1 Hello from Illumina

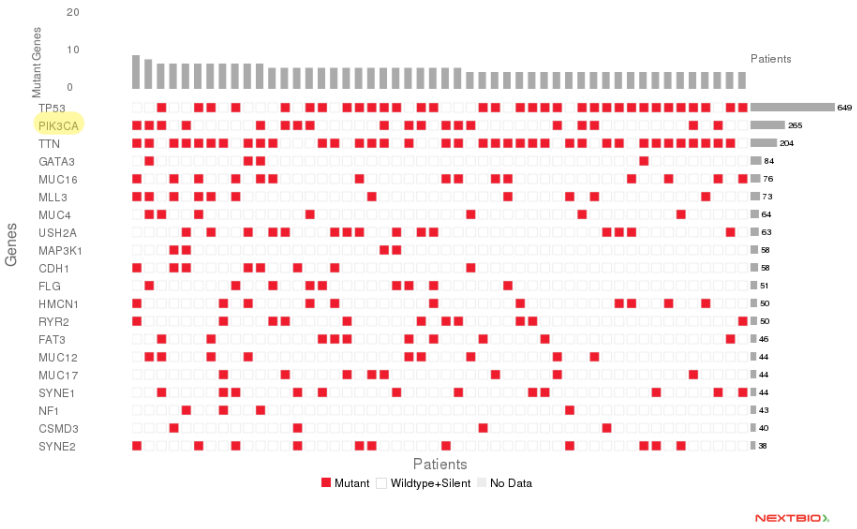
Thanks for paying Illumina a visit!

First, check out:

- the Illumina BaseSpace application platform @ basespace.com
- the Illumina NextBio platform @ nextbio.com

Second... **Ready for a challenge?**

2 CHALLENGE: Visualize patient data



2.1 Goal: make sense of clinical data

Build a web application which:

- Accepts patient data as flat text files (see next section for public example of TCGA data)
- Generates **visualizations and/or reports** of cohort (group) and individual patient data

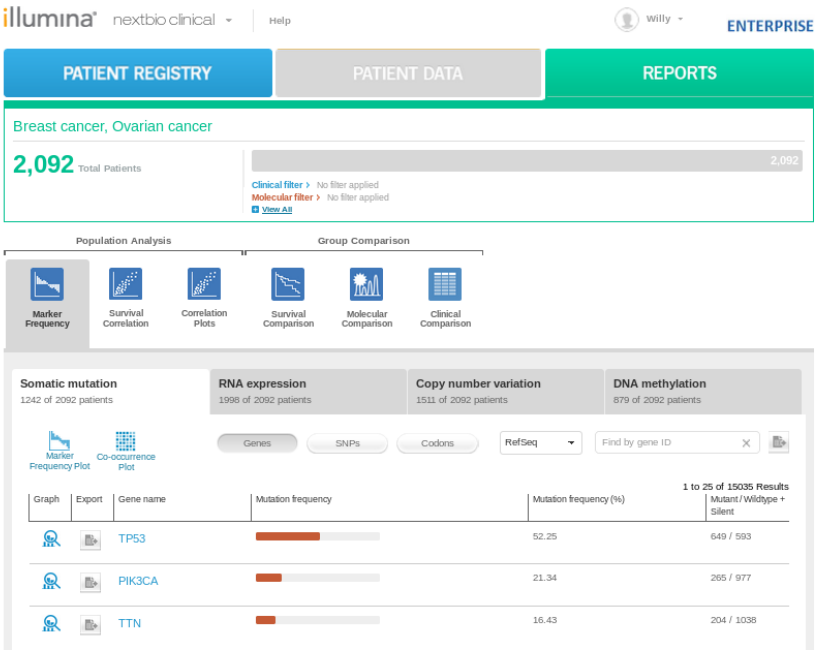
2.2 Data: Public patient data from Cbioportal

- Cbioportal.org is a great source of Clinical and Molecular data

- Many APIs can be found at `http://www.cbioportal.org/web_api.jsp`
- Specific example: you can find Clinical Attribute data for cancer patients at `http://www.cbioportal.org/webservice.do?cmd=getClinicalData&case_set_id=ov_tcga_all`

If you can accept such a file, or the URL to it, and show you any visualization you feel like building, e.g. a scatterplot, or an interactive view of a patient "graph" where links are similarity across attributes, PCA, etc. We encourage you to go wild – surprise us with the reporting/mashup of data.

3 Context: NextBio Clinical



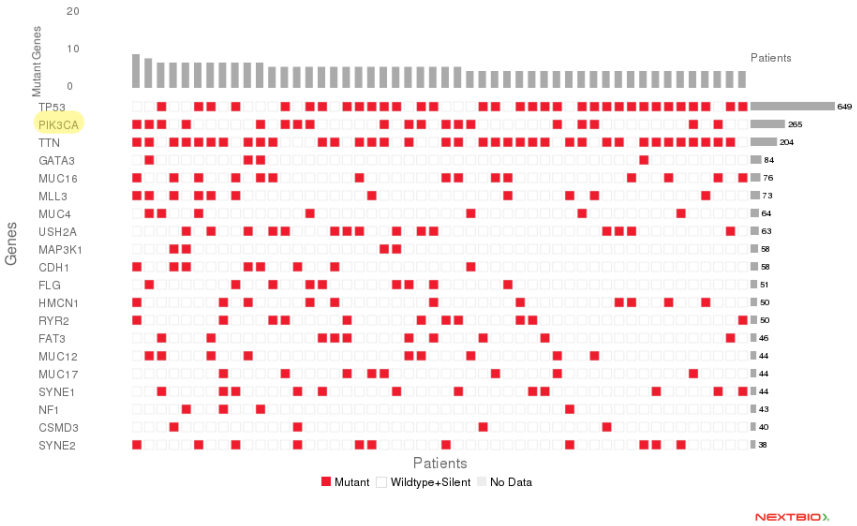
We're trying to make sense of this data too!

Here's a STORY told with data visualization (using NextBio Clinical, www.nextbio.com visualizations)

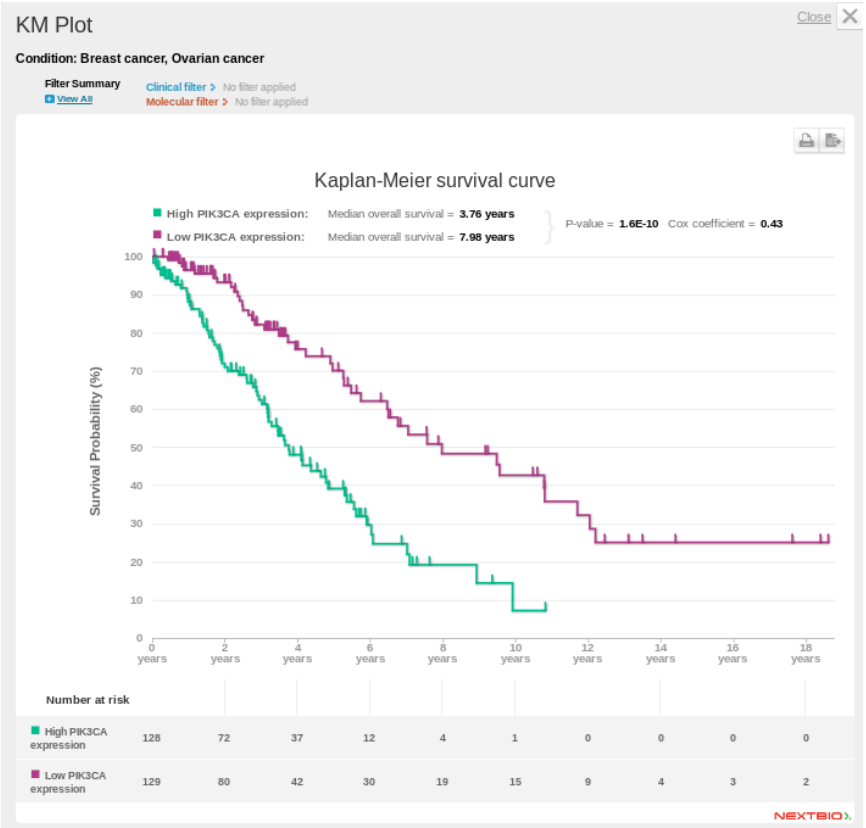
3.1 PIK3CA in Ovarian and Breast cancer

- Breast cancer and ovarian cancer afflict a huge number of women

- NextBio Clinical has a huge repository of data including public ovarian/breast cancer data
- You can easily filter public data by condition (Show Basic Patient finder) to focus on these two
- And generate reports and visualizations to answer questions, e.g.:
- What are the most mutated genes in our Ovarian/Breast cancer cohort? And how many genes are mutated per patient?

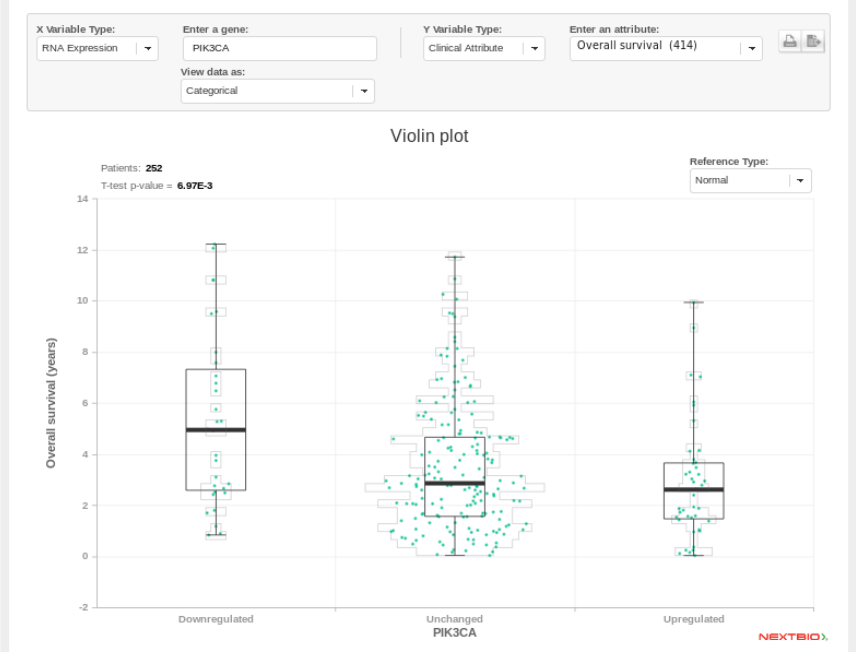


- PIK3CA looks interesting. We can ask more questions: Does the level of PIK3CA expression* predict survival? Why look at expression rather than mutation? We can't change our genome easily, but may be able to modulate expression

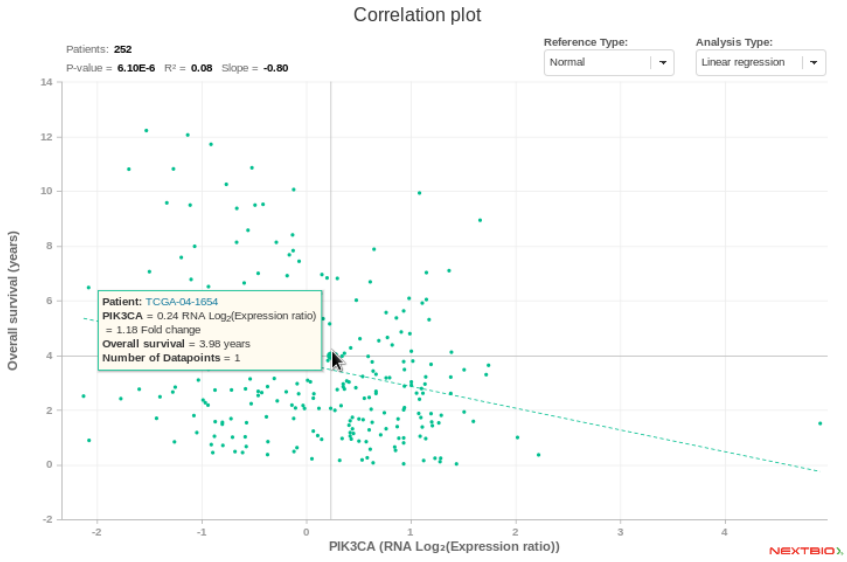


Clearly, YES - patients with high expression live about HALF as long.

- We can dive down into the correlation between expression and survival (or any other clinical or molecular attribute of the cohort) at the patient level with a "correlation scatterplot":
- Looking at the survival of patients grouped by whether PIK3CA is downregulated, unchanged, or upregulated reinforces the association found earlier:

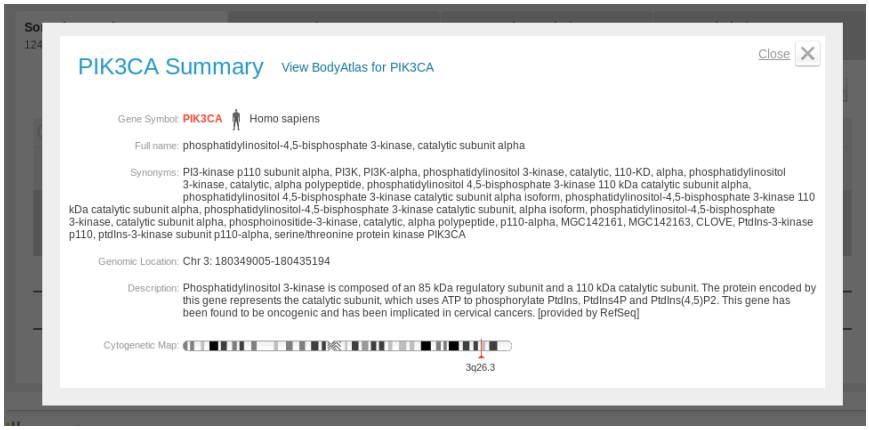


- But clearly our categories have overlapping survival distributions. A plain scatterplot may be better, and perhaps the correlation will lead to a significant linear association:



- Yes

- And we can ask what this gene is:



But we need your (their) help to do more, to produce more informative visualizations, to compute more powerful statistics, and to load information from more data providers and formats.

4 Data / APIs

`http://www.cbioportal.org/web_api.jsp`

Example flat data files:

Clinical data for Ovarian cancer: `http://www.cbioportal.org/webservice.do?cmd=getClinicalData&case_set_id=ov_tcga_all`

Mutation data for BRCA1 in TCGA Ovarian+Breast: `http://www.cbioportal.org/webservice.do?cmd=getMutationData&genetic_profile_id=ov_tcga_mutations+ucec_tcga_mutations&gene_list=BRCA1`

Thanks, Willy