Scientists use public databases to leap over scourge of publication bias
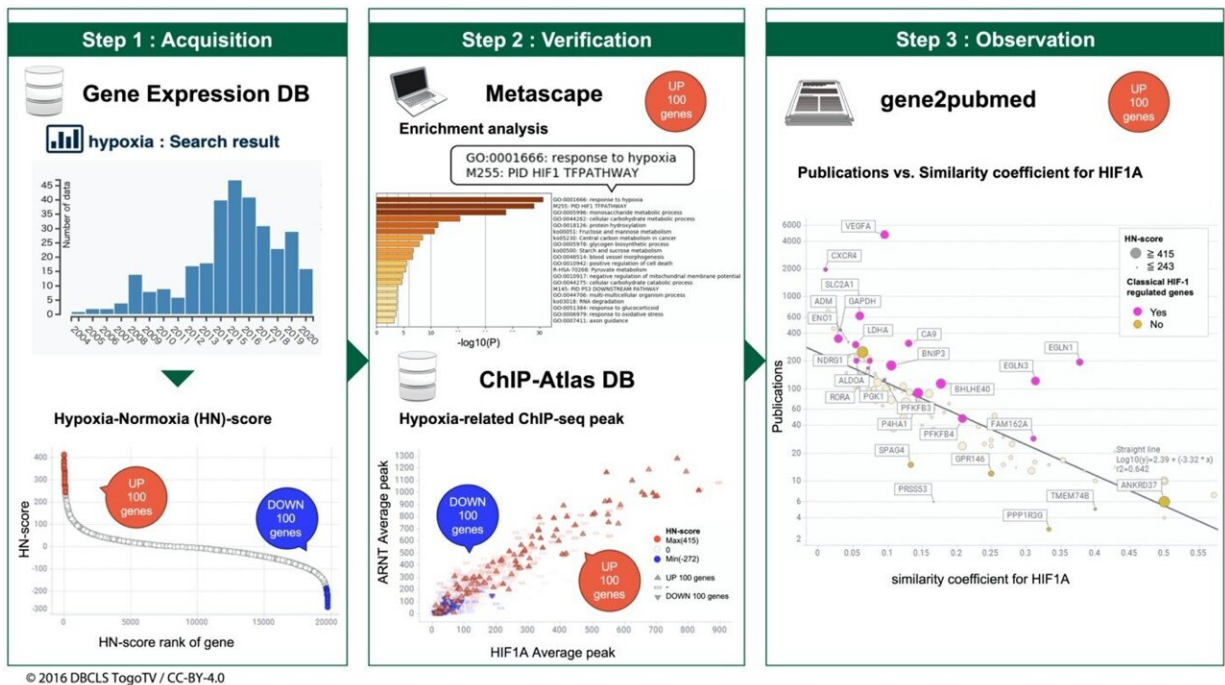
June 9 2021

Schematic view of hypoxic transcriptome meta-analysis.Image caption: Step 1. Evaluation and listing of upregulation and downregulation of hypoxia-inducible genes. Step 2. Confirmation of known hypoxic stimulation-related genes. Step 3. Discovery of novel genes related to hypoxic stimulus-response. Credit: Yoko Ono, Hiroshima University Image usage restrictions: CC-BY-4.0

Scientists have leapt over the emerging problem of publication bias within genetic research by performing a meta-analysis of publicly

available databases of 'transcriptomes," or the full range of messenger RNA molecules produced by an organism. Researchers from Hiroshima University applied the technique to their own field—the study of the genes that are activated when an organism experiences low-oxygen conditions—but it should also be applicable in any other fields that make use of the transcriptome, providing a powerful weapon against the threat posed by publication bias.

The meta-analysis technique was published in a paper appearing in the journal *Biomedicines* in May 2021.

Scientists are often held up as the pinnacle of objective, disinterested observation and investigation. But in recent years, the danger of what is called [publication bias](#), or sometimes the 'file-drawer problem' is being recognized right across the natural sciences. It describes the bias of researchers and of scientific journals toward the publication of results that support the hypotheses of researchers or otherwise show a significant finding. Both researchers and journals are frequently not very excited about experiments that do not support their hypotheses, and so the findings are left 'in the researcher's file drawer." There may be no malevolent intention behind such exclusion, but the lack of publication of these 'boring' results does skew, or bias, what exists across the published [scientific literature](#). Ironically, the more well studied the field, the greater the effect of such publication bias.

The Hiroshima University researchers had noticed that there were some 600,000 publications in [scientific journals](#) that described around 20,000 [human genes](#) that code for the building of a specific protein (as opposed to 'non-coding' [genes](#) that perform other functions but do not code for a protein). Across this enormous number of publications, there were a whopping 9,000 articles that discussed the p53 gene, but some 600 genes were not mentioned at all.

Within their own field, they found that this sort of publication bias had led to a great focus on genes that are already well known to be activated during conditions of hypoxia, or low-oxygen conditions.

Under hypoxic conditions, hypoxia-inducible transcription factors, or HIFs, are produced. Transcription factors are proteins that control the rate of transcription (copying a segment of DNA into RNA in order for the RNA to then be translated into a protein). When human cells are deprived of sufficient oxygen, there is a response system governed by the HIFs that attempts to ameliorate the effects of hypoxia by preventing cells from differentiating and by promoting the formation of blood vessels. And HIFs are known to be activated by certain genes.

"But are there other genes that are activated during hypoxia that all researchers have up till now somehow missed due to this publication bias?" asked lead author Hidemasa Bono, professor in the Graduate School of Integrated Sciences for Life at Hiroshima University.

As transcriptome researchers, Bono and his team knew that, unlike with scientific articles in journals, all transcriptome data have to be archived in public databases, not just the interesting transcriptome data.

"So we thought that if we performed a meta-analysis—or an analysis of analyses—on these transcriptome data, we might be able to identify novel hypoxia-inducible genes that had been buried by publication bias," Bono added.

They searched publicly available transcriptome databases to obtain hypoxia-related experimental data, retrieved the metadata, and then manually curated it. They then selected all the genes that are expressed during hypoxic stimulation, and evaluated their relevance in hypoxia by performing gene enrichment analyses. This latter method allows the statistical identification of groups of genes that are over-represented in a

large set of genes, and so may be associated with a particular condition.

Alongside this, the researchers performed a bibliometric analysis, a statistical method of analyzing the occurrence of particular words or phrases commonly used in library and information science. The bibliometric analysis in this case was used on the gene2pubmed—a gene literature data source that describes which genes have been discussed where in the scientific literature—to identify genes that have not been well studied in relation to hypoxia. This is a new type of analysis that within biomedical research is called the 'bibliome."

By combining the transcriptomic meta-analysis and the bibliome, the researchers were indeed able to find four genes that were not previously known to be associated with hypoxia.

The results have encouraged the researchers to keep going with their meta-analysis technique. They plan to keep using public databases to discover similar new findings, in particular to investigate gene activation (expression) of non-coding RNAs, or RNA molecules that are not translated into proteins, during hypoxic stress.