nature
genetics

npg

Second call for pan-cancer analysis

We invite proposals by 1 March for pan-cancer analyses from public data sets to be submitted by the end of 2015. We will coordinate expressions of interest from the *Nature* family of journals and index the resulting publications prominently in a web focus and special issue of the journal.

Cancer is a set of many diseases, but the commonalities and contrasts in the biology among cancers are informative. Insights from one tumor type can highlight mechanisms and molecular pathways and predict treatment outcomes for another tumor type. This type of data integration has been termed pan-cancer analysis.

In 2013, we organized publication of the initial investigations of The Cancer Genome Atlas (TCGA) Research Network's pan-cancer analysis group comprising 17 publications across 6 journals (http://www.nature.com/tcga/). Since then, the field has matured, with robust data pipelines delivering many more integrated data sets from different tumor types. The Analysis format—of new hypotheses from existing data, validated in yet more existing data sets, and *in silico* modeling—has gained in popularity as bioinformaticians and molecular geneticists have teamed up to consider this mode of publication.

However, in comparison to the data generation projects that most laboratories carry out, analytical papers are scarcer, more variable in quality and harder to improve to the point where referees endorse them for publication. Many Analyses are simply premature and too small in their scope and application. Proof of concept and testing of new methods are not enough; rather, robust sets of interlocking methods and verifications, together with useful new concepts and strategies, are very much needed. This is why consultation with a range of journals about their standards and benchmarks is desirable when embarking on such publications.

Last month, William Lee and colleagues (*Nat. Genet.* **46**, 1160–1165, 2014) identified noncoding mutations with consequences for gene expression in 863 human tumors, extending the ideas of recurrently mutated drivers and focal mutational hotspots to genes for which no coding mutations have been found. Notably, this work used three distinct analytical methods to identify variants. An implication of this work is the new ability to identify upstream regulators of these mutated genes that function in the gene expression networks responsible for cancer cell phenotypes as well as an ability to begin to understand the genetic changes driving about a quarter of tumors of various tissue origins for which no coding mutations have been identified. Now, in this issue, Erik Larsson and colleagues (page 1258) reexamine 505 tumor genomes of 14 cancer types for somatic mutations altering transcription, finding *TERT* promoter mutations in 6 types but surprisingly few promoter mutations with transcriptional consequences at the whole-tumor level. Also in this issue, Peter Sorger and colleagues generate multiscale networks to integrate biophysical and genomic data from TCGA, finding, for example, that phosphoprotein alterations result in new functions but that SH2 domain alterations cause loss of function (pp1252, 1363). The new approach may call into question the current emphasis on the consistency of somatic mutation recurrence frequency, as the diversity of mutations with calculable consequences was here found to be much greater than anticipated.

Reanalyses of TCGA and International Cancer Genome Consortium (ICGC) data should be carried out and published with consultation and in accordance with the wishes of the data producers but need not be limited to consortium members. In some cases, this may mean coordinating publication with first analyses of tissue-specific cancer genomics data sets. To aid journals with this aim, it is very important that all analyses state the accession codes and versions of any data used and include a statement by the analysis groups that they are aware of any specific data use restrictions declared by the data-producing consortia. It is our view that any analysis sufficiently creative and original to be worthy of publication is unlikely to infringe the declared aims of the data producers to publish the initial analysis of the data they have generated. ∎