# 1. Gene Aliases

Ctsd, Cathepsin D, CLN10, CPSD, Ceroid-Lipofuscinosis, Neuronal 10, EC 3.4.23.5, Epididymis Secretory Sperm Binding Protein Li 130P, Cathepsin D (Lysosomal Aspartyl Protease), Lysosomal Aspartyl Peptidase, Lysosomal Aspartyl Protease, HEL-S-130P, EC 3.4.23

[<https://www.genecards.org/cgi-bin/carddisp.pl?gene=CTSD&keywords=ctsd>].

# 2. Association with Toxicity and/or Disease at a Transcriptional Level

* In human lung cancer tumors, cathepsins B and D mRNA levels did not correlate with any tumor characteristics. However, in most analyzed tumors, expression of cathepsin D mRNA was downregulated compared with adjacent normal tissue indicating cathepsin D mRNA may be a potential lung cancer marker [PMID: 20722505].
* RNAseq on small cell lung cancer (SCLC) patients’ tumor sample and normal lung tissue showed that an apoptosis-related, gene Cathepsin D (CTSD), was identified to be up-regulated in the low-risk group, and its higher expression correlated with better overall survival in SCLC. Functional enrichment analysis showed that the low-risk group was also enriched in the apoptosis pathway and high immune infiltration of T cells [PMID: 37189142].
* In lungs of mice exposed to cigarette smoke or control air, there was a significant increase of matrix metalloproteinase (MMP)-12 and Cathepsin D mRNA, compared to air-exposed mice with Cathepsin D being predominantly expressed in macrophages [PMID: 16192742].

# 3. Summary of Protein Family and Structure

* Size: 412 amino acids
* Molecular mass: 44552 Da
* Protein Accession: P07339
* Blocks: Pepsin (A1) aspartic protease family signature
* Domains: Aspartic\_peptidase\_A1, Aspartic\_peptidase\_AS, Aspartic\_peptidase\_N, Cathepsin\_D, PEPTIDASE\_A1, Peptidase\_aspartic\_dom\_sf
* Family: Belongs to the peptidase A1 family
* Acid protease active in intracellular protein breakdown. Plays a role in APP processing following cleavage and activation by ADAM30 which leads to APP degradation. Involved in the pathogenesis of several diseases such as breast cancer and possibly Alzheimer disease [PubMed:2733303]
* The combined substitution of two noncontinuous sequences of cathepsin D (lysine 203 and amino acids 265-292) into the analogous positions of glycopepsinogen resulted in phosphorylation of the oligosaccharides of the expressed chimeric molecule [PMID: 2170024].
* The lysosomal targeting region of cathepsin D includes the putative binding site of the cis-Golgi phosphotransferase which is responsible for the initial sorting step for soluble proteins destined for lysosomes by phosphorylating the carbohydrates on these molecules [PMID: 8467789].

# 4. Proteins Known to Interact with Gene Product

## Interactions with experimental support

* **APP** Gamma-secretase C-terminal fragment 50; Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Interaction between APP molecules on neighboring cells promotes synaptogenesis. Involved in cell mobility and transcription regulation through protein-protein interactions. Can promote transcription activation through binding to APBB1-KAT5 and inhibits Notch signaling through interaction with Numb. Couples to apoptosis- inducing pathways such as those mediated by G(O) and JIP. [PMID: 10605825, PMID: 10931940, PMID: 17112520, PMID: 7523115, PMID: 8930981, PMID: 8943232]
* **PPT1** Palmitoyl-protein thioesterase 1; Removes thioester-linked fatty acyl groups such as palmitate from modified cysteine residues in proteins or peptides during lysosomal degradation. Prefers acyl chain lengths of 14 to 18 carbons ; Belongs to the palmitoyl-protein thioesterase family. [PMID: 25865307, PMID: 26217791]
* **GRHPR** Glyoxylate reductase/hydroxypyruvate reductase; Enzyme with hydroxy-pyruvate reductase, glyoxylate reductase and D-glycerate dehydrogenase enzymatic activities. Reduces hydroxypyruvate to D-glycerate, glyoxylate to glycolate oxidizes D- glycerate to hydroxypyruvate; Belongs to the D-isomer specific 2-hydroxyacid dehydrogenase family. [PMID: 26344197, PMID: 31536960]
* **NAPSA** Napsin-A; May be involved in processing of pneumocyte surfactant precursors. [PMID: 26186194, PMID: 28514442]
* **SERPINA4** Kallistatin; Inhibits human amidolytic and kininogenase activities of tissue kallikrein. Inhibition is achieved by formation of an equimolar, heat- and SDS-stable complex between the inhibitor and the enzyme, and generation of a small C-terminal fragment of the inhibitor due to cleavage at the reactive site by tissue kallikrein. Belongs to the serpin family. [PMID: 11258665, PMID: 11341921]
* **LAMP1** Lysosome-associated membrane glycoprotein 1; Presents carbohydrate ligands to selectins. Also implicated in tumor cell metastasis. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000236671 9606.ENSP00000333298](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000236671%0D9606.ENSP00000333298)]
* **ENSP00000489910** Peptidase A1 domain-containing protein; Belongs to the peptidase A1 family. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000236671 9606.ENSP00000489910](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000236671%0D9606.ENSP00000489910)]

The interactions list has been truncated to include only interactions with the strongest support from the literature.

# 5. Links to Gene Databases

* GeneCards (human): <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CTSD>
* Harmonizome (human): <https://maayanlab.cloud/Harmonizome/gene/CTSD>
* NCBI (human): <https://www.ncbi.nlm.nih.gov/gene/1509>
* NCBI (rat): <https://www.ncbi.nlm.nih.gov/gene/171293>
* Ensemble (human): <https://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000117984>
* Ensemble (rat): <https://useast.ensembl.org/Rattus_norvegicus/Gene/Summary?g=ENSRNOG00000020206>
* Rat Genome Database (rat): <https://rgd.mcw.edu/rgdweb/report/gene/main.html?id=621511>
* Uniprot (human): <https://www.uniprot.org/uniprotkb/P07339>
* Uniprot (rat): <https://www.uniprot.org/uniprotkb/P24268>
* Wikigenes (human): <https://www.wikigenes.org/e/gene/e/1509.html>
* Wikigenes (rat): <https://www.wikigenes.org/e/gene/e/171293.html>
* Alphafold (human): <https://alphafold.ebi.ac.uk/entry/P07339>
* Alphafold (rat): <https://alphafold.ebi.ac.uk/entry/P24268>
* PDB (human): <https://www.rcsb.org/structure/1LYA>, <https://www.rcsb.org/structure/1LYB>, <https://www.rcsb.org/structure/1LYW>, <https://www.rcsb.org/structure/4OBZ>, <https://www.rcsb.org/structure/4OC6>, <https://www.rcsb.org/structure/4OD9>
* PDB (mouse): none
* PDB (rat): <https://www.rcsb.org/structure/5UX4>

# 6. GO Terms, MSigDB Signatures, Pathways Containing Gene with Descriptions of Gene Sets

## **Pathways:**

* **Neutrophil degranulation**: Neutrophils are the most abundant leukocytes (white blood cells), indispensable in defending the body against invading microorganisms. In response to infection, neutrophils leave the circulation and migrate towards the inflammatory focus. They contain several subsets of granules that are mobilized to fuse with the cell membrane or phagosomal membrane, resulting in the exocytosis or exposure of membrane proteins. Traditionally, neutrophil granule constituents are described as antimicrobial or proteolytic, but granules also introduce membrane proteins to the cell surface, changing how the neutrophil responds to its environment (Borregaard et al. 2007). Primed neutrophils actively secrete cytokines and other inflammatory mediators and can present antigens via MHC II, stimulating T-cells (Wright et al. 2010). Granules form during neutrophil differentiation. Granule subtypes can be distinguished by their content but overlap in structure and composition. The differences are believed to be a consequence of changing protein expression and differential timing of granule formation during the terminal processes of neutrophil differentiation, rather than sorting (Le Cabec et al. 1996). The classical granule subsets are Azurophil or primary granules (AG), secondary granules (SG) and gelatinase granules (GG). Neutrophils also contain exocytosable storage cell organelles, storage vesicles (SV), formed by endocytosis they contain many cell-surface markers and extracellular, plasma proteins (Borregaard et al. 1992). Ficolin-1-rich granules (FG) are like GGs highly exocytosable but gelatinase-poor (Rorvig et al. 2009) [<https://reactome.org/PathwayBrowser/#/R-HSA-6798695>].
* **MHC class II antigen presentation**: Antigen presenting cells (APCs) such as B cells, dendritic cells (DCs) and monocytes/macrophages express major histocompatibility complex class II molecules (MHC II) at their surface and present exogenous antigenic peptides to CD4+ T helper cells. CD4+ T cells play a central role in immune protection. On their activation they stimulate differentiation of B cells into antibody-producing B-cell blasts and initiate adaptive immune responses. MHC class II molecules are transmembrane glycoprotein heterodimers of alpha and beta subunits. Newly synthesized MHC II molecules present in the endoplasmic reticulum bind to a chaperone protein called invariant (Ii) chain. The binding of Ii prevents the premature binding of self antigens to the nascent MHC molecules in the ER and also guides MHC molecules to endocytic compartments. In the acidic endosomal environment, Ii is degraded in a stepwise manner, ultimately to free the class II peptide-binding groove for loading of antigenic peptides. Exogenous antigens are internalized by the APC by receptor mediated endocytosis, phagocytosis or pinocytosis into endocytic compartments of MHC class II positive cells, where engulfed antigens are degraded in a low pH environment by multiple acidic proteases, generating MHC class II epitopes. Antigenic peptides are then loaded into the class II ligand-binding groove. The resulting class II peptide complexes then move to the cell surface, where they are scanned by CD4+ T cells for specific recognition (Berger & Roche 2009, Zhou & Blum 2004, Watts 2004, Landsverk et al. 2009) [<https://reactome.org/PathwayBrowser/#/R-HSA-2132295>].
* **Metabolism of Angiotensinogen to Angiotensins**: Angiotensinogen, a prohormone, is synthesized and secreted mainly by the liver but also from other tissues (reviewed in Fyhrquist and Saijonmaa 2008, Cat and Touyz 2011). Renin, an aspartyl protease specific for angiotensinogen, is secreted into the bloodstream by juxtaglomerular cells of the kidney in response to a drop in blood pressure. Renin cleaves angiotensinogen to yield a decapaptide, angiotensin I (angiotensin-1, angiotensin-(1-10)). Circulating renin can also bind the membrane-localized (pro)renin receptor (ATP6AP2) which increases its catalytic activity. After cleavage of angiotensinogen to angiotensin I by renin, two C-terminal amino acid residues of angiotensin I are removed by angiotensin-converting enzyme (ACE), located on the surface of endothelial cells, to yield angiotensin II (angiotensin-2, angiotensin-(1-8)), the active peptide that causes vasoconstriction, resorption of sodium and chloride, excretion of potassium, water retention, and aldosterone secretion. More recently other, more tissue-localized pathways leading to angiotensin II and alternative derivatives of angiotensinogen have been identified (reviewed in Kramkowski et al. 2006, Kumar et al. 2007, Fyhrquist and Saijonmaa 2008, Becari et al. 2011). Chymase, cathepsin G, and cathepsin X (cathepsin Z) can each cleave angiotensin I to yield angiotensin II. Angiotensin-converting enzyme 2 (ACE2) cleaves 1 amino acid residue from angiotensin I (angiotensin-(1-10)) to yield angiotensin-(1-9), which can be cleaved by ACE to yield angiotensin-(1-7). ACE2 can also cleave angiotensin II to yield angiotensin-(1-7). Neprilysin can cleave either angiotensin-(1-9) or angiotensin I to yield angiotensin-(1-7). Angiotensin-(1-7) binds the MAS receptor (MAS1, MAS proto-oncogene) and, interestingly, produces effects opposite to those produced by angiotensin II. Aminopeptidase A (APA, ENPEP) cleaves angiotensin II to yield angiotensin III (angiotensin-(2-8)), which is then cleaved by aminopeptidase N (APN, ANPEP) yielding angiotensin IV (angiotensin-(3-8)). Angiotensin IV binds the AT4 receptor (AT4R, IRAP, LNPEP, oxytocinase). Inhibitors of renin (e.g. aliskiren) and ACE (e.g. lisinopril, ramipril) are currently used to treat hypertension (reviewed in Gerc et al. 2009, Verdecchia et al. 2010, Alreja and Joseph 2011) [<https://reactome.org/PathwayBrowser/#/R-HSA-2022377>].
* **Collagen degradation**: Collagen fibril diameter and spatial organisation are dependent on the species, tissue type and stage of development (Parry 1988). The lengths of collagen fibrils in mature tissues are largely unknown but in tendon can be measured in millimetres (Craig et al. 1989). Collagen fibrils isolated from adult bovine corneal stroma had ~350 collagen molecules in transverse section, tapering down to three molecules at the growing tip (Holmes & Kadler 2005). The classical view of collagenases is that they actively unwind the triple helical chain, a process termed molecular tectonics (Overall 2002, Bode & Maskos 2003), before preferentially cleaving the alpha2 chain followed by the remaining chains (Chung et al. 2004). More recently it has been suggested that collagen fibrils exist in an equilibrium between protected and vulnerable states (Stultz 2002, Nerenberg & Stultz 2008). The prototypical triple-helical structure of collagen does not fit into the active site of collagenase MMPs. In addition the scissile bonds are not solvent-exposed and are therefore inaccessible to the collagenase active site (Chung et al. 2004, Stultz 2002). It was realized that collagen must locally unfold into non-triple helical regions to allow collagenolysis. Observations using circular dichroism and differential scanning calorimetry confirm that there is considerable heterogeneity along collagen fibres (Makareeva et al. 2008) allowing access for MMPs at physiological temperatures (Salsas-Escat et al. 2010). Collagen fibrils with cut chains are unstable and accessible to proteinases that cannot cleave intact collagen strands (Woessner & Nagase 2000, Somerville et al. 2003). Continued degradation leads to the formation of gelatin (Lovejoy et al. 1999). Degradation of collagen types other than I-III is less well characterized but believed to occur in a similar manner. Metalloproteinases (MMPs) play a major part in the degradation of several extracellular macromolecules including collagens. MMP1 (Welgus et al. 1981), MMP8 (Hasty et al. 1987), and MMP13 (Knauper et al. 1996), sometimes referred to as collagenases I, II and III respectively, are able to initiate the intrahelical cleavage of the major fibril forming collagens I, II and III at neutral pH, and thus thought to define the rate-limiting step in normal tissue remodeling events. All can cleave additional substrates including other collagen subtypes. Collagenases cut collagen alpha chains at a single conserved Gly-Ile/Leu site approximately 3/4 of the molecule’s length from the N-terminus (Fields 1991, Chung et al. 2004). The cleavage site is characterised by the motif G(I/L)(A/L); the G-I/L bond is cleaved. In collagen type I this corresponds to G953-I954 in the Uniprot canonical alpha chain sequences (often given as G775-I776 in literature). It is not clear why only this bond is cleaved, as the motif occurs at several other places in the chain. MMP14, a membrane-associated MMP also known as Membrane-type matrix metalloproteinase 1 (MT-MMP1), is able to cleave collagen types I, II and III (Ohuchi et al. 1997) [<https://reactome.org/PathwayBrowser/#/R-HSA-1442490>].
* **Estrogen-dependent gene expression**: Estrogens mediate their transcriptional effects through interaction with the estrogen receptors, ESR1 (also known as ER alpha) and ESR2 (ER beta). ESR1 and ESR2 share overlapping but distinct functions, with ESR1 playing the primary role in transcriptional activation in most cell types (Hah and Krauss, 2014; Haldosen et al, 2014. The receptors function as ligand-dependent dimers and can activate target genes either through direct binding to an estrogen responsive element (ERE) in the target gene promoter, or indirectly through interaction with another DNA-binding protein such as RUNX1, SP1, AP1 or NF-kappa beta (reviewed in Bai and Gust, 2009; Hah and Krause, 2014). Binding of estrogen receptors to the DNA promotes the assembly of higher order transcriptional complexes containing methyltransferases, histone acetyltransferases and other transcriptional activators, which promote transcription by establishing active chromatin marks and by recruiting general transcription factors and RNA polymerase II. ESR1- and estrogen-dependent recruitment of up to hundreds of coregulators has been demonstrated by varied co-immunoprecipitation and proteomic approaches (Kittler et al, 2013; Mohammed et al, 2013; Foulds et al, 2013; Mohammed et al, 2015; Liu et al, 2014; reviewed in Magnani and Lupien, 2014; Arnal, 2017). In some circumstances, ligand-bound receptors can also promote the assembly of a repression complex at a target gene, and in some cases, heterodimers of ESR1 and ESR2 serve as repressors of ESR1-mediated target gene activation (reviewed in Hah and Kraus, 2014; Arnal et al, 2017). Phosphorylation of the estrogen receptor also modulates its activity, and provides cross-talk between nuclear estrogen-dependent signaling and non-genomic estrogen signaling from the plasma membrane (reviewed in Anbalagan and Rowan, 2015; Halodsen et al, 2014; Schwartz et al, 2016). A number of recent genome wide studies highlight the breadth of the transcriptional response to estrogen. The number of predicted estrogen-dependent target genes ranges from a couple of hundred (based on microarray studies) to upwards of 10000, based on ChIP-chip or ChIP-seq (Cheung and Kraus, 2010; Kinnis and Kraus, 2008; Lin et al, 2004; Welboren et al, 2009; Ikeda et al, 2015; Lin et al, 2007; Carroll et al, 2006). Many of these predicted sites may not represent transcriptionally productive binding events, however. A study examining ESR1 binding by ChIP-seq in 20 primary breast cancers identified a core of 484 ESR-binding events that were conserved in at least 75% of ER+ tumors, which may represent a more realistic estimate (Ross-Innes et al, 2012). These studies also highlight the long-range effect of estrogen receptor-binding, with distal enhancer or promoter elements regulating the expression of many target genes, often through looping or other higher order chromatin structures (Kittler et al, 2013; reviewed in Dietz and Carroll, 2008; Liu and Cheung, 2014; Magnani and Lupien, 2014). Transcription from a number of estrogen-responsive target genes also appears to be primed by the binding of pioneering transcription factors such as FOXA1, GATA3, PBX1 among others. These factors bind to heterochromatin by virtue of their winged helix domains and promote chromatin opening, allowing subsequent recruitment of other transcription factors (reviewed in Zaret and Carroll, 2011; Fiorito et al, 2013; Arnal et al, 2017; Magnani et al, 2011) [<https://reactome.org/PathwayBrowser/#/R-HSA-9018519>].
* **Insulin receptor recycling**: Triggered by acidification of the endosome, insulin dissociates from the receptor and is degraded. The receptor is dephosphorylated and re-integrated into the plasma membrane, ready to be activated again by the binding of insulin molecules [<https://reactome.org/PathwayBrowser/#/R-HSA-77387>].

## GO terms:

**autophagosome assembly** [The formation of a double membrane-bounded structure, the autophagosome, that occurs when a specialized membrane sac, called the isolation membrane, starts to enclose a portion of the cytoplasm. GO:0000045]

**epithelial tube branching involved in lung morphogenesis** [The process in which a highly ordered sequence of patterning events generates the branched epithelial tubes of the lung, consisting of reiterated combinations of bud outgrowth, elongation, and dichotomous subdivision of terminal units. GO:0060441]

**insulin catabolic process** [The chemical reactions and pathways resulting in the breakdown of insulin. GO:1901143]

**insulin receptor recycling** [The process that results in the return of an insulin receptor to an active state at the plasma membrane. An active state is when the receptor is ready to receive an insulin signal. Internalized insulin receptors can be recycled to the plasma membrane or sorted to lysosomes for protein degradation. GO:0038020]

**lipoprotein catabolic process** [The chemical reactions and pathways resulting in the breakdown of any conjugated, water-soluble protein in which the covalently attached nonprotein group consists of a lipid or lipids. GO:0042159]

**positive regulation of apoptotic process** [Any process that activates or increases the frequency, rate or extent of cell death by apoptotic process. This term should only be used when it is not possible to determine which phase or subtype of the apoptotic process is positively regulated by a gene product. Whenever detailed information is available, the more granular children terms should be used. GO:0043065]

**protein catabolic process** [The chemical reactions and pathways resulting in the breakdown of a protein by the destruction of the native, active configuration, with or without the hydrolysis of peptide bonds. This term refers to the breakdown of mature proteins. For cleavage events involved in generating a mature protein from a precursor, consider instead the term ‘protein maturation ; GO:0051604’ and its children. GO:0030163]

**proteolysis** [The hydrolysis of proteins into smaller polypeptides and/or amino acids by cleavage of their peptide bonds. This term was intentionally placed under ‘protein metabolic process ; GO:0019538’ rather than ‘protein catabolic process ; GO:0030163’ to cover all processes centered on breaking peptide bonds, including those involved in protein processing. GO:0006508]

**regulation of establishment of protein localization** [Any process that modulates the frequency, rate or extent of the directed movement of a protein to a specific location. GO:0070201]

**response to nutrient levels** [Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus reflecting the presence, absence, or concentration of nutrients. GO:0031667]

## MSigDB Signatures:

**REACTOME\_NEUTROPHIL\_DEGRANULATION**: Neutrophil degranulation [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_NEUTROPHIL\_DEGRANULATION.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_NEUTROPHIL_DEGRANULATION.html)

**REACTOME\_INNATE\_IMMUNE\_SYSTEM**: Innate Immune System [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_INNATE\_IMMUNE\_SYSTEM.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_INNATE_IMMUNE_SYSTEM.html)

**HEBERT\_MATRISOME\_TNBC\_BONE\_BRAIN\_LIVER\_LUNG\_METASTASTASES**: Matrisome proteins found in significantly higher abundance in TNBC brain, bone, liver and lung metastastases compared to normal samples. [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HEBERT\_MATRISOME\_TNBC\_BONE\_BRAIN\_LIVER\_LUNG\_METASTASTASES.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HEBERT_MATRISOME_TNBC_BONE_BRAIN_LIVER_LUNG_METASTASTASES.html)

**REACTOME\_ADAPTIVE\_IMMUNE\_SYSTEM**: Adaptive Immune System [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_ADAPTIVE\_IMMUNE\_SYSTEM.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_ADAPTIVE_IMMUNE_SYSTEM.html)

**KEGG\_LYSOSOME**: Lysosome [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG\_LYSOSOME.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG_LYSOSOME.html)

**REACTOME\_COLLAGEN\_DEGRADATION**: Collagen degradation [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_COLLAGEN\_DEGRADATION.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_COLLAGEN_DEGRADATION.html)

**HEBERT\_MATRISOME\_TNBC\_BONE\_BRAIN\_LUNG\_LIVER\_METASTASTASES\_TUMOR\_CELL\_DERIVED**: Tumor cell-derived matrisome proteins found in significantly higher abundance in TNBC brain, bone, liver and lung metastastases compared to normal samples. [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HEBERT\_MATRISOME\_TNBC\_BONE\_BRAIN\_LUNG\_LIVER\_METASTASTASES\_TUMOR\_CELL\_DERIVED.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HEBERT_MATRISOME_TNBC_BONE_BRAIN_LUNG_LIVER_METASTASTASES_TUMOR_CELL_DERIVED.html)

**REACTOME\_EXTRACELLULAR\_MATRIX\_ORGANIZATION**: Extracellular matrix organization [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_EXTRACELLULAR\_MATRIX\_ORGANIZATION.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION.html)

**REACTOME\_DEGRADATION\_OF\_THE\_EXTRACELLULAR\_MATRIX**: Degradation of the extracellular matrix [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_DEGRADATION\_OF\_THE\_EXTRACELLULAR\_MATRIX.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX.html)

**REACTOME\_SIGNALING\_BY\_RECEPTOR\_TYROSINE\_KINASES**: Signaling by Receptor Tyrosine Kinases [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_SIGNALING\_BY\_RECEPTOR\_TYROSINE\_KINASES.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_SIGNALING_BY_RECEPTOR_TYROSINE_KINASES.html)

**REACTOME\_SIGNALING\_BY\_NUCLEAR\_RECEPTORS**: Signaling by Nuclear Receptors [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_SIGNALING\_BY\_NUCLEAR\_RECEPTORS.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_SIGNALING_BY_NUCLEAR_RECEPTORS.html)

**REACTOME\_SIGNALING\_BY\_INSULIN\_RECEPTOR**: Signaling by Insulin receptor [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_SIGNALING\_BY\_INSULIN\_RECEPTOR.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_SIGNALING_BY_INSULIN_RECEPTOR.html)

**REACTOME\_ESR\_MEDIATED\_SIGNALING**: ESR-mediated signaling [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_ESR\_MEDIATED\_SIGNALING.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_ESR_MEDIATED_SIGNALING.html)

**NABA\_MATRISOME\_ASSOCIATED**: Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/NABA\_MATRISOME\_ASSOCIATED.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/NABA_MATRISOME_ASSOCIATED.html)

**REACTOME\_INSULIN\_RECEPTOR\_RECYCLING**: Insulin receptor recycling [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_INSULIN\_RECEPTOR\_RECYCLING.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_INSULIN_RECEPTOR_RECYCLING.html)

**KEGG\_MEDICUS\_REFERENCE\_NUCLEAR\_INITIATED\_ESTROGEN\_SIGNALING\_PATHWAY**: Pathway Definition from KEGG: E2 -> ((ESR1/2)+(NCOA1/2/3)) => (BCL2,EBAG9,KRT19,CTSD,TFF1,PGR) [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG\_MEDICUS\_REFERENCE\_NUCLEAR\_INITIATED\_ESTROGEN\_SIGNALING\_PATHWAY.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG_MEDICUS_REFERENCE_NUCLEAR_INITIATED_ESTROGEN_SIGNALING_PATHWAY.html)

**NABA\_MATRISOME**: Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/NABA\_MATRISOME.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/NABA_MATRISOME.html)

**REACTOME\_MHC\_CLASS\_II\_ANTIGEN\_PRESENTATION**: MHC class II antigen presentation [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_MHC\_CLASS\_II\_ANTIGEN\_PRESENTATION.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_MHC_CLASS_II_ANTIGEN_PRESENTATION.html)

# 7. Gene Descriptions

**NCBI Gene Summary**: This gene encodes a member of the A1 family of peptidases. The encoded preproprotein is proteolytically processed to generate multiple protein products. These products include the cathepsin D light and heavy chains, which heterodimerize to form the mature enzyme. This enzyme exhibits pepsin-like activity and plays a role in protein turnover and in the proteolytic activation of hormones and growth factors. Mutations in this gene play a causal role in neuronal ceroid lipofuscinosis-10 and may be involved in the pathogenesis of several other diseases, including breast cancer and possibly Alzheimer’s disease. [provided by RefSeq, Nov 2015]

**GeneCards Summary**: CTSD (Cathepsin D) is a Protein Coding gene. Diseases associated with CTSD include Ceroid Lipofuscinosis, Neuronal, 10 and Neuronal Ceroid Lipofuscinosis. Among its related pathways are Peptide hormone metabolism and ESR-mediated signaling. Gene Ontology (GO) annotations related to this gene include aspartic-type endopeptidase activity. An important paralog of this gene is NAPSA.

**UniProtKB/Swiss-Prot Summary**: Acid protease active in intracellular protein breakdown. Plays a role in APP processing following cleavage and activation by ADAM30 which leads to APP degradation [PMID: 27333034]. Involved in the pathogenesis of several diseases such as breast cancer and possibly Alzheimer disease.

# 8. Cellular Location of Gene Product

Cytoplasmic expression with a granular in all tissues. Predicted location: Secreted, Intracellular (different isoforms) [<https://www.proteinatlas.org/ENSG00000117984/subcellular>]

# 9. Mechanistic Information

* Inhibition of Cathepsin D (CD) protease with Pepstatin A suppressed p53-dependent apoptosis in lymphoid cells, suggesting a possible role for CD in p53-dependent cell death. CD-/- fibroblasts were found to be more resistant to killing by adriamycin and etoposide, as compared to CD+/+ cells. These observations link CD protease to p53-dependent tumor suppression and chemosensitivity [PMID: 9619826].
* OLFM4 negatively interacts with cathepsin D and SDF-1 and inhibits prostate cancer growth and bone metastasis [PMID: 21470957].
* In human primary lung adenocarcinoma tumors, proteomic data showed that there was a higher dependency of glycolysis among the tumors with poor prognosis and an up-regulation of HIF1alpha mRNA expression in tumors with early relapse. Cathepsin D was one of the proteins that was upregulated in the poor prognosis group and was confirmed to originate from the tumor and not from a stromal or inflammatory component. HIF1alpha has previously been shown to induce expression of Cathepsin D, which was found to be upregulated in the relapse patients. The data suggests that higher glycolytic activity and HIF activation are associated with cathepsin D upregulation in tumors with poor prognosis [PMID: 23902561].
* High glucose (HG) increases CTSD expression, induces lysosomal membrane permeabilization and triggers CTSD release from the lysosomes, which collectively contributes to HG-induced cardiomyocyte injury [PMID: 31862139].
* The data suggests that treatment with a cytotoxic drug the activation of a Cathepsin D (CD)-Bax loop leads to the generalized permeabilization of lysosomes and eventually of mitochondria, thus reaching the point of no return, and culminates with the activation of the caspase cascade. The findings imply that dysfunctional permeabilization of lysosomes contributes to the development of chemoresistance [PMID: 18657225].
* LCN2 expression is significantly lower in glioblastoma multiforme (GBM) than in normal tissues and is associated with poorer GBM patient survival. In contrast, higher CTSD expression was observed in GBM tumors than in normal tissues, and higher CTSD expression was associated with poorer overall and disease-free survival. LCN2-overexpressing GBM cells showed significantly reduced proliferation and migration/invasion abilities, including lower expression of cathepsin D (CTSD) protein and mRNA in LCN2-overexpressing GBM cells than in controls. In vitro studies with MEK inhibitors suggest that overexpression of LCN2 inhibits proliferation and invasiveness of GBM cells by increasing ERK phosphorylation which may target cathepsin D expression downstream [PMID: 34062746].
* Cathepsin D (Cath-D) binds to BAT3 in ER+ BCC and they partially co-localize at the surface of lysosomes and in the nucleus. BAT3 silencing inhibits Cath-D accumulation in the nucleus, indicating that Cath-D nuclear targeting is controlled by BAT3. Data suggests that cathepsin D (Cath-D) acts as a nuclear transcriptional cofactor of TRPS1 to regulate ER+ breast cancer cell proliferation and transformation in a non-proteolytic manner [PMID: 26183398].
* Overexpression of S100P led to changes in the expression levels of several cytoskeletal proteins including increased expression of another early pancreatic cancer marker, S100A6, as well as the aspartic protease cathepsin D, both of which are involved in cellular invasion. The data suggests that these changes could contribute to the metastatic spread of pancreatic cancer [PMID: 17875703].

## Summary

CTSD, encoding the enzyme Cathepsin D, is crucial in intracellular protein breakdown and is implicated in the pathogenesis of several diseases, including lung cancer [CS: 9]. In the context of lung toxicity and diseases, the dysregulation of CTSD can be understood through its functional roles and the specific cellular responses to lung stressors [CS: 8]. For example, in response to cigarette smoke exposure, there is an upregulation of Cathepsin D mRNA in mice lungs, primarily expressed in macrophages [CS: 7]. This increase in CTSD expression may be a cellular response to the increased need for protein breakdown and turnover caused by smoke-induced damage [CS: 8]. The proteolytic activity of Cathepsin D could help in degrading damaged or misfolded proteins, a common consequence of oxidative stress from smoke, thereby aiding in cellular repair and maintenance [CS: 8].

In the case of lung cancer, different patterns of CTSD expression are observed [CS: 9]. For instance, in human lung cancer tumors, CTSD mRNA is often downregulated compared to adjacent normal tissue, suggesting a role in cancer progression [CS: 8]. This could be attributed to the enzyme’s role in protein degradation and activation of growth factors [CS: 7]. Downregulation in cancer cells may be a mechanism to reduce the breakdown of proteins that are beneficial for tumor growth and survival [CS: 8]. On the other hand, in small cell lung cancer (SCLC), higher expression of CTSD correlates with better overall survival, indicating a complex role in cancer biology [CS: 7]. This higher expression in the low-risk group could be facilitating the removal of damaged proteins and preventing the accumulation of harmful products, which is critical in maintaining cellular health and combating cancer progression [CS: 7].

# 10. Upstream Regulators

* There are two p53 DNA-binding sites located in the Cathepsin D (CD)-promoter which appear to mediate transactivation of a CD-promoter luciferase-reporter during p53-dependent apoptosis. The findings suggest that these observations link CD protease to p53-dependent tumor suppression and chemosensitivity [PMID: 9619826].
* In human breast cancer cells, cathepsin-D and pS2 mRNAs are specifically and directly induced by estrogens at the transcriptional level, and pS2 mRNA, like cathepsin-D mRNA, are rapidly induced by epidermal growth factor [PMID: 2612733]. Peptide growth factors, such as insulin, insulin-like growth factor I, and basic fibroblast growth factor, can also increase the steady state levels of cathepsin-D and pS2 mRNAs in breast cancer cell lines. Overall, cathepsin-D and pS2 genes are under complex regulation in breast cancer cells, since growth factors stimulate their expression via indirect mechanisms contrasting with the primary transcriptional effects of estrogens [PMID: 2664475].
* RhoA and RhoC play different, but clear, roles in ERalpha signaling by similarly modulated ERalpha recruitment to the vitellogenin estrogen responsive element (ERE) present in a luciferase reporter gene and to the promoters of progesterone receptor (PR), cathepsin D, and pS2 genes. RhoA up-regulated the ERE-luciferase reporter gene activity and PR mRNA expression and tended to down-regulate cathepsin D and pS2 mRNA expression. These GTPases are definitely involved, along with RhoB, in ERalpha recruitment and, to some extent, ERalpha cofactor balance [PMID: 24096540].

# 11. Tissues/Cell Type Where Genes are Overexpressed

**Tissue type enchanced**: low tissue specificity [<https://www.proteinatlas.org/ENSG00000117984/tissue>]

**Cell type enchanced**: distal enterocytes, hofbauer cells, macrophages, proximal enterocytes (cell type enhanced) [[https://www.proteinatlas.org/ENSG00000117984/single+cell+type](https://www.proteinatlas.org/ENSG00000117984/single%2Bcell%2Btype)]

# 12. Role of Gene in Other Tissues

* In mice deficient in *Ctsd,* they displayed impaired myocardial autophagosome removal, reduced autophagic flux, and restrictive cardiomyopathy. The data suggests that myocardial CTSD upregulation induced by myocardial infarction protects against cardiac remodeling and malfunction, which is at least in part through promoting myocardial autophagic flux [PMID: 28694354].
* In fibrotic specimens, proliferative epithelial formations were strongly stained with cathepsin D antibodies whereas detached, desquamated epithelial cells were weakly positive or negative, suggesting that cathepsin D plays a role in the remodeling process during fibrogenesis [PMID: 8764928].
* A study of transcriptomic signatures in peripheral blood suggests an essential role of neutrophil proteases, such as cathepsin D (CTSD), in chronic obstructive pulmonary disease patients with critical respiratory illness [PMID: 22852767].
* In experimentally derived hormone-independent and hormone-responsive variants breast cancer cell lines, results showed that these cells can invade locally from solid mammary fat pad tumors, and produce primary extensions on the surface of intraperitoneal structures including liver, pancreas, and diaphragm of nude mice. The increased metastatic potential was not associated with an increase in either the level of laminin attachment, laminin receptor mRNA expression, or secreted type IV collagenolytic activity. Results did not detect a significant decrease in the steady-state mRNA levels of the metastasis inhibitor nm23 gene. When growing without estrogen in vitro, MCF7/LCC1 cells produce elevated levels of the estrogen-inducible cathepsin D enzyme [PMID: 8380760].
* In an analysis of gene expression data used to identify the co-expression genes related to osteoporosis and atherosclerosis, six hub genes (namely, COL1A1, IBSP, CTSD, RAC2, MAF, and THBS1) were obtained *via* taking interaction of different analysis results. The results of common genes analysis showed that immune and inflammatory response may be a common feature in the pathophysiology of osteoporosis and atherosclerosis [PMID: 35937806].
* In patients, there is a significant correlation between high cathepsin D concentrations in the cytosol of primary breast cancer and development of metastasis [PMID: 1965795]. This marker is independent of other prognostic factors and appears to be particularly useful in lymph node-negative tumors. These results suggest that overexpression and possible derouting of cathepsin D plays an important role in invasion and metastasis of breast cancer [PMID: 2207345].
* Cathepsins B, D, and G are isoenzymes that catalyze the production of angiotensin peptides, hence providing bypass loops for the renin-angiotensin system (RAS). Cathepsin B and cathepsin D, and to a lesser extent cathepsin G, are expressed in WHO grade I meningioma (MG) localized to the microvessels of patients. Cathepsin B and cathepsin D are enzymatically active and are localized to the putative putative tumor stem cells population on the microvessels, whereas cathepsin G was localized to cells scattered within the interstitium. These results suggest the presence of bypass loops for the RAS, within WHO grade I MG [PMID: 30949483].
* The expression of cathepsin D is induced in inflammation-associated intestinal macrophages and the presence of cathepsin D might contribute to the mucosal damage in inflammatory bowel disease [PMID: 15030527].
* In a meta-analysis of numerous Alzheimer’s disease observation studies, molecules related to risk factors that are involved in neuroinflammation (C1q), metabolism disorder (P-S312-IRS-1), neurotrophic deficiency (HGF), vascular injury (VEGF-D), and autophagy-lysosomal system dysfunction (cathepsin D) were increased [PMID: 35535875].
* The expression of cathepsin B and cathepsin D was detected in the putative cancer stem cell subpopulations within liver metastasis from colon adenocarcinoma patient samples [PMID: 30177970].

# 13. Chemicals Known to Elicit Transcriptional Response of Biomarker in Tissue of Interest

## **Compounds that increase expression of the gene:**

* Yessotoxin [PMID: 30679557]
* benzo[a]pyrene diol epoxide I [PMID: 20382639]
* bisphenol A [PMID: 29275510]
* carbon nanotube [PMID: 25554681]
* resveratrol [PMID: 18089832]
* silicon dioxide [PMID: 32721576]

# 14. DisGeNet Biomarker Associations to Disease in Organ of Interest

Most relevant biomarkers with lower score or lower probability of association with disease or organ of interest:

* Malignant neoplasm of breast [PMID: 10481941, PMID: 11165043, PMID: 12462383, PMID: 12651610, PMID: 21311773]
* Malignant Neoplasms [PMID: 10481941, PMID: 18559512, PMID: 21148553, PMID: 22996917, PMID: 28317223]