

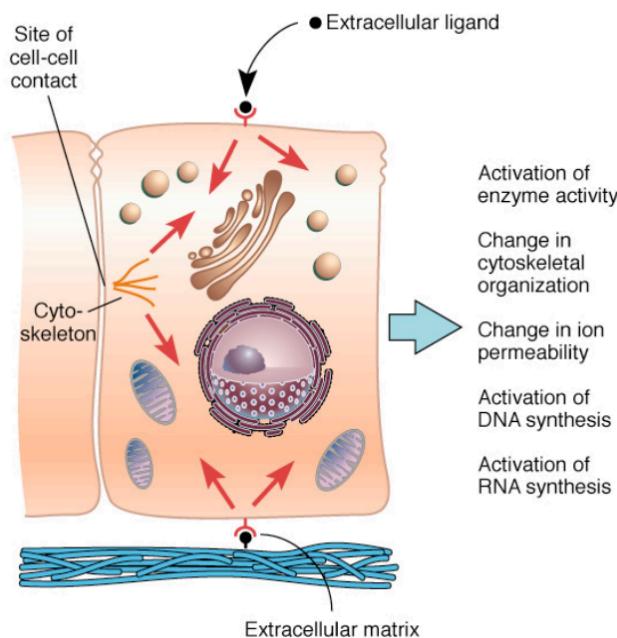
<https://www.sinobiological.com/research/signal-transduction/metabolism>

In the Avatar movie, the concept of "signal transduction" is used as a scientific explanation for how the plants on Pandora, and the Na'vi, communicate and interact, with Sigourney Weaver's character, Grace, using the term to describe this process.



Cell Signaling Overview:

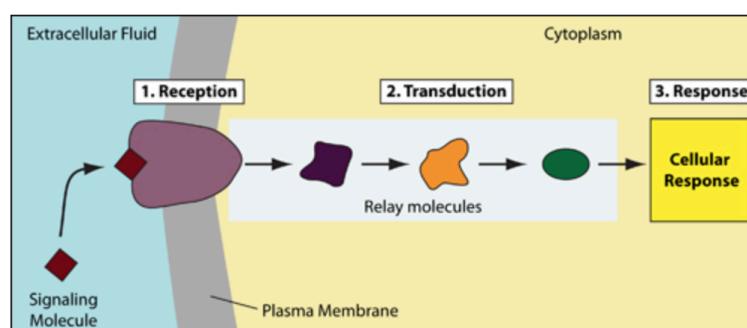
Signal transduction is how cells receive and respond to signals from their environment. These signals can come in the form of chemicals from nearby cells (paracrine), distant cells (endocrine), or even the same cell (autocrine). These signals often affect cell metabolism or gene expression.



Cell signaling is how cells detect and respond to signals from their environment. These signals can come from other cells or even from the same cell. These signals often affect cell processes like metabolism and gene expression.

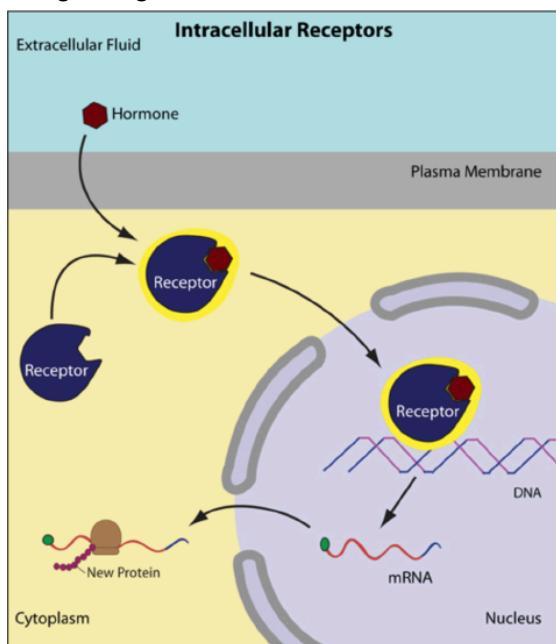
Stages of Cell Signaling:

1. Reception: A signal molecule (ligand) binds to a receptor on or inside the cell.
2. Transduction: The receptor changes shape, triggering a series of reactions inside the cell.
3. Response: The cell reacts to the signal, such as changing gene expression or metabolism.



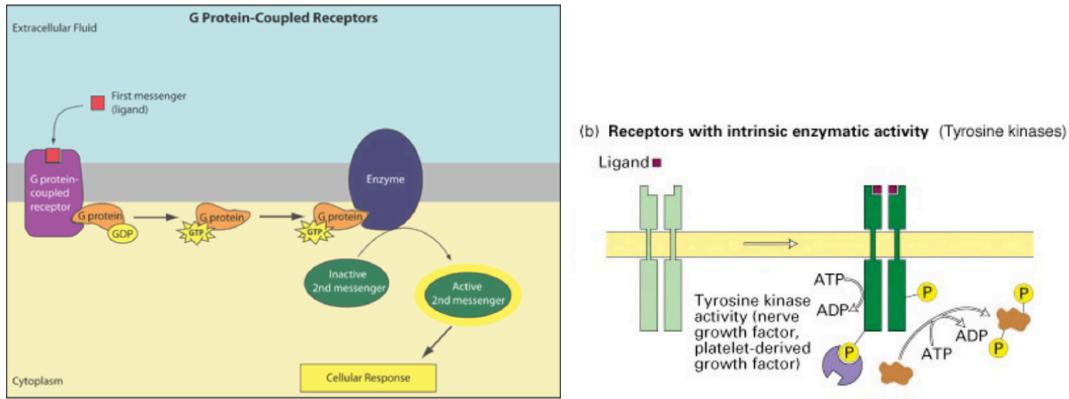
Key Components of Cell Signaling:

1. Ligands (Signals): Chemical signals like hormones or growth factors can come from nearby (paracrine), distant (endocrine), or the same cell (autocrine).
2. Receptors: These proteins bind to signals and trigger responses. Receptors can be on the cell surface or inside the cell.
3. Signaling Specificity: Different cells have different types of receptors, and signaling can be amplified through receptor clustering.
4. Second Messengers: Molecules like cAMP or calcium (Ca^{2+}) spread the signal inside the cell and help trigger responses like enzyme activation.
5. Signal Transducers: Proteins that relay the signal inside the cell, including G proteins and kinases.
6. Transcription Factors: These are proteins that regulate gene expression in response to signaling.



Types of Receptors:

- GPCRs: These are the largest group of receptors, involved in processes like vision and hormone responses.
- Ion Channels & RTKs: Receptors that open ion channels or trigger signaling pathways via phosphorylation.



Complexity of Signaling:

- Pathway Interactions: Different signaling pathways can cross-talk or interact to produce more diverse responses, such as using feedback mechanisms to either amplify or reduce the signal.
- Post-Translational Modifications: After proteins are made, modifications like phosphorylation can change their function, allowing for quick responses.
- Compartmentalized Signaling: Different cell compartments (e.g., mitochondria, nucleus) can carry out distinct signaling processes, ensuring proper function.

Overall, cell signaling involves a sophisticated system of receptors, messengers, and pathways that allow cells to respond to various signals and maintain proper function.

Translational Value of Understanding Signal Transduction:

Understanding how cells communicate through signaling pathways is key to developing drugs for diseases like cancer. When these pathways go wrong, they can cause diseases, so drugs that target signaling components can help fix abnormal signaling. For instance, drugs targeting the Ras pathway are being developed to treat cancer, and inhibitors of proteins like mTOR help treat drug-resistant cancers. Knowing these pathways helps create better, more effective drugs with fewer side effects.

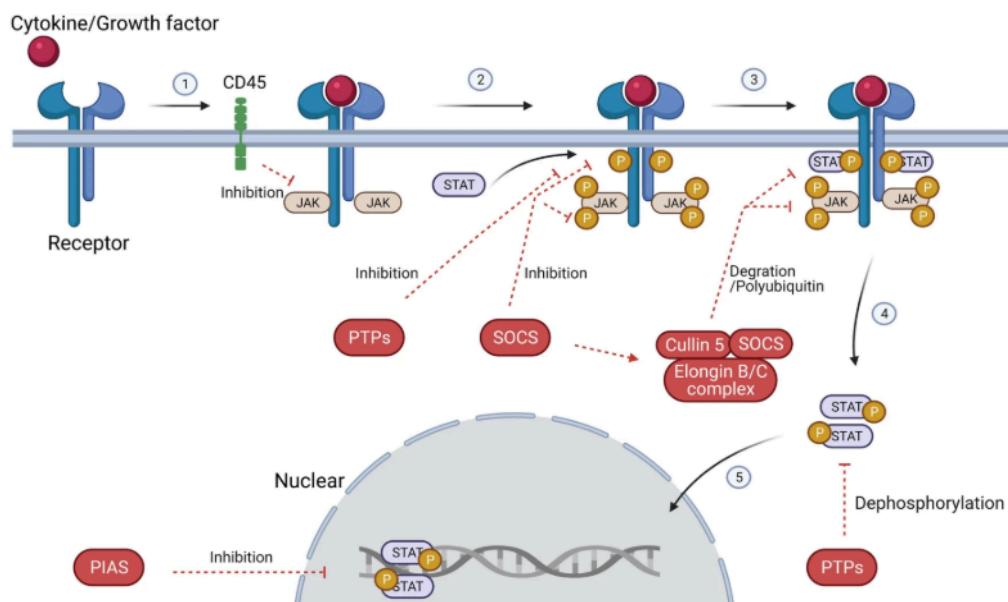
Cell signaling started simple in unicellular organisms but became more complex in multicellular organisms. Over time, more proteins and genes were added, leading to different signaling networks. For example, signaling proteins in animals differ from those in plants, creating diverse responses to the same signals. Some signaling pathways are critical for development and diversity in multicellular organisms, like the receptor tyrosine kinases (RTKs), which evolved before multicellularity. These signaling pathways are shaped by evolution and natural selection, allowing organisms to adapt. Even though signaling pathways are complex, some are conserved and essential for cell function, like RTKs, JAK/STAT, and Wnt pathways.

Intracellular Signal Transduction:

When a signal reaches a receptor on the cell's surface, it triggers a series of reactions inside the cell, often amplifying the signal. This process often involves enzymes like protein kinases that regulate gene expression and cellular responses. Key second messengers like cAMP and cGMP activate protein kinases, which modify other proteins to influence cellular functions.

For example:

- **cAMP** activates protein kinase A (PKA), which regulates processes like glycogen breakdown.
- **cGMP** is involved in blood vessel dilation and vision.
- Phospholipids, like PIP2, produce second messengers (like DAG and IP3) that activate protein kinase C and release calcium ions, impacting metabolism, survival, and gene expression.



Activation and negative regulation of JAK/STAT signaling pathways. Black arrows indicate the activation process. Red dotted arrows indicated negative regulation. Activation of the JAK/STAT signaling pathway: (1) cytokines and growth factors bind to their corresponding receptors, leading to receptor dimerization and recruitment of related JAKs; (2) JAK activation leads to tyrosine phosphorylation of the receptors and formation of docking sites for STAT; (3) STATs are phosphorylated by tyrosine; (4) STATs dissociate from the receptor to form homodimers or heterodimers; (5) STAT dimers enter the nucleus, bind to DNA, and regulate transcription. Negative regulation of the JAK/STAT signaling pathway: There are three main types of proteins involved in the negative regulation of the JAK/STAT signaling pathway: the PIAS (protein inhibitor of activated STAT), CIS/SOCS (suppressor of cytokine signaling) family, and PTPs (protein tyrosine phosphatases). PIAS mainly interacts with STAT dimers to inhibit STAT binding to DNA, thereby blocking JAK/STAT signal transduction. The CIS/SOCS family negatively regulates the JAK/STAT pathway in three ways: (1) binding to a tyrosine kinase receptor to block the recruitment of STAT; (2) binding directly to JAK to inhibit its kinase activity; (3) forming an elongin B/C-cullin5 complex that degrades JAK or STAT bound to the SOCS protein through polyubiquitination and proteasome degradation. PTPs inhibit the JAK/STAT pathway by interacting with JAK, STAT, or receptors to (1) dephosphorylate the STAT dimer; (2) interact with the receptor to dephosphorylate the related JAK; and (3) in the case of CD45 (a transmembrane PTP) inhibits the phosphorylation of JAK. Created with BioRender.com

The **ERK MAP kinase pathway**, activated by growth factor receptors, controls cell growth and survival. It begins when Ras, a small GTP-binding protein, is activated by receptor tyrosine kinases, which leads to a cascade of reactions affecting gene expression. Mutations in Ras can lead to uncontrolled cell growth, contributing to cancer.

Other pathways, like **JNK** and **p38**, respond to stress and inflammation. The **JAK/STAT** pathway directly connects receptor activation to gene transcription, playing a key role in immune responses, inflammation, and cancer development.

JAK-STAT Pathway: This pathway is vital for immune cell signaling and controlling inflammation. **Problems with JAK-STAT signaling are linked to diseases like autoimmune disorders, cancer, COVID-19, and neurodegeneration.** JAK inhibitors are being tested as treatments to reduce inflammation and abnormal cell growth in these conditions.

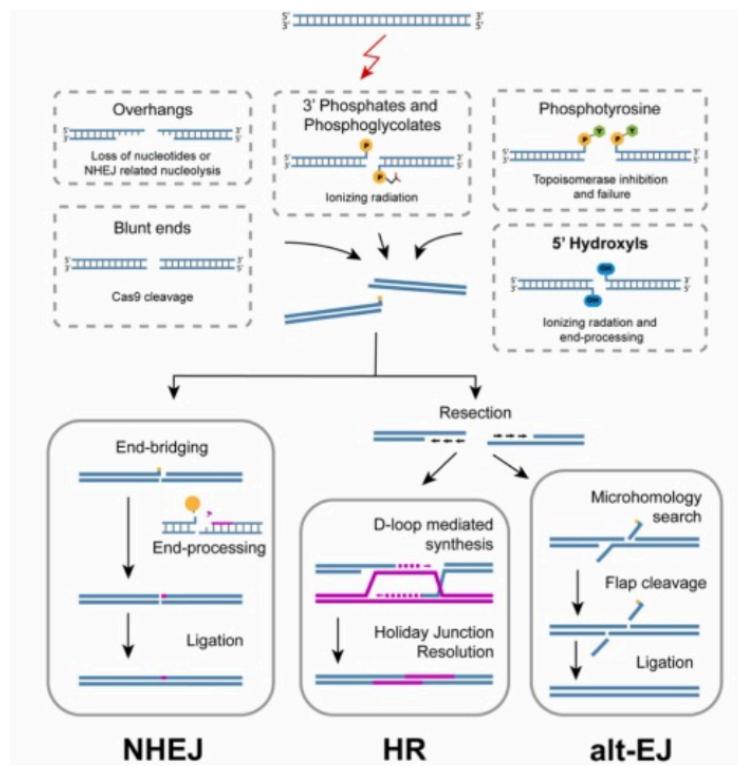
By understanding these pathways, we can develop drugs to target these specific signaling mechanisms and treat diseases more effectively.

<https://www.kegg.jp/pathway/map05171+K11217> Covid 19 & pathways involved Go down to get the genes involved, their inhibitors, targets: <https://www.kegg.jp/entry/hsa:3716>

KEGG ORTHOLOGY: K11217

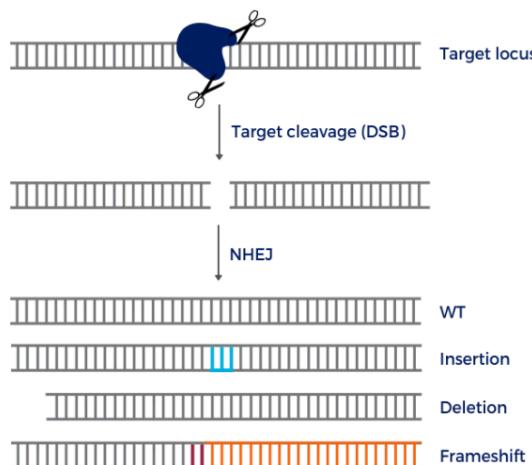
[Help](#)

Entry	K11217	KO
Symbol	JAK1	
Name	Janus kinase 1 [EC:2.7.10.2]	
Pathway	map01521	EGFR tyrosine kinase inhibitor resistance
	map04151	PI3K-Akt signaling pathway
	map04217	Necroptosis
	map04380	Osteoclast differentiation
	map04550	Signaling pathways regulating pluripotency of stem cells
	map04620	Toll-like receptor signaling pathway
	map04621	NOD-like receptor signaling pathway
	map04630	JAK-STAT signaling pathway
	map04658	Th1 and Th2 cell differentiation
	map04659	Th17 cell differentiation
	map05140	Leishmaniasis
	map05145	Toxoplasmosis
	map05152	Tuberculosis
	map05160	Hepatitis C
	map05161	Hepatitis B
	map05162	Measles
	map05163	Human cytomegalovirus infection
	map05164	Influenza A
	map05165	Human papillomavirus infection
	map05166	Human T-cell leukemia virus 1 infection
	map05167	Kaposi sarcoma-associated herpesvirus infection
	map05168	Herpes simplex virus 1 infection
	map05169	Epstein-Barr virus infection
	map05171	Coronavirus disease - COVID-19
	map05200	Pathways in cancer
	map05203	Viral carcinogenesis
	map05212	Pancreatic cancer
	map05235	PD-L1 expression and PD-1 checkpoint pathway in cancer



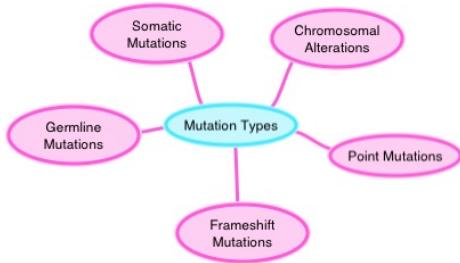
DNA double-strand breaks (deDSBs) are harmful because they can cause chromosomal rearrangements. These breaks are repaired by two main methods: homologous recombination (HR) and non-homologous end joining (NHEJ). HR requires a matching DNA sequence and begins by trimming the broken DNA ends to expose single-stranded DNA (ssDNA). On the other hand, NHEJ repairs the breaks by directly joining the broken DNA ends with little or no trimming, and involves factors like KU, 53BP1, and RIF1, which prevent the trimming process.

When there's minimal trimming of DNA ends, an error-prone repair method called alternative end-joining (Alt-EJ) may take place, using microhomologies to join the ends. During DNA replication, HR is typically used for error-free repair, but NHEJ is also active in certain cases to support cell division, especially under replication stress. NHEJ factors are found at replication forks, helping to repair DNA breaks or restart stalled replication.



CRISPR mechanism to induce loss-of-function gene mutation. The Cas9-mediated induction of double-strand break (DSB) activates endogenous mechanisms of DNA repair, aiming to fix the genomic damage produced at the locus of interest. These error-prone mechanisms can induce different outcomes in the target locus.

Types of Mutations



What Causes Albinism?

Albinism is caused by a mutation in the gene responsible for producing melanin, a protein that gives color to the skin and eyes. This mutation can result in little or no melanin production.

Mutations

A mutation is a change in the DNA or RNA sequence. While mutations are often linked to science fiction, they happen naturally in everyone. Most people have many mutations, and they are essential for evolution as they create new genetic variations. Although most mutations have no impact, some can be beneficial, and even harmful mutations rarely cause drastic changes.

Types of Mutations

- **Germline Mutations:** These occur in reproductive cells (sperm or eggs). They are significant because they can be passed on to offspring, affecting every cell in the next generation.
- **Somatic Mutations:** These happen in other body cells and are not passed to offspring. They usually have little effect, as they only affect the individual where they occur.

Chromosomal Alterations

These mutations change the structure of chromosomes, often when a part of a chromosome breaks off and reattaches incorrectly. They can be serious and may result in the death of an organism or lead to abnormalities, like Down Syndrome, which is caused by a chromosomal duplication.

Point Mutations

A point mutation changes a single nucleotide in the DNA sequence. These mutations are typically less serious than chromosomal alterations. There are three types of point mutations:

- **Silent Mutation:** No effect; the change doesn't alter the protein.
- **Missense Mutation:** Changes one amino acid, which may or may not affect the protein's function.

- **Nonsense Mutation:** A change that creates a stop codon, usually leading to a nonfunctional protein.

Type	Description	Example	Effect
Silent	mutated codon codes for the same amino acid	CAA (glutamine) → CAG (glutamine)	none
Missense	mutated codon codes for a different amino acid	CAA (glutamine) → CCA (proline)	variable
Nonsense	mutated codon is a premature stop codon	CAA (glutamine) → UAA (stop)	usually serious

Frameshift Mutations

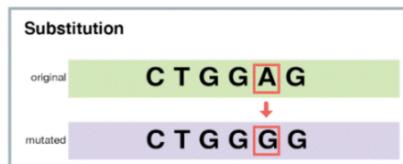
A frameshift mutation occurs when nucleotides are added or removed from the DNA sequence, shifting the reading frame. This alters how the codons are read and can drastically change the protein produced.

- Germline mutations affect reproductive cells and can be passed to offspring.
- Somatic mutations affect body cells and cannot be inherited.
- Chromosomal alterations change chromosome structure and can cause severe effects.
- Point mutations alter a single nucleotide, which can have varying effects.
- Frameshift mutations change the reading frame of the sequence, leading to major changes in the protein.

DNA can change in several ways, leading to different types of mutations. Here's a quick overview:

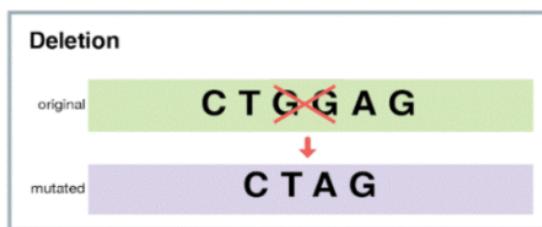
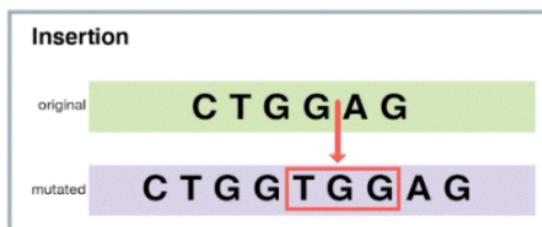
1. **Substitution:** One base is replaced with another (e.g., changing A to G). This can:

- Change a codon to code for a different amino acid, altering the protein (e.g., sickle cell anemia).
- Change a codon to one that still codes for the same amino acid, causing no change (silent mutation).
- Change a codon to a “stop” codon, leading to an incomplete protein that may not work.

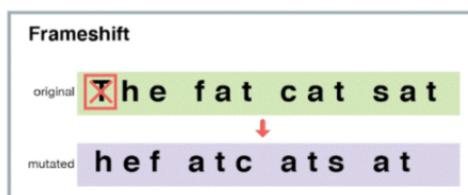


2. **Insertion:** Extra base pairs are added into the DNA sequence.

Deletion: A section of DNA is lost or removed.



3. **Frameshift:** Insertions or deletions can shift the reading of the gene, causing the codons to be read incorrectly. This often results in a non-functional or incomplete protein.

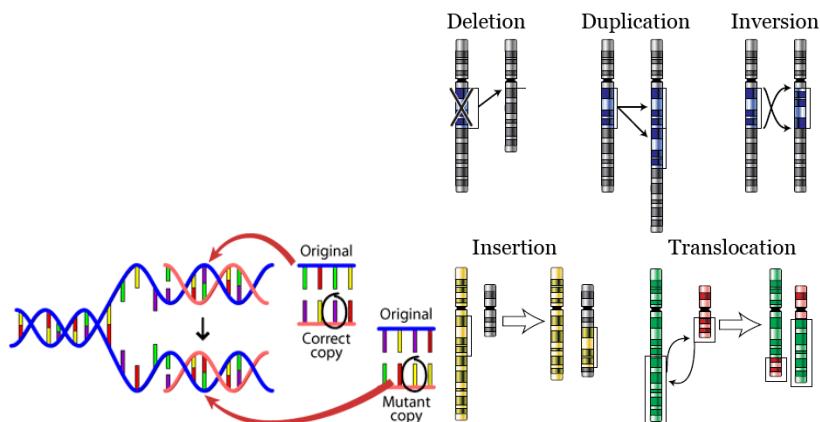


AUG-AAU-ACG-GCU = start-asparagine-threonine-alanine

Now, assume an insertion occurs in this sequence. Let's say an **A** nucleotide is inserted after the start codon **AUG**:

AUG-AAA-UAC-GGC-U = start-lysine-tyrosine-glycine

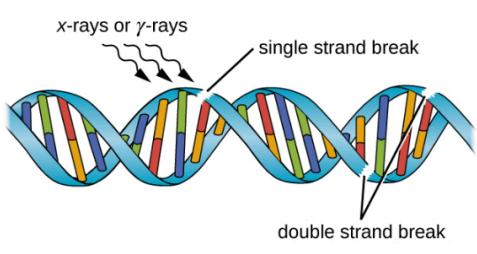
While there are other types of mutations, these are the main ones to understand.



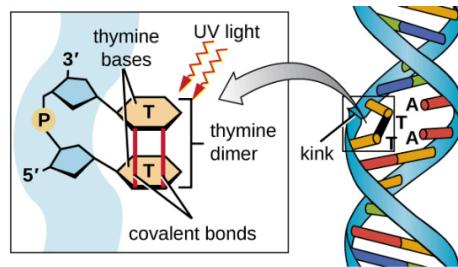
Causes of Mutations

Mutations can happen for several reasons:

- DNA copying errors:** When a cell divides, it copies its DNA, but sometimes the copy is imperfect, leading to a mutation.
- External influences:** Exposure to chemicals or radiation can damage DNA. The cell tries to fix this, but the repair might not be perfect, causing a mutation.



(a) Ionizing radiation



(b) Non-ionizing radiation

Effects of Mutations

- Somatic mutations:** These occur in non-reproductive cells and do not get passed to offspring. For example, a mutation causing a tulip's petal to have two colors won't affect the plant's seeds. Somatic mutations can also cause diseases like cancer.

Germ line mutations: These happen in reproductive cells (eggs and sperm) and can be passed to offspring, affecting evolution.

Google most researched disease

 PharmaVoice
https://www.pharmavoice.com › news › breast-cancer-p... ::

Breast cancer tops list of most studied diseases

24 Jan 2024 — Breast cancer tops list of most studied diseases. The most studied diseases last year included three types of cancer as research into COVID-19 ...

In 2023, **breast cancer** was the most studied disease for the third year in a row, followed by solid tumors and prostate cancer. Cancer research remained a top priority, fueled by rising demand for new treatments. **Breast cancer alone has 94 drugs in development, despite declining death rates.** Overall, oncology trials are growing, with over 1,600 treatments and vaccines in development. Meanwhile, COVID-19 research declined sharply post-pandemic, while stroke research rose in importance due to increasing death rates. Though trial activity is recovering from the pandemic, lingering impacts may slow drug development and raise R&D costs through 2025.

<https://www.cancer.gov/search/results?swKeyword=breast+cancer+genes+involved>

Results for: breast cancer genes involved

Results 1–20 of 20911 for: breast cancer genes involved

Genetics of Breast and Gynecologic Cancers (PDQ®)

Genetics of Breast and Gynecologic Cancers includes the hereditary cancer syndromes BRCA1/BRCA2 (hereditary breast and ovarian cancer), syndrome, ATM, PALB2, CHEK2 and other genes. Get comprehensive information on these syndromes in this clinician summary.

<https://www.cancer.gov/types/breast/hp/breast-ovarian-genetics-pdq>

A family history of these cancers, especially in multiple relatives, can suggest a genetic risk, often inherited in an autosomal dominant pattern. **BRCA1** and **BRCA2** are the main genes linked to high breast and ovarian cancer risk, while Lynch syndrome genes are linked to endometrial cancer. Screening and risk-reducing surgeries like mastectomy or ovary removal can lower cancer risk. Genetic testing decisions are influenced by emotional, psychological, and family factors. While testing often causes little long-term distress, communication and decision-making remain complex in high-risk families.

Breast Cancer Risk Factors:

Risk factors can apply to all women but may have a **stronger or different effect** in women with high genetic risk (e.g., BRCA mutation carriers).

Key Risk Factors:

Age: Risk increases with age, especially after 50. Women with genetic mutations may develop it earlier.

Family History:

- Having a mother, sister, or multiple relatives with breast cancer increases risk. The risk is even higher if relatives were diagnosed **young**, had **bilateral breast cancer**, or were **male**. Studies show a **higher chance** of developing cancer when multiple relatives are affected.
- 2. **Benign Breast Disease (BBD):** Non-cancerous changes in breast tissue can raise risk, especially in women already at high genetic risk.
- 3. **Breast Density:** Women with **dense breast tissue** (as seen on mammograms) have a higher risk. This holds true even for **BRCA1/BRCA2 carriers**.
- 4. **Background Parenchymal Enhancement (BPE):** Like breast density, **BPE seen on MRI** can also signal higher risk.

Contraceptives and Breast Cancer Risk

Contraceptives help with birth control and other health issues like irregular bleeding. Some studies show that **long-term use of oral contraceptives (OCs)** may slightly raise the risk of breast cancer, but this seems to be temporary.

- Women with a **family history of breast cancer** or those with **BRCA1/BRCA2 gene mutations** may not face higher risk from using OCs, based on some studies.
- However, for **BRCA2 carriers**, using OCs may increase the risk (especially in certain studies).
- **Older types of OCs** (before 1975) and starting use **before age 20** may increase risk more in BRCA1/2 carriers.
- Other contraceptive methods, like **hormonal IUDs (LNG-IUS)**, may slightly increase breast cancer risk in the general population, but studies are not yet clear for high-risk women.

<https://www.genome.jp/entry/hsa:672> and <https://www.kegg.jp/pathway/hsa05224+672>

Entry	K10605	KO
Symbol	BRCA1	
Name	breast cancer type 1 susceptibility protein [EC:2.3.2.27]	
Pathway	map01524 Platinum drug resistance map03440 Homologous recombination map03460 Fanconi anemia pathway map04120 Ubiquitin mediated proteolysis map04151 PI3K-Akt signaling pathway map05206 MicroRNAs in cancer map05224 Breast cancer	
Disease	H00019 Pancreatic cancer H00027 Ovarian cancer H00031 Breast cancer H00238 Fanconi anemia H01554 Fallopian tube cancer H01665 Primary peritoneal carcinoma H02531 Familial breast-ovarian cancer	

What are BRCA genes?

Everyone has BRCA1 and BRCA2 genes. These genes help stop cells from growing too fast. But if there's a change (mutation) in them, they may not work properly—this can lead to cancer.

Do BRCA gene mutations mean you will get cancer?

No. A mutation increases your *risk* but doesn't mean you *will* get cancer. It can make it harder for your cells to repair damage, which raises the chances of cancer over time.

Why BRCA Matters for Pancreatic Cancer

BRCA mutations are well known in breast and ovarian cancers. But they're also linked to a higher risk of pancreatic cancer—especially BRCA2.

Gene breast cancer

Create RSS Save search Advanced

Tabular 20 per page Sort by Relevance

Send to:

Search results

Items: 1 to 20 of 28305 << First < Prev Page 1 of 1416 Next > Last >>

 See also 573 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> BRCA2 ID: 675	BRCA2 DNA repair associated [<i>Homo sapiens</i> (human)]	Chromosome 13, NC_000013.11 (32315077..32400268)	BRCC2, BROVCA2, FACD, FAD, FAD1, FANCD, FANCD1, GLM3, PNCA2, XRCC11	600185
<input type="checkbox"/> BRCA1 ID: 672	BRCA1 DNA repair associated [<i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (43044295..43170327, complement)	BRCA1, BRCC1, BROVCA1, FANCS, IRIS, PNCA4, PPP1R53, PSCP, RNF53	113705

BRCA1: <https://www.ncbi.nlm.nih.gov/gene/672>

Role of BRCA1 Gene

The BRCA1 gene makes a large protein found in the nucleus of cells (about 190 kD in size). This protein is essential for keeping our DNA stable and protects cells from turning cancerous, making it a tumor suppressor.

Key Functions:

- DNA Repair: It helps fix double-stranded breaks in DNA.
- Gene Regulation: It interacts with RNA polymerase II and histone deacetylases, influencing gene transcription.
- Cell Surveillance: Forms a complex (called BASC) with other proteins to monitor and respond to DNA damage.

Structure:

- The gene spans ~110,000 base pairs and has 22 exons. The C-terminal part of the protein is important for interacting with other proteins involved in gene expression and repair.

Mutations and Cancer:

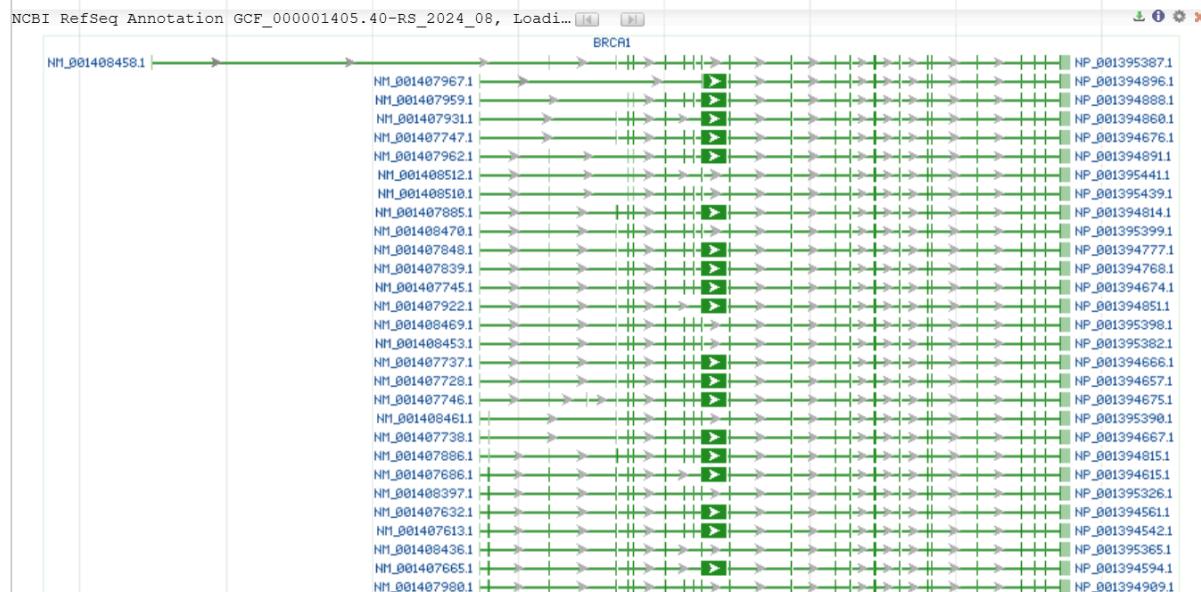
- Faulty BRCA1 can't repair DNA properly, leading to genomic instability. ~40% of inherited breast cancers and >80% of inherited breast + ovarian cancers are linked to mutations in BRCA1.

Variants:

- The gene undergoes alternative splicing, producing many versions of the BRCA1 protein. Some variants are linked to disease, but not all are fully understood yet.

There is also a pseudogene (non-functional copy) related to BRCA1 on the same chromosome (chromosome 17).

genomic neighborhood around the BRCA1 gene



Downloaded the neighborhood

NCBI RefSeq Annotation GCF_000001405.40-RS_2024_08						
Accession	Start	Stop	Gene sym	Strand	NCBI Gene ID	Name
NC_000017.11	43025231	43032041	RND2	plus	8153	Rho family GTPase 2
NC_000017.11	43044295	43170327	BRCA1	minus	672	BRCA1 DNA repair associated
NC_000017.11	43079261	43079816	RPL21P4	plus	140660	ribosomal protein L21 pseudogene 4
NC_000017.11	43125557	43153671	NBR2	plus	10230	neighbor of BRCA1 IncRNA 2
NC_000017.11	43144845	43145323	HMGN1P2	plus	100885865	high mobility group nucleosome binding domain 1 pseudogene 29
NC_000017.11	43148366	43170403	LOC10192	minus	101929767	uncharacterized LOC101929767
NC_000017.11	43160288	43170964	LOC12490	minus	124900391	uncharacterized LOC124900391
NC_000017.11	43168070	43169954	BRCA1P1	minus	394269	BRCA1 pseudogene 1
NC_000017.11	43170409	43211688	NBR1	plus	4077	NBR1 autophagy cargo receptor

Are there variants reported in this gene?

Molecular consequence clear

Search results

Display options ▾ Sort by Location ▾ Download ▾ Items: 47

Filters activated: Missense, Indel. Clear all to show 15252 items.

Variation	Gene (Protein Change)	Type (Consequence)	Condition	Classification, Review status
<input type="checkbox"/> NM_007294.4(BRCA1):c.5529_553_0delinsCA (p.Leu184Ile)	BRCA1 (L1865+80 more)	Indel (missense variant +2 more)	Heredity cancer-predisposing syndrome +2 more	 Conflicting classifications of pathogenicity ★
<input type="checkbox"/> NM_007294.4(BRCA1):c.5510_551_1delinsCT (p.Trp1837Ter)	BRCA1 (W1790S+80 more)	Indel (missense variant +2 more)	not provided +1 more	 Likely pathogenic ★★
<input type="checkbox"/> NM_007294.4(BRCA1):c.5488_548_9delinsAG (p.Ala1830Arg)	BRCA1 (A1851R+80 more)	Indel (missense variant +2 more)	Heredity breast ovarian cancer syndrome	 Uncertain significance ★
<input type="checkbox"/> NM_007294.4(BRCA1):c.5359_536_3delinsAGTGA (p.Cys1787_Gly1788delinsSerAsp)	BRCA1	Indel (missense variant +2 more)	Breast-ovarian cancer, familial, susceptibility to, 1	 Pathogenic ★★★
<input type="checkbox"/> NM_007294.4(BRCA1):c.5354_535_0del	BRCA1	Indel	Heredity cancer-predisposing	 Uncertain significance ★

<https://www.ncbi.nlm.nih.gov/clinvar/?term=%22BRCA1%22%5BGENE%5D&redir=gene>

Summary of the Numbers (Impact Level Estimate):

Category	Number of Variants	Likely Impact
High Impact: Frameshift, Nonsense, Splice site	~4,000	High pathogenic potential
Moderate Impact: Missense	~6,000+	Variable impact – requires functional assessment
Regulatory Impact: UTR, ncRNA	~5,500	Can affect gene expression
Near gene	0	None found outside gene body in this dataset

Filtered the SNPs based on their IDs

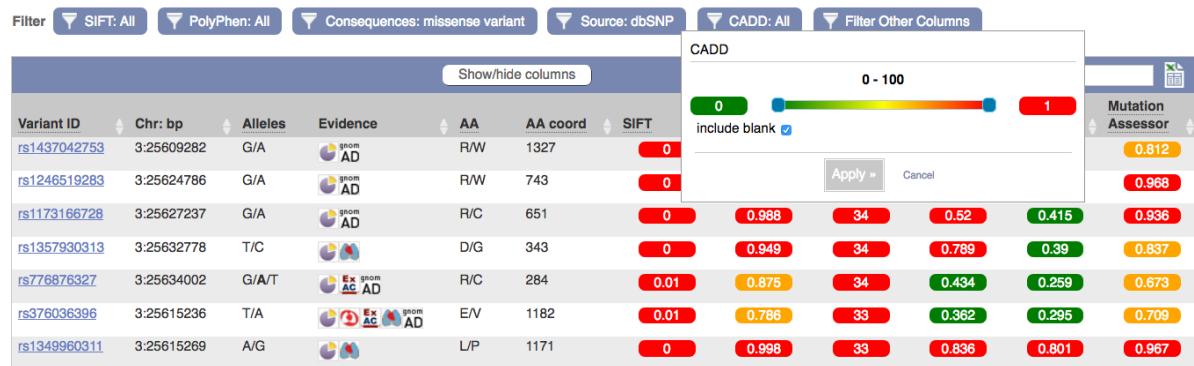
<https://www.ncbi.nlm.nih.gov/snp/>

Variant Effect Predictor - Upload the list of SNP IDs

https://asia.ensembl.org/Homo_sapiens/Tools/VEP/Results?tl=d1PkxYXPnDhrGBQR-11107125

SIFT/Polyphen

SIFT (Sorting Intolerant From Tolerant) and PolyPhen (Polymorphism Phenotyping) are tools used to **predict the potential impact of amino acid substitutions on protein function**. SIFT assesses sequence conservation and amino acid properties to classify variants as tolerated or deleterious—with scores below 0.05 indicating likely functional impact. PolyPhen-2 uses multiple features including sequence homology, protein 3D structure, and functional annotations to categorize variants as benign, possibly damaging, or probably damaging, where scores closer to 1 suggest higher probability of damage. Additional tools like MetaLR (Meta Logistic Regression) and MutationAssessor offer further predictions by integrating various scores or using evolutionary conservation. Variability in predictions is expected due to differences in methods and databases used by each tool.



Polyphen value	Qualitative prediction	Website display example
greater than 0.908	"Probably Damaging"	0.95
greater than 0.446 and less than or equal to 0.908	"Possibly Damaging"	0.5
less than or equal to 0.446	"Benign"	0.25
unknown	"Unknown"	unknown

SIFT value	Qualitative prediction	Website display example
Less than 0.05	"Deleterious"	0.01
	"Deleterious - low confidence"	0.01
Greater than or equal to 0.05	"Tolerated"	0.8
	"Tolerated - low confidence"	0.8

🧬 Summary: From Genetic Variant to Protein Structure Comparison using RMSD

1. Fetching Exon Sequence:

- Retrieved the **nucleotide sequence** of the target **exon** from the **Ensembl Genome Browser**.
- This served as the **reference (wild-type)** coding sequence.
- [https://asia.ensembl.org/Multi/Search/Results?q=ENSG0000012048%20\(BRCA1\);site=ensembl;page=1;facet_feature_type=Transcript](https://asia.ensembl.org/Multi/Search/Results?q=ENSG0000012048%20(BRCA1);site=ensembl;page=1;facet_feature_type=Transcript)

2. Generating Variant Sequences:

- Incorporated **specific SNP variants (rsIDs)** into the exon sequence.

- Created **multiple nucleotide variant sequences**, each reflecting a different **single-nucleotide change**.

3. Translating to Protein: Python

- Each modified nucleotide sequence was **translated into the corresponding protein sequence**.
- Ensured frame and codon integrity to avoid mis-translation.

4. Protein Structure Prediction:

- Used **AlphaFold2** to predict the **3D structure (PDB format)** for each protein variant.
- Ensured that all predictions were generated using consistent model parameters (e.g., relaxed/unrelaxed, seed, model rank).

5. Extracting Key Residues:

- Selected a **26-residue region of interest** from each PDB file, believed to be functionally or structurally important.

6. Superimposing Structures: Alphafold 2

- Alignment was based specifically on **C_α atoms** of the 26-residue segment to ensure uniformity.

7. Calculating RMSD:

- Computed **Root Mean Square Deviation (RMSD)** values to measure structural differences.
- Observed **very low RMSD values (mostly <0.1 Å)**, suggesting **high structural similarity** across variants.

8. Interpretation:

- Low RMSD values indicate **minimal structural impact** due to the variants.
- Functional effects, if any, would depend on whether the altered residues lie in **critical regions** (e.g., active sites, binding pockets), not just overall structure.

https://docs.google.com/spreadsheets/d/1AOeAo5Gfh8ob6bkQHkUsexp6bho5IB7mZTbrf_AQ_bk/edit?gid=0#gid=0

<https://drive.google.com/drive/folders/1sZb4p4j4JMQBVzn9EkJTfQ9MoMSPB3mj>

https://colab.research.google.com/drive/12_ZKi0CSVCyTd3-YvRlth3HOL_jevGvD

https://colab.research.google.com/drive/1CxwtSDlcfdZgeWXvlRbeVSL9ldzE7a22?usp=drive_link