Gene fusion in human cancer
## DISCLOSURES

<table>
<thead>
<tr>
<th>Commercial Interest</th>
<th>Relationship(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMS</td>
<td>Advisory Board</td>
</tr>
<tr>
<td>Guardant Health</td>
<td>Advisory Board</td>
</tr>
<tr>
<td>Bayer</td>
<td>Advisory Board</td>
</tr>
<tr>
<td>Sysmex</td>
<td>Consultant</td>
</tr>
</tbody>
</table>
A fusion gene is a hybrid formed from two previously separate genes. It can occur as a result of: translocation, interstitial deletion, or chromosomal inversion.
The first specific translocation identified in human neoplasia was t(9;22)(q34;q11), resulting in the Philadelphia chromosome.


The number of cytogenetically abnormal neoplasms reported in the literature has increased exponentially since

Mittelman Database:
Total number of cases 69,136
Total number of gene fusions 21,584
Total number of genes involved 12,089

How do gene fusions arise?
A fusion gene is a hybrid gene formed from two previously separate genes. It can occur as a result of: translocation, interstitial deletion, or chromosomal inversion.
Chromosomal translocations

Translocations generally result from swapping of chromosomal arms between heterologous chromosomes and most of them are reciprocal or balanced by nature.

Ex. *BCR-ABL*
Interstitial Deletion

Deletion that does not involve the terminal parts of a chromosome.

Ex. *TMPRSS2-ERG*
Chromosomal Inversion

Ex. EML4-ALK
Consequences of chromosomal translocations

A chromosomal translocation can broadly result in either juxtaposition of oncogenes near promoter/enhancer elements (A) or gene fusions (B). Both result in the deregulation of the expression of genes affecting various cellular and physiological processes like proliferation, differentiation, motility and apoptosis. Arrows at the genes depict the sites for double-strand break formation. Normally, the reciprocal joining does not lead to any functional product.
What are the molecular mechanisms behind gene fusions?
How does DNA break during chromosomal translocations?

There have been many attempts to decipher the reasons for fragility of chromosomes during translocations. However, the exact mechanism of most of the translocations is still elusive, except in very few cases.

V(D)J RAG-mediated recombination

V(D)J recombination is responsible for the immense diversity of antibodies and T-cell receptors (TCRs) generated during the development of B and T lymphocytes.

This mechanism contributes to chromosomal translocations in lymphoid cells.
V(D)J RAG-mediated recombination

The recombination activating genes, RAG1 and RAG2 (RAG complex), whose cellular expression is restricted to lymphocytes during their developmental stages, recognize the recombination signal sequences (RSSs) flanking the V, D and J subexons.

After binding to the 12/23 RSS, RAGs induce nicks which are later converted to DSBs by DNAPKcs–Artemis complex through a hairpin intermediate.

Mechanism of AID-mediated DSBs

The \textit{AICDA} gene encodes a DNA-editing deaminase member of the cytidine deaminase family. Member of the APOBEC family.

AID is involved in somatic hypermutation that lead to antibody diversity and class-switch recombination of immunoglobulin genes in B cells of the immune system. The same process may lead to B-cell lymphoma.

The nicks are usually repaired by the base excision repair mechanism. However, two unrepaired nicks in close vicinity can act as a double-strand break and thereby a substrate for chromosomal translocations.

Class switch recombination (CSR) is instigated by activation-induced cytidine deaminase, which converts cytosines in switch regions to uracils. The uracils are subsequently removed by two DNA-repair pathways, resulting in mutations, single-strand DNA breaks, and the double-strand breaks required for CSR.

Mechanism of AID-mediated DSBs

The level of AID activity in B cells is tightly controlled by modulating AID expression.
Activation-induced cytidine deaminase (AID)

Several factors influence the breakage rate including transcription, DNA sequence, and topological tension. These factors favor non-B DNA structures that permit formation of transient single-stranded DNA (ssDNA), making the DNA more vulnerable to agents such as the enzyme activation-induced cytidine deaminase (AID) and reactive oxygen species (ROS).

AID can act to deaminate either C’s or $\text{m}^6$C’s within preferred single-stranded target sequences (WRC). The T’s generated from deamination of $\text{m}^6$C’s are removed more slowly than U’s and are thus more long-lived lesions. After the DNA resumes its duplex conformation, the resulting T:G mismatch is vulnerable to DNA repair enzymes, such as activated Artemis, that may convert these to DSBs.


Chromosomal Translocation Fragile Zones

A subset of chromosomal translocations related to B cell malignancy in human patients arises due to DNA breaks occurring within defined 20-600 base pair (bp) zones, around the genes *CCND1* on chr 11 and *BCL2* on chr 18. CG motifs in each of the fragile sites is a hotspot for human translocation.

Fragile Zones in Human Lymphoid Chromosomal Translocations

![Diagram of fragile zones in human lymphoid chromosomal translocations](image)

AID-mediated

Relative contribution of RAG and AID in hematopoietic chromosomal translocations
How does DNA rejoin during chromosomal translocations?

Joining phase

Non-homologous end joining

Features of NHEJ:

**Canonical**: can seal double-strand ends, even distal and non-fully complementary ends, in a conservative fashion

**Alternative**: nucleotide loss from the DNA ends and alignment of microhomology (1-2 bp) between the two ends to guide rejoining

González-Marín et al., Int. J. Mol. Sci. 2012, 13(11), 14026-14052
Homology-directed Repair (HDR)
Intermingling of chromosome territories.

The formation of chromatin contacts is promoted by chromatin-binding proteins that can bind two or more genomic regions simultaneously. Such proteins include transcription factors, RNA and DNA polymerases, Polycomb repressive complexes and chromosomal scaffold proteins such as cohesin.
**Intermingling of chromosome territories.**

One of the basic features of genome organization in the interphase nucleus is that chromosomes occupy discrete chromosome territories.

An important functional consequence of chromosome intermingling is that it seems to promote preferential rearrangements between specific chromosomes, depending on their physical proximity.

The closeness of certain pairs of gene loci in the nucleus has been found to be positively correlated with their probability of translocations.

![Image](image.png)

**Fig. 1.** Representative examples of two-dimensional (2D) images of territories of chromosomes 9 (A), 22 (B) and 8 (C) in non-stimulated human lymphocytes. Chromosomes 9 and 22 are located near the center of the nucleus, in contrast to chromosome 8, which are located close to the nuclear membrane. *Bar* represents 5 μm.

**Fig. 3.** A representative example of the positions of the ABL (green signals) and BCR (red signals) genes in nuclei of G0-phase lymphocytes (A). Both ABL and BCR genes are often found in the middle of the cell nucleus and form clusters of all four genes. Definition of CA, CB, and AB distances (B). CA is the center of nucleus-to-ABL distance \( (x) \) normalized to the local radius \( (y) \): CA = \( x/y \); similarly CB = \( x/y \). Minimal ABL–BCR distance \( (AB_{min}) \) is also shown. *Bar* represents 5 μm.


*Cromosoma.* 1999 Dec;108(7):426-35.
Transcription (and splicing)-mediated gene fusion

Chimeras, not due to a genomic pathological-associated rearrangement, may originate from two separate events: intergenic splicing and transgenic splicing.

An intergenic splicing event combines exons from two adjacent genes of the same chromosome, while a transgenic splicing event combines exons from two gene locate on different chromosomes.

Prevalent in normal tissues and cancer.

Ex. CTNNB1P1-CLSTN1

Akiva et al., Genome Res. 2006 Jan; 16(1): 30–36.
Nomenclature
Translocations are designated in the format "t(X;4)(p21.2;q34)", followed by the usual description, placed between brackets, indicating the exact translocation breakpoint.
Deletion

Deletions are designated by "del" after a description of the deleted segment, i.e. the first (and last) nucleotide(s) deleted

\texttt{g.390\_1458del  (g.390\_1458del1069)}
Inversion

Inversions are designated by "inv" after the nt number of the nucleotides inverted.

g.1458_oXYZ:457inv
At the RNA level:

Gene Name - 5' partner gene involved in the fusion.
Last Observed Exon - The last exon observed before the fusion breakpoint. In the 5' gene it can be ‘UTR’ or have ‘?’ if the breakpoint is unknown.

Gene Name - The 3' partner gene involved in the fusion.
First Observed Exon - The first exon observed after the fusion breakpoint in the 3' gene. It can be ‘UTR’ or have ‘?’ if the breakpoint is unknown.

\[ r.-30\_12delinsAB001235.1:r.-124\_1298 \]

It should be noted that the descriptions at RNA level, like those at protein level, are often deduced and not based on experimental evidence.
Fusion partners
**Most common chromosomal translocations in cancer**

<table>
<thead>
<tr>
<th>Translocation</th>
<th>Genes</th>
<th>Type of cells</th>
<th>Name of cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>t(14;18)</td>
<td>BCL2</td>
<td>B-cells</td>
<td>Follicular lymphoma</td>
</tr>
<tr>
<td>t(8;14)</td>
<td>c-MYC</td>
<td>B-cells</td>
<td>Burkitt's lymphoma</td>
</tr>
<tr>
<td>t(3;14)</td>
<td>BCL6</td>
<td>B-cells</td>
<td>Diffuse large B-cell lymphoma</td>
</tr>
<tr>
<td>t(9;22)</td>
<td>BCR and ABL</td>
<td>Myeloid cells</td>
<td>Chronic myeloid leukemia</td>
</tr>
<tr>
<td>t(11;14)</td>
<td>CCND1</td>
<td>B-cells</td>
<td>Mantle cell lymphoma</td>
</tr>
<tr>
<td>t(10;14)</td>
<td>HOX11</td>
<td>T-cells</td>
<td>T-Acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>t(2;13)</td>
<td>NPM and ALK</td>
<td>B-cells</td>
<td>Anaplastic large cell lymphoma</td>
</tr>
<tr>
<td>t(21;7)</td>
<td>TMPRSS2 and ETS</td>
<td>Epithelial cells</td>
<td>Prostate carcinoma</td>
</tr>
<tr>
<td>t(12;15)</td>
<td>ETV6 and NTRK3</td>
<td>Epithelial cells</td>
<td>Breast carcinoma</td>
</tr>
<tr>
<td>t(11;22)</td>
<td>EWS and FLI-1</td>
<td>Mesenchymal cells</td>
<td>Ewing's sarcoma</td>
</tr>
</tbody>
</table>

Currently COSMIC includes information on fusions involved in solid tumours and leukaemias

https://cancer.sanger.ac.uk/cosmic/fusion
Chromosomal translocation mediated via V(D)J recombination or class switch recombination (CSR) in lymphoma

Many reciprocal chromosomal translocations that occur in early human B cells involve one DNA double-strand break (DSB) generated by the physiological process of V(D)J recombination at the immunoglobulin heavy chain (IGH) locus on chromosome 14, and a second pathological DSB at a non-IGH locus by a different mechanism that has yet to be elucidated.

Ex. BCL2-IGH
MYC-IGH
CCND1-IGH

BCL2-IGH translocation described in 1985

The fusion generates an overexpression of Bcl-2

Reed JC, Blood. 2008 Apr 1; 111(7): 3322–3330.
**BCR-ABL** in Chronic Myeloid Leukemia

**BCL2-IGH** translocation described in 1973

*ABL1-BCR* translocation also occurs and may express but is of no clinical significance.

---

<table>
<thead>
<tr>
<th>Synonym</th>
<th>Major Cluster</th>
<th>Minor Cluster</th>
<th>Micro-Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>exons 12-16</td>
<td>Between alternative exon 2, e2 and e2</td>
<td>Between exons e19 and e20</td>
</tr>
<tr>
<td>Protein</td>
<td>p210</td>
<td>p190</td>
<td>p230</td>
</tr>
<tr>
<td>Associated Leukemia</td>
<td>CML, e14q2 shown to have thrombocytosis in some studies, Ph+ ALL</td>
<td>Ph + ALL, CML that tends to have monocytosis and an aggressive course</td>
<td>Chronic Neutrophilic Leukaemia, Small reports describing patients with a course resembling classical CML</td>
</tr>
</tbody>
</table>


BCR-ABL

The BCR-ABL oncprotein has constitutive kinase activity that promotes the growth advantage of leukemic cells. BCR-ABL is localized to the cytoplasm (instead of shuttling nucleus-cytoplasm) and, due to its constitutive activation, generates constitutive activation of:

- RAS pathway
- STAT1 and 5
- PIK3CA pathway

Successful treatment with imatinib, dasatinib, nilotinib, bosutinib and ponatinib

**TMPRSS2-ERG or ETV1**

Present in 50% of prostate cancer patients

**TMPRSS-ERG** translocation described in 2005

- **TMPRSS2**: promoter sensitive to androgen regulation, highly expressed in normal prostate. 5’ UTR of **TMPRSS2**. Hence, overexpression of **ERG** or **ETV1**.

- **ETV/ERG**: TF of the ETS family

---

TMPRSS2-ERG or ETV1

Treatment based on androgen deprivation

Mosquera et al., *Clinical Cancer Research* 14, 3380-3385.
EWSR1-FLI1

N-terminal transactivating domain (TAD) of EWSR1 fused to the C-terminal (ETS) DNA-binding domain of the weak TF FLI1 in Ewing sarcoma.

<table>
<thead>
<tr>
<th>Translocation</th>
<th>Fusion</th>
<th>Frequency in ESFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>t(11;22)(q24;q12)</td>
<td>EWSR1–FLI1</td>
<td>~85%–90% of cases</td>
</tr>
<tr>
<td>t(21;22)(q22;q12)</td>
<td>EWSR1–ERG</td>
<td>~10% of cases</td>
</tr>
<tr>
<td>t(7;22)(p22;q12)</td>
<td>EWSR1–ETV1</td>
<td>Rare</td>
</tr>
<tr>
<td>t(17;22)(q12;q12)</td>
<td>EWSR1–ETV4</td>
<td>Rare</td>
</tr>
<tr>
<td>t(2;22)(q35;q12)</td>
<td>EWSR1–FEV</td>
<td>Rare</td>
</tr>
<tr>
<td>t(16;21)(p11;q22)</td>
<td>FUS–ERG</td>
<td>Rare</td>
</tr>
<tr>
<td>t(2;16)(q35;p11)</td>
<td>FUS–FEV</td>
<td>Rare</td>
</tr>
</tbody>
</table>

This results in a highly expressed TF that regulates both gene and protein-regulatory networks to drive tumor initiation and maintenance.

EWSR1-FLI1

Model of Ewing cells dissemination based on cell-to-cell heterogeneity of EWSR1-FLI1 expression.

<table>
<thead>
<tr>
<th>NCT no.</th>
<th>Patient population</th>
<th>Phase</th>
<th>Treatment</th>
<th>Mechanism of action</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT02306161</td>
<td>Newly diagnosed, metastatic disease</td>
<td>III</td>
<td>Standard chemotherapy ± ganitumab</td>
<td>IGF-1R mAb</td>
</tr>
<tr>
<td>NCT03495921</td>
<td>Recurrence after one prior therapy</td>
<td>III</td>
<td>Temozolomide and irinotecan ± vigil</td>
<td>bi-shRNA(furin) and GM-CSF augmented autologous tumor cell immunotherapy</td>
</tr>
<tr>
<td>NCT01858168</td>
<td>Recurrence after &gt;1 prior therapy</td>
<td>I</td>
<td>Olaparib and temozolomide ± irinotecan</td>
<td>PARP inhibitor</td>
</tr>
<tr>
<td>NCT02736565</td>
<td>Relapsed or refractory disease</td>
<td>I</td>
<td>pbi-shRNA™ EWS/FLI1 Type 1 LPX</td>
<td>Functional plasmid DNA construct that targets EWS/FLI1 mRNA</td>
</tr>
<tr>
<td>NCT03514407</td>
<td>Relapsed or refractory disease</td>
<td>Ib</td>
<td>INCB059872</td>
<td>LSD1 inhibitor</td>
</tr>
<tr>
<td>NCT03600649</td>
<td>Relapsed or refractory disease</td>
<td>I</td>
<td>SP-2577 (seclidemstat)</td>
<td>LSD1 inhibitor</td>
</tr>
<tr>
<td>NCT02657005</td>
<td>Relapsed or refractory disease</td>
<td>I</td>
<td>TK216</td>
<td>ETS-family transcription inhibitor</td>
</tr>
</tbody>
</table>

Current clinical trials for advanced Ewing sarcoma

**Abbreviations:** GM-CSF, granulocyte-macrophage colony-stimulating factor; IGF-1R, insulin-like growth factor one receptor; mAb, monoclonal antibody.

Acute promyelocytic leukemia (APL) is a hematological malignancy driven by a chimeric oncprotein containing the C terminus of the retinoic acid receptor-a (RARa) fused to an N-terminal partner, most commonly promyelocytic leukemia protein (PML).

RARa is a nuclear receptor that in the absence of its ligand retinoic acid (RA) represses the transcription of target genes. Physiological levels of RA convert RARa from a transcriptional repressor into a potent activator, driving expression of genes involved in myeloid differentiation.

PML-RARa acts as a transcriptional repressor of RARa and non-RARa target genes and antagonizes the formation and function of PML nuclear bodies that regulate numerous signaling pathways, hence blocking differentiation.

Responsive to all-trans retinoic acid (ATRA)
Among normal tissues, ALK protein expression is prominent in the brain and peripheral nervous system of developing embryos and decreases rapidly after birth (3). In adults, the ALK protein is expressed at low levels and only in the central nervous system, while it has not been detected in other tissues.

Specific inhibitors of the kinase activity of ALK have been developed as therapeutic drugs for EML4-ALK–positive lung cancer, three of which (crizotinib, ceritinib, and alectinib) are approved for clinical use.

Frequencies and distributions of RET aberrations

1990: described in papillary thyroid carcinoma
2012: described in Lung ADC

Inversion in Chr 10
Highly prevalent in Lung ADC
Highly prevalent in Papillary thyroid

The term fusion refers to RET rearranged with a known fusion partner (e.g., KIF5B-RET).


Drilon et al. Cancer Discov; 2013 3(6); 630-5
RET Fusion is a predictive biomarker for use of vandetanib and cabozantinib in patients.

**ROS1**

ROS1 has been described as an ‘orphan’ receptor tyrosine kinase and has no known ligand.

Gene rearrangements involving the ROS1 gene were first detected in glioblastoma tumors and cell lines in 1987.

In 2007 a *ROS1* rearrangement was identified in a cell line derived from a lung adenocarcinoma patient.
Cabozantinib, ceritinib, crizotinib, and lorlatinib are currently approved for use in patients with ROS1 Fusion in non-small cell lung carcinoma.


FIG-ROS1 fusions are localized in the Golgi, whereas other ROS1 fusion variants have been reported to the plasma membrane or cytosol.
**NTRK fusions prevalence and distribution**

In 1998, the ETV6–NTRK3 gene fusion was discovered in congenital fibrosarcoma tumours by Sorensen and colleagues.


NTRK fusions structure

NTRK gene fusion partners often contain oligomerization domains

TRK signaling

TRKs are detected in several tissue types, including neuronal cells in CNS.

TRK fusion proteins can signal through the same downstream pathways activated by full-length TRK proteins upon neutrophin binding.

Larotrectinib was the first ‘tissue agnostic’ drug approved by the FDA (Dec ‘18) and in September ‘19 by the EMA.

To conclude...

Gene fusions arise from chromosomal aberrations: translocations, inversions or deletions

The underlying mechanisms of chromosomal DSB and repair are still not fully understood

Oncogenic potential derives from overexpression and/or constitutive activation of proto-oncogenes as a result of the gene fusion. In some cases, gene fusions display novel activities compared to their non-fused counterparts.

Some gene fusions are particularly sensitive to targeted therapy, which makes them extremely relevant in the clinics