Executive summary

- Spleen tyrosine kinase (SYK) plays critical roles in B-cell and T-cell development, the maintenance of vascular integrity, and proper partitioning of the blood vascular and lymphatic vascular system.
- SYK is a highly conserved gene across mammals and its orthologs have been characterized in dog, mouse, rat, and pig.
- There are two major protein isoforms of SYK, a longer canonical isoform SYK(L) of 635 aa and a shorter isoform SYK(S) of 615 aa.
- SYK(L) is expressed in hematopoietic cells as well as in non-hematopoietic cells. The expression of SYK(S) in normal tissues is uncommon.
- 873 polymorphisms are reported for SYK, of which 15 are located in the coding region. SYK polymorphisms are not likely to be associated with any disease.
- There are no reports of protein damaging or disease causing somatic mutations in SYK. Hypermethylation of SYK is reported in many cancers. Large insertions/deletions within the SYK locus have caused exon skipping leading to generation of splice variants of SYK.
- The canonical SYK transcript has two potential transcription start sites. A significant increase in SYK expression was observed in breast cancer mediated by p53 acetylation of the SYK promoter.
- SYK and ZAP-70 are members of the Syk family of cytosolic protein kinases implicated in antigen and Fc receptor signaling.
- The SYK protein contains two tandem SH2 domains and a tyrosine kinase domain. There are eleven autophosphorylation sites of which five are in the inter domain linking the SH2 domains and the kinase domain, rendering this inter domain crucial for immunomodulatory signal regulation by SYK.
- Aberrant expression of SYK has been reported in a large number of cancers, allergic disorders, antibody mediated immune diseases, rheumatoid arthritis, asthma and neurodegenerative diseases.
3. Alternative transcription of SYK

Sources of information: UniGene, Entrez, Ensembl, Aceview, ASTD, ECGene, Genome Channel, NCBI-PubMed, GoogleScholar

- The canonical SYK(L) transcript NM_003177 encoding an isoform of 635 aa and the major alternate SYK(S) transcript NM_001135052 encoding the 615 aa isoform are represented in RefSeq, Uniprot, multiple databases and literature.
- The SYK(L) differs from the SYK(S) variant due to the presence of a 23 aa sequence in the interdomain-B region. This is responsible for the nuclear sub-cellular distribution of SYK(L) and contributes to some of the functional differences between the two variants.
- Additionally four transcripts have been reported in databases, with multiple clones as evidence. Two of these contain only the SH2 domain and the other two span the tyrosine kinase domain.
- SYK(L) is expressed in hematopoietic cells such as macrophages, mast cells, leukocytes, platelets and erythrocytes. In humans it is also expressed at lower levels in non-hematopoietic cells like fibroblasts, epithelial cells, neuronal cells and hepatocytes.
- The expression of SYK(S) in normal tissues is uncommon. SYK(S) was found to be expressed in the primary breast tumors, but not in the pathologically normal mammary tissues.

3A. Alternative transcripts in humans

There are two major SYK variants both containing two Src homology 2 (SH2) domains and one tyrosine kinase domain but having different interdomain-B regions. Both isoforms are reported in the RefSeq (NP_003168.2, NP_001128524.1) and Uniprot (P43405-1, P43405-2) databases as well as in literature (SykA/Syk 41, SykB/Syk 11). In addition, multiple databases report an additional 4 alternative transcripts that include only the SH2 domains or only the tyrosine kinase domain. SYK alternate transcripts are shown in figure 3.1.
Target analysis of human SYK

Table 3.1: Mapping major transcript IDs across databases

<table>
<thead>
<tr>
<th>NCBI</th>
<th>Ensembl</th>
<th>Aceview</th>
<th>ASTD</th>
<th>ECGene</th>
<th>Uniprot</th>
<th>Genome Chanel</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_003177 (Isoform 1)</td>
<td>ENST0000375754</td>
<td>aApr07</td>
<td>ENST00000326723</td>
<td>H9C8289.5</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ENST00000375746</td>
<td></td>
<td>ENST00000375746</td>
<td>H9C8289.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_001135052 (Isoform 2)</td>
<td>ENST0000037574</td>
<td>bApr07</td>
<td>ENST0000037574</td>
<td>H9C8289.11</td>
<td>P43405-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENST000000037403</td>
<td></td>
<td>TRAN000000037403</td>
<td>H9C8289.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENST00000375747</td>
<td>cApr07</td>
<td>ENST00000340029</td>
<td>H9C8289.6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENST00000375747</td>
<td></td>
<td>ENST00000375747</td>
<td>H9C8289.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENST00000375751</td>
<td>dApr07</td>
<td>ENST00000375751</td>
<td>H9C8289.9</td>
<td>P43405-2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ENST00000375751</td>
<td></td>
<td>TRAN000000037405</td>
<td>H9C8289.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Canonical long isoform NM_003177/ SykA / SYK(L) / Syk41**

The RefSeq canonical transcript NM_003177 encodes the 635 aa isoform-1 (NP_003168.2) and is identical to the Uniprot longer isoform (P43405-1) protein. Various papers have designated the longer isoform as SykA/SYK(L)/Syk41. The interdomain B (IDB) of Syk(L) contains a bipartite nuclear localization signal that is responsible for Syk(L) nuclear translocation within the DEL region. The DEL sequence (TWSAGGIISRIKSYSFPKPGHRK) contains five basic residues (R292, K294, K300, R304, and K305) that are critical for the Syk(L) nuclear translocation.

**Short isoform NM_001135052/ SykB/ SYK(S)/ Syk11**

A second major splice variant is the RefSeq transcript NM_001135052 that encodes the 612 aa isoform-2 (NP_001128524.1) is identical to the Uniprot shorter isoform (P43405-2) protein. Various publications have designated the shorter isoform as SykB/ SYK(S)/Syk11. This transcript has a deletion of 69 bp that skips exon 7 (Goodman et al 2001). The alternatively spliced in-frame transcript creates a Syk protein isoform that lacks a 23-residue DEL sequence within the 80-100 aa intervening domain IDB.

**Other Splice variants**

There are several other transcripts reported in the different databases, with multiple clones as evidence. Four of these could be potential splice variants of SYK.

- A third potential splice variant translates only the SH2 domains of SYK. This is reported by both Aceview (iApr07) and ASTD (TRAN00000037412) databases and encodes a 180 aa long predicted protein that aligns with 100% identity (aligning to amino acids 1-139) of Refseq isoform SYK (L).
Target analysis of human SYK

This alternative transcript is supported by a 5’cap (DB106658) and validated 3’polyA tail (AA872010) evidence, indicating that it could represent a stable mRNA (Aceview).

- **A fourth** potential splice variant also translates *only the SH2 domains* of SYK. This variant is reported by Aceview (eApr07), ASTD (TRAN00000037408) and ECGene (H9C8289.4) and encodes a 403 aa long predicted protein that aligns with 100% identity to the RefSeq SYK(L) (amino acids 1-305). This transcript has multiple EST evidences - CB305749, CB250400, AW086131 for 3’ polyA tail (Aceview, ECGene).

- **A fifth** possible transcript predicted by Aceview (jApr07) and ECGene (H9C8289.2) translates *only part of the protein tyrosine kinase domain of SYK*. This transcript encodes a 175 aa long predicted protein that aligns with 100% identity to the RefSeq SYK(L) (amino acids 469-635). This transcript is represented by multiple clones with evidences for 5’cap (DA994348) and 3’ polyA tail (BG397437, BU687107, AI949249).

- **A sixth** possible isoform also translates *only the protein tyrosine kinase domain* of SYK. This transcript is reported by ASTD (TRAN00000037407) and ECGene (H9C8289.2), and encodes a 293 aa long predicted protein that aligns with 100% identity to the RefSeq SYK(L) (amino acids 343-635).

The figure 3.1 below summarizes the main splice variants of SYK reported in databases. We illustrate the variants which have multiple evidences as alternatively spliced isoforms and are reported across multiple databases.

![Figure 3.1: Graphical representations of alternate transcripts of human SYK](image)

© 2010 Strand Life Sciences Pvt. Ltd.
Target analysis of human SYK

Disease associated splice variants of SYK

Goodman et al (2001) report four aberrant coding sequences in leukemic patients (pro-B leukemia) that provide evidence for a deficiency of SYK in leukemic cells from children with pro-B acute lymphoblastic leukemia. These aberrations are summarized in the table 3.2. Wang et al (2003) report that 50% of the patients, the short alternate spliced SYK variant was found in the primary breast tumors, but not in the pathologically normal mammary tissues. SYK(S) was found expressed along with SYK(L) in most of the breast cancer cell lines examined but was missing in the neighboring normal tissues. The tumor-specific expression of SYK(S) in primary breast cancers suggests its contributory roles for mammary tumor initiation or progression.

Table 3.2: SYK mRNA variants in pro-B leukemia cells

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Splicing event</th>
<th>Genetic alteration</th>
<th>Protein alteration</th>
<th>Effect on protein structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>B#1</td>
<td>Skipping of exon 13</td>
<td>113bp Del (G1870-A1982)</td>
<td>Frameshift at Arg574 of catalytic domain, leading to addition of 19 novel amino acids until stop codon</td>
<td>N-terminal 62 amino acids are replaced by 19 new ones thereby missing 3 helixes that are important for enzymatic form and structural stability</td>
</tr>
<tr>
<td>B#2</td>
<td>Insertion of exon 14</td>
<td>177bp Ins (after A1982)</td>
<td>Truncating frameshift at Tyr611 near C-terminal end of kinase domain, leading to addition of 23 novel amino acids until stop codon</td>
<td>Affects last 24 amino acids of SYK kinase domain which does not affect the kinase activity directly but may compromise structural stability</td>
</tr>
<tr>
<td>B#3</td>
<td>Skipping of exon 12</td>
<td>141bp Del (G1729-G1869)</td>
<td>Inframe deletion of 47 amino acids (ala528-Arg574) in catalytic domain</td>
<td>Missing section of C-lobe domain would likely lead to unstable protein</td>
</tr>
<tr>
<td>B#4</td>
<td>Mis-splicing at 5’end of exon 5</td>
<td>4bp Del (G865-G868)</td>
<td>Frameshift at Trp239 in second SH2 domain, leading to addition of 11 novel amino acids until stop codon; 15 buried hydrophobic residues are exposed.</td>
<td>Deletion includes last beta strand of SH2 domains and kinase domains leading to lack of kinase activity</td>
</tr>
</tbody>
</table>

3B. SYK expression in normal tissues

Yagi et al (1994) isolated 2 different cDNA clones (Syk11 and Syk41) from a cDNA library of human basophilic leukemia cell line (Jurkat). The long isoform is dominant in the Jurkat cell lines and human peripheral leukocytes. Wang et al (2006) confirmed the presence of both these isoforms at the mRNA and protein levels, in primary human airway epithelial cells and in several human pulmonary epithelial cell lines.

SYK protein-tyrosine kinase is expressed in a variety of hematopoietic cells and transmits signals from the B-Cell and T-Cell receptors in response to immunoreceptor signaling events that mediate diverse cellular functions including proliferation, differentiation, and phagocytosis. B-cell activation
Target analysis of human SYK

both in vivo and in vitro appears to increase SYK mRNA. SYK(L) is expressed in hematopoietic cells such as macrophages, mast cells, leukocytes, platelets and erythrocytes. SYK is also expressed in non-hematopoietic cells like fibroblasts, epithelial cells, breast tissue, hepatocytes, neuronal cells, and vascular endothelial cells and has been shown to be functionally important in these cell types (Yanagi et al., 2001). Increased expression of SYK in human umbilical vein epithelial cells has been reported during cell growth and serum deprivation (Inatome et al., 2001). The abundance of Syk(S) was reportedly lower than 1% of Syk(L) in lymphocyte lines examined (Law et al., 1994).

SYK isoforms show different subcellular localization patterns. Syk(L) is expressed in cytoplasm and nucleus whereas the SYK(S) is predominantly expressed in the cytoplasm (Wang et al., 2003). Sada et al (2000) reported that SykB isoform is inefficient in coupling immune receptor FcεRI aggregation to protein tyrosine phosphorylation and degranulation.

We have looked for evidences supporting the tissue of expression of the SYK canonical transcript. Unigene describes several ESTs mapping to the canonical transcript. These ESTs are derived from multiple tissues of origin from normal and pathological conditions and at various developmental stages that have been summarized in table 3.2 below.

<table>
<thead>
<tr>
<th>Normal tissue of expression</th>
<th>Expression in pathological condition</th>
<th>Developmental stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ascites, bladder, blood, bone, bone marrow, brain, cervix, connective tissue, eye, intestine, kidney, liver, lung, lymph, lymph node, mammary gland, mouth, muscle, nerve, ovary, pancreas, parathyroid, pharynx, placenta, prostate, skin, spleen, stomach, testis, thymus, thyroid, tonsil, uterus</td>
<td>bladder carcinoma, breast (mammary gland) tumor, chondrosarcoma, colorectal tumor, germ cell tumor, gastrointestinal tumor, glioma, head and neck tumor, kidney tumor, leukemia, liver tumor, lung tumor, lymphoma, non-neoplasia, ovarian tumor, prostate cancer, retinoblastoma, uterine tumor.</td>
<td>fetus, infant, juvenile, adult</td>
</tr>
</tbody>
</table>

Based on EST sources we found that SYK has multiple tissues of expression while the major splice variant (short isoform) cDNA was expressed in testis and mast cells. The unique cDNAs supporting the SYK canonical transcript and the major splice variants are summarized in table 3.3.

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>cDNA evidence</th>
<th>Tissue of expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_003177</td>
<td>DB458719.1, AK290927.1, BC011399.1, BC001645.2</td>
<td>Testis, neutrophils, Eye-retinoblastoma, Lymph- Burkitt lymphoma</td>
</tr>
<tr>
<td>NM_001135052</td>
<td>DB458719.1, Z29630.1, W33104</td>
<td>Testis, mast cells, parathyroid tumor</td>
</tr>
</tbody>
</table>
Target analysis of human SYK

In normal cells, the SYK protein was detected in freshly isolated tonsillar B-cells with different densities. SYK was also expressed in thymocytes. Higher levels of SYK protein on a per cell basis were found in low density in vivo activated B-cells compared with dense resting B cells. No SYK protein could be identified either in resting T cells or in activated T cells (Law et al., 1994). The SYK protein was strongly expressed in large and small airway epithelium, type-I pneumocytes and macrophages in human lung tissue as demonstrated using immunohistochemistry. Weaker but clearly positive staining was seen in airway smooth muscles. In the airway epithelium, the strongest staining was detected in the apical cytoplasm in ciliated columnar EC. Focal nuclear staining was also observed in the basal cells. In the type-I pneumocytes, both cytoplasmic and nuclear staining was observed. SYK was expressed in the cytoplasmic, perinuclear, and nuclear areas in PBEC and HS-24 bronchial epithelial cell lines (Ulanova et al., 2004). In situ hybridization experiments indicated high levels of SYK mRNA in normal mammary epithelial cells, lower levels in ductal carcinoma in situ, and none in invasive breast carcinoma (Coopman et al., 2002).

SYK(S) displayed intrinsic enzymatic activity comparable to that of SYK(L) but exhibited reduced ability to bind phosphorylated ITAM motif. Thus the linker domain is crucial for the ability of SYK to regulate immunomodulatory signals. The expression of Syk(S) in normal tissues is uncommon (Wang et al., 2003).

3C. Alternate transcripts of SYK in mammalian model organisms

Rowley et al, (1995) defined the longer isoform as SykB, and the shorter isoform, that typically misses 23 amino acids present in the canonical isoform, as SykA. (Note: This nomenclature is opposite to that reported in other papers). They also report that the 23 amino acid region that differentiates the 2 isoforms are exactly identical and conserved in human, rat and porcine species. This conservation is also reported in mouse, by Latour et al (1998). In addition, Latour et al (1998) showed that in comparison with SykA, the shorter isoform SykB was much less effective at mediating immunoreceptor signaling (in RBL-2H3 rat basophilic leukemia cells). They suggested that this defect seemed to be related to the reduced capacity of SykB to bind tyrosine phosphorylated ITAMs. The Syk proteins reported by Yagi et al (1994), are the naturally occurring isoforms in rats (Rowley et al, 1995) and in mice (Latour et al, 1996). Latour et al (1996) showed that ~10% of Syk RNAs in the mouse thymus, spleen, and a variety of hemopoietic cell lines were of the SykB (short isoform) configuration. They also found that SykB represented ~50% of all SYK transcripts and polypeptides present in normal mouse bone marrow cells. They also report that SykA (long isoform) predominates in B-cell lineage cells and represents a genuine isoform of SYK. It may have a predominant role in bone marrow signaling. Syk protein was detected in rat, and mouse bronchial epithelium similar to humans (Ulanova et al., 2004).
Strand Life Sciences

Strand Life Sciences (Strand) is a technology innovation company that has pioneered the practice of Scientific Intelligence in health sciences. Leveraging its unmatched interdisciplinary scientific expertise with its decision enabling technologies, Strand empowers biomedical and research scientists gain deep insights from raw data. Strand offers products and services in the areas of omics-based technologies, drug discovery, predictive systems modeling and knowledge management.

Seven of the top ten pharmaceutical companies, three of the top six biotechnology companies, and numerous academic institutions across the globe, are part of Strand’s esteemed customer profile. Strand is a portfolio company of UTI Venture Funds, Sequoia Capital, and MediBIC group (Japan).

Our Presence

INDIA
5th Floor, Kirloskar Business Park, Bellary Road, Hebbal, Bangalore 560024
T +91 80 40 STRAND (40787263), F +91 80 40787299

USA
548 Market Street, Suite 82804, San Francisco, CA 94104
T +1-650-288-4735, Toll Free +1-866-866-9326

Email: starsales@strandls.com
Web: www.strandls.com

© Strand Life Sciences
All brand names and/ or trademarks are of their respective holders.