Greedy Community Detection Synopsis

Based on script: community\_detection.py

Manually annotated results: greedy-community-detection-manual-annotations.xlsx

Using greedy modularity community detection there are 14 communities in MyGene2. Community 0 is a seemingly meaningless hair ball containing 1788 / 1992 genes & HPO terms and for that reason is ignored.

Of the other communities 7 contains related HPO terms and genes with no readily obvious previous connections to diseases seemingly associated with those HPOs. Of these 7, 4 (communities 6,8,9 and 10) have more than one occurrence of the HPO terms.

**Follow up ideas:**

Alternate community detection algorithms. The implementation I used does not account for edge weight (the number of times two terms co-appeared) I suspect this could be a useful piece of information.

Shift up the HPO tree: Community 3 contained multiple related HPO terms and genes that were not obviously already related to those HPOs, but most of the terms were used in describing just one patient so I excluded the community. But there are multiple patients in this community with unique HPO terms that are semi-related. Shift these one-off terms up the tree until they are no longer one-off could allow for more meaning full community detection.

**Community 6** - several terms about unmovable joints and a gene with function in myosin (important in muscle contraction) and genes with unknown function

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Community | Occurence\_count | Degree | Node | Description |
| 6 | 9 | 6 | HP:0001989 | Fetal akinesia sequence: Decreased fetal activity associated with multiple joint contractures |
| 6 | 5 | 5 | HP:0000282 | Facial edema: Facial puffiness |
| 6 | 7 | 7 | MYLK4 | Myosin Light Chain Kinase Family Member 4 |
| 6 | 7 | 7 | TMEM39B | Transmembrane Protein 39B |
| 6 | 7 | 7 | MYH2 | myosin heavy chain 2: Myosins are actin-based motor proteins, functions in skeletal muscle contraction |
| 6 | 7 | 7 | DNAH7 | Dynein Axonemal Heavy Chain 7: component of the inner dynein arm of ciliary axonemes |
| 6 | 6 | 6 | HP:0002803 | Congenital contracture: bent joint that cannot be straightened actively or passively |
| 6 | 6 | 6 | HP:0002202 | Pleural effusion: Fluid around lungs |

**Community 8** - multiple types of polydactyl (multiple copies of a limp) and a gene associated with bone growth but not previously associated with polydactyl!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Community | Occurence\_count | Degree | Node | Description |
| 8 | 1 | 1 | HP:0000456 | Bifid nasal tip: splitting of the nasal tip |
| 8 | 1 | 1 | HP:0001829 | Foot polydactyly: supernumerary toe |
| 8 | 2 | 2 | HP:0002984 | Hypoplasia of the radius: nderdevelopment of the radius |
| 8 | 7 | 7 | BGN | small leucine-rich proteoglycan: plays a role in bone growth, muscle development and regeneration, and collagen fibril assembly in multiple tissues |
| 8 | 1 | 1 | HP:0010503 | Fibular duplication |
| 8 | 4 | 3 | HP:0010442 | Polydactyly |
| 8 | 1 | 1 | HP:0009556 | Absent tibia |

**Community 9** – cancer & tumor types, some genes do not have known cancer association

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Community | Occurence\_count | Degree | Node | Description |
| 9 | 6 | 6 | SAE1 | SUMO1 Activating Enzyme Subunit 1: assocaited with Adult Dermatomyositis. |
| 9 | 4 | 4 | HP:0002665 | Lymphoma: Cancer of lymphatic system |
| 9 | 4 | 4 | HP:0002671 | Basal cell carcinoma |
| 9 | 6 | 6 | PLK3 | Polo Like Kinase 3: Polo-like kinases are important regulators of cell cycle progression: regulates TP53 |
| 9 | 6 | 6 | UHRF1BP1L | UHRF1 Binding Protein 1 Like: Associated with Leukodystrophy, Hypomyelinating. Known cancer association |
| 9 | 6 | 6 | TBCB | Tubulin Folding Cofactor B: Hypoparathyroidism-Retardation-Dysmorphism Syndrome and Pleural Disease. No 100% clear association with cancer |
| 9 | 5 | 5 | HP:0030430 | Neuroma: Nerve tumor |

**Community 10** – light sensitivity HPOs and DNA repair genes with no association to visual light sensitivity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Community | Occurence\_count | Degree | Node | Description |
| 10 | 2 | 2 | HP:0000613 | Photophobia: excessive sensitivity to light with the sensation of discomfort or pain in the eyes |
| 10 | 4 | 4 | HP:0000563 | Keratoconus: one-shaped deformity of the cornea characterized by the presence of corneal distortion secondary to thinning of the apex |
| 10 | 2 | 2 | HP:0000992 | Cutaneous photosensitivity: increased sensitivity of the skin to light. |
| 10 | 6 | 6 | UVRAG | UV Radiation Resistance Associated: associated with Heterotaxy (left right swap) and colon cancer |
| 10 | 2 | 2 | HP:0030127 | Endometriosis: rowth of endometrial tissue outside the uterus |
| 10 | 6 | 6 | ERCC3 | ERCC excision repair 3: associated with trichothiodystrophy (short brittle hair) xeroderma pigmentosum (DNA repair problem, cancer risk), Cockayne syndrome (DNA repair broken) |