but with subsequent divergence there is the potential for subfunctionalization (a division of existing molecular functions), which in metazoans may be achieved by dividing duties within a cell or by dividing the pattern of gene expression such that ohnologs are expressed in different cell types or at different developmental stages. There is also the possibility of neofunctionalization (the acquisition of novel functions by one or both duplicates), which may temporally coincide with the emergence of novel molecular pathways or more subtle innovations. These would be adaptive changes, but in the absence of strong selection or in species with small effective population size genetic drift may promote substitutions culminating in pseudogenization of a gene duplicate. Such occurrences may be informative with respect to the functional redundancy of DUBs (Vlasschaert et al. 2015).

Ubiquitin-mediated proteolysis plays a central role in ancient eukaryotic systems (e.g., the cell cycle), but for increasingly elaborate developmental and homeostatic pathways there may have been a requirement for an expanded DUB repertoire. This point is convincingly made by a recent survey of DUB expression in the mouse retina, documenting subfunctionalization of DUBs in this tissue (Esquerdo et al. 2016). We have sought to determine how many of the vertebrate DUBs can be considered ohnologs, and whether there are clear examples of DUB subfunctionalization and/or neofunctionalization. We chose to concentrate on the roles of DUBs in two pathways that predate 2R-WGD but whose regulation over the course of metazoan evolution has become increasingly baroque: innate immunity and DNA repair.

## Materials and Methods

### Determination of the DUB Repertoires

The Database of Ubiquitinating and Deubiquitinating Enzymes (DUDE-db) v. 1.0 (Hutchins et al. 2013) was used to derive the complete DUB repertoires of several animal, plant, and fungal genomes.

### Derivation of Homologous Relationships

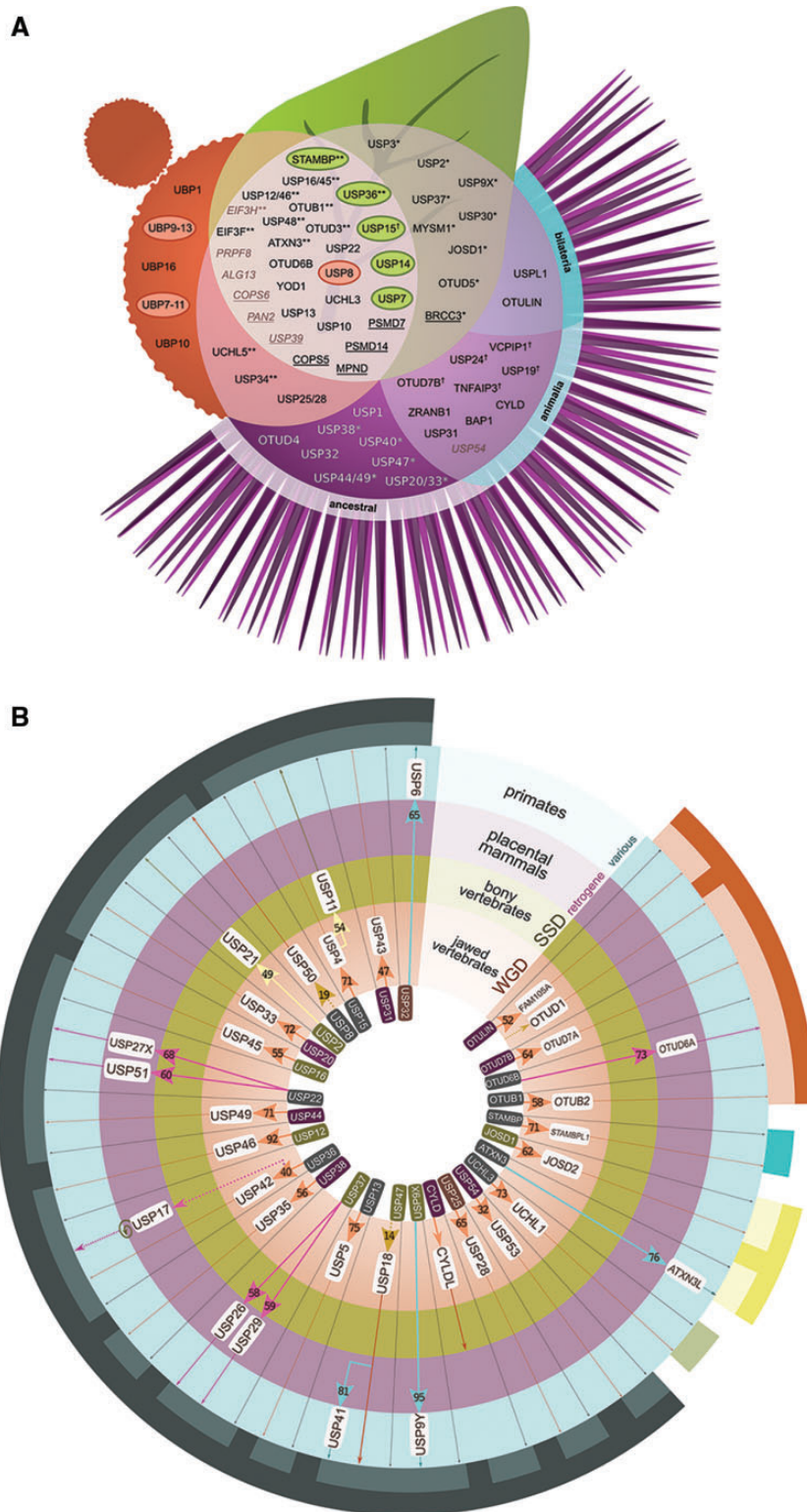
Paralogs are defined as genes within a genome sharing a common duplicative origin (e.g., whole genome duplication (WGD), small-scale duplication, retrogenic duplication). We have inferred paralogous relationships from the time of vertebrate WGD (>480 MYA) onward in figures 1B, 4, and 5. Genes present in the agnathostome ancestor are labeled “ancestral” (fig. 1A) whereas their derivatives are qualified based on the mode of duplication. The timing of these duplications is approximated by the divergence time of the earliest branching group of animals where the new paralog is present (whereas verifying its absence in the syntenic region of more earlier diverging animals). DUBs that are predicted to be paralogs by the EnsemblCompara GeneTrees pipeline (Vilella et al. 2009), including ancestral DUBs, are grouped in figure 1B. Paralogous pairs within these groups were inferred by reciprocal best BLAST (RBB) of new paralogs to determine their most likely ancestor and their similarity was quantified by pairwise global alignment using MUSCLE (Edgar 2004) (fig. 1B).

Paralog pairs are classified as ohnologs (generated by whole genome duplication) when there is only one copy in animals diverging before the vertebrate WGD events, e.g., lancelet (*Branchiostoma floridae*), sea squirt (*Ciona intestinalis*), and sea lamprey (*Petromyzon marinus*), and two copies in cartilaginous fish (elephant shark (*Callorhynchus milii*), little skate (*Leucoraja erinacea*), small spotted catshark (*Scyliorhinus canicula*) via SkateBase (Wyffels et al. 2014)), the earliest-diverging postWGD organisms (Venkatesh et al. 2014). We also verified that the two chondrichthyan paralogs are RBB matches with the human paralogs to ensure that they do not represent an independent duplication event in that lineage.

In cases where the parent gene clearly more closely resembles one of its paralogous progeny (on the basis of RBB and comparing domain structure), these are accordingly further classified as “ancestral” and “novel” ohnologs, labels which serve as contextual information in the evaluation of neofunctionalization.

### Tissue-Specific Expression Analysis

GTEX Analysis V4 gene RPKM values, along with accompanying sample ID annotations were downloaded from the GTEx portal (<http://www.gtexportal.org/home/>, last accessed 7 January 2016). The data was imported into R (<https://www.R-project.org/>, last accessed 4 January 2016) and DUB expression values were subsetted using a manually curated list of



**Fig. 1.**—Evolutionary expansion of the entire eukaryotic deubiquitinating enzyme superfamily. (A) PreWGD: Expansion from the set present in the opisthokont-archaeplastid common ancestor to that of craniata (630 MYA (Hedges et al. 2006)). All ancestral DUBs present in the human genome are categorized in this modified Venn diagram according to when they appeared evolutionarily (based on common ancestor sharing). Intersecting yeast, leaf, and sea urchin represent fungi, plants, and animals, respectively. Single asterisks (\*) indicate genes found in several Amoebozoal genomes. The human genome retains 61 DUBs from the preWGD ancestor, eight of which are constitutive protein complex members (underlined). Six have lost their isopeptidase activity

gene IDs. Each of the 2,923 sample IDs were assigned to their source tissue using the sample annotation file and the median expression for each DUB, within each tissue, was calculated. This value was used as the representative expression value for each tissue in downstream analysis.

To determine the tissue specificity of each DUB, the representative expression values from each tissue were standardized using a Z-score transformation. The standardized expression values of each DUB were then combined into a matrix, and tissues (rows) and DUBs (columns) were independently clustered by using complete linkage hierarchical clustering on the pairwise Euclidean distance values. These data were represented in a heatmap (fig. 2A) using the heatmap.2 function of the gplots package for R (<https://CRAN.R-project.org/package=gplots>, last accessed 22 January 2016). Supplementary figure S1, Supplementary Material online, was produced in a similar manner, except the Z-score standardization was performed across DUBs, but within each tissue to effectively rank DUBs by their expression values.

For pairwise comparison of tissue enrichment between paralogous DUBs (fig. 2B), a tissue specificity score was calculated by dividing the representative expression value of a DUB in a given tissue by the median expression of that DUB across all tissues. The log<sub>2</sub> ratio of the specificity scores of related DUBs was then calculated to produce a log<sub>2</sub> fold enrichment value, representing the extent to which the expression of one DUBs within the pair is specific to a given tissue, compared with its related ohnolog. This was performed for each tissue and each DUB pair.

### Embryogenesis Expression Plots

Single-cell RNA-Seq expression values from human oocytes and embryos at specific stages of development (zygote, 2-cell, 4-cell, 8-cell, morula, and blastocyst) were produced by Yan et al. Processed RPKM values were acquired (GSE36552) and DUB expression values were subsetted using a manually curated list of DUBs. The heatmap in figure 3A was produced using the same approach as described above for figure 2A. To identify clusters of distinct expression patterns throughout embryo development, the genes comprising distinct clusters (see dendrogram above

the heatmap in fig. 2A) were pooled and the expression values of each gene were normalized to the median expression value of that gene throughout embryo development. These normalized values were then log<sub>2</sub> transformed and the average log<sub>2</sub>(normalized value) for each stage was plotted.

### USP18 and ISG15 Phylogenies

We aligned USP18 and ISG15 codon sequences in three steps, first translating the sequences into amino acids, aligning the amino acid sequences with MUSCLE (Edgar 2004) and finally aligning the codon sequences against aligned amino acid sequences. These three steps are automated in DAMBE (Xia 2013). Phylogenetic trees were reconstructed with PhyML (Guindon and Gascuel 2003) using the TN93 model (Tamura and Nei 1993) as well as the maximum likelihood methods and distance-based methods (with simultaneously estimated MLCompositeTN93 distance) implemented in DAMBE. For subtrees that are not strongly supported with bootstrap values, we also consulted well corroborated species phylogenies in the Tree of Life web project (<http://tolweb.org/Vertebrata>, last accessed 26 August 2016). Our objective is to obtain a well-corroborated species tree to evaluate the synonymous and nonsynonymous substitution rate (designated K<sub>s</sub> and K<sub>a</sub>, respectively) along the branches of the tree. K<sub>a</sub>/K<sub>s</sub> ratios were computed by (1) reconstruction of ancestral sequences for each internal node of the trees by using CODEML in the PAML package (<http://abacus.gene.ucl.ac.uk/software/paml.html>, last accessed 15 August 2016), (2) pairwise comparison of codon sequences between neighboring nodes along the tree using the Li93 method (Li 1993) implemented in DAMBE. The numbers represent predictions as one cannot observe the ancestral sequences.

## Results

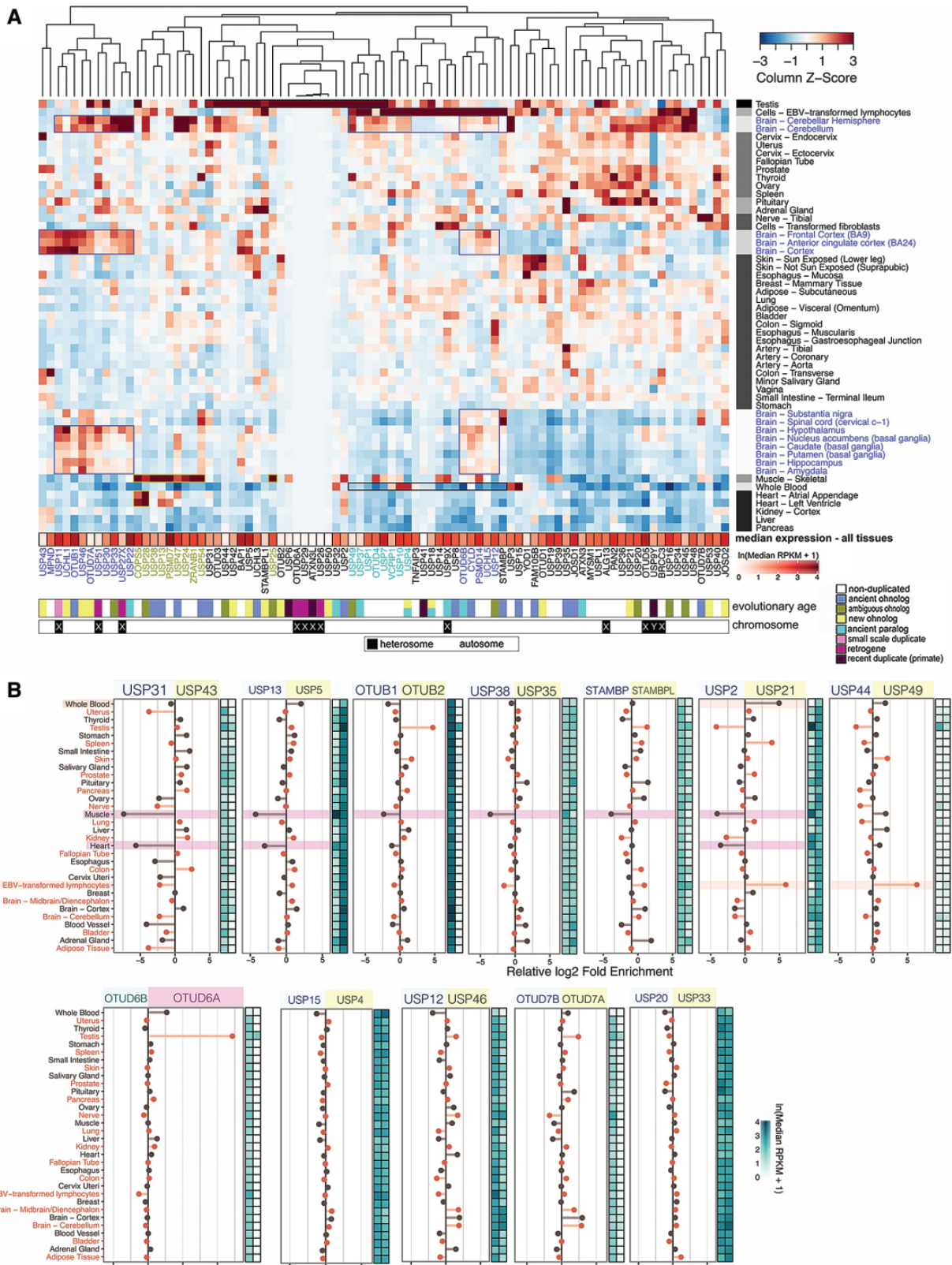
### Evolutionary History of the Deubiquitinating Enzyme Class

We first sought to delineate the evolutionary history of the entire superfamily of deubiquitinating enzymes, or DUBs. The

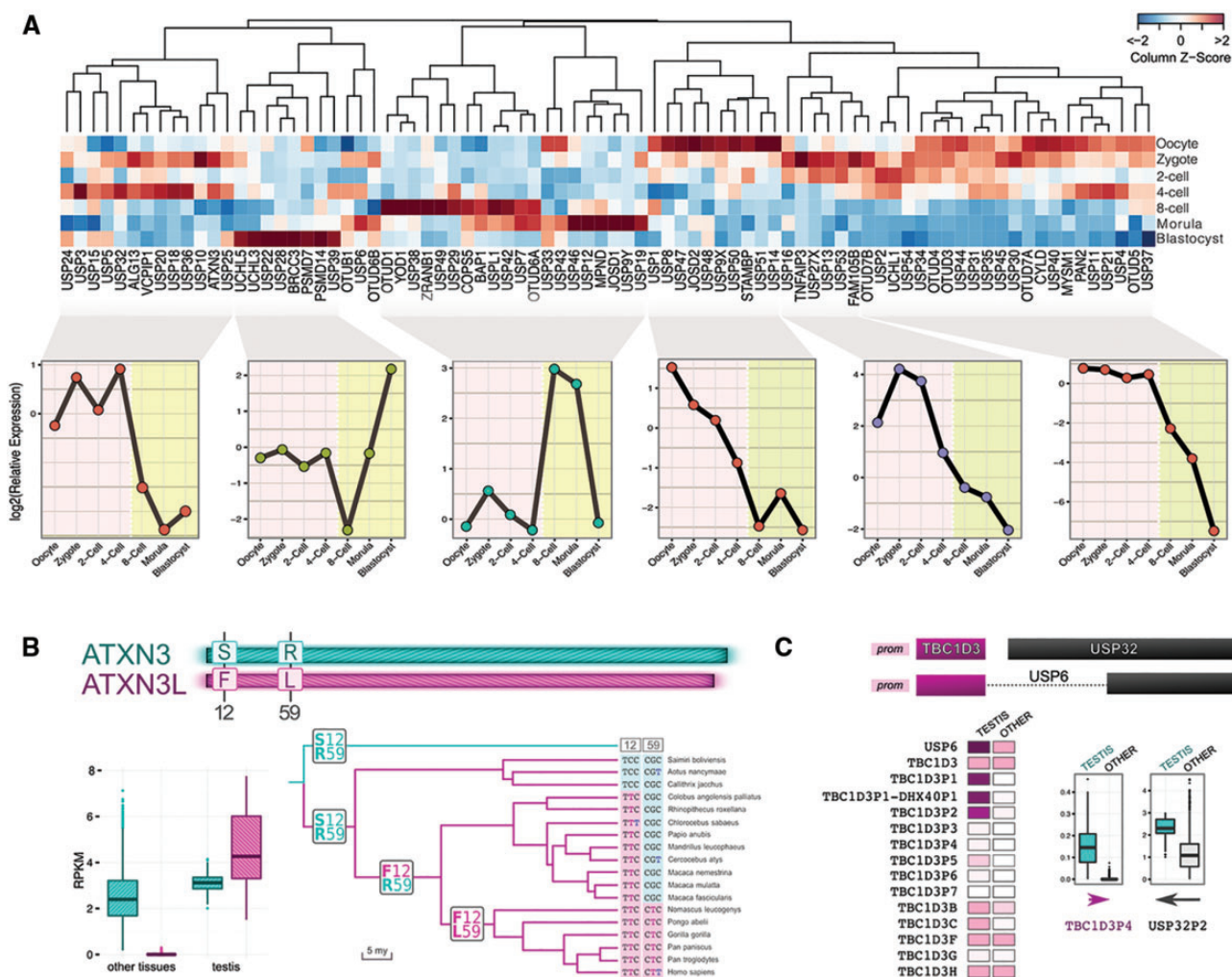
#### FIG. 1.—Continued

(italicized), and another six are notably absent in the genomes of several orders of insects (†). The chronology of evolutionary events that that yielded the yeast (*Saccharomyces cerevisiae*) DUB complement is also shown, where double asterisks (\*\*) demarcate ancestral DUBs that are present in other fungi but have been lost in yeast. Circled genes indicate DUBs that were subsequently duplicated in the fungal (orange) or plant (green) lineages. Human genome DUB nomenclature is used except in the cases of yeast-specific genes. (B) PostWGD: Expansion from the gnathostome ancestor (>480 MYA (Hedges et al. 2006)) to the human genome. Arrows connect ancestral paralogs in the centre of the circle to their duplicates, which are stratified based on age of duplication. The percent similarity between globally aligned (Edgar 2004) human duplicated DUB protein sequences is indicated on the arrowheads. Ancestral paralogs are coloured according to (A). Percentage similarity of each globally aligned pair is indicated on the arrowheads. The orange stratum represents ohnologs derived by whole genome duplication (WGD) roughly half a billion years ago. Paralogs derived by small-scale duplication (SSD) in the bony vertebrate ancestor, USP11 and USP21, are identified in the green stratum. The purple stratum groups functional retrotransposed DUBs incurred in mammals after the divergence of marsupials. Finally, the blue stratum indicates human DUBs acquired more recently by retrotransposition (ATXN3L), chimerization (USP6), and other means (USP41, USP9Y). The exterior ring groups the five families within the DUB superfamily: ubiquitin-specific proteases (USPs; grey), ubiquitin C-terminal hydrolases (UCHs; green), Machado-Joseph Disease protein domain proteases (MJDs; yellow), JAMM motif proteases (JAMM; blue), and the ovarian tumour proteases (OTUs; orange). Subgroupings within the USP, MJD, and OTU groups indicate paralogous groups as predicted by the EnsemblCompara GeneTrees pipeline (Vilella et al. 2009).





**FIG. 2.**—Tissue specificity of human deubiquitinating enzymes in an evolutionary context. (A) Clustered heatmap of standardized expression values from the GTEx project. The dendrogram above the heatmap represents DUB clustering, whereas tissue clustering is represented by the shaded boxes to the right of



**Fig. 3.**—Selectivity of deubiquitinating enzyme expression in gametes and embryogenesis. (A) Single-cell RNA-seq data was used to generate a clustered matrix of expression enrichment at different stages of human embryogenesis. Enrichment levels of several clusters drastically shifts around the 4–8 cell stage, a time when oocytic mRNA reserves become depleted and embryonic transcription begins in humans. Many of the oocyte-enriched DUBs are also testis-enriched (fig. 2A). (B) Fitness-enhancing changes in the coding sequence of testis-specific paralog, ATXN3L, throughout primate evolution. As reported by Weeks et al. (2001), ATXN3L is the most efficient isopeptidase of the MJD family owing to two hydrophobic acid substitutions. The S12F and R59L substitutions were sequentially fixed in the catarrhine and hominoid ancestors, respectively. (C) Permissive expression of the chimera USP6, its parents and their pseudogenes in the testis. Domain structures of USP6, TBC1D3, and USP32 are illustrated. Most of the known pseudogenes of *Tbc1d3* (which forms the N-terminal end of the USP6 chimera) as well as multiple *Usp32* pseudogenes are especially expressed in the testis.

DUB repertoire at the origin of two major eukaryotic supergroups, the opisthokonts and archaeplastids, is presented in figure 1A in the form of a Venn diagram with intersecting

budding yeast, sea urchin, and leaf, representing fungi, animals, and plants, respectively. The SAR supergroup, Excavata, and Amoebozoa repertoires are not explicitly represented

#### Fig. 2.—Continued

the heatmap. Notable clusters are demarcated by boxes and their relevance is discussed in the text. Median expression level for all tissues, evolutionary age (from fig. 1) and chromosomal locus are indicated for each DUB beneath the matrix. The two rows below the DUB names summarize DUB evolutionary age, as inferred from figure 1, and chromosomal loci, that is, whether a DUB is on an autosome or a sex chromosome. Integration of transcriptomic and evolutionary trends in this way permits visualization of correlations between the two. (B) Pairwise comparison of tissue expression enrichment ( $\log_2$  Fold Enrichment; see methods for details) and raw expression values (RPKM; bluescale shaded boxes) for select DUB paralogs. Highlighted in pink, many ancestral paralogs display more than 5.6-fold enrichment (2.5 on  $\log_2$  scale) in muscle tissues (heart and skeletal muscle) compared with their derivatives.

though they are used in our analysis to support inferred ancestral ages of DUBs. Amoebozoa are thought to have emerged median to opisthokonts and archaeplastids, whereas the other supergroups diverged earlier (Burki 2014). Twenty-three DUBs and six DUB-related genes, which have homology with DUBs but do not have isopeptidase properties are inferred to be at the base of the opisthokonts and archaeplastids. Three additional DUBs are shared among opisthokonts (fungi and animals) and seven are yeast-specific, amounting to a total of 39 fungal DUBs. This number is greater than the 24 DUBs reported in a Hidden Markov Model scan of the *Saccharomyces cerevisiae* genome (Hutchins et al. 2013) because it includes those not found in the yeast genome but present in more than five other fungal genomes as well as in Amoebozoa genomes (e.g., *Acanthamoeba*, *Dictyostelium*, *Actyostelium*, and *Polyspondylium*) and because four DUB members were overlooked in the original study (PAN2, ALG13, COPS5, and MPND). An additional nine DUBs are shared between plants and animals thus were also likely present in the opisthokont-archaeplastid common ancestor. DUBs that underwent subsequent duplication in plants or yeast and those that are constitutive subunits of complexes are identified in figure 1A. As the figure indicates 20 ancestral DUBs are unique to Animalia, though five of these are also found in Amoebozoa genomes. Some DUBs were found to be notably absent in the genomes of several orders of Insecta (including Diptera, which contains *Drosophila melanogaster*) and in some cases were absent from the entire class. As previously noted, all five families of DUBs are ancestrally represented (Hutchins et al. 2013), though DUBs within these families do not all share sufficient similarity to be considered paralogs (Vilella et al. 2009) (fig. 1B).

The working set of 59 DUBs common to Animalia is expansive, and the repertoire increases by more than half that number in the human genome. Two DUBs with particular cleavage targets, OTULIN (linear ubiquitin chain specificity (Keusekotten et al. 2013)) and USPL1 (SUMO specificity (Schulz et al. 2012)), emerged in bilateria (fig. 1A) whereas the remainder arose during vertebrate radiation. Figure 1B presents the sequential expansion of DUBs from the bilaterian set to the distinct human genetic repertoire by means of whole genome duplication, small-scale duplication, retrogenic duplication, tandem duplication, and chimerization. The first major expansion of DUBs coincides with the emergence of jawed vertebrates approximately 450 million years ago (Venkatesh et al. 2014). Two rounds of whole genome duplication (WGD) in the basal vertebrate (Dehal and Boore 2005) provided extensive genetic fodder for network rewiring and organismal remodeling in the subsequent radiation of vertebrates. At least 21 DUBs were derived and maintained from this WGD event (fig. 1B).

Given the method of duplication, ohnologs should only be located adjacent to one another if significant chromosomal rearrangement occurred. The 3' tail of the *Usp50* gene, which

first appears in cartilaginous fish, overlaps with the 3'UTR of its paralog, *Usp8*, in both the human and elephant shark (*Callorhynchus milii*) genomes. USP8 is the ancestral DUB with which USP50 retains highest coding sequence similarity (in shark and human) and thus is most likely to be its parent. Though USP50 arose concordant with WGD timing, its arrangement with USP8 suggests that their relationship may not be ohnologous.

From the analysis presented in figure 1B, it will be evident that whereas some ohnologous pairs have diverged drastically (e.g., USP47/USP18 or USP8/USP50) most have retained more than 50% sequence similarity. In some of these cases, both equally resemble protein sequences from the single ancestral copies in invertebrate Animalia. High coding sequence similarity does not always equate to limited subfunctionalization. For example, JOSD1 is a membrane-bound DUB that requires allosteric activating ubiquitination and regulates cell motility and endocytotic processes, whereas its ohnolog JOSD2 is cytoplasmic and retains innate deubiquitinase activity (Seki et al. 2013, p. 1). UCHL1 and UCHL3 are 73% similar but display different patterns of tissue-specific gene expression: an upstream neuron-restrictive silencing element (NRSE) drives neuron-specific expression in the former (Barrachina et al. 2007) and renders it a critical player in neuronal homeostasis (UCHL1 deficiency in neurodegeneration is not physiologically rescued by UCHL3). Other ohnologous DUB pairs, such as USP4 and USP15 are interchangeable with respect to organismal viability but have subtle yet evolutionarily-stable properties that are distinctive (Vlasschaert et al. 2015). Functional redundancy has however not been evaluated for many WGD-derived pairs, including USP12-USP46 whose similarity is 92%.

Following the WGD mass expansion, small-scale chromosomal duplications of the genomic regions encoding USP2 and USP4 gave rise to USP21 and USP11, respectively, in bony vertebrates (Vlasschaert et al. 2015). All other changes contributing to the human DUB repertoire occurred in the mammalian lineage. Several retrogenes, mostly incurred in the mammalian ancestor, were added whereas the CYLD duplicate (named CYLD-like in fig. 1B, or CYLDL, for convenience) was specifically deleted in all mammals. It has been established that the mammalian X chromosome has a superabundance of functional processed retro-pseudogenes (Drouin 2006) (Potrzebowski et al. 2010), which may reflect the smaller effective population size of the X chromosome compared with autosomes, and a resulting reduction in the efficiency of purifying selection. USP29 is an autosomal exception, but it is poorly conserved: human and gorilla protein identity is only 80% and the gene is absent in multiple species. Its reported role in stabilizing p53 (Liu et al. 2011) is therefore somewhat counterintuitive. USP17 is a mammalian retrogene with clusters of variably transcribed pseudogenes tandemly arranged on the chromosomes 4 and 8 (Burrows et al. 2005; Burrows et al. 2010). USP9Y is a Y-linked, nonretrogenic copy of USP9X that originated in the common ancestor of



euarchontoglires (a clade that includes primates and rodents) and encompasses several proximal pseudogenes. USP32 similarly has an array of neighboring, nonprocessed pseudogenes. As with the USP17 and USP32 pseudogenes, the X-chromosome retrogene ATXN3L, the exon-bearing tandem duplicate USP18 and the chimera USP6 were more recently acquired at various times during primate evolution. The detailed divergence and functional discrepancies for the majority of paralogous pairs in figure 1B has yet to be formally addressed.

### Related Deubiquitinating Enzymes Evolve Discrete Spatiotemporal Occupations

Retained duplicate genes are thought to diverge either in terms of coding sequence or by varying their expression to accomplish similar functions in discrete spatiotemporal domains (Nguyen Ba et al. 2014). To illustrate the latter, figure 2 presents the tissue specificities of all known human DUBs using RNA-Seq expression data from 2923 samples of 53 human tissues obtained from the Genotype-Tissue Expression (GTEx) Project (Lonsdale et al. 2013). A row-normalized counterpart of figure 2A that ranks absolute DUB expression for each tissue and provides an estimate of cellular mRNA distributions is provided in [supplementary figure S1A](#), [Supplementary Material](#) online. A few clusters of tissue-specificity are observable in figure 2A. For example, there is a cluster of DUBs with relative enrichment in brain tissues encompassing known neuron-specific DUBs UCHL1 (Day and Thompson 2010) and USP11 (Vlasschaert et al. 2015) as well as USP43, MPND, OTUB1, USP46, OTUD7A, USP30, USP33, USP22, USP27X, and USP51. In fact, UCHL1, USP11, USP22, and OTUB1, in that order, are the four DUBs with highest raw expression values in the brain ([supplementary fig. S1B](#), [Supplementary Material](#) online). It is peculiar that USP27X and USP51, X-linked retrogenes derived from a USP22 ancestor, maintain their source gene's expression pattern. There are documented cases of retrogenes derived from aberrant transcripts where the promoter sequence is retained (McCarrey 1987); it remains to be seen whether this is the source of the conserved expression pattern in these paralogs. There exists a second set of DUBs with moderately brain-enriched expression that form a subcluster within a large group of DUBs enriched in lymphocytes transformed with Epstein-Barr virus (a standard method lymphocyte immortalization). A third set exists with enrichment in both these immortalized lymphocytes and in the cerebellum. Concordant with reports that EBV transformation results in enriched expression of genes related to cell cycle and immunity (Çalışkan et al. 2011), DUBs with integral roles in these processes are highly enriched in EBV-induced lymphocytes. For example, USP4 regulates the stability of multiple innate immunity proteins (e.g., TAK1 (Fan et al. 2011, p. 1), TRAFs 2 & 6 (Xiao et al. 2012, p. 2), RIG-I (Wang et al. 2013)), as well as cell cycle checkpoint regulators pRb (Blanchette et al. 2001), and ARF-BP1 (Zhang

et al. 2011), a p53 antagonist. Most DUBs enriched in the immortalized lymphocytes are contrastingly depleted in whole blood cell sample (fig. 2A) from which the immortalized cell lines were derived (Lonsdale et al. 2013).

Several DUBs are enriched in muscle cells as indicated in figure 2A. All of these except USP25 form a cluster, within which USP28 and COPS5 form a heart-specific subcluster. USP25, the ohnolog of USP28, is also testis-enriched, precluding its integration into the muscle cluster. The molecular basis of this tissue bias is known: USP25 has muscle- and testis-specific isoforms whereas USP28 has a heart- and brain-specific isoform that is preferentially expressed (Valero et al. 2001). Building of their established evolutionary relationships (fig. 1), figure 2B features pairwise plots of tissue specificity differences and absolute expression values for several evolutionarily-related DUBs.

Perhaps most intriguing in figure 2A is the pervasive enrichment of DUBs in the testis, which defines one large cluster though is also observed in several other DUBs. This group includes all X-chromosomal retrogenes except the two derived from USP22 as well as other DUBs incurred recently during the subspecialisation of primate branches, though is not restricted to these. However, the “young DUBs”, namely USP6, OTUD6A, USP29, ATXN3L, and USP26, along with USP50, form a subcluster characterized by virtual absence of expression in other tissues. It is thought that the permissive chromatin architecture in meiotic spermatocytes and postmeiotic spermatids permits widespread genetic expression in these cells (Soumillon et al. 2013); as such, many new genes are birthed with facilitated expression in germ cells, which comprise a large fraction of testicular samples (Baran et al. 2015). Comparison of open chromatin marks at OTUD6B and OTUD6A loci across several mouse tissues suggests DUB retrogenic testis-specificity is conserved ([supplementary fig. S1C](#), [Supplementary Material](#) online).

Deubiquitinating enzymes exhibit clustered temporal expression patterns during embryonic development (Yan et al. 2013), many of which appear to become activated or deactivated at the time of embryonic activation (fig. 3A). Of note, several DUBs from the testis cluster in figure 2A are also expressed in oocytes, where transcripts for all except ATXN3L and USP26 are detectable at some point during early embryonic development (fig. 3A; “expression” of testis-specific DUBs in the early embryo may however reflect residual male germ cell transcripts [Johnson et al. 2011]). ATXN3L was duplicated from ATXN3 in the simian ancestor (figs. 1B and 3B) and is the most effective ubiquitin cleaver of all Josephin domain-containing proteins, attributable to the optimization of two amino acid sites (Weeks et al. 2011). The hydrophobic substitutions, S12F and R59L, were acquired in a step-wise manner during the relatively brief evolutionary lifespan of ATXN3L (fig. 3B) and only synonymous substitutions are observed at these sites in organisms diverging after the substitutions for hydrophobic residues. USP26 equally retains



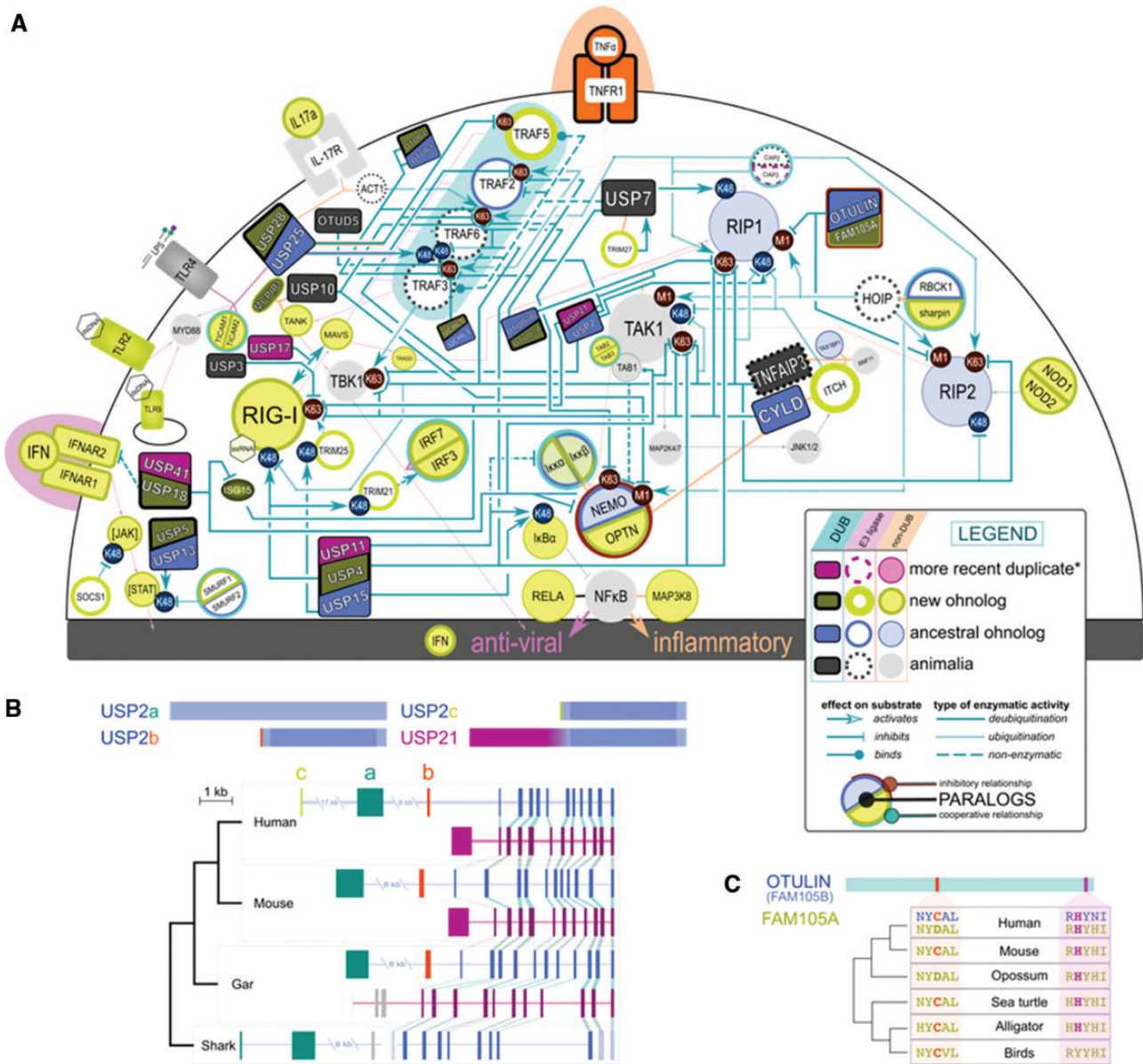
significant conservation across mammals. Though silenced in all other cell types including fertilized zygotes, these X-linked genes may be important for testis development as they are subject to evolutionary constraints. USP6, a nonretrogenic germ cell-specific DUB on chromosome 17 (Chr17) in humans, is a chimera derived from the fusion of the N-terminus of a *Tbc1d3* paralog and the C-terminus of a *Usp32* paralog (Paulding et al. 2003). In addition to the protein-coding genes themselves, there are multiple USP2 and TBC1D3 pseudogenes annotated on Chr17 that display varied levels of expression. Of note, several *Tbc1d3* and *Usp32* pseudogene copies have testis-specific expression whereas others are not expressed (fig. 3C). Transcription at these genetic loci, whether broadly or specific to the testis, supports the idea that the “pseudogene” label does not necessarily indicate absolute absence of expression and may in fact be erroneous.

### Continuous Rewiring of the Deubiquitinating Enzyme System in Immunity

The ubiquitin-proteasome system (UPS) plays critical roles in innate immunity, where many deubiquitinating enzymes terminate immune responses to prevent chronic inflammation (Sun 2008). Owing to their extensive modulatory roles, CYLD and TNFAIP3 (A20) are often integrated into immune pathway schematics (Kanehisa et al. 2006) whereas many others have reported roles in immunity (fig. 4A). Cartilaginous fish represent the earliest diverging clade with an adaptive immune system (Flajnik and Rummel 2000; Venkatesh et al. 2007) and incorporate several novel innate immunity genes (Venkatesh et al. 2014). WGD in the gnathostome ancestor concomitantly generated 21 DUBs (fig. 1B); the retention of some of these ohnologs may be driven by immune pathway regulation. Figure 4A depicts a snapshot of the evolutionary emergence of immune genes—as is evident WGD has been responsible for generation most of the components of the antiviral pathway as well as the UPS components regulating it, such as DUBs USP18 (Goldmann et al. 2015), USP4 (Wada et al. 2006; Wang et al. 2013) and UCHL1 (Karim et al. 2013) and E3 ligases SOCS1 (Ungureanu et al. 2002, p. 1), TRIM21 (Higgs et al. 2008) and TRIM25 (Gack et al. 2007, p. 25). Innate immune rewiring is apparent in nonohnologous DUBs (fig. 1A) such as USP3 (Cui et al. 2014), USP10 (Niu et al. 2013; Wang et al. 2015), USP7 (Zapata et al. 2001; Zaman et al. 2013), and OTUD5 (DUBA) (Kayagaki et al. 2007), as well as in ancestral paralogs that have different roles in immunity than their progeny (e.g., USP2; fig. 4B). Although USP2a is an ancestral gene, the two short USP2 isoforms with alternative 5' exons arose over the course of evolution. The appearance of USP2b coincides with the gene duplication event which gave rise to USP21 in bony vertebrates (fig. 4B). The other short USP2 isoform (USP2c) acts to preserve cell viability in inflammation because, contrary to USP2a, it is not inhibited by TRAF2

(Mahul-Mellier et al. 2012) (fig. 4A). The distinct 5' exon of human USP2c is only predicted to be protein-coding in certain other primates and rodents (excluding mouse). USP2c, and its distinct immune functions, may thus represent a novel innovation in the euarchontoglires ancestor. The NF- $\kappa$ B Essential Modulator (NEMO) is the regulatory subunit ( $\gamma$ ) of the IKK complex that activates the NF- $\kappa$ B pathway. An ohnolog of NEMO, optineurin (OPTN), maintains structural homology though evolved to negatively regulate NEMO signaling competitively (Zhu et al. 2007) and in association with DUBs (Nagabhushana et al. 2011). Conversely, FAM105A is a conserved ohnolog of FAM105B (commonly referred to as OTULIN for “OTU deubiquitinase with linear linkage specificity”) that bears inactivating substitutions in human. FAM105A may competitively inhibit OTULIN's roles in immunity, akin to OPTN and NEMO, an hypothesis supported by the fact that FAM105A retains ubiquitin-binding abilities (Oshikawa et al. 2012). The substitutions that lead to the FAM105A inactivation were, however, only acquired during mammalian radiation (fig. 4C); thus, in other vertebrates, FAM105A may still act as a DUB.

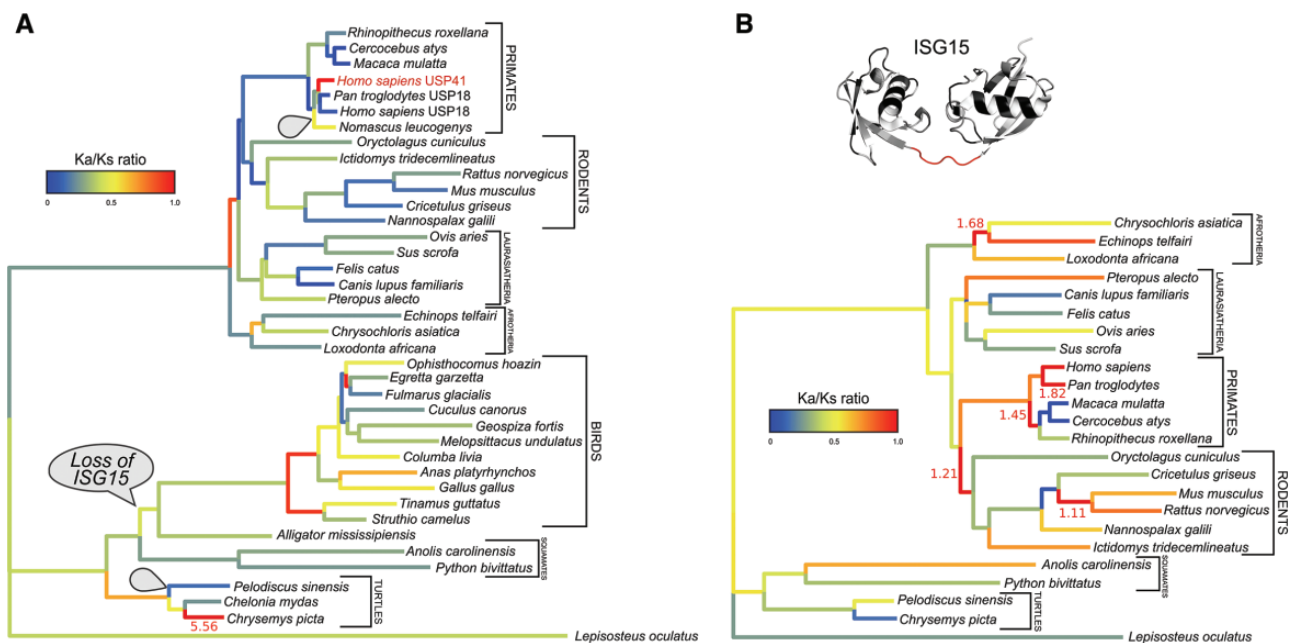
Interferon-stimulated gene 15 (ISG15) is a ubiquitin-like modifier that emerged in gnathostomes (Loeb and Haas 1992) and mediates species-specific roles in immunity. As its name suggests, ISG15 is stimulated by interferon (IFN) signaling, as are the host enzymes that mediate its conjugation and removal from target proteins. In humans, ISG15 increases viral susceptibility and is a critical allosteric regulator of USP18 (a terminator of IFN signaling), whereas in mice it plays an antiviral role and does not stabilize USP18 (Speer et al. 2016). Five modified amino acids in human ISG15 relative to mice enable the NS1 protein of influenza B to target this ohnolog and is the molecular basis of species-specific infection (Yuan and Krug 2001, p. 15; Guan et al. 2011). Further reflective of its divergence, coding sequences of zebrafish and human ISG15 are both more similar to each species' di-ubiquitin gene than to each other, though they are syntenic, suggesting a common origin. However divergent, ISG15 removal is nevertheless exclusively mediated by USP18 in humans, mice, and zebrafish (Chen et al. 2015). We investigated whether there are changes in the selective pressure acting on USP18 correlative to changes in ISG15. We performed a Ka/Ks analysis, which calculates the ratio of nonsynonymous nucleotide substitutions per nonsynonymous codon site (Ka) to synonymous substitutions per synonymous site (Ks) per for each branch of a phylogenetic tree. Strongly conserved genes have low Ka/Ks values (near 0), because greater ratios indicate that more changes to the coding sequence were retained over the course of evolution. Conventionally, it was thought that Ka/Ks = 1 is indicative of neutral evolution, Ka/Ks < 1 is indicative of purifying selection, and Ka/Ks > 1 is indicative of positive selection. Although such interpretation has many problems due to mutation bias and amino acid usage bias (Xia and Kumar 2006),



**FIG. 4.**—Evolutionary network of ubiquitin proteasomal system involvement in innate immunity. (A) Comprehensive network of currently known DUB interactions in innate immunity. The roles of E3 ubiquitin ligase enzymes that interregulate with DUBs are also depicted. All components of the schematic are coloured according to their evolutionary age and the nature of their interactions within the system are indicated by line and arrow type (see Legend). Detailed information about the depicted interactions is available in [supplementary tables S1 and S2, Supplementary Material](#) online. (B) Evolution and conservation of USP2 isoforms and paralogs. USP2 has three isoforms (a, b, c) and one conserved paralog, USP21. Schematic depicts homology of exons throughout evolution and the emergence of new 5' exons, which are unique in each isoform and paralog. USP2a is ancestral; USP2b and USP21 were derived in bony vertebrates whereas USP2c is present in some but not all euarchontoglires. (C) FAM105A, an OTULIN ohnolog, retains catalytic residues necessary for linear ubiquitin cleavage in several vertebrates but not humans.

the relative change in Ka/Ks ratio along a phylogenetic tree may still shed light on the changing selection intensity of USP18 and ISG15. Inferences as to strength and type of selection can be more reliably drawn for longer tree branches and when the number of sites (length of aligned sequence) is larger because these lower the standard error.

The phylogenetic trees in figure 5A and B, respectively, illustrate rates of change in the coding sequences of USP18 and ISG15 over the course of vertebrate evolution. The USP18 phylogeny in figure 5A includes 14 more species than the ISG15 tree in figure 5B because the ISG15 gene was deleted in the ancestor of Archosauria (birds and crocodilians) as well



**FIG. 5.**—Changing intensity of purifying selection, measured by the ratio of nonsynonymous substitution rate over synonymous substitution rate (Ka/Ks), during the evolution and functional diversification of (A) USP18 and (B) ISG15. Branches are coloured according to the inset scale (blue-to-red for 0-to-1), and associated numbers are indicated near the branches when Ka/Ks values exceed 1.

as specifically from the soft-shelled turtle (*Pelodiscus sinensis*) and gibbon (*Nomascus leucogenys*) genomes. Consistent with the hypothesis that USP18 and ISG15 co-evolve, figure 5A illustrates that there is decreased selective pressure (increased Ka/Ks values) on the USP18 genes of birds relative to mammals. However, the evolution of these genes does not appear to be correlated in the turtle lineage: ISG15, deleted in the soft-shelled turtle, is highly conserved in painted turtles and sea turtles, whereas USP18 is present in soft-shelled turtles but dramatically altered relative to the inferred ancestor in sea turtles.

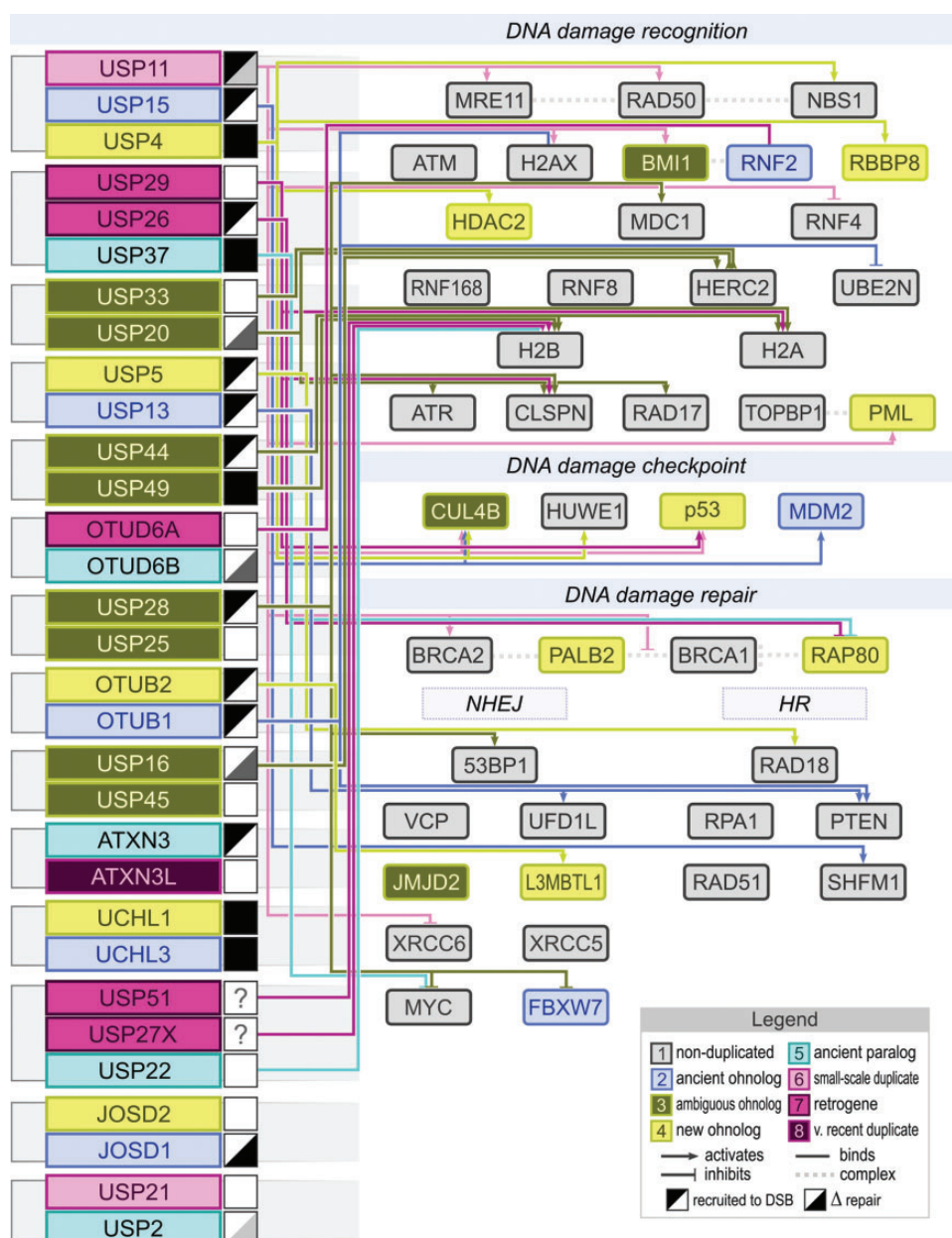
USP18 is relatively well conserved among mammals (fig. 5A). USP18 undergoes an appreciably higher amount of coding sequence change during boreoeutheria divergence (Ka/Ks value of 0.84). In addition, duplication of USP18 in *Homininae* gave rise to USP41 (fig. 1B), which maintains expression pattern homology and moderate sequence identity (globally aligned human proteins: 79%) though it also has a high Ka/Ks value (0.96). In contrast, ISG15 experiences relatively high rates of evolutionary change among mammals (fig. 5B). Though this may suggest that selection does not act to preserve USP18-ISG15 concordance, the sequence of ISG15 is much shorter; lower absolute Ka and Ks values increase the representation of nonsynonymous mutations on the Ka/Ks ratios coded in the tree. Thus, ISG15 may be one of many interactors influencing the selective pressure acting on USP18 and USP18-derived sequences. For example, Although the 14-amino acid tail required for IFNAR2 inhibition (Goldmann et al.

2015) is deleted in USP41, the catalytic residues required for ISG15 cleavage are conserved.

### Neofunctionalization of Ohnologs in Double-Strand Break Repair

Several deubiquitinating enzymes critically regulate the DNA damage response (DDR), including factors involved in the recognition of double-strand breaks (DSBs) and subsequent signaling to arrest the cell cycle for either apoptosis or DNA repair by nonhomologous end joining (NHEJ) or homologous repair (HR) (Nishi et al. 2014; Citterio 2015; Kee and Huang 2015). Figure 6 illustrates DUB paralog involvement in DSB repair from an evolutionary perspective. Several instances of postduplicative innovations are apparent. Of note, the interaction between USP4 and the DNA endonuclease CtIP (RDDP8), critical to DSB end resection in HR (Liu et al. 2015; Wijnhoven et al. 2015), represents a postWGD neofunctionalization. Both proteins are novel ohnologs and the interaction domains on USP4 and RDDP8 do not share homology in the ancestor. That is, the insert region of USP4 is significantly different from that of USP15 (Vlasschaert et al. 2015), and the N-terminal domain of RDDP8 (pfam10482) is found only in gnathostomes (postWGD organisms) whereas its C-terminal domain (pfam08573) bears homology with many earlier-diverging eukaryotes including yeast. USP4's interaction with HDAC2 likely also emerged after WGD as a small region of USP4 (a.a. 188–302), which encompasses its unique alternatively spliced exon





**FIG. 6.**—Rewiring of the deubiquitinating enzyme network for new roles in DNA repair. Groups of paralogous DUBs with known involvements in double-strand break (DSB) repair (Nishi et al. 2014) are shown with their interaction partners in an evolutionary context. Paralogy is indicated by boxes to the left of the figure. Shaded boxes next to each DUB name indicate whether they are recruited to sites of DNA damage and whether they exert a quantitative effect on DNA repair, as indicated in the Legend. Detailed information about the depicted interactions is available in [supplementary table S3, Supplementary Material online](#).

(Vlasschaert et al. 2015; Vlasschaert et al. 2016), interacts with this HDAC1-derived ohnolog.

Other examples of ohnologous innovation include OTUB1/OTUB2, USP16/45, and USP25/28. OTUB1 binds Ubc13 (UBE2N) to inhibit K63-linked-ubiquitination of chromatin in DSB repair, whereas its ohnolog, OTUB2, shows much lower affinity for Ubc13 (Sato et al. 2012) and instead promotes HR

by interacting with another novel ohnolog, L3MBTL1 (Kato et al. 2014; Citterio 2015). USP16 pairs with HERC2 to remove H2A K15-linked ubiquitin conjugates and downregulate DSB repair (Zhang et al. 2014) whereas its ohnolog, USP45, has no known involvements in the DNA damage response. The N-termini of human USP16 and USP45 have high sequence identity (42% for the first 400 residues), but differ

markedly in exon 5, which codes for the coiled-coil HERC2 interaction domain in USP16. The USP16/45 ancestor in invertebrates carries traits of both of its progeny, though is predicted to lack the N-terminal coiled-coil domain using the COILS server (Lupas et al. 1991). Consistent with the hypothesis that USP16-HERC2 is a novel WGD-derived interaction, the coiled-coil domain first emerges in gnathostome USP16 (supplementary fig. S2, Supplementary Material online). Finally, highly similar ohnologs USP28 and USP25 are each singularly involved in DNA repair and immune responses, respectively. USP28 associates with and stabilizes 53BP1, Claspin and MDC1 in DSB repair (Zhang et al. 2006) (fig. 6), whereas USP25 is not known to regulate any DDR players (Zhang et al. 2006; Breuer et al. 2013). Contrarily, USP25 downregulates cellular responses to bacterial infection and autoimmunity by removing K63-linked ubiquitin on TRAF5 and TRAF6 when in complex with ACT1A (fig. 4A). USP25 also stabilizes TRAF2 (Zhong et al. 2013), TRAF3 and TRAF6 (Lin et al. 2015) to promote antiviral immunity; USP28 has no known immune interactors (Sowa et al. 2009; Breuer et al. 2013). These ohnolog-specific roles in two important cellular processes could represent subfunctionalization as the USP25/28 common ancestor in invertebrates resembles both its derivatives equally and all of the aforementioned substrates are ancestral and nonduplicated (with the exception of TRAF5, which was derived from TRAF2). Thus, USP4, OTUB2, USP16, and USP28 appear to be WGD-derived paralogs with innovative interactions in DNA repair relative to their ancestors.

## Discussion

Tracing the genealogy and radiation of the 93 deubiquitinating enzymes present in the human genome can help answer questions relating to their redundant and distinct roles. The collection of DUBs is indeed a superfamily composed of five functionally-related families, all of which are represented in the eukaryotic common ancestor. Members of these families cooperate to accomplish the common goal of meeting cells' dynamic deubiquitination needs in various processes, such as immune reactions and DNA repair, and often converge to serve functionally redundant roles. Duplications at several points in evolution, most notably the whole genome duplications preceding gnathostome emergence, provided evolutionary fodder for DUB system rewiring. We set out to detangle the DUB network and distinguish redundancy that results from postduplicative conservation of interaction domains from neofunctionalization events, such as the WGD-derived USP4-CtIP interaction in DSB repair. We also sought to predict whether novel paralogs whose cellular roles have not yet been extensively studied (e.g., USP41, ATXN3L, FAM105A) might be important nodes in these networks.

Twenty DUBs have been retained in the human genome from the two rounds of whole-genome duplication of the ancestral vertebrate genome, which contained 61 DUBs.

This means that more than 160 duplicates have become non-functional over the course of evolution: after the second round of WGD there would be 183 new ohnologs (244 minus the original 61), of which 20 were actually retained, meaning 163 were lost. Several selective and stochastic factors determine whether genes are maintained. According to the nearly neutral theory, stochastic mutations, including deleterious ones, can become more easily fixed in species with small effective population sizes. In addition, the rates of different types of mutations vary across different genomic regions and between species. Thus, in addition to selection for beneficial mutations (or against deleterious ones), mutational biases and genetic drift play an appreciable role in losses or changes in gene function over time. These factors may explain why, for example, CYLDL is absent in mammals and ISG15 has been lost in several species including birds, gibbon and soft-shelled turtle (fig. 5B) despite possible benefit to the organism. Loss of nodes within a system due to drift can affect the selective pressure acting on other genes that regulated that node ancestrally, as is seen in the case of USP18 and ISG15 (fig. 5).

Whether they occur by ohnologous, retrogenic, or segmental means, all gene duplication events originate in single organisms. Retention of paralogs along a branch of the tree of life, for example in the mammalian lineage, requires that duplicates be prevalent enough to be fixed within the effective population at its root. Purifying selection against deleterious mutations presumably enables duplicates to attain fixation whereas retaining function. An evolutionary conundrum known as Ohno's dilemma (Bergthorsson et al. 2007) sequentially follows: if selective pressure conserves the protein-coding sequence of paralogs, how do new functions emerge? The innovation, amplification, and divergence (IAD) model (Bergthorsson et al. 2007) proposes that new paralogs are subject to continuous positive selection to exploit beneficial side-functions of the parent gene. The utility of this side function is "recognized" by the system via amplification of the parent gene, which facilitates fixation of a duplicate with mutations that enhance this side-function. Selection would then operate to maintain both paralogs in their divergent roles. This mechanism may explain the evolution of the USP16-HERC2 interaction (fig. 6, supplementary fig. S2, Supplementary Material online): the USP16 coil-coiled interaction domain at ~200 a.a. is absent in lancelet USP16/45 gene and in the human copy of USP45, though traces of a coiled-coil domain are present in shark USP45. It is possible that a USP16/45 copy with a rudimentary coil-coiled domain was amplified preceding gnathostome divergence and its capacity to interact with WGD-derived HERC2 was "recognized" as advantageous. Thus, selection for the coil-coiled domain of USP16 may have enabled its fixation and allowed divergence from USP45 over subsequent vertebrate evolution.

Some DUB-derived genes have lost their deubiquitinase function, either ancestrally (italicized in fig. 1A) or during the course of vertebrate evolution (e.g., FAM105A in fig. 4C), and

have adopted other cellular roles. They may have also lent domains to other nonDUBs: USP7 is one of three proteins containing a mepirin and TRAF homology (MATH) domain in the ancestral eukaryote (the two others are E3 ubiquitin ligases TRIM37 and SPOP), a domain that is prevalent in many vertebrate immune signalling proteins (Zapata et al. 2001), especially in plants (Liu et al. 2009). Thus, the influence of ubiquitin-conjugating system expansion directly extends beyond that which is portrayed in figure 1.

An emerging theory in molecular evolution may offer a model of paralog retention that is complementary to the IAD model. The prerequisite amplification step in the IAD model poses an additional hurdle for retrogene retention, because these, by virtue of their method of duplication, would not normally be expressed and systemic “recognition” of their important subfunction would not be possible. In our present work, we noted that several human DUB genes that were derived relatively recently (along the mammalian lineage) are expressed specifically in the testis. The “out of the testis” gene birth model (Kaessmann 2010) hypothesizes that the permissive open chromatin structure in meiotic spermatocytes and postmeiotic spermatids allows the expression of sequences that lack efficient promoters (e.g., retrogenes) in these cells. Testis-specific expression is reportedly associated with demethylation of CpG-rich promoters (Soumillon et al. 2013) and/or endogenous retrovirus LTR exaptation (Melé et al. 2015). In the absence of a conventional promoter, the potentially deleterious impact of genetic drift on gene expression would be reduced in the testis, and its potential role in neofunctionalization would subsequently be enhanced because mutations are then statistically more likely to be adaptive in this scenario. Testis-specific expression would then substitute for the “amplification” step in the IAD model in some cases, and selection for stronger promoters would enable wider expression profiles for these paralogs. Several DUBs and DUB pseudogenes with testis-specific expression may have been (or may be in the process of being) birthed out of the testis (fig. 2).

The 2R-WGD events proposed by Ohno lie in the distant past, and given the complexity of biological networks deciphering the later subfunctionalization, neofunctionalization, and nonfunctionalization of resulting ohnologs at a global level is a formidable task. Receptor tyrosine kinases comprise a large, highly networked family of proteins encoded by genes recently reported to have been generated largely through 2R-WGD (Brunet et al. 2016). In a conceptually similar manner, we have investigated deubiquitinating enzymes as a microcosm of gene evolution following 2R-WGD, and in addition to illuminating the evolutionary relationships of DUB genes have identified specific instances in which subfunctionalization, neofunctionalization, and nonfunctionalization has occurred in the context of innate immunity and DNA repair. There is accumulating evidence in the literature that members of the same repertoire of actors play important roles in other

cellular systems (developmental pathways involving Wnt or TGF- $\beta$ , or tumour suppressor mechanisms). These added roles may impart added selective constraints on DUB gene evolution, which in the examples just given would be limited to metazoans in which the pathways are operative. This may be a fruitful direction for future research, but with regard to gene evolution it is clear that within the DUB superfamily there is more to be learned.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) in the form of Discovery Grants RGPIN/2013-261252 to X.X. and RGPIN/2015-05879 to D.A.G.

## Literature Cited

- Baran Y, et al. 2015. The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* 25:927–936.
- Barrachina M, et al. 2007. Target genes of neuron-restrictive silencer factor are abnormally up-regulated in human myotilinopathy. *Am J Pathol.* 171:1312–1323.
- Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci.* 104:17004–17009.
- Blanchette P, Gilchrist CA, Baker RT, Gray DA. 2001. Association of UNP, a ubiquitin-specific protease, with the pocket proteins pRb, p107 and p130. *Oncogene* 20:5533–5537.
- Breuer K, et al. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* 41:D1228–D1233.
- Brunet FG, Volff J-N, Scharl M. 2016. Whole genome duplications shaped the receptor tyrosine kinase repertoire of jawed vertebrates. *Genome Biol Evol.* 8:1600–1613.
- Burke F. 2014. The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harb Perspect Biol.* 6(5): a016147.
- Burrows JF, McGrattan MJ, Johnston JA. 2005. The DUB/USP17 deubiquitinating enzymes, a multigene family within a tandemly repeated sequence. *Genomics* 85:524–529.
- Burrows JF, Scott CJ, Johnston JA. 2010. The DUB/USP17 deubiquitinating enzymes: a gene family within a tandemly repeated sequence, is also embedded within the copy number variable Beta-defensin cluster. *BMC Genomics* 11:250.
- Çalışkan M, Cusanovich DA, Ober C, Gilad Y. 2011. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum Mol Genet.* 20:1643–1652.
- Chen C, Zhang Y-B, Gui J-F. 2015. Expression characterization, genomic structure and function analysis of fish ubiquitin-specific protease 18 (USP18) genes. *Dev Comp Immunol.* 52:112–122.
- Citterio E. 2015. Fine-tuning the ubiquitin code at DNA double-strand breaks: deubiquitinating enzymes at work. *Front Genet.* 6:282.
- Cui J, et al. 2014. USP3 inhibits type I interferon signaling by deubiquitinating RIG-I-like receptors. *Cell Res.* 24:400–416.
- Day INM, Thompson RJ. 2010. UCHL1 (PGP 9.5): neuronal biomarker and ubiquitin system protein. *Prog Neurobiol.* 90:327–362.



- Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* [Internet] 3. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1197285/>.
- Drouin G. 2006. Processed pseudogenes are more abundant in human and mouse X chromosomes than in autosomes. *Mol Biol Evol.* 23:1652–1655.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Esquerdo M, et al. 2016. Expression atlas of the deubiquitinating enzymes in the adult mouse retina, their evolutionary diversification and phenotypic roles. *PLoS One* 11:e0150364.
- Fan Y-H, et al. 2011. USP4 targets TAK1 to downregulate TNF $\alpha$ -induced NF- $\kappa$ B activation. *Cell Death Differ.* 18:1547–1560.
- Flajnik MF, Rummel LL, 2000. The immune system of cartilaginous fish. In: Du Pasquier L, editors *Origin and evolution of the vertebrate immune system. current topics in microbiology and immunology.* Berlin, Heidelberg: Springer. p. 249–270. Available from: [http://link.springer.com.proxy.bib.uottawa.ca/chapter/10.1007/978-3-642-59674-2\\_11](http://link.springer.com.proxy.bib.uottawa.ca/chapter/10.1007/978-3-642-59674-2_11).
- Gack MU, et al. 2007. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* 446:916–920.
- Goldmann T, et al. 2015. USP18 lack in microglia causes destructive interferonopathy of the mouse brain. *EMBO J.* 34:1612–1629.
- Grau-Bové X, Sebé-Pedrós A, Ruiz-Trillo I. 2015. The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. *Mol Biol Evol.* 32:726–739.
- Guan R, et al. 2011. Structural basis for the sequence-specific recognition of human ISG15 by the NS1 protein of influenza B virus. *Proc Natl Acad Sci U S A.* 108:13468–13473.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.
- Higgs R, et al. 2008. The E3 Ubiquitin Ligase RO52 negatively regulates IFN- $\beta$  production post-pathogen recognition by polyubiquitin-mediated degradation of IRF3. *J Immunol.* 181:1780–1786.
- Hutchins AP, Liu S, Diez D, Miranda-Saavedra D. 2013. The repertoires of ubiquitinating and deubiquitinating enzymes in eukaryotic genomes. *Mol Biol Evol.* 30:1172–1187.
- Johnson GD, et al. 2011. The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction* 141:21–36.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20:1313–1326.
- Kanehisa M, et al. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34:D354–D357.
- Karim R, et al. 2013. Human papillomavirus (HPV) upregulates the cellular deubiquitinase UCHL1 to suppress the keratinocyte's innate immune response. *PLoS Pathog* 9:e1003384.
- Kato K, et al. 2014. Fine-tuning of DNA damage-dependent ubiquitination by OTUB2 supports the DNA repair pathway choice. *Mol Cell.* 53:617–630.
- Kayagaki N, et al. 2007. DUBA: a deubiquitinase that regulates type I interferon production. *Science* 318:1628–1632.
- Kee Y, Huang TT. 2015. The role of deubiquitinating enzymes in DNA repair. *Mol Cell Biol.* 36(4):524–544.
- Keusekotten K, et al. 2013. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell* 153:1312–1326.
- Komander D, Rape M. 2012. The ubiquitin code. *Annu Rev Biochem.* 81:203–229.
- Koonin EV. 2015. Archaeal ancestors of eukaryotes: not so elusive any more. *BMC Biol.* 13:1–7.
- Li WH. 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. *Mol Evol.* 36:96–99.
- Lin D, et al. 2015. Induction of USP25 by viral infection promotes innate antiviral responses by mediating the stabilization of TRAF3 and TRAF6. *Proc Natl Acad Sci U S A.* 112:11324–11329.
- Liu H, et al. 2015. The deubiquitylating enzyme USP4 cooperates with CtIP in DNA double-strand break end resection. *Cell Rep.* 13:93–107.
- Liu J, et al. 2011. JTV1 co-activates FBP to induce USP29 transcription and stabilize p53 in response to oxidative stress. *EMBO J.* 30:846–858.
- Liu J, Elmore JM, Coaker G. 2009. Investigating the functions of the RIN4 protein complex during plant innate immune responses. *Plant Signal Behav.* 4:1107–1110.
- Loeb KR, Haas AL. 1992. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J Biol Chem.* 267:7806–7813.
- Lonsdale J, et al. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 45:580–585.
- Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* 252:1162–1164.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Mahul-Mellier A-L, et al. 2012. De-ubiquitinating proteases USP2a and USP2c cause apoptosis by stabilising RIP1. *Biochim Biophys Acta.* 1823:1353–1365.
- Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. *Phil Trans R Soc B.* 370:20140330.
- McCarrey JR. 1987. Nucleotide sequence of the promoter region of a tissue-specific human retroposon: comparison with its housekeeping progenitor. *Gene* 61:291–298.
- Melé M, et al. 2015. The human transcriptome across tissues and individuals. *Science* 348:660–665.
- Nagabhushana A, Bansal M, Swarup G. 2011. Optineurin is required for CYLD-dependent inhibition of TNF $\alpha$ -induced NF- $\kappa$ B activation. *PLoS ONE* 6:e17477.
- Nguyen Ba AN, et al. 2014. Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. *PLoS Comput Biol.* [Internet] 10. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256066/>.
- Nishi R, et al. 2014. Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nat Cell Biol.* 16:1016–1026.
- Niu J, et al. 2013. USP10 inhibits genotoxic NF- $\kappa$ B activation by MCPIP1-facilitated deubiquitination of NEMO. *EMBO J.* 32:3206–3219.
- Nunoura T, et al. 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39:3204–3223.
- Oshikawa K, Matsumoto M, Oyama K, Nakayama KI. 2012. Proteome-wide identification of ubiquitylation sites by conjugation of engineered lysine-less ubiquitin. *J Proteome Res.* 11:796–807.
- Paulding CA, Ruvolo M, Haber DA. 2003. The Tre2 (USP6) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci U S A.* 100:2507–2511.
- Potrzebowski L, Vinckenbosch N, Kaessmann H. 2010. The emergence of new genes on the young therian X. *Trends Genet.* 26:1–4.
- Sato Y, et al. 2012. Molecular basis of Lys-63-linked polyubiquitination inhibition by the interaction between human deubiquitinating enzyme OTUB1 and ubiquitin-conjugating enzyme UBC13. *J Biol Chem.* 287:25860–25868.
- Schulz S, et al. 2012. Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep.* 13:930–938.
- Seki T, et al. 2013. Jsd1, a membrane-targeted deubiquitinating enzyme, is activated by ubiquitination and regulates membrane dynamics, cell motility, and endocytosis. *J Biol Chem.* 288:17145–17155.
- Soumilion M, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 3:2179–2190.

- Sowa ME, Bennett EJ, Gygi SP, Harper JW. 2009. Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138:389–403.
- Speer SD, et al. 2016. ISG15 deficiency and increased viral resistance in humans but not mice. *Nat Commun*. 7:11496.
- Sun S-C. 2008. Deubiquitylation and regulation of the immune response. *Nat Rev Immunol*. 8:501–511.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 10:512–526.
- Ungureanu D, Saharinen P, Junttila I, Hilton DJ, Silvennoinen O. 2002. Regulation of Jak2 through the ubiquitin-proteasome pathway involves phosphorylation of Jak2 on Y1007 and interaction with SOCS-1. *Mol Cell Biol*. 22:3316–3326.
- Valero R, et al. 2001. Characterization of alternatively spliced products and tissue-specific isoforms of USP28 and USP25. *Genome Biol*. 2:RESEARCH0043.
- Varshavsky A. 2012. The ubiquitin system, an immense realm. *Annu Rev Biochem*. 81:167–176.
- Venkatesh B, et al. 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol*. 5:e101.
- Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179.
- Vilella AJ, et al. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 19:327–335.
- Vlasschaert C, Xia X, Coulombe J, Gray DA. 2015. Evolution of the highly networked deubiquitinating enzymes USP4, USP15, and USP11. *BMC Evol Biol*. 15:230.
- Vlasschaert C, Xia X, Gray DA. 2016. Selection preserves Ubiquitin Specific Protease 4 alternative exon skipping in therian mammals. *Sci Rep*. 6:20039.
- Wada K, Tanji K, Kamitani T. 2006. Oncogenic protein UnpEL/Usf4 deubiquitinates Ro52 by its isopeptidase activity. *Biochem Biophys Res Commun*. 339:731–736.
- Wang L, et al. 2013. USP4 positively regulates RIG-I-mediated antiviral response through deubiquitination and stabilization of RIG-I. *J Virol*. 87:4507–4515.
- Wang W, et al. 2015. TRAF Family Member-associated NF- $\kappa$ B Activator (TANK) inhibits genotoxic nuclear factor  $\kappa$ B activation by facilitating deubiquitinase USP10-dependent deubiquitination of TRAF6 ligase. *J Biol Chem*. 290:13372–13385.
- Weeks SD, Grasty KC, Hernandez-Cuevas L, Loll PJ. 2011. Crystal structure of a Josephin-ubiquitin complex: evolutionary restraints on ataxin-3 deubiquitinating activity. *J Biol Chem*. 286:4555–4565.
- Wijnhoven P, et al. 2015. USP4 auto-deubiquitylation promotes homologous recombination. *Mol Cell*. 60:362–373.
- Wolfe K. 2000. Robustness: it's not where you think it is. *Nat Genet*. 25:3–4.
- Wyffels J, King B L, et al. 2014. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. F1000Research [Internet] 3. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4184313/>.
- Xiao N, et al. 2012. Ubiquitin-specific protease 4 (USP4) targets TRAF2 and TRAF6 for deubiquitination and inhibits TNF $\alpha$ -induced cancer cell migration. *Biochem J*. 441:979–986.
- Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol*. 30:1720–1728.
- Xia X, Kumar S. 2006. Codon-based detection of positive selection can be biased by heterogeneous distribution of polar amino acids along protein sequences. *Comput Syst Bioinformatics Conf*. 2006:335–340.
- Yan L, et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol*. 20:1131–1139.
- Yuan W, Krug RM. 2001. Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein. *EMBO J*. 20:362–371.
- Zaman MM-U, et al. 2013. Ubiquitination-deubiquitination by the TRIM27-USP7 complex regulates tumor necrosis factor alpha-induced apoptosis. *Mol Cell Biol*. 33:4971–4984.
- Zapata JM, et al. 2001. A diverse family of proteins containing tumor necrosis factor receptor-associated factor domains. *J Biol Chem*. 276:24242–24252.
- Zhang D, Zaugg K, Mak TW, Elledge SJ. 2006. A role for the deubiquitinating enzyme USP28 in control of the DNA-damage response. *Cell* 126:529–542.
- Zhang X, Berger FG, Yang J, Lu X. 2011. USP4 inhibits p53 through deubiquitinating and stabilizing ARF-BP1. *EMBO J*. 30:2177–2189.
- Zhang Z, Yang H, Wang H. 2014. The histone H2A deubiquitinase USP16 interacts with HERC2 and fine-tunes cellular response to DNA damage. *J Biol Chem*. 289:32883–32894.
- Zhong H, et al. 2013. Ubiquitin-specific proteases 25 negatively regulates virus-induced type I interferon signaling. *PLoS One* 8:e80976.
- Zhu G, Wu C-J, Zhao Y, Ashwell JD. 2007. Optineurin negatively regulates TNF $\alpha$ -induced NF- $\kappa$ B activation by competing with NEMO for ubiquitinated RIP. *Curr Biol*. 17:1438–1443.

Associate editor: Bill Martin