From bewilderment to enlightenment: Logic in cancer research

Wong Limsoon

Based on work with Wilson Goh and Sriganesh Srihari

Two bewilderments

- **Breast cancer survival signatures are no better than random signatures**

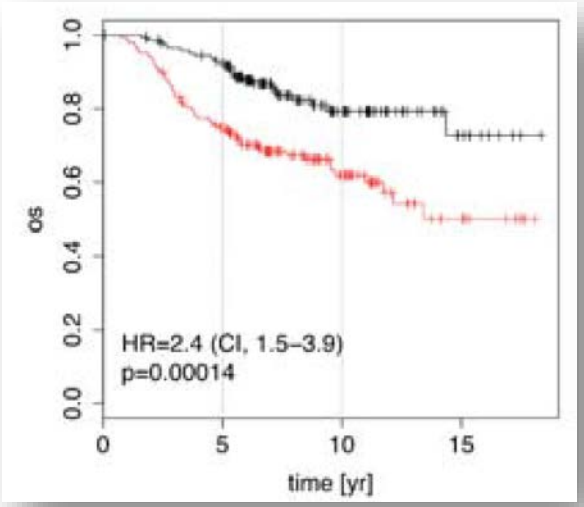
- **Mutation mutual exclusivity are not associated with synthetic lethality**

And maybe
some enlightenment at the end….

Story #1

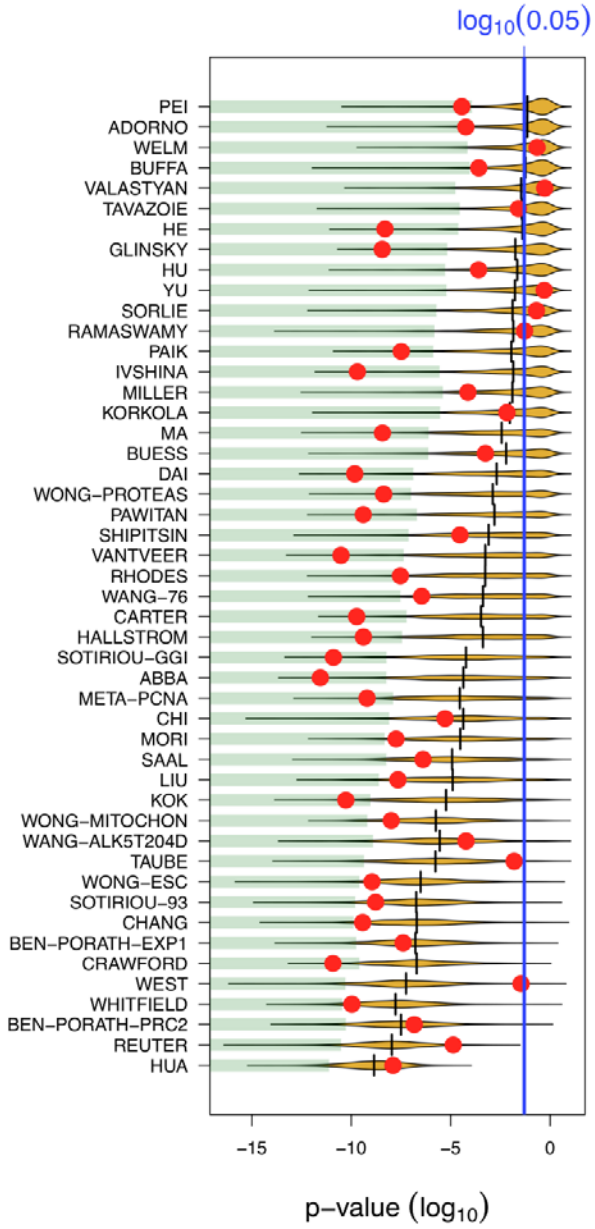
BREAST CANCER-SURVIVAL SIGNATURES

Venet et al., *PLOS Comput Biol*, 2011



A seemingly obvious conclusion

- **A multi-gene signature (social defeat in mice) is claimed as a good biomarker for breast cancer survival**
  - Cox's survival model p-value << 0.05
- **A straightforward Cox's analysis. Anything wrong?**



Venet et al., *PLOS Comput Biol*, 2011

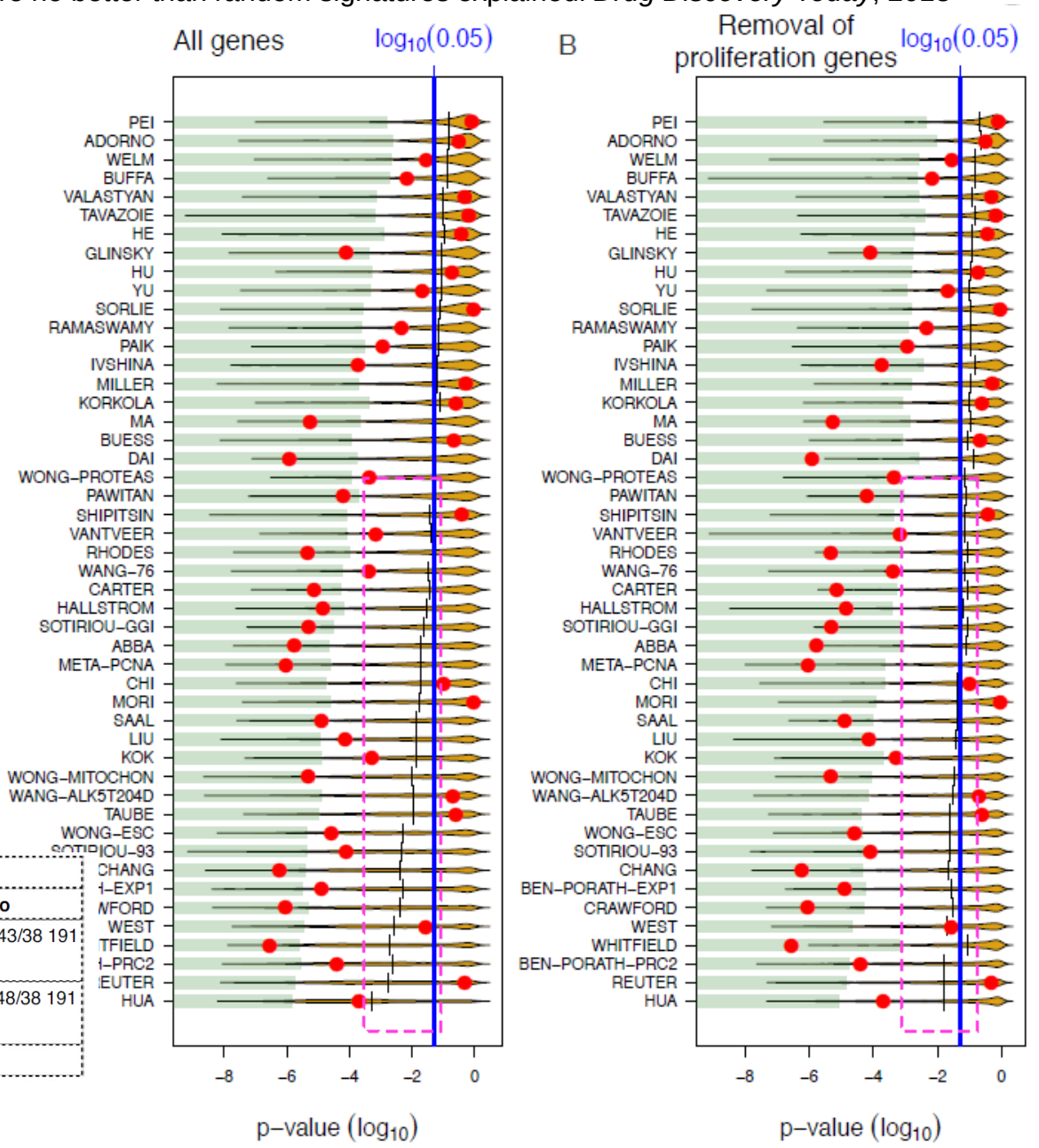
In fact, almost all random signatures also have p-value < 0.05

And the larger a random signature is, the more likely this happens

- **Perhaps instead of asking whether a signature is significant, ask what makes a signature (random or otherwise) significant**

Goh & Wong, Why breast cancer signatures are no better than random signatures explained. *Drug Discovery Today*, 2018

- **Proliferation is a hallmark of cancer**
- **Hypothesis: proliferation-associated genes make a signature significant**



# of random signatures w/ ≥1 prolif gene

Cutoffs	Counts			
	NP	P	Marginals	Odds ratio
Above 0.05	7043	19 043	26 086	7043/9809)/(19 043/38 191 =1.44x
Below 0.05	2766	19 148	21 914	2766/9809)/(19 148/38 191 =0.56x
Marginals	9809	38 191	48 000	

But …

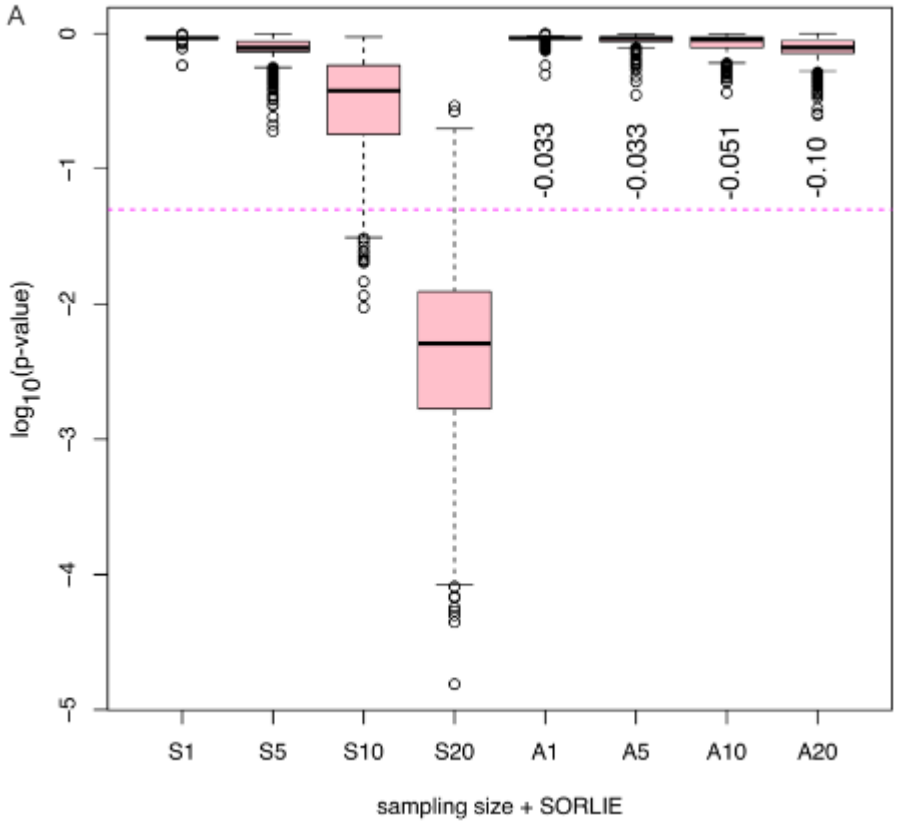
- **Many random signatures with proliferation genes are not significant**
- **Which proliferation genes make many random signatures significant?**
- **What other key factors make many random signatures significant?**

- **Some helpful analytical practices**
  - Leverage existing data and knowledge
  - Careful and systematic evaluation of gene sets
  - Rigorous testing against as many published datasets as possible

Leverage background knowledge

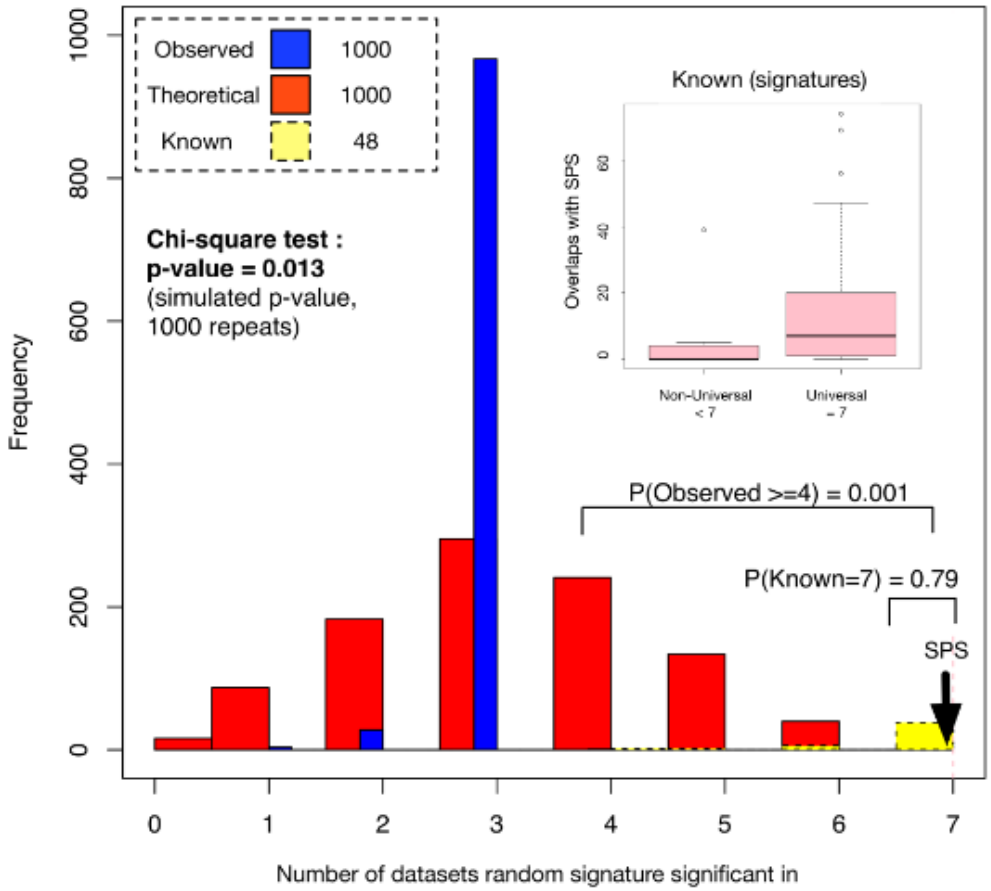
- **Background knowledge**
  - Proliferation is a cancer hallmark
- **Good signatures with high diff in p-values before vs after removing proliferation genes**
  - GLINSKY, DAI, RHODES, ABBA, WHITFIELD
- **SPS = { genes appearing in at least two of these good signatures }**
  - 83 genes in total
  - 81 of these are proliferation associated

Systematic evaluation



- **SPS genes show additive effect, other proliferation genes don't**

Test on many datasets



- **SPS is universally significant on 7 breast cancer datasets**
- **Random signatures (same size as SPS) are hardly universal, even though they get better p-values than known signatures on some datasets**

Story #2

SYNTHETIC LETHALS

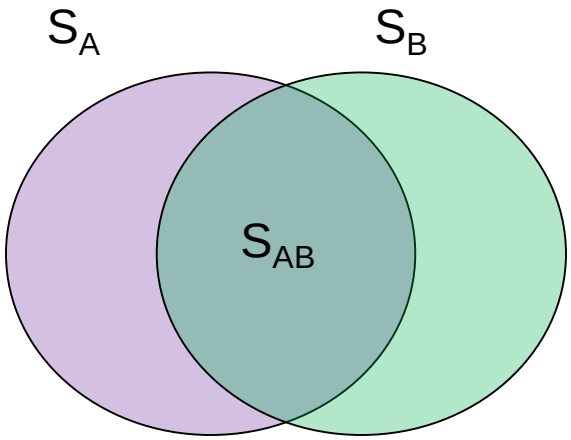
Srihari et al. Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. *Biology Direct*, 10:57, 2015.

Synthetic lethal pairs

- **Fact/postulate**
  - When a pair of genes is synthetic lethal, mutations of these two genes avoid each other
- **Observation**
  - Mutations in genes (A,B) are seldom observed in the same subjects
- **Conclusion by abduction**
  - Genes (A,B) are synthetic lethal

- **Why interested in synthetic lethality?**
  - Synthetic-lethal partners of frequently mutated genes in cancer are likely good treatment targets

A seemingly obvious approach based on hypergeometric test



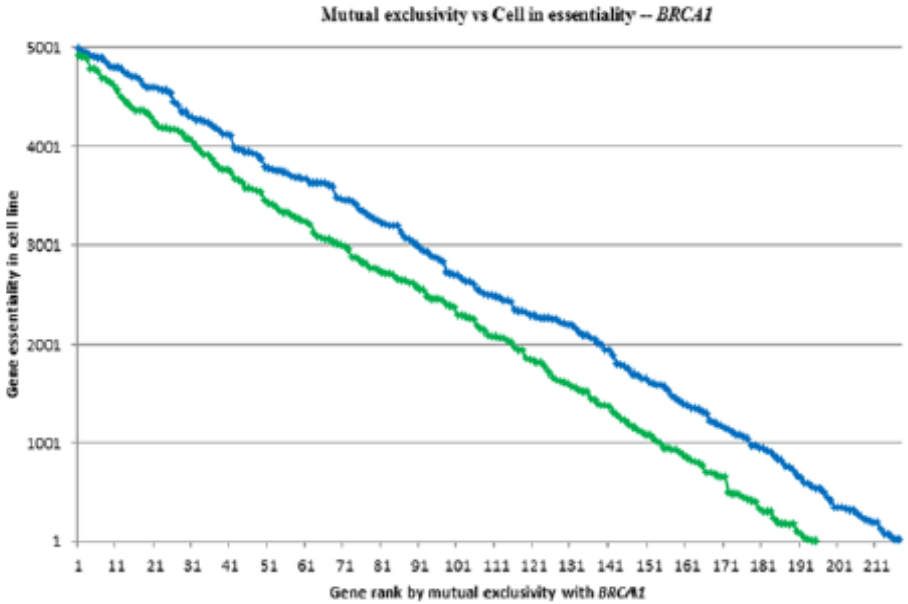
$$P[X \leq |S_{AB}|] = 1 - P[X > |S_{AB}|], \quad (1)$$

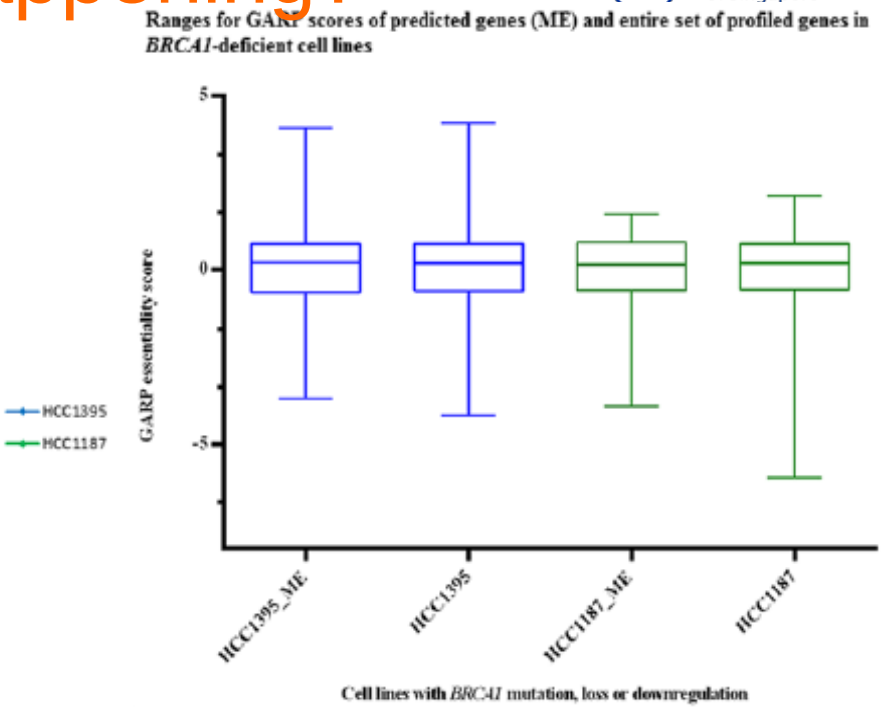
where $P[X > |S_{AB}|]$ is computed using the hypergeometric probability mass function for $X = k > |S_{AB}|$:

$$P[X > |S_{AB}|] = \sum_{k=|S_{AB}|+1}^{|S_B|} \frac{\binom{|S_A|}{k}\binom{|S|-|S_A|}{|S_B|-k}}{\binom{|S|}{|S_B|}}$$

- **Mutations of genes (A,B) avoid each other if $P[X \leq S_{AB}] \leq 0.05$**
- **Anything wrong with this?**

What is happening?





- **Among top ME-genes, GARP score ranks correlate with mutual exclusion ranks**

- **But GARP scores of ME-genes (i.e. have significantly mutually exclusive mutations to BRCA1) are similar to other genes**

Srihari et al. *Biology Direct*, 10:57, 2015.

The hypergeometric distribution does not reflect real-world mutations

$$P[X \leq |S_{AB}|] = 1 - P[X > |S_{AB}|], \quad (1)$$

where $P[X > |S_{AB}|]$ is computed using the hypergeometric probability mass function for $X = k > |S_{AB}|$:

$$P[X > |S_{AB}|] = \sum_{k=|S_{AB}|+1}^{|S_B|} \frac{\binom{|S_A|}{k}\binom{|S|-|S_A|}{|S_B|-k}}{\binom{|S|}{|S_B|}}$$

- **The Hypergeometric distribution assumes**
  - Mutations are independent
  - Mutations have equal chance to appear in a subject

- **Real-life mutations**
  - Inherited in blocks; those close to each other are correlated
  - Some subjects have more mutations than others, e.g. those with defective DNA-repair genes

⇒ **Null distribution is not hypergeometric, binomial, etc.**

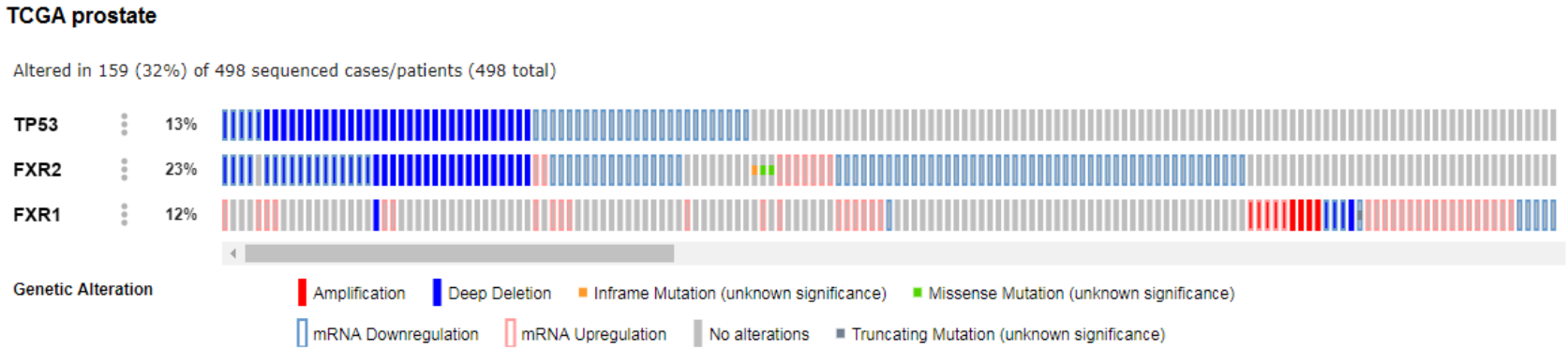
Solution?

- **Group genes into genomic clusters**
- **Test genes in far-apart genomic clusters for mutually exclusive mutations**
- **Mutually exclusive clusters should contain synthetic-lethal & collateral-lethal gene pairs**

Illustrative example

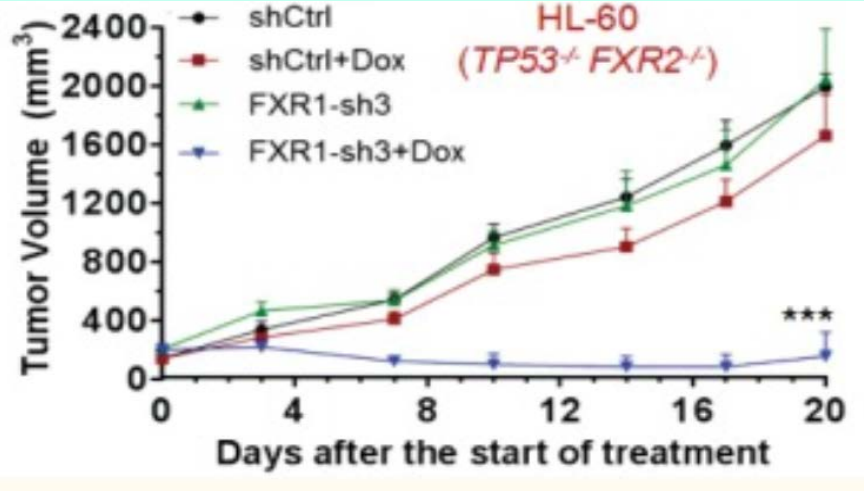
- **FXR2 is located near TP53**
- **FXR1 and FXR2 are paralogs that buffer each other's function**

- **Do FXR1 and TP53 deletions avoid each other?**

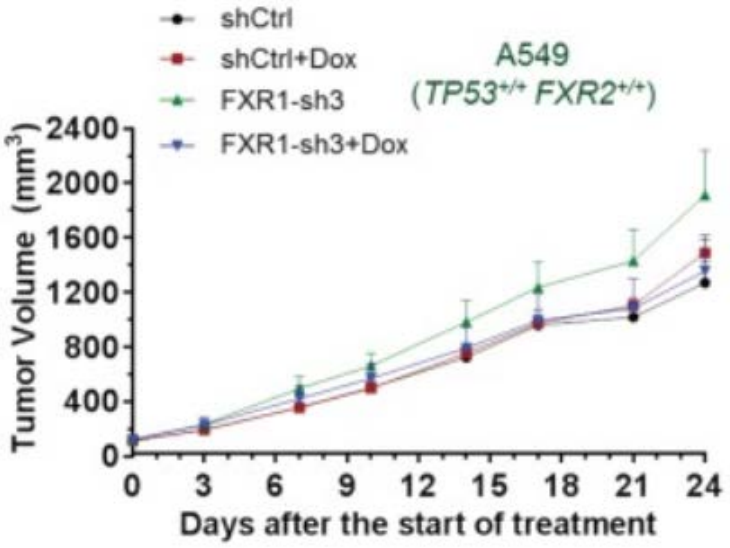


- **Is FXR1 synthetic lethal to TP53?**
- **Does inhibiting FXR1 lead to cell death for TP53-deleted cell lines?**

Tumour bearing homozygous TP53/FXR2 co-deletion shrinks upon doxycycline-induced FXR1 knock down



Fan et al., eLife, 6:e26129, 2017



Summary

- Bewilderment: **Breast cancer survival signatures are no better than random signatures**
- Enlightenment: **SPS genes**

- Bewilderment: **Mutation mutual exclusivity are not associated with synthetic lethality**
- Enlightenment: **Collateral lethality**