*Identification of MIEs as subnetworks and hub genes using the Bayesian SBM.* In order to obtain the list of key genes and pathways associated with the exposure to EE2, we utilized lists of top ranking genes and pathways that are reported to be associated with EE2 in the *Comparative Toxicogenomics Database (CTD)* (http://ctd.mdibl.org/). We chose three significantly associated KEGG pathways, including *Metabolic pathways*, *PI3K-Akt signaling pathway*, and *Cell cycle*, which correspond to the first, third, and 13th top ranking pathways in CTD (corrected *p*-values = 4.2e-253, 2.1e-79, and 2.0e-47). Among the genes belonging to these pathways, we further selected those corresponding to the top 500 genes that are reported to be associated with EE2 in CTD. Finally, we applied the Bayesian SBM to the genes belonging to these pathways. One of the key parameters of this model is *K*, the number of subnetworks we want to identify. The Bayesian SBM has an interesting property such that it can identify both subnetworks and hub genes within a unified framework, by

|  |
| --- |
| A B C D    **Figure 4. Subnetworks and hub genes identified using the Bayesian SBM, for the genes belonging to *Metabolic pathways* (circle) and *PI3K-Akt signaling pathway* (triangle) in (A, B) and for the genes belonging to *Metabolic pathways* (circle) and *Cell cycle* (triangle) in (C, D). Colors of nodes in each figure indicate different gene groups identified by the Bayesian SBM. In the Bayesian SBM, the number of gene groups (*K*) was set to 2 for (A) and (C) and set to 4 for (B) and (D).** |
| using different *K* values. Specifically, for small *K*, the Bayesian SBM first identifies hub genes (nodes linked to large numbers of edges) and general subnetworks. Then, as *K* increases, this model sequentially identifies more specific subnetworks and finer sets of hub genes. |

**Fig. 4** displays subnetworks and hub genes identified by the Bayesian SBM. First, we considered the genes belonging to Metabolic pathways (circles in **Fig. 4A and 4B**) and PI3K-Akt signaling pathway (triangles in **Fig. 4A and 4B**). When (**Fig. 4A**), hub genes across two pathways were identified as a group (red: *EGFR*, *GAPDH*, *HSP90AA1*, *ALDOA*, *PGK1*, *FN1*, *LDHA*, *ENO1*), which includes genes important for HIF-1 signaling pathway and glycolysis/gluconeogenesis pathway. Non-hub genes across two pathways were identified as another group (green), which includes genes related to PI3K-Akt and FoxO signaling pathways as well as response to nitrogen/hormone/organic cyclic compound biological processes and positive regulation of cell proliferation. When we increased (**Fig. 4B**), the Bayesian SBM found a group of non-hub genes across two pathways (green) and three groups of hub genes. The three groups of hub genes include 1) hub genes specific to Metabolic pathways (light blue: *ENO1*, *ENO3*), which is related to phenylalanine/tyrosine/ tryptophan biosynthesis pathway and glycolysis/gluconeogenesis pathway, NADH regeneration biological process; 2) hub genes specific to PI3K-Akt signaling pathway (red: *EGFR*, *HSP90AA1*), which is related to regulation of nitric oxide metabolic process and ERBB signaling pathway, ERBB receptor binds tightly to other EGF receptor to form a heterodimer, enhancing kinase-mediated activation of downstream signaling pathways, such as those involving MAPK and PI3K; and 3) hub genes linking two pathways (dark blue: *GAPDH*, *ALDOA*, *FN1*, *PGK1*, *LDHA*), which is related to glycolysis/gluconeogenesis and HIF-1 pathways.

Next, we considered the genes belonging to Metabolic pathways (circles in **Fig. 4C and 4D**) and Cell cycle (triangles in **Fig. 4C and 4D**). Interestingly, this time, when (**Fig. 4C**), the Bayesian SBM nicely separates out the genes belonging to the Metabolic pathways (red; related to glucose/NADH/ hexose/monosaccharide metabolic processes and glycolysis/gluconeogenesis and HIF-1 signaling pathways) from those belonging to Cell cycle (green; related to regulation of cell cycle, p53 and FoxO signaling pathways). When we increased (**Fig. 4D**), the Bayesian SBM identifies the hub genes belonging to Cell cycle (dark blue: *CDK1*, *CDK4*, *CDKN1A*, *GADD45A*, *PCNA*) while each of genes belonging to Metabolic pathways (red; related to glycolysis/gluconeogenesis and HIF-1 signaling pathways), non-hub genes belonging to Cell cycle (green; related to regulation of cell cycle, p53 and FoxO signaling pathways), and outlying genes (light blue; related to metabolic processes) make a separate group. The hub genes belonging to Cell cycle includes fundamental cell cycle and nucleotide excision repair regulators: 1) cyclin-dependent kinases (CDKs) are a family of [sugar](https://en.wikipedia.org/wiki/Sugar) [kinases](https://en.wikipedia.org/wiki/Kinases) well-known for their regulatory role in the [cell cycle](https://en.wikipedia.org/wiki/Cell_cycle) (42); 2) cyclin dependent kinase inhibitor 1A (CDKN1A) inhibits the activity of CDK2 and CDK4 complexes, and thus functions as a regulator of cell cycle progression at G1 (43); 3) growth arrest- and DNA damage-inducible 45a (GADD45a) encodes a predominantly nuclear protein that influences G2/M cell cycle progression, access to DNA on damaged chromatin, genomic stability, nucleotide excision repair, apoptosis, and signaling through the mitogen-activated protein (MAP) kinases p38 and c-Jun N-terminal kinase (JNK) (44); 4) proliferating cell nuclear antigen (PCNA) is found in the nucleus and is a cofactor of DNA polymerase delta that acts as a homotrimer and helps increase the processivity of leading strand synthesis during DNA replication (45). Interestingly, PCNA-CDKN1A and PCNA-GADD45A interactions are essential for nucleotide excision repair (46, 47). This result indicates that the Bayesian SBM has the potential to identify MIEs as subnetworks and hub genes within a unified framework.