

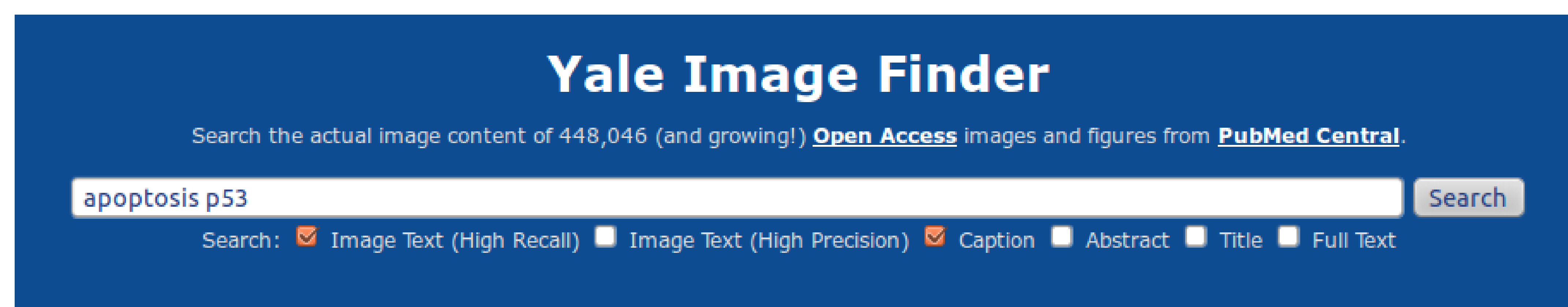
# Image Retrieval in Controlled English

Michael Krauthammer and Tobias Kuhn  
Department of Pathology, Yale University

## Yale Image Finder

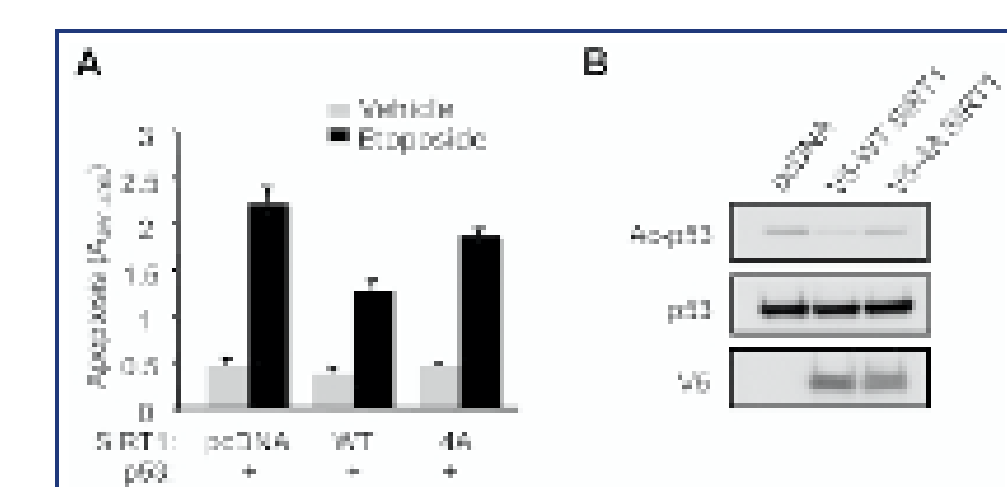
The Yale Image Finder (YIF):

- is a search engine for **biomedical images**
- is **keyword-based**
- indexes the images of the open access articles of **PubMed Central**
- can search the image caption and **text within images**

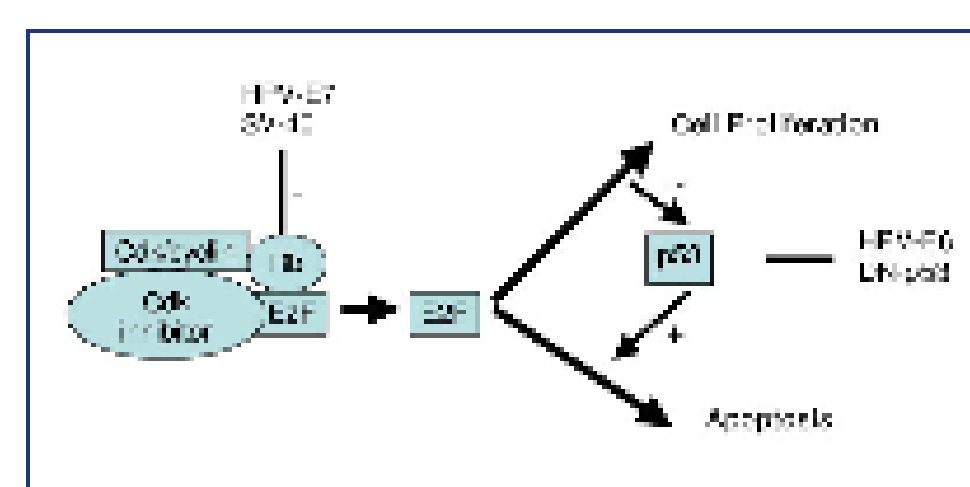


177 results found.

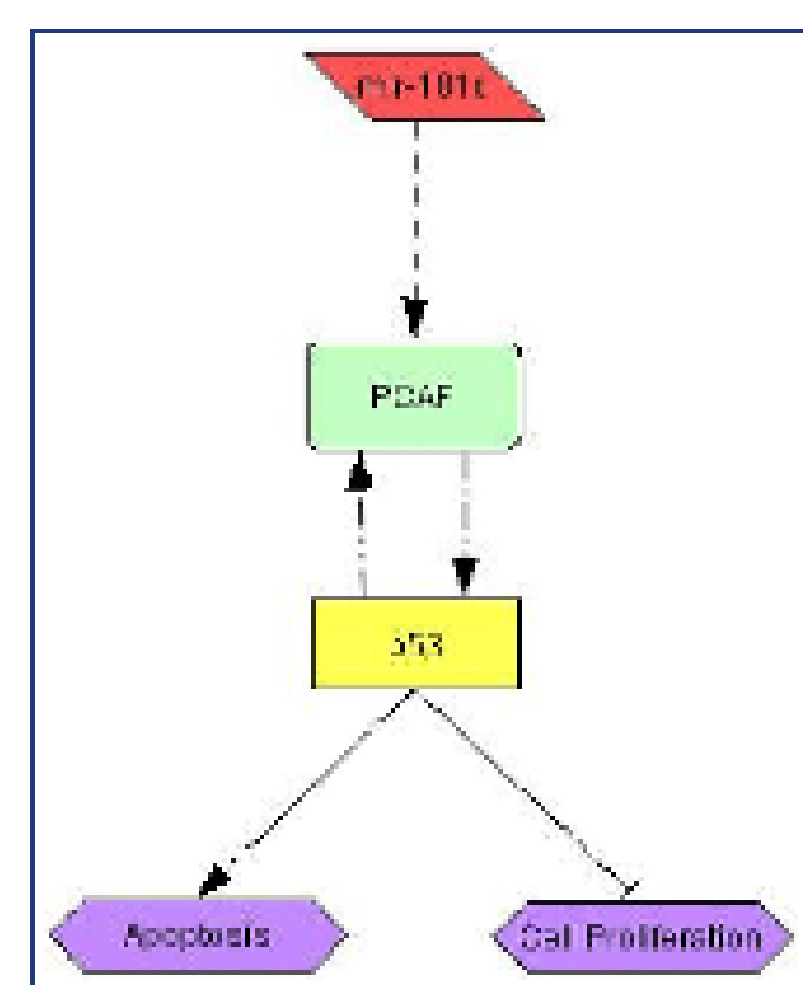
<< < 1 2 3 4 5 >>



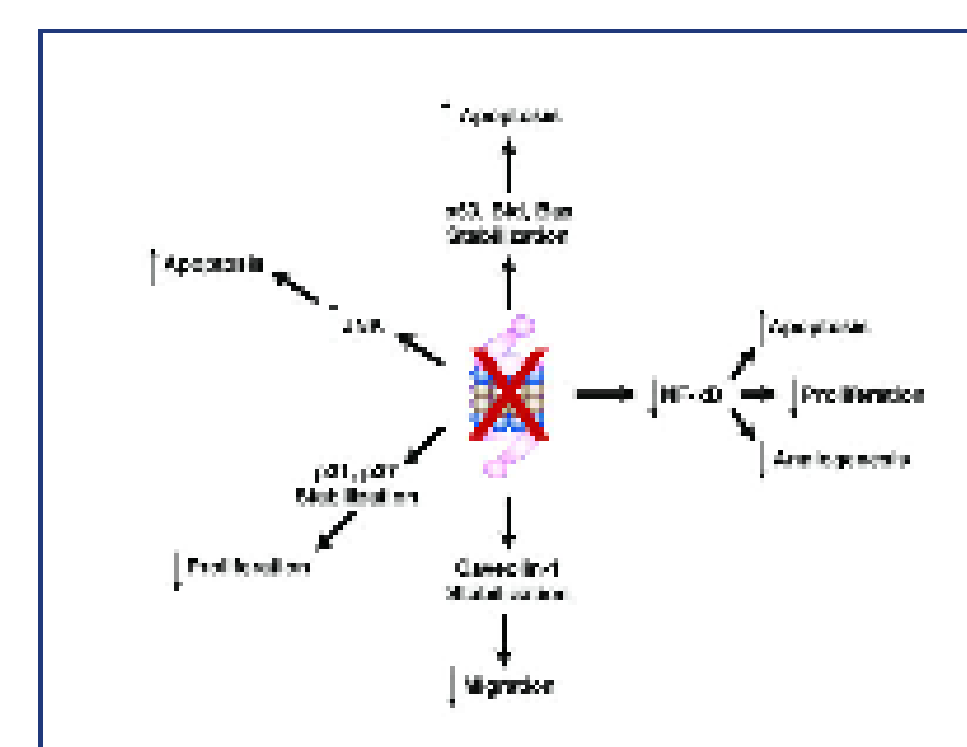
**Figure 6** CK2 phosphorylation confers increased SI...



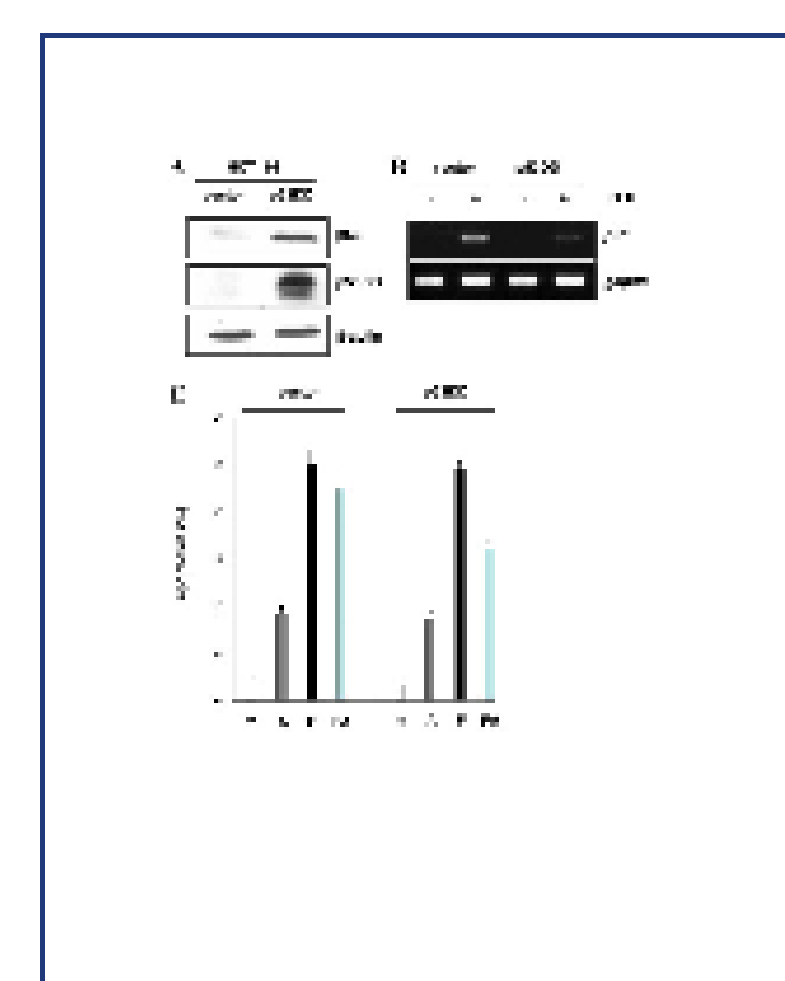
**Figure 1** A schematic representation of RB/p53 int...



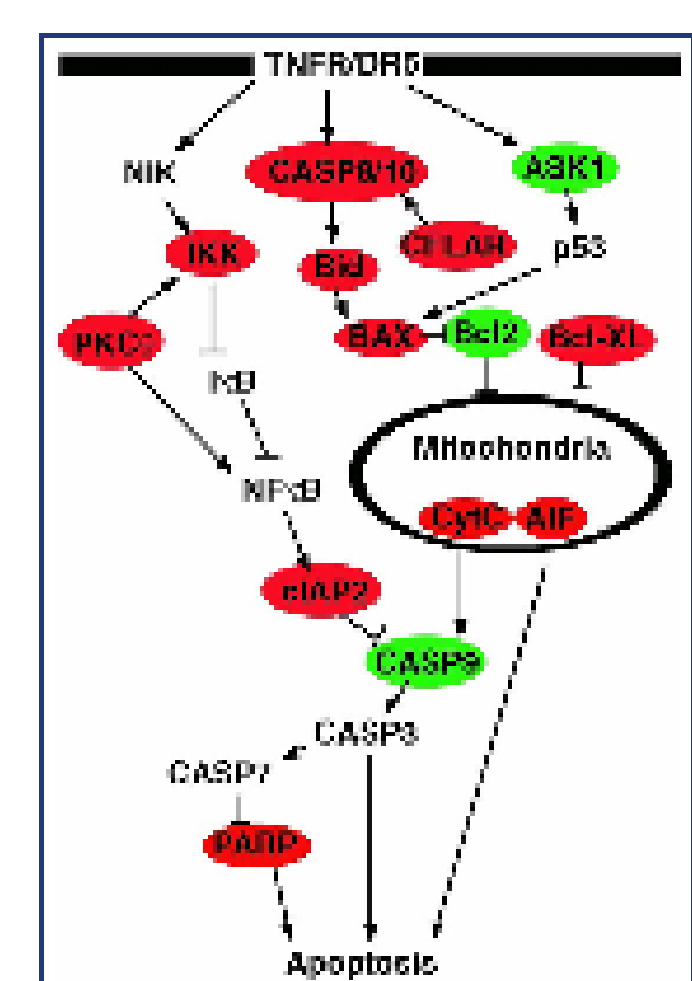
**Figure 8** Tumor suppressive pathway involving miR-...



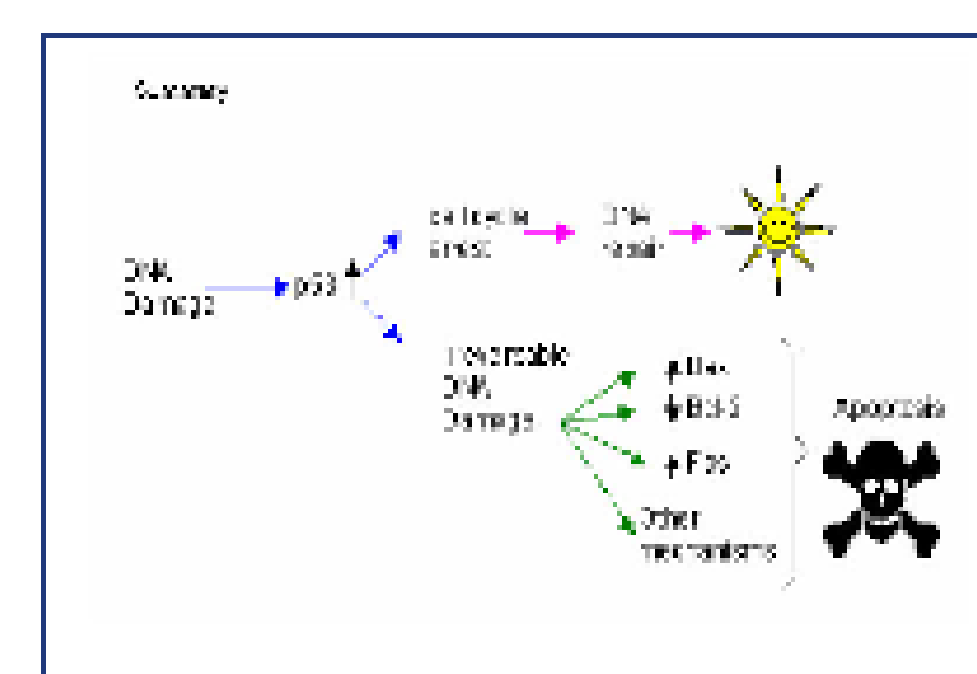
**Figure 2** Inhibition of the proteasome by bortezom...



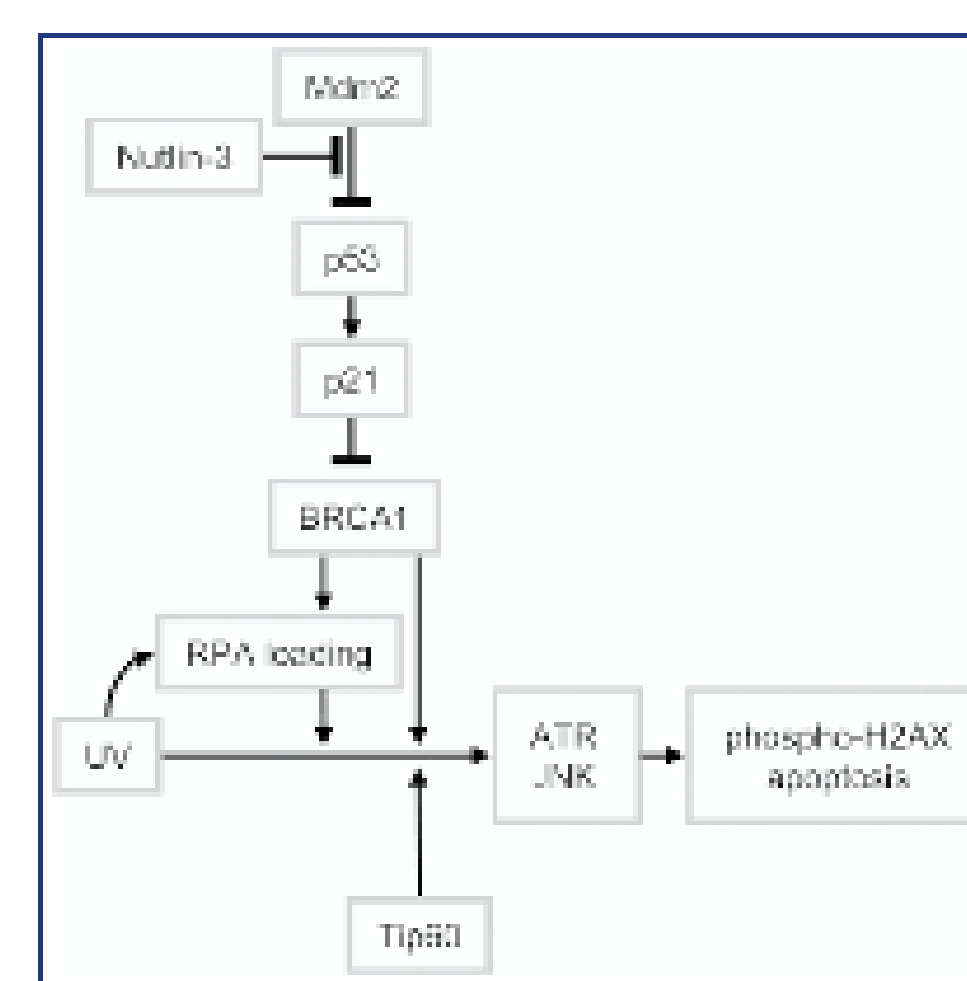
**Figure 4** Dominant inhibitory effect of p53DD on t...



**Figure 5** Apoptosis pathway with genes displaying ...



**Figure 2** Summary of the mechanism of action of th...

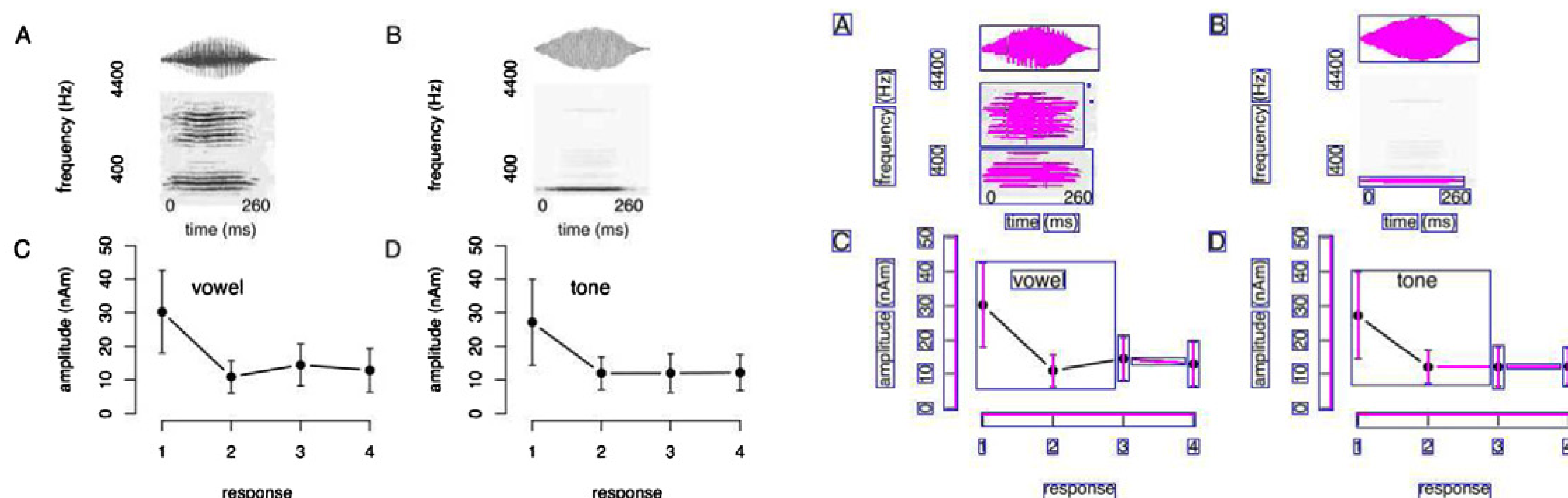


**Figure 9.** p53 and Tip60 each modulate the response...

## Image Analysis and Classification

Images are analyzed and classified:

- **Segments** of the image are recognized
- **OCR** is applied to segments containing text
- Images are **classified** as gels, graphs, plots, etc. (work in progress)
- **Semantic properties and relations** are extracted from the image (work in progress)



## Controlled English Interface

Retrieving Images in Controlled English (Rice):

- is a **prototype of a semantic search engine** for biomedical images
- uses **controlled English** as its query language

Controlled English queries in Rice:

- **look natural** but have a **precise mapping to logic**
- are **matched with the extracted image models** to find results (work in progress)

## Rice

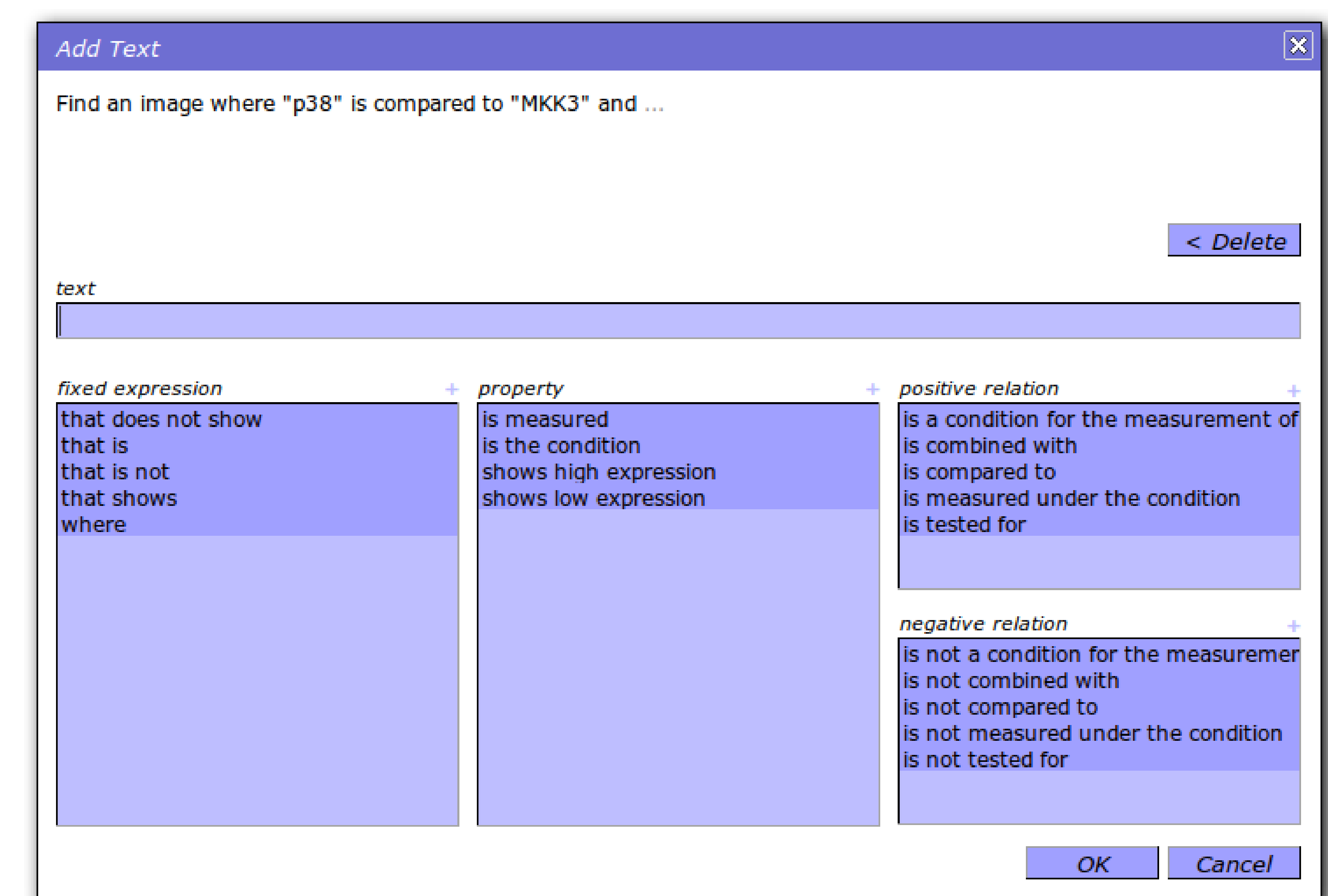
Retrieving Images through Controlled English

Find an image  
that is a Northern blot ;  
where "BCR/ABL" is combined with "mcl-1" ;

submit debug  
Find an image  
that is a Western blot ;  
+ where "p38" is compared to "MKK3" ;  
+ where "RU486" is combined with "DEX" ;

Writing queries:

- Controlled English is easy to read and understand but **hard to write**
- **Predictive editors** can solve this problem
- Users construct syntactically correct sentences in an **iterative and guided way**



Benefits:

- Existing query interfaces are **either very simple** (i.e. keyword-based) **or too complex** to be usable without training
- With Rice, complex and precise queries can be written in a **natural and intuitive way**
- Rice is **immediately accessible** to researchers interested in the results represented in images of the biomedical literature