# 1. Gene Aliases

Cathepsin K, CTSO2, PKND, CTSO, Cathepsin O2, EC 3.4.22.38, Cathepsin O, Cathepsin X, PYCD, Cathepsin K (Pycnodysostosis), Cathepsin O1, EC 3.4.22, CTS02, CTSO1 [<https://www.genecards.org/cgi-bin/carddisp.pl?gene=CTSK&keywords=Ctsk>]

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

* Enhanced gene transcription for CTSK and increase cathepsin K protein expression were observed in human pulmonary fibrosis. Lung specimens obtained from patients with lung fibrosis fibroblasts expressed larger amounts of catK than those obtained from normal lungs indicating a role of CTSK in lung matrix homeostasis. Primary lung fibroblasts derived from CTSK(-/-) mice showed a decreased collagenolytic activity indicating the role of catK in collagen degradation and lung fibrosis. [PMID: 15161653].

# 3. Summary of Protein Family and Structure

* Protein Accession: P43235
* Size: 329 amino acids
* Molecular mass: 36966 Da
* Domains: [Papain-like\_cys\_pep\_sf](http://www.ebi.ac.uk/interpro/entry/IPR038765), [Pept\_cys\_AS](http://www.ebi.ac.uk/interpro/entry/IPR000169), [Pept\_asp\_AS](http://www.ebi.ac.uk/interpro/entry/IPR025661), [Pept\_his\_AS](http://www.ebi.ac.uk/interpro/entry/IPR025660), [Peptidase\_C1A](http://www.ebi.ac.uk/interpro/entry/IPR013128), [Peptidase\_C1A\_C](http://www.ebi.ac.uk/interpro/entry/IPR000668), [Peptidase\_C1A\_papain-like](http://www.ebi.ac.uk/interpro/entry/IPR039417), [Prot\_inhib\_I29](http://www.ebi.ac.uk/interpro/entry/IPR013201)
* Blocks: Papain cysteine protease (C1) family signature, Proteinase inhibitor I29, cathepsin propeptide
* Family: Belongs to the peptidase C1 family (also called papain family protein, is a papain-like cysteine peptidase that catalyzes the hydrolysis of peptide bonds in substrates using a catalytic dyad of Cys and His residues)
* Cathepsin K (Ctsk) is a lysosomal cysteine protease that plays a significant role in the skeletal system, primarily related to bone remodeling and resorption. This class of enzymes is known for their ability to break down proteins using a cysteine residue in their active site [PMID: 28651365].
* Cathepsin K is capable of cleaving native type I collagen at the triple helical regions at PH values between 4.5 and 6.6. Cathepsin K degrades collagens at different sites in the N-terminus region [PMID: 21670768]. By breaking down collagen, Ctsk contributes to the process of bone resorption, allowing for the remodeling of bone tissue. [Kelley and Firestein’s Textbook of Rheumatology (Tenth Edition)], Volume 1, 2017, Pages 106-125, [<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/cathepsin-k>]. Ctsk has specific domains and endoprotease activity. These attributes enable it to cleave proteins like fibrinogen under acidic pH conditions.

# 4. Proteins Known to Interact with Gene Product

## Interactions with experimental support

* **KNG1** . [PMID: 12492488, PMID: 7492318, PMID: 9276160]
* **ARID4B** AT-rich interactive domain-containing protein 4B; Acts as a transcriptional repressor. May function in the assembly and/or enzymatic activity of the Sin3A corepressor complex or in mediating interactions between the complex and other regulatory complexes. Plays a role in the regulation of epigenetic modifications at the PWS/AS imprinting center near the SNRPN promoter, where it might function as part of a complex with RB1 and ARID4A. [PMID: 26186194, PMID: 28514442]
* **C18orf25** Uncharacterized protein C18orf25; Chromosome 18 open reading frame 25. [PMID: 26186194, PMID: 28514442]
* **CTSV** Cathepsin L2; Cysteine protease. May have an important role in corneal physiology; Belongs to the peptidase C1 family. [PMID: 26186194, PMID: 28514442]
* **ERBIN** Erbin; Acts as an adapter for the receptor ERBB2, in epithelia. By binding the unphosphorylated ‘Tyr-1248’ of receptor ERBB2, it may contribute to stabilize this unphosphorylated state. Inhibits NOD2-dependent NF-kappa-B signaling and proinflammatory cytokine secretion ; Belongs to the LAP (LRR and PDZ) protein family. [PMID: 26186194, PMID: 28514442]
* **CAMP** Cathelicidin antimicrobial peptide; Binds to bacterial lipopolysaccharides (LPS), has antibacterial activity. [PMID: 25884905]
* **CGAS** Cyclic GMP-AMP synthase; Nucleotidyltransferase that catalyzes the formation of cyclic GMP-AMP (cGAMP) from ATP and GTP and plays a key role in innate immunity. Catalysis involves both the formation of a 2’,5’ phosphodiester linkage at the GpA step and the formation of a 3’,5’ phosphodiester linkage at the ApG step, producing c[G(2’,5’)pA(3’,5’)p]. Acts as a key cytosolic DNA sensor, the presence of double-stranded DNA (dsDNA) in the cytoplasm being a danger signal that triggers the immune responses. [PMID: 30471916]
* **COL1A2** Collagen alpha-2(I) chain; Type I collagen is a member of group I collagen (fibrillar forming collagen); Belongs to the fibrillar collagen family. [PMID: 19010413]
* **CTSL** Cathepsin L1 heavy chain; Thiol protease important for the overall degradation of proteins in lysosomes (Probable). Involved in the solubilization of cross-linked TG/thyroglobulin and in the subsequent release of thyroid hormone thyroxine (T4) by limited proteolysis of TG/thyroglobulin in the thyroid follicle lumen (By similarity). [PMID: 28514442]
* **FGFR3** Fibroblast growth factor receptor 3; Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays an essential role in the regulation of cell proliferation, differentiation and apoptosis. Plays an essential role in the regulation of chondrocyte differentiation, proliferation and apoptosis, and is required for normal skeleton development. Regulates both osteogenesis and postnatal bone mineralization by osteoblasts. Promotes apoptosis in chondrocytes, but can also promote cancer cell proliferation. Required for normal development of the inner ear. [PMID: 16169070]
* **LIG4** DNA ligase 4; Efficiently joins single-strand breaks in a double-stranded polydeoxynucleotide in an ATP-dependent reaction. Involved in DNA non- homologous end joining (NHEJ) required for double-strand break repair and V(D)J recombination. The LIG4-XRCC4 complex is responsible for the NHEJ ligation step, and XRCC4 enhances the joining activity of LIG4. Binding of the LIG4-XRCC4 complex to DNA ends is dependent on the assembly of the DNA-dependent protein kinase complex DNA-PK to these DNA ends. [PMID: 22990118]
* **RARRES2** Retinoic acid receptor responder protein 2; Adipocyte-secreted protein (adipokine) that regulates adipogenesis, metabolism and inflammation through activation of the chemokine-like receptor 1 (CMKLR1). Its other ligands include G protein-coupled receptor 1 (GPR1) and chemokine receptor-like 2 (CCRL2). Positively regulates adipocyte differentiation, modulates the expression of adipocyte genes involved in lipid and glucose metabolism and might play a role in angiogenesis, a process essential for the expansion of white adipose tissue. [PMID: 21715684]
* **SERPINB13** Serpin B13; May play a role in the proliferation or differentiation of keratinocytes; Belongs to the serpin family. Ov-serpin subfamily. [PMID: 12504904]
* **SERPINB3** Serpin B3; May act as a papain-like cysteine protease inhibitor to modulate the host immune response against tumor cells. Also functions as an inhibitor of UV-induced apoptosis via suppression of the activity of c-Jun NH(2)-terminal kinase (JNK1). [PMID: 9548757]
* **SPARC** SPARC; Appears to regulate cell growth through interactions with the extracellular matrix and cytokines. Binds calcium and copper, several types of collagen, albumin, thrombospondin, PDGF and cell membranes. There are two calcium binding sites; an acidic domain that binds 5 to 8 Ca(2+) with a low affinity and an EF-hand loop that binds a Ca(2+) ion with a high affinity. [PMID: 8647860]

## Interactions with text mining support

* **ACP5** Tartrate-resistant acid phosphatase type 5; Involved in osteopontin/bone sialoprotein dephosphorylation. Its expression seems to increase in certain pathological states such as Gaucher and Hodgkin diseases, the hairy cell, the B-cell, and the T- cell leukemias; Belongs to the metallophosphoesterase superfamily. Purple acid phosphatase family. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000271651 9606.ENSP00000218758](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000271651%0D9606.ENSP00000218758)]
* **NFATC1** Nuclear factor of activated T-cells, cytoplasmic 1; Plays a role in the inducible expression of cytokine genes in T-cells, especially in the induction of the IL-2 or IL-4 gene transcription. Also controls gene expression in embryonic cardiac cells. Could regulate not only the activation and proliferation but also the differentiation and programmed death of T-lymphocytes as well as lymphoid and non-lymphoid cells. Required for osteoclastogenesis and regulates many genes important for osteoclast differentiation and function (By similarity). [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000271651 9606.ENSP00000389377](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000271651%0D9606.ENSP00000389377)]
* **TNFSF11** Tumor necrosis factor ligand superfamily member 11, membrane form; Cytokine that binds to TNFRSF11B/OPG and to TNFRSF11A/RANK. Osteoclast differentiation and activation factor. Augments the ability of dendritic cells to stimulate naive T-cell proliferation. May be an important regulator of interactions between T-cells and dendritic cells and may play a role in the regulation of the T-cell-dependent immune response. May also play an important role in enhanced bone-resorption in humoral hypercalcemia of malignancy. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000271651 9606.ENSP00000381775](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000271651%0D9606.ENSP00000381775)]
* **MMP9** 67 kDa matrix metalloproteinase-9; May play an essential role in local proteolysis of the extracellular matrix and in leukocyte migration. Could play a role in bone osteoclastic resorption. Cleaves KiSS1 at a Gly-|-Leu bond. Cleaves type IV and type V collagen into large C-terminal three quarter fragments and shorter N-terminal one quarter fragments. Degrades fibronectin but not laminin or Pz-peptide. Belongs to the peptidase M10A family. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000271651 9606.ENSP00000361405](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000271651%0D9606.ENSP00000361405)]
* **DCSTAMP** Dendritic cell-specific transmembrane protein; Probable cell surface receptor that plays several roles in cellular fusion, cell differentiation, bone and immune homeostasis. Plays a role in TNFSF11-mediated osteoclastogenesis. Cooperates with OCSTAMP in modulating cell-cell fusion in both osteoclasts and foreign body giant cells (FBGCs). Participates in osteoclast bone resorption. Involved in inducing the expression of tartrate-resistant acid phosphatase in osteoclast precursors. Plays a role in haematopoietic stem cell differentiation of bone marrow cells toward the myeloid lineage. [[https://string-db.org/newstring\_cgi/show\_edge\_details.pl?identifiers=9606.ENSP00000271651 9606.ENSP00000297581](https://string-db.org/newstring_cgi/show_edge_details.pl?identifiers=9606.ENSP00000271651%0D9606.ENSP00000297581)]

# 5. Links to Gene Databases

* GeneCards (human): <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CTSK>
* Harmonizome (human): <https://maayanlab.cloud/Harmonizome/gene/CTSK>
* NCBI (human): <https://www.ncbi.nlm.nih.gov/gene/1513>
* NCBI (rat): <https://www.ncbi.nlm.nih.gov/gene/29175>
* Ensemble (human): <https://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000143387>
* Ensemble (rat): <https://useast.ensembl.org/Rattus_norvegicus/Gene/Summary?g=ENSRNOG00000021155>
* Rat Genome Database (rat): <https://rgd.mcw.edu/rgdweb/report/gene/main.html?id=61810>
* Uniprot (human): <https://www.uniprot.org/uniprotkb/P43235>
* Uniprot (rat): <https://www.uniprot.org/uniprotkb/O35186>
* Wikigenes (human): <https://www.wikigenes.org/e/gene/e/1513.html>
* Wikigenes (rat): <https://www.wikigenes.org/e/gene/e/29175.html>
* Alphafold (human): <https://alphafold.ebi.ac.uk/entry/P43235>
* Alphafold (rat): <https://alphafold.ebi.ac.uk/entry/O35186>
* PDB (human): <https://www.rcsb.org/structure/1ATK>, <https://www.rcsb.org/structure/1AU0>, <https://www.rcsb.org/structure/1AU2>, <https://www.rcsb.org/structure/1AU3>, <https://www.rcsb.org/structure/1AU4>, <https://www.rcsb.org/structure/1AYU>, <https://www.rcsb.org/structure/1AYV>, <https://www.rcsb.org/structure/1AYW>, <https://www.rcsb.org/structure/1BGO>, <https://www.rcsb.org/structure/1BY8>, <https://www.rcsb.org/structure/1MEM>, <https://www.rcsb.org/structure/1NL6>, <https://www.rcsb.org/structure/1NLJ>, <https://www.rcsb.org/structure/1Q6K>, <https://www.rcsb.org/structure/1SNK>, <https://www.rcsb.org/structure/1TU6>, <https://www.rcsb.org/structure/1U9V>, <https://www.rcsb.org/structure/1U9W>, <https://www.rcsb.org/structure/1U9X>, <https://www.rcsb.org/structure/1VSN>, <https://www.rcsb.org/structure/1YK7>, <https://www.rcsb.org/structure/1YK8>, <https://www.rcsb.org/structure/1YT7>, <https://www.rcsb.org/structure/2ATO>, <https://www.rcsb.org/structure/2AUX>, <https://www.rcsb.org/structure/2AUZ>, <https://www.rcsb.org/structure/2BDL>, <https://www.rcsb.org/structure/2R6N>, <https://www.rcsb.org/structure/3C9E>, <https://www.rcsb.org/structure/3KW9>, <https://www.rcsb.org/structure/3KWB>, <https://www.rcsb.org/structure/3KWZ>, <https://www.rcsb.org/structure/3KX1>, <https://www.rcsb.org/structure/3O0U>, <https://www.rcsb.org/structure/3O1G>, <https://www.rcsb.org/structure/3OVZ>, <https://www.rcsb.org/structure/4DMX>, <https://www.rcsb.org/structure/4DMY>, <https://www.rcsb.org/structure/4N79>, <https://www.rcsb.org/structure/4N8W>, <https://www.rcsb.org/structure/4X6H>, <https://www.rcsb.org/structure/4X6I>, <https://www.rcsb.org/structure/4X6J>, <https://www.rcsb.org/structure/4YV8>, <https://www.rcsb.org/structure/4YVA>, <https://www.rcsb.org/structure/5TUN>, <https://www.rcsb.org/structure/6HGY>, <https://www.rcsb.org/structure/6QBS>, <https://www.rcsb.org/structure/6QL8>, <https://www.rcsb.org/structure/6QLM>, <https://www.rcsb.org/structure/6QLW>, <https://www.rcsb.org/structure/6QLX>, <https://www.rcsb.org/structure/6QM0>, <https://www.rcsb.org/structure/7NXL>, <https://www.rcsb.org/structure/7NXM>, <https://www.rcsb.org/structure/7PCK>, <https://www.rcsb.org/structure/7QBL>, <https://www.rcsb.org/structure/7QBM>, <https://www.rcsb.org/structure/7QBN>, <https://www.rcsb.org/structure/7QBO>
* PDB (mouse): <https://www.rcsb.org/structure/5T6U>
* PDB (rat): none

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

## **Pathways:**

* **Activation of Matrix Metalloproteinases:** The matrix metalloproteinases (MMPs), previously known as matrixins, are classically known to be involved in the turnover of extracellular matrix (ECM) components. However, recent high throughput proteomics analyses have revealed that ~80% of MMP substrates are non-ECM proteins including cytokines, growth factor binding protiens, and receptors. It is now clear that MMPs regulate ECM turnover not only by cleaving ECM components, but also by the regulation of cell signalling, and that some MMPs are beneficial and may be drug anti-targets. Thus, MMPs have important roles in many processes including embryo development, morphogenesis, tissue homeostasis and remodeling. They are implicated in several diseases such as arthritis, periodontitis, glomerulonephritis, atherosclerosis, tissue ulceration, and cancer cell invasion and metastasis. All MMPs are synthesized as preproenzymes. Alternate splice forms are known, leading to nuclear localization of select MMPs. Most are secreted from the cell, or in the case of membrane type (MT) MMPs become plasma membrane associated, as inactive proenzymes. Their subsequent activation is a key regulatory step, with requirements specific to MMP subtype. [<https://reactome.org/PathwayBrowser/#/R-HSA-1592389>].
* **Degradation of the extracellular matrix:** Matrix metalloproteinases (MMPs), previously referred to as matrixins because of their role in degradation of the extracellular matrix (ECM), are zinc and calcium dependent proteases belonging to the metzincin family. They contain a characteristic zinc-binding motif HEXXHXXGXXH (Stocker & Bode 1995) and a conserved Methionine which forms a Met-turn. Humans have 24 MMP genes giving rise to 23 MMP proteins, as MMP23 is encoded by two identical genes. All MMPs contain an N-terminal secretory signal peptide and a prodomain with a conserved PRCGXPD motif that in the inactive enzyme is localized with the catalytic site, the cysteine acting as a fourth unpaired ligand for the catalytic zinc atom. Activation involves delocalization of the domain containing this cysteine by a conformational change or proteolytic cleavage, a mechanism referred to as the cysteine-switch (Van Wart & Birkedal-Hansen 1990). Most MMPs are secreted but the membrane type MT-MMPs are membrane anchored and some MMPs may act on intracellular proteins. Various domains determine substrate specificity, cell localization and activation (Hadler-Olsen et al. 2011). MMPs are regulated by transcription, cellular location (most are not activated until secreted), activating proteinases that can be other MMPs, and by metalloproteinase inhibitors such as the tissue inhibitors of metalloproteinases (TIMPs). MMPs are best known for their role in the degradation and removal of ECM molecules. In addition, cleavage of the ECM and other cell surface molecules can release ECM-bound growth factors, and a number of non-ECM proteins are substrates of MMPs (Nagase et al. 2006). MMPs can be divided into subgroups based on domain structure and substrate specificity but it is clear that these are somewhat artificial, many MMPs belong to more than one functional group (Vise & Nagase 2003, Somerville et al. 2003). [<https://reactome.org/PathwayBrowser/#/R-HSA-1474228&PATH=R-HSA-1474244>].
* **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>].
* **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>].
* **RNA Polymerase II Transcription:** RNA polymerase II (Pol II) is the central enzyme that catalyses DNA- directed mRNA synthesis during the transcription of protein-coding genes. Pol II consists of a 10-subunit catalytic core, which alone can elongate the RNA transcript, and a complex of two subunits, Rpb4/7, that is required for transcription initiation. The transcription cycle is divided in three major phases: initiation, elongation, and termination. Transcription initiation includes promoter DNA binding, DNA melting, and initial synthesis of short RNA transcripts. The transition from initiation to elongation is referred to as promoter escape and leads to a stable elongation complex that is characterized by an open DNA region or transcription bubble. The bubble contains the DNA-RNA hybrid, a heteroduplex of eight to nine base pairs. The growing 3-end of the RNA is engaged with the polymerase complex active site. Ultimately transcription terminates and Pol II dissociates from the template. [<https://reactome.org/PathwayBrowser/#/R-HSA-73857&PATH=R-HSA-74160>].
* **RUNX1 regulates transcription of genes involved in differentiation of keratinocytes:** The RUNX1:CBFB complex directly inhibits transcription of the SERPINB13 gene (Nomura et al. 2005), a gene involved in keratinocyte differentiation that is frequently down-regulated in head and neck cancers (Boyapati et al. 2011). RUNX1 also inhibits transcription of STAT3 inhibitors SOCS3 and SOCS4, resulting in elevated STAT3 activity. RUNX1-mediated increase in STAT3 activity, first discovered in keratinocytes, is thought to be involved in the maintenance of epithelial stem cells and contributes to development of epithelial cancers, including squamous cell carcinoma (SCC) of the skin (Scheitz et al. 2012). [<https://reactome.org/PathwayBrowser/#/R-HSA-8939242>].
* **Toll-like Receptor Cascades:** In human, ten members of the Toll-like receptor (TLR) family (TLR1-TLR10) have been identified (TLR11 has been found in mouse, but not in human). All TLRs have a similar Toll/IL-1 receptor (TIR) domain in their cytoplasmic region and an Ig-like domain in the extracellular region, where each is enriched with a varying number of leucine-rich repeats (LRRs). Each TLR can recognize specific microbial pathogen components. The binding pathogenic component to TLR initializes signaling pathways that lead to induction of Interferon alpha/beta and inflammatory cytokines. There are two main signaling pathways. The first is a MyD88-dependent pathway that is common to all TLRs, except TLR3; the second is a TRIF(TICAM1)-dependent pathway that is peculiar to TLR3 and TLR4. TLR4-mediated signaling pathway via TRIF requires adapter molecule TRAM (TRIF-related adapter molecule or TICAM2). TRAM is thought to bridge between the activated TLR4 complex and TRIF. (Takeda & Akira 2004; Akira 2003; Takeda & Akira 2005; Kawai 2005; Heine & Ulmer 2005). This pathway is organized as trafficking and processing of TLR, various TLR cascades (TLR10, TLR3, TLR5, TLR7/8, TLR9, TLR4, TLR2) and their regulation. [<https://reactome.org/PathwayBrowser/#/R-HSA-168898>].
* **Trafficking and processing of endosomal TLR:** Mammalian TLR3, TLR7, TLR8, TLR9 are endosomal receptors that sense nucleic acids that have been released from endocytosed/phagocytosed bacteria, viruses or parasites. These TLRs have a ligand-recognition domain that faces the lumen of the endosome (which is topologically equivalent to the outside of the cell), a transmembrane domain, and a signaling domain that faces the cytosol. Under normal conditions, self-nucleic acids are not recognized by TLRs due to multiple levels of regulation including receptor compartmentalization, trafficking and proteolytic processing (Barton GM et al 2006, Ewald SE et al 2008). At steady state TLR3, TLR7, TLR8, TLR9 reside primarily in the endoplasmic reticulum (ER), however, their activation by specific ligands only occurs within acidified endolysosomal compartments (Hacker H et al 1998, Funami K et al 2004, Gibbard RJ et al 2006). Several chaperon proteins associate with TLRs in the ER to provide efficient translocation to endolysosome. Upon reaching endolysosomal compartments the ectodomains of TLR7 and TLR9 are proteolytically cleaved by cysteine endoproteases. Both full-length and cleaved C-terminus of TLR9 bind CpG-oligodeoxynucleotides, however it has been proposed that only the processed receptor is functional. Although similar cleavage of TLR3 has been reported by Ewald et al 2011, other studies demonstrated that the N-terminal region of TLR3 ectodomain was implicated in ligand binding, thus TLR3 may function as a full-length receptor (Liu L et al 2008, Tokisue T et al 2008). There is no data on TLR8 processing, although the cell biology of TLR8 is probably similar to TLR9 and TLR7 (Gibbard RJ et al 2006, Wei T et al 2009). [<https://reactome.org/PathwayBrowser/#/R-HSA-1679131>].

## GO terms:

**apoptotic process** [A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathway phase) which trigger an execution phase. The execution phase is the last step of an apoptotic process, and is typically characterized by rounding-up of the cell, retraction of pseudopodes, reduction of cellular volume (pyknosis), chromatin condensation, nuclear fragmentation (karyorrhexis), plasma membrane blebbing and fragmentation of the cell into apoptotic bodies. When the execution phase is completed, the cell has died. GO:0006915]

**bone resorption** [The process in which specialized cells known as osteoclasts degrade the organic and inorganic portions of bone, and endocytose and transport the degradation products. GO:0045453]

**cellular response to transforming growth factor beta stimulus** [Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a transforming growth factor beta stimulus. GO:0071560]

**cellular response to tumor necrosis factor** [Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a tumor necrosis factor stimulus. GO:0071356]

**cellular response to zinc ion starvation** [Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of deprivation of zinc ions. GO:0034224]

**collagen catabolic process** [The proteolytic chemical reactions and pathways resulting in the breakdown of collagen in the extracellular matrix, usually carried out by proteases secreted by nearby cells. GO:0030574]

**immune response** [Any immune system process that functions in the calibrated response of an organism to a potential internal or invasive threat. GO:0006955]

**intramembranous ossification** [Direct ossification that occurs within mesenchyme or an accumulation of relatively unspecialized cells.|An instance of intramembranous ossification may also be classified as metaplastic; the former classifies based on tissue type location, and the latter based on mechanism/cell division. GO:0001957]

**mononuclear cell differentiation** [The process in which a relatively unspecialized cell acquires the specialized features of a mononuclear cell. GO:1903131]

**negative regulation of cartilage development** [Any process that decreases the rate, frequency, or extent of cartilage development, the process whose specific outcome is the progression of the cartilage over time, from its formation to the mature structure. Cartilage is a connective tissue dominated by extracellular matrix containing collagen type II and large amounts of proteoglycan, particularly chondroitin sulfate. GO:0061037]

**positive regulation of apoptotic signaling pathway** [Any process that activates or increases the frequency, rate or extent of apoptotic signaling pathway. GO:2001235]

**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]

**proteolysis involved in protein catabolic process** [The hydrolysis of a peptide bond or bonds within a protein as part of the chemical reactions and pathways resulting in the breakdown of a protein by individual cells. GO:0051603]

**response to ethanol** [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 an ethanol stimulus. GO:0045471]

**response to insulin** [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 an insulin stimulus. Insulin is a polypeptide hormone produced by the islets of Langerhans of the pancreas in mammals, and by the homologous organs of other organisms. GO:0032868]

**response to organic cyclic compound** [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 an organic cyclic compound stimulus. GO:0014070]

**thyroid hormone generation** [The formation of either of the compounds secreted by the thyroid gland, mainly thyroxine and triiodothyronine. This is achieved by the iodination and joining of tyrosine molecules to form the precursor thyroglobin, proteolysis of this precursor gives rise to the thyroid hormones.|Note that this term does not fall under the general GO definition for biosynthetic processes, which is ‘The chemical reactions and pathways resulting in the formation of…’, because thyroid hormones can only be formed by the proteolysis of a larger molecule (see term definition). The word ‘generation’ is therefore used in place of biosynthesis. GO:0006590]

## MSigDB Signatures:

**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)

**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)

**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)

**NABA\_MATRISOME\_ASSOCIATED**: Ensemble of genes encoding ECM-associated proteins including ECM-affilaited 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\_TRANSCRIPTIONAL\_REGULATION\_BY\_RUNX1**: Transcriptional regulation by RUNX1 [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_TRANSCRIPTIONAL\_REGULATION\_BY\_RUNX1.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX1.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)

**WP\_OSTEOCLAST\_SIGNALING**: Osteoclast signaling [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/WP\_OSTEOCLAST\_SIGNALING.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/WP_OSTEOCLAST_SIGNALING.html)

**REACTOME\_RNA\_POLYMERASE\_II\_TRANSCRIPTION**: RNA Polymerase II Transcription [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_RNA\_POLYMERASE\_II\_TRANSCRIPTION.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION.html)

**REACTOME\_ACTIVATION\_OF\_MATRIX\_METALLOPROTEINASES**: Activation of Matrix Metalloproteinases [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_ACTIVATION\_OF\_MATRIX\_METALLOPROTEINASES.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_ACTIVATION_OF_MATRIX_METALLOPROTEINASES.html)

**WP\_IL\_26\_SIGNALING\_PATHWAYS**: IL 26 signaling pathways [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/WP\_IL\_26\_SIGNALING\_PATHWAYS.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/WP_IL_26_SIGNALING_PATHWAYS.html)

**CARRILLOREIXACH\_MRS3\_VS\_LOWER\_RISK\_HEPATOBLASTOMA\_DN**: Genes significantly down-regulated in the high-risk Molecular Risk Stratification (MRS-3) hepatoblastoma (HB) as compared with intermediate-risk (MRS-2) and low-risk (MRS-1) molecular HBs, assessed by Human Transcriptome Array (HTA). [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/CARRILLOREIXACH\_MRS3\_VS\_LOWER\_RISK\_HEPATOBLASTOMA\_DN.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/CARRILLOREIXACH_MRS3_VS_LOWER_RISK_HEPATOBLASTOMA_DN.html)

**REACTOME\_RUNX1\_REGULATES\_TRANSCRIPTION\_OF\_GENES\_INVOLVED\_IN\_DIFFERENTIATION\_OF\_KERATINOCYTES**: RUNX1 regulates transcription of genes involved in differentiation of keratinocytes [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME\_RUNX1\_REGULATES\_TRANSCRIPTION\_OF\_GENES\_INVOLVED\_IN\_DIFFERENTIATION\_OF\_KERATINOCYTES.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_RUNX1_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_DIFFERENTIATION_OF_KERATINOCYTES.html)

**KEGG\_TOLL\_LIKE\_RECEPTOR\_SIGNALING\_PATHWAY**: Toll-like receptor signaling pathway [[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG\_TOLL\_LIKE\_RECEPTOR\_SIGNALING\_PATHWAY.html]](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY.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**: The protein encoded by this gene is a lysosomal cysteine proteinase involved in bone remodeling and resorption. This protein, which is a member of the peptidase C1 protein family, is predominantly expressed in osteoclasts. However, the encoded protein is also expressed in a significant fraction of human breast cancers, where it could contribute to tumor invasiveness. Mutations in this gene are the cause of pycnodysostosis, an autosomal recessive disease characterized by osteosclerosis and short stature. [provided by RefSeq, Apr 2013]

**GeneCards Summary**: CTSK (Cathepsin K) is a Protein Coding gene. Diseases associated with CTSK include Pycnodysostosis and Nail Disorder, Nonsyndromic Congenital, 9. Among its related pathways are Gene expression (Transcription) and Innate Immune System. Gene Ontology (GO) annotations related to this gene include cysteine-type endopeptidase activity and collagen binding. An important paralog of this gene is CTSS.

**UniProtKB/Swiss-Prot Summary**: Thiol protease involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling. Displays potent endoprotease activity against fibrinogen at acid pH. May play an important role in extracellular matrix degradation. Involved in the release of thyroid hormone thyroxine (T4) by limited proteolysis of TG/thyroglobulin in the thyroid follicle lumen [PMID: 11082042].

# 8. Cellular Location of Gene Product

Mainly localized to vesicles. Predicted location: Membrane, Intracellular (different isoforms) [<https://www.proteinatlas.org/ENSG00000143387/subcellular>]

# 9. Mechanistic Information

* Overexpression of cathepsin K reduced lung collagen deposition and improved lung function parameters in the lungs of transgenic mice, thereby providing at least partial protection against bleomycin-induced pulmonary fibrosis [PMID: 18638383].
* The expression of CTSK in the osteoclastogenesis-supporting cells, including dental pulp stem cells, gingival fibroblasts, and periodontal ligament fibroblasts (PDLFs) was significantly elevated by treatment with inflammatory cytokines such as TNFalpha and IL-1beta [PMID: 33445732].
* In conditions like periodontal diseases (PD), which involve increased cytokine production and inflammatory cell infiltration, CTSK expression is elevated. This is often linked to the role of CTSK in bone remodeling, where it participates in bone resorption during inflammatory responses. In PD, CTSK contributes to the destruction of periodontal tissues and deficiency of Ctsk results in decreased immune-cell infiltration and osteoclast numbers. [PMID: 25896020].
* Genetic deletion of Cathepsin K (Ctsk) in RA, caused a radical reduction in inflammation and bone erosion within RA joint capsules and periodontal lesions, a drastic decrease in immune-cell infiltration, and a significant reduction in osteoclasts, macrophages, dendritic and T-cells. Deficiency of Ctsk greatly decreased the expression of TLR-4, 5, and 9 and their downstream cytokines in periodontal gingival epithelial lesions and synovial RA lesions. [PMID: 25896020].

## Summary

Ctsk, encoding Cathepsin K, is dysregulated in lung diseases and toxicities due to its role in extracellular matrix degradation and response to inflammation [CS: 8]. In lung fibrosis, a condition characterized by excessive collagen deposition and fibrous tissue formation, the overexpression of Ctsk and increased Cathepsin K protein expression are observed [CS: 7]. This elevation in Ctsk activity aids in breaking down excessive collagen, as Cathepsin K can degrade collagen, particularly under acidic conditions [CS: 9].

Inflammatory conditions in the lung, like those induced by toxic exposures, often result in elevated levels of cytokines such as TNFalpha and IL-1beta [CS: 8]. These cytokines, in turn, increase the expression of Ctsk in various cells, including lung fibroblasts [CS: 7]. The upregulation of Ctsk under these conditions might serve to counteract the effects of inflammation and excessive extracellular matrix deposition [CS: 6]. For example, in pulmonary fibrosis induced by agents like bleomycin, overexpression of cathepsin K has been shown to reduce lung collagen deposition and improve lung function, suggesting a compensatory mechanism to maintain lung matrix homeostasis and function amid toxic stress [CS: 7].

# 10. Upstream Regulators

* The expression of CTSK in the osteoclastogenesis-supporting cells, including dental pulp stem cells, gingival fibroblasts, and periodontal ligament fibroblasts (PDLFs) was significantly elevated by treatment with inflammatory cytokines such as TNFalpha and IL-1beta [PMID: 33445732].
* Pathological conditions can disrupt the balance between bone resorption and formation. CTSK plays a crucial role in bone resorption, and in situations where bone resorption exceeds bone formation, such as in osteoporosis, CTSK expression could be induced [PMID: 28651365].
* The activity of cathepsin K is modulated by several factors: RANKL, NFAT, and microphthalmia transcription factor (MiTF) enhance cathepsin K expression, and therefore osteoclast formation and bone resorption. On the other hand, interferon (IFN)-gamma, estradiol, calcitonin, and calcium reduce it [PMID: 32582709].

# 11. Tissues/Cell Type Where Genes are Overexpressed

**Tissue type enchanced**: cervix (tissue enhanced) [<https://www.proteinatlas.org/ENSG00000143387/tissue>]

**Cell type enchanced**: fibroblasts, langerhans cells, leydig cells, peritubular cells (cell type enhanced) [<https://www.proteinatlas.org/ENSG00000143387/single+cell+type>]

# 12. Role of Gene in Other Tissues

* The CTSK gene, which codes for cathepsin K, is highly overexpressed in both glioblastoma (GBM) tissues and GBM cells. This suggests a significant involvement of cathepsin K in the pathology of GBM [PMID: 25356585].
* Cathepsin K is secreted by osteoclasts to degrade collagen and other matrix proteins during bone resorption. Global deletion of Ctsk is known to have an impact on bone physiology [PMID: 23321671]. High expression of Cathepsin K in osteoclasts and its capacity to efficiently degrade Type I collagen suggest that Cathepsin K may play a major role in human osteoclastic bone resorption [PMID: 8567669].
* Microarray data has shown that CTSK is upregulated in patients with periodontal disease compared to healthy individuals, indicating its role in periodontal disease [PMID: 33445732].
* Osteoclast Differentiation: Deletion of CTSK in osteoclasts enhances bone formation in vivo by increasing the generation of osteoclast-derived sphingosine-1-phosphate (S1P) [PMID: 23321671].
* Elevated gene and protein expression of CTSK were strongly associated to lymph node metastasis and perineural invasion of oral squamous cell carcinoma [PMID: 29618339].
* Deficiency of Cathepsin K prevents inflammation and bone erosion in rheumatoid arthritis and periodontitis [PMID: 25896020].
* Cathepsin K mRNA is abundantly expressed in giant cell tumor (GCT) of bone implicating cathepsin K as the principal protease in GCT [PMID: 15277232].
* CTSK was up-regulated in white adipose tissue (WAT) of overweight/obese subjects, and it had a significant positive correlation with body mass index [PMID: 16912123].

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

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

* C60 fullerene [PMID: 20471445]
* benzo[a]pyrene diol epoxide I [PMID: 20382639]
* carbon nanotube [PMID: 25554681, PMID: 24911292, PMID: 19836432]
* paraquat [PMID: 26345256, PMID: 32680482]
* perfluorooctane-1-sulfonic acid [PMID: 22237054]
* quartz [PMID: 19836432]
* serpentine asbestos [PMID: 16251409]
* silicon dioxide [PMID: 22431001, PMID: 26345256]
* titanium dioxide [PMID: 23409001, PMID: 23557971, PMID: 27760801]
* tremolite asbestos [PMID: 29279043]

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

* chloropicrin [PMID: 28476498]

# 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:

* Tumor Cell Invasion [PMID: 10693863, PMID: 16025436, PMID: 23152410, PMID: 28117540, PMID: 28216213]
* Neoplasms [PMID: 10923921, PMID: 16912171, PMID: 17683065, PMID: 18053985, PMID: 21146373]
* Neoplasm Metastasis [PMID: 17683065, PMID: 25249554, PMID: 27709599, PMID: 31112710]
* Rheumatoid Arthritis [PMID: 10693863, PMID: 11733367, PMID: 11920402, PMID: 29431119]
* Breast Carcinoma [PMID: 25249554, PMID: 28463564, PMID: 31112710]