Liat1, an arginyltransferase-binding protein whose evolution among primates involved changes in the numbers of its 10-residue repeats

Christopher S. Brower[1], Connor E. Rosen, Richard H. Jones[2], Brandon C. Wadas, Konstantin I. Piatkov, and Alexander Varshavsky[3]

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Alexander Varshavsky, October 13, 2014 (sent for review October 6, 2014; reviewed by Wolfgang Baumeister and Avram Hershko)

The arginyltransferase Ate1 is a component of the N-end rule pathway, which recognizes proteins containing N-terminal degradation signals called N-degrons, polyubiquitylates these proteins, and thereby causes their degradation by the proteasome. At least six isoforms of mouse Ate1 are produced through alternative splicing of *Ate1* pre-mRNA. We identified a previously uncharacterized mouse protein, termed Liat1 (ligand of Ate1), that interacts with Ate1 but does not appear to be its arginylation substrate. Liat1 has a higher affinity for the isoforms Ate1^{1A7A} and Ate1^{1B7A}. Liat1 stimulated the in vitro N-terminal arginylation of a model substrate by Ate1. All examined vertebrate and some invertebrate genomes encode proteins sequelogous (similar in sequence) to mouse Liat1. Sequelogs of Liat1 share a highly conserved ∼30-residue region that is shown here to be required for the binding of Liat1 to Ate1. We also identified non-Ate1 proteins that interact with Liat1. In contrast to *Liat1* genes of nonprimate mammals, *Liat1* genes of primates are subtelomeric, a location that tends to confer evolutionary instability on a gene. Remarkably, Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, whereas Liat1 proteins of other mammals contain a single copy of this motif. Quantities of these repeats are, in general, different in Liat1 of different primates. For example, there are 1, 4, 13, 13, 17, and 17 repeats in the gibbon, gorilla, orangutan, bonobo, neanderthal, and human Liat1, respectively, suggesting that repeat number changes in this previously uncharacterized protein may contribute to evolution of primates.

Ate1 | arginylation | N-end rule | tandem repeats | primate evolution

The N-end rule pathway recognizes proteins containing N-terminal degradation signals called N-degrons, polyubiquitylates these proteins, and thereby causes their degradation by the proteasome (Fig. 1 *A* and *B*) (1–9). The main determinant of an N-degron is a destabilizing N-terminal residue of a protein. Recognition components of the N-end rule pathway are called N-recognins. In eukaryotes, N-recognins are E3 ubiquitin (Ub) ligases that can target N-degrons (Fig. 1). Bacteria also contain a (Ub-independent) version of the N-end rule pathway (10, 11).

Regulated degradation of proteins or their fragments by the N-end rule pathway mediates a strikingly broad range of functions, including: the sensing of heme, nitric oxide, oxygen, and short peptides; control of protein quality and subunit stoichiometries, including the elimination of misfolded proteins; regulation of G proteins; repression of neurodegeneration; regulation of apoptosis, chromosome cohesion/segregation, transcription, and DNA repair; control of peptide import; regulation of meiosis, autophagy, immunity, fat metabolism, cell migration, actin filaments, cardiovascular development, spermatogenesis, and neurogenesis; the functioning of adult organs, including the brain, muscle and pancreas; and the regulation of many processes in plants (4–9) (Fig. 1 *A* and *B*; see *Supporting Information* for an expanded legend and references to this figure).

In eukaryotes, the N-end rule pathway consists of two branches. One branch, called the Ac/N-end rule pathway, targets proteins for degradation through their N$^{\alpha}$-terminally acetylated (Nt-acetylated) residues (Fig. 1*B*) (2, 3, 12). Degradation signals and E3 Ub ligases of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively. Nt-acetylation of cellular proteins is apparently irreversible, in contrast to acetylation-deacetylation of internal Lys residues. Approximately 90% of human proteins are cotranslationally Nt-acetylated by ribosome-associated Nt-acetylases (13). Posttranslational Nt-acetylation occurs as well. Many, possibly most, Nt-acetylated proteins contain Ac/N-degrons (Fig. 1*B*) (2–4, 12).

The pathway's other branch, called the Arg/N-end rule pathway, targets specific unacetylated N-terminal residues (Fig. 1*A*) (3, 14–16). The "primary" destabilizing N-terminal Arg, Lys, His, Leu, Phe, Tyr, Trp, and Ile, are directly recognized by N-recognins. The unacetylated N-terminal Met, if it is followed by a bulky hydrophobic (Φ) residue, also acts as a primary destabilizing residue (Fig. 1*A*) (3). In contrast, unacetylated N-terminal Asn, Gln, Asp, and Glu (as well as Cys, under some metabolic conditions) are destabilizing because of their preliminary enzymatic modifications, which include N-terminal deamidation (Nt-deamidation) of Asn and Gln and Nt-arginylation of Asp, Glu, and oxidized Cys (Fig. 1*A*) (4, 6, 7, 17).

Nt-arginylation is mediated by the *Ate1*-encoded arginyltransferase (Arg-tRNA-protein transferase; R-transferase), a

Significance

We describe the discovery and analyses of a previously uncharacterized protein, termed Liat1 (ligand of Ate1), which has been identified because of its interactions with the Ate1 arginyltransferase, a component of the N-end rule pathway of protein degradation. All vertebrates contain proteins similar to mouse Liat1. Remarkably, Liat1 proteins of some primates contain tandem repeats of a 10-residue sequence, whereas Liat1 proteins of other mammals contain a single copy of this motif. Quantities of these repeats are, in general, different in Liat1 of different primates. For example, there are 4, 13, 13, and 17 repeats in the gorilla, orangutan, bonobo, and human Liat1, respectively, suggesting that repeat number changes in this previously uncharacterized protein may contribute to evolution of primates.

Author contributions: C.S.B., C.E.R., and A.V. designed research; C.S.B., C.E.R., R.H.J., B.C.W., and K.I.P. performed research; B.C.W. and K.I.P. contributed new reagents/analytic tools; C.S.B., C.E.R., and A.V. analyzed data; and C.S.B., C.E.R., B.C.W., and A.V. wrote the paper.

Reviewers included: W.B., Max Planck Institute of Biochemistry; and A.H., Technion Israel Institute of Technology.

The authors declare no conflict of interest.

[1]Present address: Department of Biology, Texas Woman's University, Denton, TX 76204.

[2]Present address: Department of Radiology, University of Minnesota, Minneapolis, MN 55455.

[3]To whom correspondence should be addressed. Email: avarsh@caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1419587111/-/DCSupplemental.

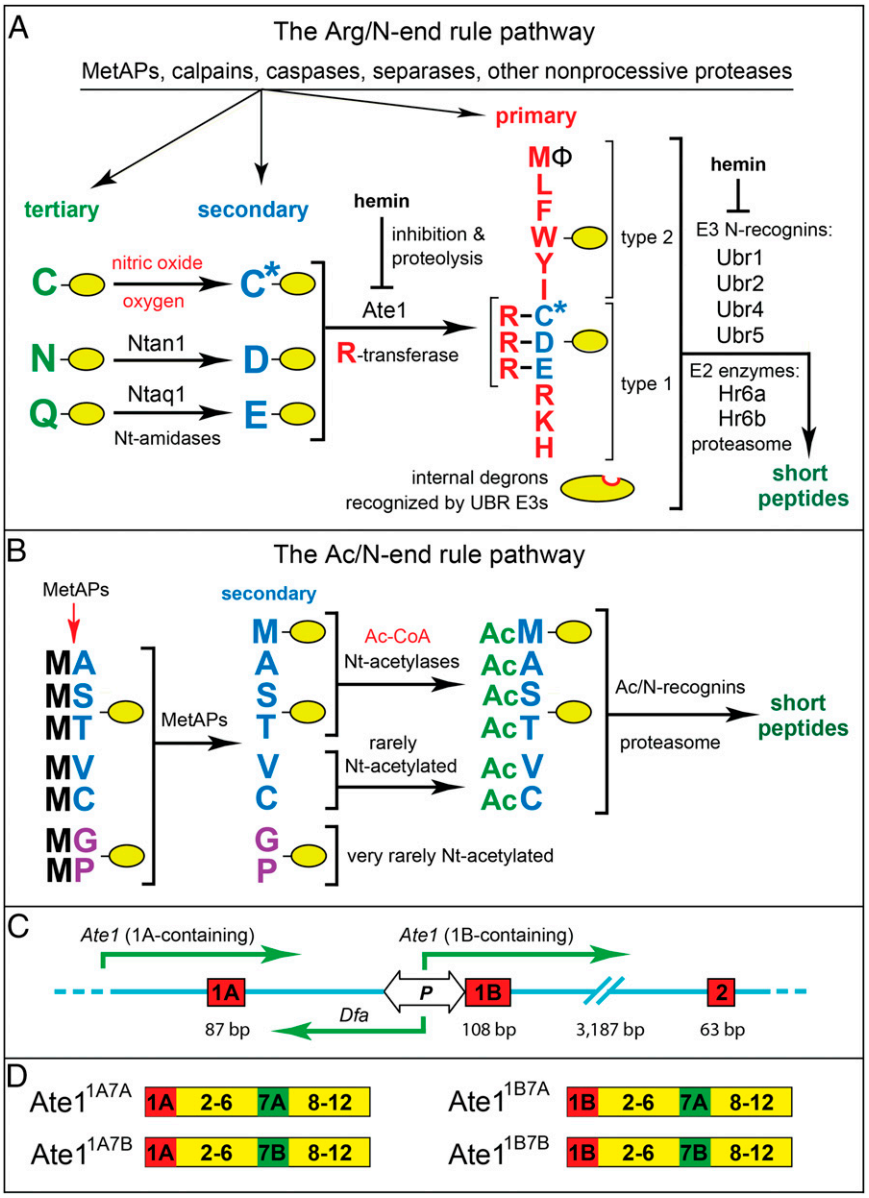

Fig. 1. The mammalian N-end rule pathway and the Ate1 arginyltransferase (R-transferase). See introductory paragraphs for a summary of the pathway's mechanistic aspects and biological functions. N-terminal residues are denoted by single-letter abbreviations. A yellow oval denotes the rest of a protein substrate. (*A*) The Arg/N-end rule pathway. It recognizes proteins through their unacetylated N-terminal residues and contains the N-terminal arginylation (Nt-arginylation) branch, mediated by the arginyltransferase (R-transferase) Ate1 and by the Nt-amidases Ntan1 and Ntaq1 that act upstream of Ate1 (4–6). (*B*) The Ac/N-end rule pathway. It recognizes proteins through their N^{α}-terminally acetylated (Nt-acetylated) residues (2, 3, 12). Red arrow on the left indicates the removal of the N-terminal Met residue by Met-aminopeptidases (MetAPs). N-terminal Met is retained if a residue at position 2 is larger than Val. (*C*) The bidirectional $_{Dfa}P_{Ate1}$ promoter upstream of exon 1B of the mouse *Ate1* gene (4, 21, 27). (*D*) Four major mouse Ate1 R-transferase isoforms and their designations (4, 21). See *Supporting Information* for a detailed legend and supplementary references to this figure.

component of the Arg/N-end rule pathway and a subject of the present study (Fig. 1 *A* and *D*) (18–23). Alternative splicing of mouse *Ate1* pre-mRNAs yields at least six R-transferase isoforms, which differ in their Nt-arginylation activity (Fig. 1 *C* and *D*) (18, 21). R-transferases are sequelogous (similar in sequence) (24) throughout most of their ∼60-kDa spans from fungi to mammals (4). R-transferase can arginylate not only N-terminal Asp and Glu, but also N-terminal Cys, if it has been oxidized to Cys-sulfinate or Cys-sulfonate, through reactions mediated by NO, oxygen, and N-terminal Cys-oxidases (6, 7, 20). The resulting circuits can act as sensors of NO and oxygen in a cell through reactions that start with a conditional oxidation of N-terminal Cys in proteins such as the Rgs4, Rgs5, and Rgs16 regulators of G proteins in mammals (4, 20, 25) and specific transcriptional regulators in plants (reviewed in refs. 4, 6, and 7).

There are dozens of either identified Nt-arginylated proteins (including natural protein fragments) or proteins that are predicted to be Nt-arginylated. Many, possibly most, of these proteins are conditionally or constitutively short-lived substrates of the Arg/N-end rule pathway (Fig. 1*A*) (14, 15, 20, 26, and references therein). In contrast, there were, until now, no analyzed protein ligands of R-transferase that did not appear to be its substrates. We describe a previously uncharacterized mouse protein, termed Liat1 (ligand of Ate1) that binds to the mouse Ate1 R-transferase, is apparently not arginylated by it, and has a higher affinity for specific splicing-derived Ate1 isoforms. We

also identified proteins other than R-transferase that appear to interact with Liat1. Remarkably, Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, in contrast to a single copy of this motif in Liat1 of other mammals. Quantities of repeats in Liat1 are, in general, different among different primates, suggesting that repeat number changes in this previously uncharacterized protein may contribute to evolution of primates.

Results

Alternative Ate1 Exons and Their Sequelogies (Sequence Similarities). The 59-kDa mouse R-transferase is encoded by *Ate1*. This gene, located on chromosome 7 (7-F3), contains 14 protein-coding exons that encompass, together with introns, ~128 kb of *Ate1* DNA (Fig. 1 *C* and *D*) (18, 21, 23, 27). Mouse *Ate1* encodes at least six splicing-derived Ate1 isoforms. Fig. 1*D* shows designations and exon compositions of the four major isoforms; they are enzymatically active R-transferases whose levels of expression vary among different mouse tissues (21). The splicing of *Ate1* pre-mRNAs involves transcription from two distinct *Ate1* promoters and the alternative utilization of two exon pairs: (*i*) either exon 1A or 1B, and (*ii*) either exon 7A or 7B (Fig. 1 *C* and *D*). The alternative exons 1A and 1B encode two sequelogous (24) ~30-residue N-terminal regions of R-transferase (Fig. 1*D* and Fig. S1) (21). The alternative exons 7A and 7B encode two sequelogous 43-residue regions of R-transferase (Fig. 1*D* and Fig. S1).

Two-Hybrid Detection of a Protein That Interacts with Ate1. In a search for mouse proteins that specifically interact with mouse R-transferase without being arginylation substrates, we used a yeast-based two-hybrid (Y2H) assay, using a mouse testis cDNA library and a Gal4 DNA-binding domain (DBD)–Ate1^{1B7A} fusion as bait. This screen identified nine independent positive DNA clones encoding different (overlapping) segments of a previously uncharacterized mouse protein (LOC74230; Acc. NP_941039) (Fig. 2*A*). This 228-residue protein, encoded by a single-copy mouse gene, was termed Liat1 (Fig. 3*A*).

To verify and expand these results, we expressed, in *Saccharomyces cerevisiae* of appropriate genetic backgrounds, Gal4 DBD fusions to three mouse Ate1 isoforms, Ate1^{1B7A}, Ate1^{1A7A} or Ate1^{1B7B}, and a Gal4–activation domain (AD) fusion to the full-length mouse Liat1. These strains were assayed for their ability to grow on minimal (SD) media lacking Leu, Trp, His, and Ade, thereby indicating the presence or absence of interactions among the examined protein fusions. The results were in agreement with the findings by the initial two-hybrid screen in that the Ate1^{1B7A} R-transferase isoform interacted with Liat1 (Figs. 1*D* and 2*A*).

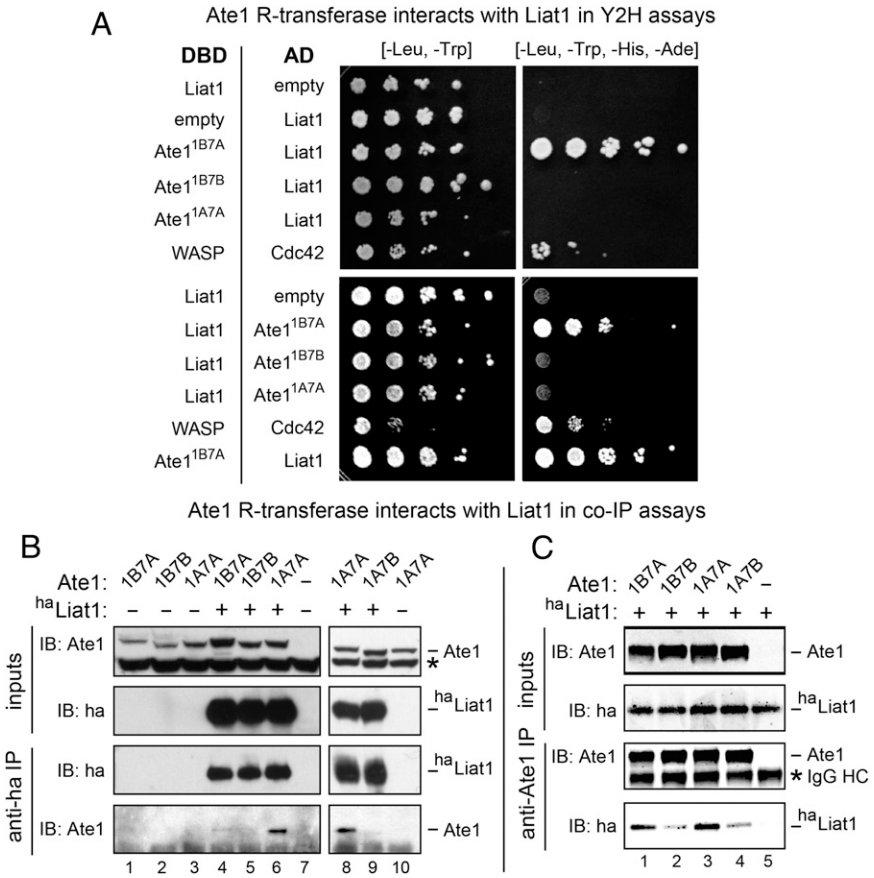

Fig. 2. Detection and characterization of interactions between the mouse Ate1 R-transferase and a previously uncharacterized mouse protein termed Liat1. (*A*) Detection of Ate1–Liat1 interactions using a Y2H assay. A Gal4 DBD–Ate1^{1B7A} fusion and other DBD-based fusions (e.g., a control DBD fusion containing the WASP protein) vs. coexpressed Gal4–AD fusions to full-length Liat1 and to other proteins, including specific Ate1 isoforms as well as the control (WASP-interacting) Cdc42 protein. The ability to grow on minimal media lacking Leu, Trp, His, and Ade was used to detect interactions among the examined protein fusions. See *SI Experimental Procedures* for additional information. (*B*) Detection of Ate1–Liat1 interactions using co-IP assays with mouse cell extracts. Plasmids expressing ^{3ha}Liat1 and specific Ate1 isoforms were transiently expressed in mouse *Ate1*$^{-/-}$ EF cells, followed by immunoprecipitations with anti-ha antibody, SDS/PAGE of immunoprecipitates, and immunoblotting with both anti-ha and anti-Ate1 antibodies. (*C*) Same as in *B* but using purified mouse ^{3ha}Liat1 and purified mouse Ate1 isoforms, with immunoprecipitation by anti-Ate1 antibody.

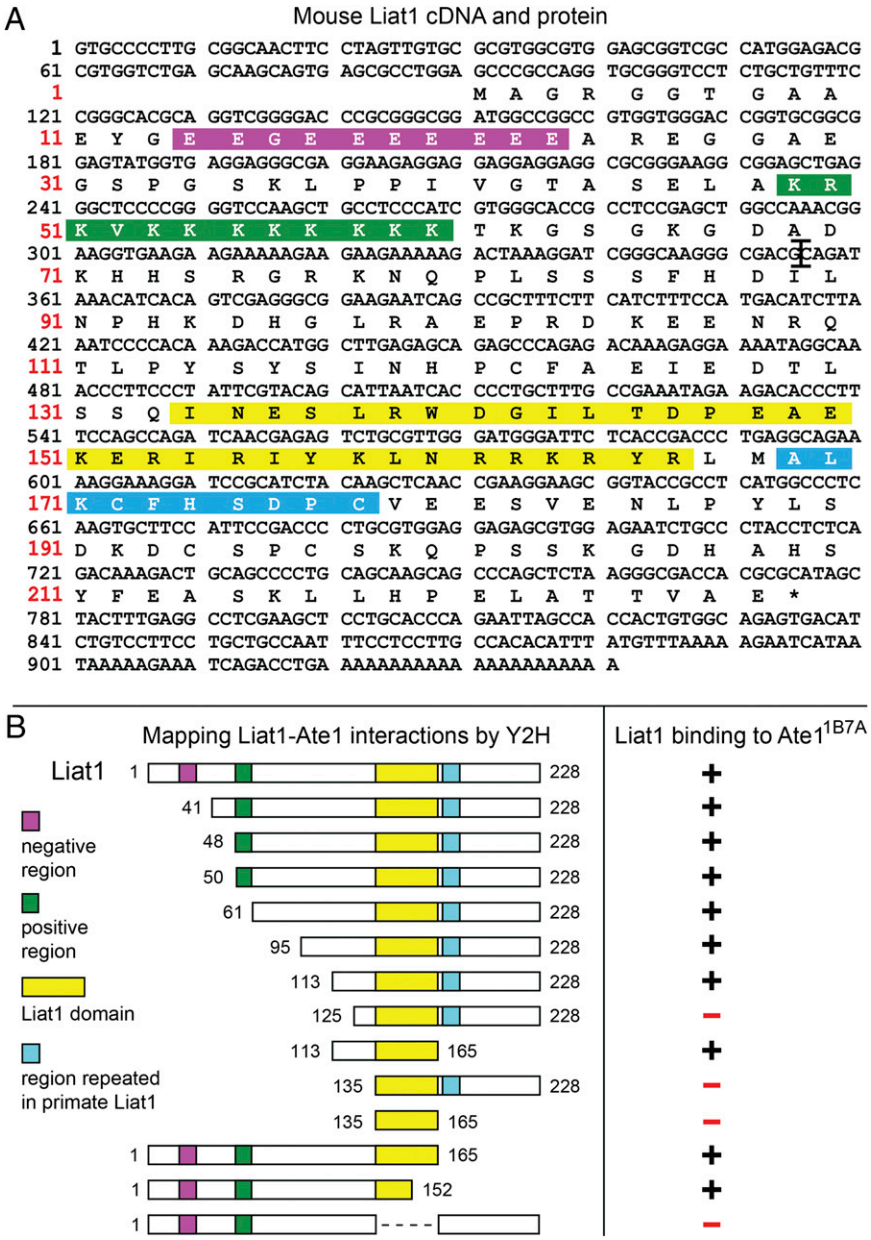

Fig. 3. Mouse Liat1 and the mapping of its Ate1-binding region. (*A*) Nucleotide sequence of mouse *Liat1* cDNA and amino acid sequence features of mouse Liat1 (NP_941039). A large black "I" after the Lys-rich region denotes the position of a single 2,737-bp intron between two protein-coding exons in the *Liat1* genomic DNA. The purple, green, yellow, and blue rectangles denote, respectively, the negatively charged region of mouse Liat1, its positively charged region, its particularly highly conserved ~30-residue domain, termed the Liat1 domain, and a 10-residue region that becomes tandemly repeated in Liat1 of some primates, including humans, but is a single-copy sequence in Liat1 of other mammals, including mouse Liat1. The same color scheme is used to denote these regions of Liat1 in other figures of this paper. Black and red numbers on the left indicate nucleotide and amino acid residue numbers, respectively. See also the main text and Figs. S2–S4. (*B*) Mapping Liat1–Ate1 interactions using Y2H assays. The Gal4 DBD–Ate1^{1B7A} fusion was examined, using Y2H, for interactions with the coexpressed Gal4–AD fusion to either full-length Liat1 or to the indicated AD fusions encoding Liat1 fragments. The results of this assay are summarized on the right, and the color coding of Liat1 domains (the same as in *A*) is indicated on the left.

However, the two-hybrid assay did not detect interactions of Liat1 with the Ate1^{1A7A} and Ate1^{1B7B} isoforms, in contrast to the Ate1^{1B7A} isoform (Figs. 1*D* and 2*A*). This result was unexpected because differences among Ate1 isoforms are significant but not large, given sequelogies (24) between the encoded sequences of the Ate1 exons 1A vs. 1B (36% identity) and 7A vs. 7B (30% identity), respectively (Fig. 1*D* and Fig. S1). Although the affinity of Ate1^{1A7A} and Ate1^{1B7B} for Liat1 was too low for detection by two-hybrid assays (in contrast to the affinity of Ate1^{1B7A}) (Figs. 1*D* and 2*A*), an immunoprecipitation assay, described below, did detect a complex between Liat1 and the Ate1^{1A7A} isoform. One ramification of these two-hybrid results was a high likelihood that the interaction between Liat1 and Ate1^{1B7A} was specific, given the ability of Liat1 to distinguish, in its binding patterns, among three similar Ate1 isoforms (Figs. 1*D* and 2*A*, and Fig. S1).

A Region of Liat1 Required for Interaction with Ate1. The Y2H assay was also used, using the mouse Ate1^{1B7A} isoform and either N-terminal/C-terminal truncations or internal deletions of mouse Liat1, to delineate a region of Liat1 that was required for

the observed interaction (Figs. 2*A* and 3*B*). Liat1$^{113\text{-}165}$, a 53-residue internal segment of the 228-residue Liat1 that contained a particularly strongly conserved ∼30-residue region (termed the Liat1 domain; see below and Fig. S2) was sufficient for the Liat1–Ate1^{1B7A} interaction, whereas the Liat1 domain alone did not suffice (Fig. 3*B*). It was also found that only a part of the Liat1 domain was required for the binding of a C-terminally truncated Liat1 to Ate1^{1B7A}, provided that a region immediately upstream of the Liat1 domain was present as well (Fig. 3*B*).

Coimmunopreciptation Assays. We also transiently coexpressed, in mouse *Ate1*$^{-/-}$ embryonic fibroblasts (EFs) lacking R-transferase (19), the N-terminally triple ha-tagged mouse Liat1 (^{3ha}Liat1) and one of the four untagged mouse R-transferases, Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, or Ate1^{1B7B} (Figs. 1*D* and 2*B*). ^{3ha}Liat1 in cell extracts was immunoprecipitated with anti-ha antibody, followed by SDS/PAGE and immunoblotting with either anti-ha or the previously characterized, affinity-purified antibody to mouse Ate1 (20). The results of these coimmunoprecipitation (co-IP) assays were reproducible in independent assays and showed an efficacious co-IP of ^{3ha}Liat1 with Ate1^{1A7A} (Fig. 2*B*, lanes 6 and 8). A less efficacious but still detectable co-IP with Ate1^{1B7A} was also observed (Fig. 2*B*, lane 4 vs. lanes 6 and 8), whereas co-IP with the isoforms Ate1^{1A7B} and Ate1^{1B7B} was negligible (Fig. 2*B*).

We also carried out co-IP assays with recombinant proteins that had been expressed in *Escherichia coli* using the Ub fusion technique and were purified by affinity chromatography that included the removal of Ub moiety (28, 29). These procedures yielded purified ^{3ha}Liat1 and purified, untagged Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B} or Ate1^{1B7B}. Equal amounts of Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, or Ate1^{1B7B} were incubated with an ∼fourfold molar excess of purified full-length ^{3ha}Liat1, followed by immunoprecipitation with anti-Ate1, SDS/PAGE, and immunoblotting with either anti-ha or anti-Ate1 (Fig. 2*C*). These "carrier-free" co-IP assays with purified proteins (instead of cell extracts containing these proteins) detected interactions between Liat1 and all four Ate1 isoforms (Fig. 2*C*). Note, however, that co-IPs of purified ^{3ha}Liat1 with purified Ate1^{1B7A} and Ate1^{1A7A} were significantly more efficacious than co-IPs with the other two isoforms, Ate1^{1A7B} and Ate1^{1B7B} (Fig. 2*C*, lanes 1 and 3 vs. lanes 2 and 4). These results were in qualitative agreement with the findings by both two-hybrid assays and co-IP assays with cell extracts, neither of which could detect Liat1 interactions with the isoforms Ate1^{1A7B} and Ate1^{1B7B} (Fig. 2 *B* and *C*). In addition, co-IPs with purified Liat1 and purified Ate1^{1B7A} or Ate1^{1A7A} (Fig. 2*C*) indicated that Liat1–Ate1 interactions did not require other proteins.

We do not understand the cause of the reproducible difference in which Ate1^{1B7A} was the only Liat1-interacting Ate1 isoform in two-hybrid assays (Figs. 2*A* and 3*B*), whereas Ate1^{1A7A} was the apparently highest-affinity Liat1-interacting Ate1 isoform in co-IP assays with cell extracts (Fig. 2*B*). A plausible possibility is a different pattern of modifications (e.g., phosphorylation) of Ate1^{1B7A} vs. Ate1^{1A7A} that influences the outcomes of two different kinds of binding experiments (protein fusions in Y2H assays in vivo vs. epitope-tagged ^{3ha}Liat1 and untagged Ate1 in co-IP assays with mammalian cell extracts in vitro). The other two isoforms, Ate1^{1A7B} and Ate1^{1B7B}, were negative in regard to interactions with Liat1 in both two-hybrid assays and co-IP assays with cell extracts (Figs. 2*B* and 3*B*). The noninteracting (or weakly interacting) Ate1^{1A7B} and Ate1^{1B7B} isoforms both contained the 7B exon, in contrast to the other two Ate1 isoforms (Fig. 1*D* and Fig. S1).

Sequence Features of the Mouse Liat1 Protein. The mouse *Liat1* cDNA (1700016K19Rik; Acc. NM_198637) encodes a 228-residue (25.5 kDa), previously uncharacterized protein (NP_941039) with a deduced pI of 7.45 (Fig. 3*A*). The *Liat1* gene, located on the mouse chromosome 11-B5, contains a single 2,737-bp intron between two protein-coding exons, and no alternative mouse *Liat1* cDNA isoforms could be detected in databases (Figs. 3*A* and 4). Amino acid sequence features of mouse Liat1 include a negatively charged 10-residue region (9 of 10 residues are Glu) and a positively charged 12-residue region (11 of 12 residues are basic, largely Lys) (Fig. 3*A*). These regions are present in Liat1 of all examined mammals, but Liat1 of other vertebrates can lack either one of the two charged regions or both of them (Fig. 4*B*).

An aspect of Liat1 that is particularly conserved in evolution is a ∼30-residue region termed the Liat1 domain (Figs. 3 and 4, and Figs. S2–S4). The 32-residue mouse and human Liat1 domains are 96% identical (Figs. S2 and S3). Genomes of all examined vertebrates—and of some invertebrates as well—encode proteins sequelogous to mouse Liat1 (Fig. 4 and Figs. S2–S4). All sequelogs of Liat1 share at least the ∼30-residue Liat1 domain. In fact, among vertebrates other than mammals and in Liat1-containing invertebrates as well, the Liat1 domain is often the only region that identifies a protein as a sequelog of mammalian Liat1 (Figs. 3 and 4 and Figs. S2–S4).

The extent of conservation of the Liat1 domain among mammals (>95%) and a weaker but still considerable conservation of this domain between, for example, mammals and invertebrates, such as sea anemone or acorn worm (Fig. 4 and Figs. S2–S4), indicate a purifying selection that maintained this Ate1-interacting domain (of unknown function) in the course of animal evolution. At the same time, large clades of organisms, including plants and fungi, lack proteins that can be identified as Liat1 through sequelogies (sequence similarities) alone, as distinguished from still possible spalogies (spatial similarities) (24).

Tandem Repeats of a 10-Residue Motif in Liat1 of Primates. Another feature of mammalian Liat1 is a 10-residue sequence immediately downstream of the Liat1 domain (Figs. 3 and 4 and Fig. S3). This 10-residue motif is present as a single copy in nonprimate mammalian Liat1 proteins, but in Liat1 of some primates, including humans, this sequence is tandemly repeated (Fig. 4 and Fig. S3). For example, there are 4, 13, 13, 17, 17, and 18 repeats in the, gorilla, orangutan, bonobo, neanderthal, human, and baboon Liat1, respectively (Fig. 4 and Fig. S3). At the same time, some primates, such as gibbon and bushbaby, contain just one copy of the 10-residue motif, similarly to nonprimate mammals (Fig. 4 and Fig. S3). The specific sequences of 10-residue repeats in a tandem array are either identical or nearly identical to each other both within a given Liat1 and among Liat1 proteins of different primate species (e.g., Fig. S3).

The probability of mutations that alter the quantity of repeats in a gene can be orders of magnitude higher than the probability of, for example, missense mutations (*Discussion*) (30, 31). As a result, genetic variation that stems from repeat number changes in specific proteins can greatly exceed variation caused by other changes. This difference would be even higher if a gene in question is subtelomeric (*Discussion*) (31). *Liat1* genes are subtelomeric in all examined primates, including humans (32), and most (although not all) primate Liat1 proteins contain tandem repeats of the 10-residue motif (Fig. 4 and Fig. S3). In contrast, *Liat1* genes of nonprimate mammals are not subtelomeric, and all such Liat1 proteins contain one copy of the 10-residue motif (Fig. 4*A*).

The possibility that differences, among primates, in the quantities of their 10-residue repeats (Fig. 4) signify a role for these repeats in primate evolution is considered in the *Discussion*. In contrast to nonprimate *Liat1*, still unresolved aspects of subtelomeric primate *Liat1* genes are complicated enough to preclude definitive conclusions about expression patterns of, for example, human *Liat1* (called *C17orf97* in databases) until its

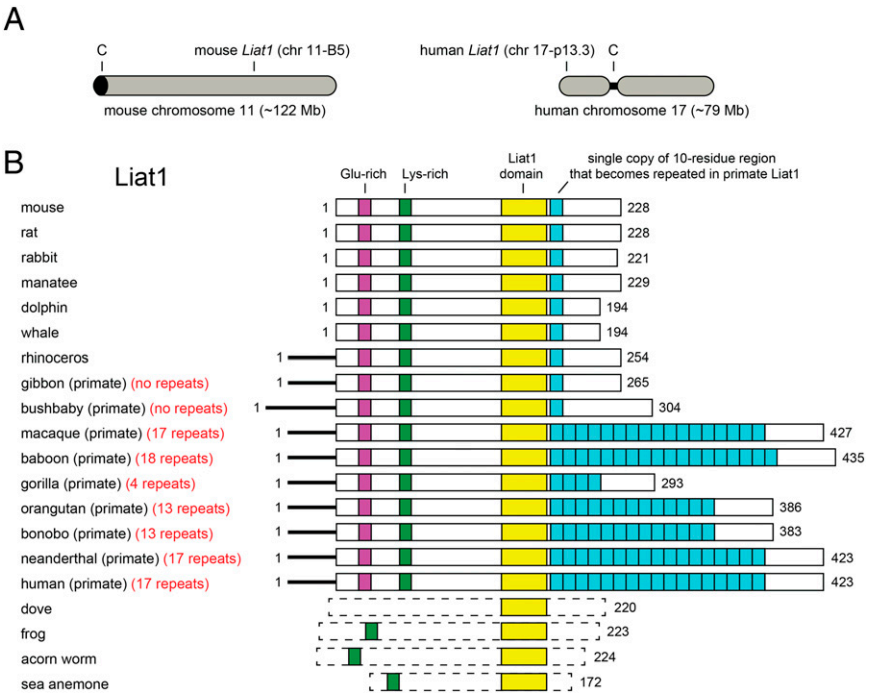

Fig. 4. Chromosomal locations of *Liat1* genes and evolution of Liat1 proteins. (*A*) *Liat1* genes are located away from telomeres in both mice (chromosome 11-B5) and other nonprimate mammals. In contrast, *Liat1* genes are subtelomeric in all examined primates (e.g., human chromosome 17-p13.3), indicating a translocation of *Liat1* to a subtelomeric site in an ancestor of modern primates, from gibbons and bushbabies to macaques and humans. (*B*) Evolution of Liat1 proteins. The colors of rectangles (the same as in Fig. 3*A*) denote, respectively, the negatively charged region of Liat1 (purple), its positively charged region (green), its particularly highly conserved ~30-residue domain (the Liat1 domain) (yellow), and a 10-residue segment that becomes tandemly repeated in Liat1 of some primates (blue). In all examined Liat1 proteins that contain these regions they are present in the same (indicated) order. Although each of these regions, in a given Liat1, is unambiguously recognizable upon inspection of its amino acid sequence, the "linear" distances between these regions are not identical even among mammals and are particularly variable among nonmammalian Liat1 proteins (Figs. S3 and S4). This aspect of Liat1 is largely bypassed in these diagrams, the chief aim of which is to highlight the presence or absence of specific regions and the emergence of tandem 10-residue repeats in Liat1 of some primates, including humans and great apes. Dashed and shifted rectangles in Liat1 of nonmammalian vertebrates and invertebrates signify a nonconservation of distances between specific domains of Liat1 in the indicated organisms, in comparison with a significant (although incomplete) conservation of interregion distances among mammalian Liat1 proteins. See also the main text and Figs. S2–S4. For the accession numbers of specific Liat1 proteins and for Latin names of the cited animal species, see the legend to Fig. S2.

DNA, pre-mRNAs, and mature mRNAs are extensively characterized. For example, current descriptions of human *Liat1* (*C17orf97*) (www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=C17orf97) are consistent with the existence of at least one additional intron in the region of human *Liat1* repeats, in contrast to mouse *Liat1*, which lacks both a second intron and the tandem repeats (Figs. 3*A* and 4 and Fig. S3) (*Discussion*).

Expression of Liat1 in Mouse Tissues and in Mouse or Human Cell Lines. A commercial antibody to human C17orf97 (Liat1) was raised in rabbits against a 93-residue internal segment that included the highly conserved 32-residue human Liat1 domain, which is 96% identical to its counterpart in mouse Liat1 (Fig. 3*A* and Figs. S2 and S3). In immunoblots of extracts from mouse tissues or NIH 3T3 mouse cell line, this anti-Liat1 antibody detected largely a single band at the expected ~26 kDa M_r of mouse Liat1 in the heart, kidney, liver, spleen, testis, lung, thymus, pancreas, and brown adipose tissue, and in NIH 3T3 cells as well (Fig. 5*B*, lanes 1–5, and Fig. 5*C*, lanes 1–9). Immunoblotting with decreasing amounts of purified recombinant (untagged) mouse Liat1 indicated that this antibody to human Liat1 could detect down to ~50 ng of mouse Liat1 per lane (Fig. 5*A*). At this (moderate) sensitivity, little if any Liat1 was detected in the total mouse brain, hippocampus, cerebellum, and white adipose tissue (Fig. 5*B*, lane 6, and Fig. 5*C*, lanes 1–3). We note that histochemical data in the Human Protein Atlas (HPA023583), produced through the use of the same antibody, suggested the presence of Liat1 (C17orf97) in the human brain (www.proteinatlas.org/ENSG00000187624/tissue). In situ hybridization of a mouse 1700016K19Rik (*Liat1*) cDNA to sections of mouse brain indicated the presence of *Liat1* mRNA at least in the cerebellum (mouse.brain-map.org/experiment/show?id=69114594).

Immunoblotting with extracts from human HeLa and HEK293T cells detected two Liat1 species of ~55 kDa and ~50 kDa (Fig. 5*C*, lanes 10 and 11). Human Liat1 is predicted to be 423 residues long (46.4 kDa) because of 17 tandem repeats in human Liat1 of a 10-residue sequence, in contrast to a single copy of this sequence in the 228-residue (25.5 kDa) mouse Liat1 and other nonprimate Liat1 proteins (Fig. 3, 4, and 5*C*, lanes 10 and 11). A smaller of two human Liat1 species in Fig. 5*C* (lanes 10 and 11) may be either a cleavage product of the larger human Liat1 or a pre-mRNA splicing-derived Liat1 isoform. Remarkably, this human Liat1 immunoblotting pattern also contained an ~26-kDa band that comigrated with the ~26-kDa band recognized by the same antibody in mouse tissues (Fig. 5*C*, lanes 10 and 11 vs. lanes 4–9). These results are consistent with the existence of human Liat1 isoforms in which quantities of the 10-residue repeat can vary between ~17 and 1, a functionally remarkable possibility that will be followed up through a more detailed understanding of primate *Liat1* genes.

Effect of Liat1 on Ate1-Mediated Nt-Arginylation. The previously characterized in vitro Nt-arginylation assay (21) used [^{14}C]-L-Arg,

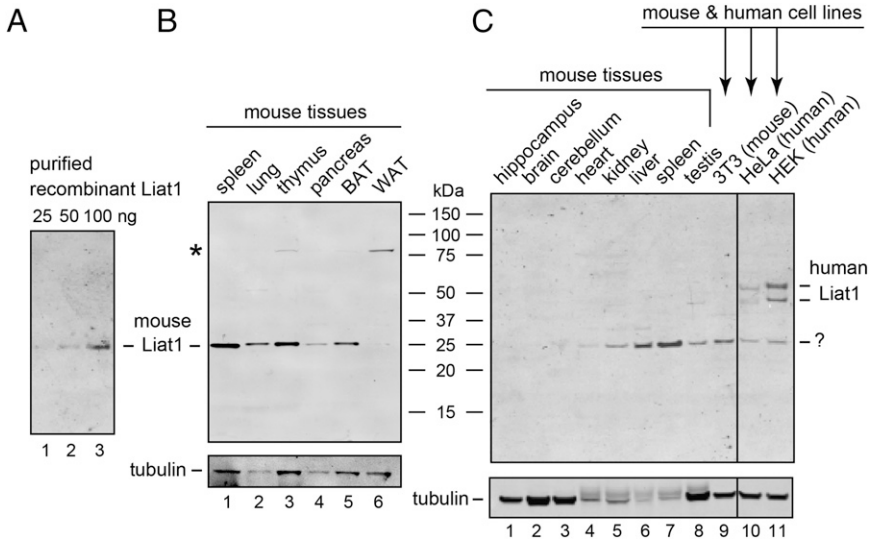

Fig. 5. Analyses of Liat1 proteins by immunoblotting. (*A*) Purified recombinant (untagged) mouse Liat1 was fractionated by SDS/PAGE, following by immunoblotting with an antibody to a highly conserved region of human Liat1 (C17orf97) (see the main text). Lanes 1–3: 25, 50, and 100 ng of mouse Liat1 per lane, respectively. (*B*) Same as in *A*, but with extracts from the indicated mouse tissues that had been fractionated by SDS/PAGE. An asterisk indicates a larger, apparently unrelated (cross-reacting) protein species in extracts from thymus and white adipose tissue (WAT). The results of immunoblotting with antitubulin antibody are shown as well. (*C*) Lanes 1–8, same as in *B*, with extracts from indicated mouse tissues. Lane 9, same as in lanes 1–8 but an extract from mouse NIH 3T3 cells. Lanes 10, 11, same as lane 8 but extracts from human HeLa and HEK293T cells. Larger human Liat1 proteins, presumably corresponding to species with multiple tandem repeats (see Fig. 4 and the main text), are indicated on the right. A question mark denotes a putative Liat1 species from human cell lines that comigrates with the much smaller (26 kDa) mouse Liat1 and may contain just one copy of the 10-residue motif that is tandemly repeated in other species of human Liat1 (see the main text).

purified mouse Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, and Ate1^{1B7B} isoforms, an Arg-tRNA–generating system, and either bovine α-lactalbumin (it bears the Nt-arginylatable N-terminal Glu residue) (Fig. S5*B*) or purified C-terminally tagged recombinant reporters X-DHFRbt (X = Asp, Cys or Arg-Cys) based on the mouse dihydrofolate reductase [DHFR (the "bt" tag is described in ref. 17)]. X-DHFRbt contained the six-residue sequence KGLAGL immediately after the N-terminal X residue (X = Asp, Cys or Arg-Cys), followed by the DHFRbt moiety. KGLAGL is the sequence of mouse Rgs4 (a physiological Arg/N-end rule substrate; see introductory paragraphs) immediately after its wild-type N-terminal Cys (20).

Initially, identical amounts of the individual purified Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, and Ate1^{1B7B} isoforms were incubated for 60 min at 37 °C with the rest of assay's components, including [^{14}C]-L-Arg and X-DHFRbt (X = Asp, Cys, or Arg-Cys), and either with or without Ate1-lacking extract from mouse *Ate1*$^{-/-}$ EFs, followed by SDS/PAGE and autoradiography (Fig. S5*A*). As expected (21), all Ate1 R-transferase isoforms could Nt-arginylate the N-terminal Asp residue of Asp-DHFRbt, with Ate1^{1B7B} being most active, and with Ate1^{1B7A}, Ate1^{1A7A}, and Ate1^{1A7B}, exhibiting respectively, ∼66%, ∼65%, and ∼12% of the activity of Ate1^{1B7B} (Fig. S5*A*). No Nt-arginylation of the "pre-arginylated" Arg-Asp-DHFRbt was observed under any conditions (Fig. S5*A*), as expected, given the specificity of R-transferase (Fig. 1*A*) (4). There was no detectable Nt-arginylation of Cys-DHFRbt in the absence of extract from *Ate1*$^{-/-}$ EF cells, but this reporter was Nt-arginylated in the presence of extract, presumably because of the oxidation of N-terminal Cys by compounds in cell extracts (7, 20).

The Liat1-supplemented version of this assay (without extract from *Ate1*$^{-/-}$ EF cells) used X-DHFRbt (X = Asp, Cys). Each of the Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, and Ate1^{1B7B} isoforms was preincubated for 30 min at 37 °C with an ∼fourfold molar excess of purified untagged mouse Liat1 or with buffer alone before their addition to the arginylation assay. The Nt-arginylation of Asp-DHFRbt by Ate1 was reproducibly enhanced in the presence of Liat1, from 1.3-fold to 2.3-fold (Fig. S5*C*). This effect is unlikely to stem from increased macromolecular crowding upon the addition of Liat1, because the addition of equal or larger amounts of BSA did not alter the efficacy of Nt-arginylation. The unknown mechanistic cause of moderate but reproducible effects of Liat1 on Nt-arginylation is a subject for future studies.

Significantly, we could not detect—despite attempts to do so—the conjugation of ^{14}C-Arg to Liat1 itself in this assay, in either the presence or absence of test proteins, such as α-lactalbumin or X-DHFRbt. (Upon SDS/PAGE, the untagged mouse Liat1 migrates significantly below the band of X-DHFRbt.) Even overexposures of autoradiograms of electrophoretically fractionated proteins after ^{14}C-arginylation in the presence of purified Liat1 did not reveal any significant ^{14}C in the vicinity of the ∼26-kDa Liat1 band, indicating that Liat1 is not an Ate1 substrate.

Identification of Liat1-Binding Proteins Other than Ate1. We also carried out a search for mouse proteins other than Ate1 that interact with mouse Liat1 (Fig. 3*A*), using a library of mouse brain cDNAs fused to the Gal4 AD vis-á-vis *Liat1* cDNA fused to the Gal4 DBD. This screen identified ∼40 different mouse proteins that appeared to bind to mouse Liat1, with the corresponding cDNA isolates having passed standard controls of the Y2H assay (Fig. S6*C*). In addition, an independent Y2H screen for interactions among protein methyltransferases and proteins that bind to them (including their substrates) identified the human Jmjd6 methyltransferase as a putative ligand of human C17orf97 (33) (i.e., the human Liat1 protein) (Fig. S6*C*).

We also searched for protein interactions with Liat1 using GST-pulldowns with GST-Liat1 and extracts from mouse EF cells (Fig. S6*A*). These MS-based assays added the ribosomal proteins S14 and S19 as well as three specific histones to the Y2H-based list of other Liat1 ligands (Fig. S6). Thus, identified putative ligands of Liat1 encompassed a broad range of functional classes, including components of the translation, transcription, and Ub-proteasome systems. Our ongoing co-IP assays that independently verify the

binding of Liat1 to its putative ligands [which were detected by Y2H or GST-pulldowns (Fig. S6*C*)] have recently confirmed that Jmjd6 and the ribosomal protein S14 can be coimmunoprecipitated with mouse Liat1 from a cell extract (Fig. S6*B*). These and other binding assays with putative ligands of Liat1 will continue to verify the current preliminary list (Fig. S6), thereby making possible systematic analyses of confirmed Liat1 ligands in regard to specific functions of their interactions with Liat1.

Discussion

We identified a 26-kDa mouse protein, termed Liat1, by detecting its binding to the 59-kDa mouse Ate1 R-transferase, a component of the Arg/N-end rule pathway (Figs. 1–3). The Liat1–Ate1 interaction was shown to require at least a part of the highly conserved ∼30-residue region of Liat1 (Fig. 3*B*). Biological functions of Liat1–Ate1 interactions (Figs. 2 and 3) remain to be understood, in part because Liat1 is a previously uncharacterized protein. (In current databases, human Liat1 is called C17orf97.) We also identified putative Liat1-binding proteins other than Ate1 (Fig. S6), but the biological function of Liat1 remains to be discovered.

Further analyses showed that Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, in contrast to a single copy of this motif in Liat1 of other mammals, including rodents. Quantities of these repeats are, in general, different in Liat1 of different primates. For example, there are 4, 13, 13, 17, and 17 repeats in the gorilla, orangutan, bonobo, neanderthal, and human Liat1, respectively (Fig. 4 and Fig. S3), suggesting that repeat number changes in Liat1 might play a role in evolution of primates. As evidence, these differences in the quantities of Liat1 repeats do not rise above a correlational argument, and the current disposition is further complicated by our finding of intraspecies variability in the number of repeats. For example, the Liat1 proteins encoded by an anonymous human genome (NM_001013672) and by the genome of James D. Watson (www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=C17orf97) were found to contain 17 tandem repeats in each, whereas the genome of J. Craig Venter encodes Liat1 that contains 18 repeats (NP_001013694).

Analyses of primate genomes have so far identified relatively few alterations that could be demonstrated to contribute to phenotypic differences among these species. The "causative" alterations include variants of developmental enhancer DNA sequences as well as specific alleles of genes that influence circadian rhythms, human brain size, language, other human traits, and differences in body sizes among primates (34–37). On evolutionary timescales the emergence of anatomical and behavioral features that distinguish humans from great apes was remarkably fast, because the last common ancestor of humans and chimpanzees lived about 6 million y ago.

The efficacy of selection pressure is limited by the extent of relevant genetic variation within a breeding population. We suggest, therefore, that the rapidity of evolution that led to humans may have involved proteins containing tandemly repeated sequences, given frequent changes in the numbers of such repeats. Analogous arguments have been made both in general and to account for the strikingly rapid (in less than 1,000 y) emergence of modern dog breeds, in response to selection pressures imposed by breeders (30, 38). Morphological differences among dog breeds, and even within a breed, were found to correlate with variations in the numbers of one-residue or two-residue repeats in proteins that regulate embryonic development and postnatal growth (30).

Tandem repeats of amino acid sequences that range in size from 1 to more than 100 residues are a feature of many cellular proteins (30, 36, 38, 39). Heritable changes in the number of repeat units are known to underlie a significant fraction of phenotypic variability both among different species and within a species (36, 38). The previously demonstrated genetic instability of repeats stems from several causes, including unequal crossover, DNA replication slippage, and double-strand break repair (30, 31). In an evolving species, genetic variation that results from repeat number changes in specific proteins can greatly exceed variation that is caused, for example, by missense mutations. This difference would be even higher if a gene in question is subtelomeric, because proximity to a telomere tends to increase the frequency of recombination-mediated changes in the quantity of repeats in a gene (31).

Remarkably, *Liat1* genes are subtelomeric in all examined primates, including humans, and most primate Liat1 proteins contain tandem repeats of a 10-residue motif (Fig. 4 and Fig. S3). Potentially telling exceptions include the gibbon and bushbaby Liat1 proteins, which contain a single copy of the 10-residue motif that is repeated in other examined primates (Fig. 4*B*) (37). In contrast to primate *Liat1* genes, their counterparts in other mammals, including the mouse, are not subtelomeric (Fig. 4*A*), and all predicted nonprimate Liat1 proteins in databases contain one copy of the 10-residue motif. The function of Liat1 repeats (Fig. 4) and their possible relevance to anatomic and phenotypic evolution of primates remain to be addressed.

One subtelomeric gene encoding a protein that contains tandem repeats is *Drd4* (40). A human dopamine receptor encoded by *Drd4* contains varying numbers of tandem repeats of a 16-residue sequence that forms the third cytosolic loop of the receptor. Both the quantity of 16-residue repeats and specific sequences within each repeat are often altered among individual humans, with repeat numbers varying between 2 and 10 (40). Analogous polymorphisms of these repeats were also observed in Drd4 of other primates (40). Human Drdr4 receptors with different repeat numbers were reported to differ in functional properties, and were also differentially regulated in diseases such as schizophrenia. Nevertheless, there is still no definitive evidence that the observed frequencies of *Drd4*-encoded isoforms containing different quantities of repeats had been caused, at least in part, by positive selection, as distinguished from a quasi-neutral drift (41).

In sum, a major unanswered question is whether changes in repeat quantities among Liat1 proteins of different primates stemmed, at least in part, from specific selection pressures. The alternative scenario is that most variation in Liat1 repeats (Fig. 4 and Fig. S3) may be the result of genetic drift and unselected fixations of altered repeat quantities, because of a small effective population size of an evolving species, including occasional population bottlenecks. A smaller population is characterized by correspondingly weaker forces of natural selection. Near-neutral evolution of proteins under such conditions is discussed by Lynch (42). Because small changes in Liat1 repeat numbers may have modest phenotypic effects, a near-neutral drift would be likely to at least contribute to evolution of repeats in Liat1. If so, it is the selection-based, adaptation-centered hypothesis about a functional role of repeat number changes in Liat1 that must be experimentally verified vis-á-vis the competing null hypothesis, in which these changes would be caused by a near-neutral genetic drift.

The understanding of both Liat1 itself and the functional significance of its binding to specific isoforms of the Ate1 R-transferase would be advanced by answers to at least the following questions. What are the composition and functions of in vivo complexes that contain Ate1, Liat1, and other macromolecular components? What is the biological function of Liat1? How does this function relate to the known role of the Ate1 R-transferase in the Arg/N-end rule pathway (Fig. 1*A*)? Does a divergent Liat1 protein—for example, the one of sea anemone [it is identifiable as Liat1 solely through its ∼30-residue Liat1 domain (Figs. S2 and S4)—specifically bind to the sea anemone Ate1? [The binding of mouse Liat1 to mouse Ate1 requires the Liat1 domain and apparently does not involve the 10-residue

motif that forms repeats in Liat1 of primates (Figs. 2, 3*B*, and 4, and Figs. S2–S4).]

Furthermore, do repeats of the 10-residue motif in, for example, human Liat1 (Fig. 4 and Fig. S3) interact with each other? Do these repeats specifically bind to any human protein? If they do, does the binding of Liat1 repeats to that protein also involve other regions of human Liat1? Is there a counterpart of a human repeat-binding protein in, for example, mouse cells? (Mouse Liat1 contains one copy of the sequence that forms repeats in human Liat1.) As mentioned in *Results*, it would also be essential to understand, in detail, a primate (e.g., the human or macaque) *Liat1* gene, given its more complex organization (including the presence of Liat1 isoforms) than the structure of repeat-lacking *Liat1* genes of nonprimate mammals (Figs. 3*A* and 4).

The understanding of Liat1 would also benefit from constructing and characterizing mouse strains that either lack Liat1 or contain its counterpart in which the 10-residue motif had been amplified to yield primate-like tandem repeats. It would also be informative to determine—with a primate such as, for example, macaque—the phenotypic effects of strongly increased or strongly decreased quantities of 10-residue repeats in Liat1 (Fig. 4). [Gene-specific alterations of the macaque genome through the clustered regularly interspaced short palindromic repeats (CRISPR) technology have already been achieved (43).] Given the preferential binding of Liat1 to specific isoforms of the Ate1 R-transferase (Fig. 2), further studies of Liat1 may also advance the functional understanding of Ate1 isoforms. In addition, the obscurity—until the present study—of tandem repeats in primate Liat1, and still incomplete descriptions of subtelomeric loci among the sequenced metazoan genomes (because of complexities of dealing with telomere-proximal microsatellite DNA repeats), suggest that it may be informative to further explore subtelomeric regions of the 23 human chromosomes for the presence of other uncharacterized genes that might encode repeats analogous to those of Liat1.

Experimental Procedures

Y2H Assays. Y2H assays were carried out with *S. cerevisiae*, using the BD Matchmaker kit (BD Biosciences), the plasmid pCB132, which expressed the $Gal4^{DBD}$-$ATE1^{1B7A}$ fusion (Table S1), and a mouse testis cDNA pACT library (Clontech).

Construction and Expression of Recombinant Proteins in BL21 (DE3) *E. coli*. The untagged mouse $Ate1^{1B7A}$, $Ate1^{1B7B}$, $Ate1^{1A7A}$, and $Ate1^{1A7B}$, the untagged mouse Liat, and ^{3ha}Liat1 were expressed as Ub fusions in *E. coli*, followed by the removal of Ub and purification of recombinant proteins by Mono-S chromatography (*SI Experimental Procedures*).

In Vitro Arginylation Assay. The Nt-arginylation assay was performed essentially as described previously (21).

Tissue Extracts and Immunoblotting. Extract preparation and immunoblotting were carried out essentially as described previously (3, 12, 21).

GST Pulldown Assay and Immunoprecipitations. Mouse EF cells were transiently transformed with a plasmid expressing either GST or a GST–Liat1 fusion, and GST-pulldown assays were carried out as described in *SI Experimental Procedures*. Immunoprecipitations with anti-flag or anti-ha antibodies were performed as previously described (3, 12).

Additional information regarding experimental procedures is given in *SI Experimental Procedures*.

ACKNOWLEDGMENTS. We thank E. Udartseva for excellent technical assistance, and other members of the A.V. laboratory for their help and advice. This study was supported by the National Institutes of Health Grants DK039520 and GM031530 (to A.V.).

1. Bachmair A, Finley D, Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234(4773):179–186.
2. Hwang C-S, Shemorry A, Varshavsky A (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 327(5968):973–977.
3. Kim H-K, et al. (2014) The N-terminal methionine of cellular proteins as a degradation signal. *Cell* 156(1-2):158–169.
4. Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* 20:1298–1345.
5. Kim JM, Hwang CS (2014) Crosstalk between the Arg/N-end and Ac/N-end rule. *Cell Cycle* 13(9):1366–1367.
6. Tasaki T, Sriram SM, Park KS, Kwon YT (2012) The N-end rule pathway. *Annu Rev Biochem* 81:261–289.
7. Gibbs DJ, Bacardit J, Bachmair A, Holdsworth MJ (2014) The eukaryotic N-end rule pathway: Conserved mechanisms and diverse functions. *Trends Cell Biol* 24(10): 603–611.
8. Dougan DA, Micevski D, Truscott KN (2012) The N-end rule pathway: From recognition by N-recognins, to destruction by AAA+proteases. *Biochim Biophys Acta* 1823(1): 83–91.
9. Varshavsky A (2008) Discovery of cellular regulation by protein degradation. *J Biol Chem* 283(50):34469–34489.
10. Tobias JW, Shrader TE, Rocap G, Varshavsky A (1991) The N-end rule in bacteria. *Science* 254(5036):1374–1377.
11. Rivera-Rivera I, Román-Hernández G, Sauer RT, Baker TA (2014) Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. *Proc Natl Acad Sci USA* 111(37):E3853–E3859.
12. Shemorry A, Hwang C-S, Varshavsky A (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol Cell* 50(4): 540–551.
13. Starheim KK, Gevaert K, Arnesen T (2012) Protein N-terminal acetyltransferases: When the start matters. *Trends Biochem Sci* 37(4):152–161.
14. Brower CS, Piatkov KI, Varshavsky A (2013) Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. *Mol Cell* 50(2): 161–171.
15. Piatkov KI, Oh J-H, Liu Y, Varshavsky A (2014) Calpain-generated natural protein fragments as short-lived substrates of the N-end rule pathway. *Proc Natl Acad Sci USA* 111(9):E817–E826.
16. Yamano K, Youle RJ (2013) PINK1 is degraded through the N-end rule pathway. *Autophagy* 9(11):1758–1769.
17. Wang H, Piatkov KI, Brower CS, Varshavsky A (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. *Mol Cell* 34(6):686–695.
18. Kwon YT, Kashina AS, Varshavsky A (1999) Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol Cell Biol* 19(1):182–193.
19. Kwon YT, et al. (2002) An essential role of N-terminal arginylation in cardiovascular development. *Science* 297(5578):96–99.
20. Hu R-G, et al. (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* 437(7061):981–986.
21. Hu R-G, et al. (2006) Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. *J Biol Chem* 281(43):32559–32573.
22. Saha S, Kashina A (2011) Posttranslational arginylation as a global biological regulator. *Dev Biol* 358(1):1–8.
23. Brower CS, Varshavsky A (2009) Ablation of arginylation in the mouse N-end rule pathway: Loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. *PLoS ONE* 4(11):e7757.
24. Varshavsky A (2004) 'Spalog' and 'sequelog': Neutral terms for spatial and sequence similarity. *Curr Biol* 14(5):R181–R183.
25. Lee MJ, et al. (2005) RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. *Proc Natl Acad Sci USA* 102(42):15030–15035.
26. Varshavsky A (2012) Augmented generation of protein fragments during wakefulness as the molecular cause of sleep: A hypothesis. *Protein Sci* 21(11):1634–1661.
27. Brower CS, Veiga L, Jones RH, Varshavsky A (2010) Mouse Dfa is a repressor of TATA-box promoters and interacts with the Abt1 activator of basal transcription. *J Biol Chem* 285(22):17218–17234.
28. Catanzariti A-M, Soboleva TA, Jans DA, Board PG, Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci* 13(5):1331–1339.
29. Varshavsky A (2005) Ubiquitin fusion technique and related methods. *Methods Enzymol* 399:777–799.
30. Fondon JW, 3rd, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci USA* 101(52):18058–18063.
31. Brown CA, Murray AW, Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 20(10):895–903.
32. Zody MC, et al. (2006) DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* 440(7087):1045–1049.
33. Weimann M, et al. (2013) A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat Methods* 10(4):339–342.
34. Enard W (2011) FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr Opin Neurobiol* 21(3):415–424.
35. Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, Pollard KS (2013) Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci* 368(1632):20130025.
36. Sabino FC, et al. (2014) Evolutionary history of the PER3 variable number of tandem repeats (VNTR): Idiosyncratic aspect of primate molecular circadian clock. *PLoS ONE* 9(9):e107198.
37. Carbone L, et al. (2014) Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513(7517):195–201.

38. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 44:445–477.
39. Verstrepen KJ, Jansen A, Lewitter F, Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* 37(9):986–990.
40. Livak KJ, Rogers J, Lichter JB (1995) Variability of dopamine D4 receptor (DRD4) gene sequence within and among nonhuman primate species. *Proc Natl Acad Sci USA* 92(2): 427–431.
41. Hattori E, et al. (2009) Variable number of tandem repeat polymorphisms of DRD4: Re-evaluation of selection hypothesis and analysis of association with schizophrenia. *Eur J Hum Genet* 17(6):793–801.
42. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA).
43. Niu Y, et al. (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156(4):836–843.