Into the blue...Using mouse models to uncover genes driving tumorigenesis and therapy resistance in human breast cancer

Ruiter, J.R. de

Propositions

1. Insertional mutagenesis is an efficient approach for identifying candidate cancer genes and therapy resistance mechanisms using targeted sequencing, but it should be combined with complementary sequencing approaches for comprehensive mutation analyses (chapters 3 and 4 of this thesis).

2. RNA-sequencing-based approaches for detecting transposon insertions improve on DNA-based methods by focusing on insertions that are actually expressed, and by allowing identification of additional mutation types (e.g., expression changes, gene fusions, single nucleotide variants) from a single dataset (chapter 4 of this thesis).

3. If the presence of even a single driver gene can strongly influence tumor evolution, it is crucial to study tumor development, progression and therapy response in mice with the appropriate genetic backgrounds to have any real relevance for human patients (chapter 5 of this thesis).

4. Somatic engineering approaches greatly enhance our capability to generate complex mouse models containing combinations of different driver mutations, whilst retaining the same characteristics as models generated through germline engineering. (chapter 5 of this thesis)

5. To improve the relevance of computational analyses, computational biologists should be well versed in the biological background and should critically evaluate their results from a biological perspective[1–3].

6. To ensure reproducibility of published results, computational biologists should be required to provide supplementary documents and scripts that unambiguously define the data, workflow, execution environment and results of the computational analyses, in order to generate a verifiable re-executable publication[4,5].

7. To improve the accessibility and reproducibility of sequencing analyses, community efforts should focus on creating ‘gold-standard’ sequencing pipelines that can easily be applied to new or existing datasets[6–8].

8. One of the biggest challenges in the current field of computational biology, is collecting (sequencing) data in centralized repositories or portals in a manner that they can easily be queried, combined, filtered and visualized by fellow researchers[9,10].

9. All models are wrong, but some are useful. (George Box)

10. If you want to go quickly, go alone. If you want to go far, go together. (African proverb)

References

1. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Medicine* **2,** 0696–0701 (2005).

2. Begley, C. G. & Ellis, L. M. Drug development: Raise standards for preclinical cancer research. *Nature* **483,** 531–533 (2012).

3. Vidgen, B. & Yasseri, T. P-values: misunderstood and misused. **4,** 10–14 (2016).

4. Ince, D. C., Hatton, L. & Graham-Cumming, J. The case for open computer programs. *Nature* **482,** 485–488 (2012).

5. Check Hayden, E. Rule rewrite aims to clean up scientific software. *Nature* **520,** 276–277 (2015).

6. Köster, J. & Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts480

7. Fisch, K. M. *et al.* Omics Pipe: A community-based framework for reproducible multi-omics data analysis. in *Bioinformatics* **31,** 1724–1728 (2015).

8. Han, B. W., Wang, W., Zamore, P. D. & Weng, Z. PiPipes: A set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome-and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* **31,** 593–595 (2015).

9. Klonowska, K. *et al.* Oncogenomic portals for the visualization and analysis of genome-wide cancer data. *Oncotarget* **7,** 176–192 (2016).

10. Newberg, J. Y., Mann, K. M., Mann, M. B., Jenkins, N. A. & Copeland, N. G. SBCDDB: Sleeping Beauty Cancer Driver Database for gene discovery in mouse models of human cancers. *Nucleic Acids Res.* 1–7 (2017). doi:10.1093/nar/gkx956