

## Call for data analysis papers

**Community standards for data access, interoperability and metadata only make sense if data are creatively reused to further research. We are therefore inviting the submission of Analysis papers that reformat and integrate existing data sets to generate substantial novel insights into gene expression in cell differentiation transitions and different cell fates.**

The journal offers to consider for peer review and subsequent publication any Analysis paper described in a presubmission inquiry by 1 May and submitted by 1 September of this year. Building on the successful model of our focuses on iCOGS (<http://www.nature.com/icogs/>) and The Cancer Genome Atlas (TCGA) pan-cancer analysis (<http://www.nature.com/tcga/>), we will solicit and coordinate editorial interest in these presubmission inquiries from the Nature research journals, *Nature Communications*, *Scientific Reports* and *Scientific Data* and from competing journals, with a view to organizing and indexing this set of parallel publications later in the year. We will also apply advice on data interoperability from a range of experts, including but not limited to BioSharing (<http://www.biosharing.org/>), the Global Alliance for sharing data (<http://www.ebi.ac.uk/about/news/press-releases/Global-Alliance>), ELIXIR (<http://www.elixir-europe.org/>) and the US National Institutes of Health (NIH) Big Data to Knowledge initiative (<http://www.bd2k.nih.gov/>), and offer guidance to authors on adding value to proposed analyses.

We invite new research in any area of gene expression and cell identity (including genetics, epigenetics, development, cell reprogramming, synthetic biology and cancer). Analysis that builds and tests citable falsifiable models and semantic predictions is very much encouraged. An example of the type and quality of work we are inviting is shown in a recent publication that integrates the epigenetic signatures of *cis*-regulatory sites with the transcriptional activity attributable to cell type-specific transcription factors and disease-associated regulatory variation (*Nat. Genet.* **46**, 136–143, 2014). Another example combines data on mutation profiles and structural genome rearrangements from different cancers to infer perturbation of cell signaling pathways and drug targets (*Nat. Genet.* **45**, 1127–1133, 2013).

The publications we are looking for will be Analysis papers containing new hypotheses, predictions and insights from existing data. No new experimental work need be done, but submissions integrating new data with existing data sets will also be considered for publication as Articles. Editors will look primarily for novelty and conceptual advances as well as for value added by reusing previously unconnected data sets. The value may come from adding metadata or making data sets accessible in standard semantic formats supporting machine reasoning or from applying new software or analysis. The resource value of reformatted or opened data sets will be considered.

Peer referees will look for separation of hypothesis generation and validation, so predictions should be tested by simulation or on a comparable data set. Referees often tell us that network models need to show at least three types of robustness, as is well illustrated in Thread 2 of the TCGA focus (<http://www.nature.com/ng/2013/131017/full/ng.2787.html>) under the subheading “Testing the robustness of network assignment,” where we show examples of the technical performance of an analytical program on a gold standard data set and the concordance of its technical performance on missing data. Finally, the authors should compare the results generated with an unseen data set.

Open data, open-source software and open publication are encouraged but will not be mandatory, provided the case can be made that the proprietary resources used are completely described, sufficiently transparent and add value. Publications accepted after peer review and the application of normal editorial criteria at this journal may be published under an open license if they offer a community standard. No special consideration will be made for the business model of the ensuing publications, as data reuse is more important to the future of research than open access. ■