A.1 Chemical-gene relationships

Here we describe the major clusters in Figure 2 of the main paper (reproduced below). We use "C" to represent a chemical, and "G" to represent a gene.

The first major cluster, cluster 3, refers to inhibition (the chemical, C, is an inhibitor of the protein, G). This is reported mostly in a static context in patterns such as “C, a G inhibitor” and “G inhibition by C”. The mechanism behind the inhibition is usually unclear from these descriptions. Is C inhibiting the activity of the protein G or the expression of G’s mRNA? It’s difficult to tell.

Clusters 5 and 6 specifically describe effects on protein activity with 6, the larger cluster, referring mainly to situations where C is an agonist or antagonist of G. Antagonists are often referred to as “blockers” or “inhibitors”, while agonists are referred to as “activators” or “ligands”.

Clusters 8, 9, 10 and 11a all describe effects on mRNA and protein levels, rather than protein activity. Cluster 10 specifically refers to inhibition, while the effects in clusters 8 and 9 are mixed: some positive, some negative, some neutral. Cluster 11a sometimes refers to a treatment response, as though C is administered as a therapy or the paper is investigating G’s response to C.

The dependency paths in clusters 14-16 all describe the binding of C to a protein, G, which is usually a receptor for C. In the associated sentences, C is often an endogenous compound, such as an amino acid or hormone.

Clusters 11c and 19-21 contain relationships of reverse directionality from the rest of the dendrogram. The relationships we have described
so far relate to situations where the chemical, C, acts on the protein, G, perhaps by inhibiting it, inducing its activity, or raising/lowering its expression/synthesis. Instead, clusters 11c and 19-21 describe situations where the protein acts on the chemical: enzymes that modify chemical structures, transporters that shuttle chemicals across cell membranes, and a variety of other pharmacokinetic (PK) relationships.

Cluster 11c contains most of the PK relationships, including effects of G on C’s metabolism and situations where C is actually a metabolite produced by G after acting on some other chemical. Some transport relationships are also found here, though most of these are in clusters 19 and 21.

Cluster 20 refers to enzymatic modification of C by G. Usually G is an enzyme that specifically targets C and contains C’s name within its own name.

While cluster 11c contained some fine-grained local structure – dependency paths specifically referring to metabolism or secretion tended to cluster close together, for example – it was surprisingly difficult to distinguish different classes of PK relationships within this cluster.

We did not assign themes to the last major group of clusters in the dendrogram (clusters 23-30) because these reflected a major class of errors where part of a protein, such as an amino acid or specific binding domain like a zinc finger, was misidentified as a chemical. While amino acids and elements like zinc are chemicals, the relationships reflected here are whole-part, not interactions between distinct entities.
Table A1: Cluster descriptions for chemical (C) – gene (G) interactions, following the cluster numbers illustrated in Figure 2 in the main text.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Size</th>
<th>Theme</th>
<th>Selected Descriptive Patterns</th>
<th>Entity Pair with Pattern (C / G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>36</td>
<td>inhibition</td>
<td>“C, a G inhibitor”</td>
<td>ARRY-614 / p38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G specific inhibitor, C”</td>
<td>naringenin / Smad3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C, an inhibitor of G”</td>
<td>PSC_833 / P-glycoprotein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G inhibition by C”</td>
<td>NVP-AUY922 / Hsp90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effects of the G inhibitor, C, on…”</td>
<td>SCH_34826 / enkephalinase</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>effect on protein activity</td>
<td>“[chemical]-dependent effects of C on G activity”</td>
<td>fenfluramine / renin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effect of C on G activity”</td>
<td>donepezil / acetylcholinesterase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“inhibition of G activity by C”</td>
<td>plumbagin / Nox-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“study on interaction of C with G”</td>
<td>caffeine / myoglobin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G activity in patients on C”</td>
<td>tacrolimus / CYP3A4</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>agonism / antagonism</td>
<td>“effect of C, a selective G antagonist”</td>
<td>MTEP / mGluR5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C, a G agonist”</td>
<td>roxindole / 5-HT1A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“inactivation of G by C”</td>
<td>mitomycin_C / DT-diaphorase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G agonist, C, …”</td>
<td>cigitazone / PPAR-gamma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“study of a G antagonist, C, …”</td>
<td>CI-988 / CCK-B_receptor</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>secretion, production, synthesis</td>
<td>“effect of G on C metabolism”</td>
<td>dopamine / cholecystokinin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C inhibits G secretion”</td>
<td>Dasatinib / TNF-alpha</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effects of G on C metabolism”</td>
<td>steroid / angiotensin_ii</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G stimulates C production”</td>
<td>prostaglandin_E2 / interleukin-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“upregulation of C synthesis by G”</td>
<td>prostaglandin_E2 / interleukin-1beta</td>
</tr>
<tr>
<td>9</td>
<td>76</td>
<td>affects expression</td>
<td>“C inhibits G expression”</td>
<td>AG490 / NFATc1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effect of C on G production”</td>
<td>neopterin / erythropoietin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C induces the expression of G”</td>
<td>Nicotine / C-reactive_protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C upregulates G expression”</td>
<td>Dexamethasone / Kv1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effects of C on the expression of G”</td>
<td>letrozole / HOXA10</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>inhibition of activity / expression</td>
<td>“the new G inhibitors, C and…”</td>
<td>rofecoxib / COX-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“the effect of G inhibition by C”</td>
<td>tolcapone / COMT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“inactivation of G by C”</td>
<td>carbodiimides / thrombin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“C effects on G: …”</td>
<td>Naloxone / beta-endorphin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“the effect of C on G activity in…”</td>
<td>aspartame / acetycholinesterase</td>
</tr>
<tr>
<td>11a</td>
<td>62</td>
<td>response to treatment</td>
<td>“G responses to C”</td>
<td>cimetidine / Prolactin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effect of C on G”</td>
<td>gossypol / LDH-X</td>
</tr>
</tbody>
</table>
| 11c | 96 | metabolism, secretion/uptake | “effect of G on C metabolism”
“effects of G on C formation”
“effect of G on the secretion of C”
“control of G by C”
“G stimulates C uptake”
“[chemical] may reduce G concentration via C”
“G stimulates C transport”
“effect of C on G release” | ribavirin / interferon_alpha_2b
clotfibrate / insulin
amines / renin
dopamine / beta-endorphin
cyclic_AMP / adrenomedullin
omeprazole / intrinsic_factor
retinoic_acid / c-jun
phenylalanine / Insulin
catecholamines / leptin
calcium / Prolactin
fenfluramine / growth_hormone |
| 11d | 23 | binding (uptake/release) | “C release from G”
“binding of C to G”
“C uptake by G”
“enhancement of action of G by C”
“controlled release of G by C” | iron / transferrin
calcium / troponin_C
potassium / HKT1
glucose / insulin-like_growth_factor_1
polyurethane / IGF-1 |
| 13 | 18 | modulation of expression, substrates | “C modulates [event] through G”
“C binding to G”
“C induced by G”
“C is a G substrate”
“C mediates [event] by G” | Metformin / SIRT1
cyanide / myeloperoxidase
nitric_oxide / iNOS
caffeine / cytochrome_p450_1a2
superoxide / c-Src |
| 14 | 24 | receptor binding | “antagonist of the G C receptor”
“effect of C receptors, G and...”
“[chemical] antagonism of a G C agonist”
“interaction of G with C receptors...”
“a new selective G agonist, C...” | tachykinin / NK1
steroid / pS2
dopamine / D-2
estrogen / DYX1C1
procaterol / beta_2-adrenoceptor |
| 15 | 14 | receptors | “G, a C receptor...”
“deletion of the C G gene...”
“G, a major C receptor...”
“the C domain of G”
“analysis of G C channels...” | free_fatty_acid / GPR40
adenosine / A1_receptor
somatostatin / SSTR4
zinc / SIP1
potassium / KCNQ2 |
| 16 | 16 | receptor [subunit] | “the C carrier subunit (G) of...”
“increased expression of C receptor (G)”
“the G subunit of the C receptor”
“human C receptor subunit (G)”
“C receptor G subunits” | acyl / NDUFAB1
benzodiazepine / PBR
NMDA / GluN2B
acetylcholine / CHRNA4
AMPA / GluR1 |
| 19 | 38 | channels / transporters | “regulation of G transporters C and...”
“G is a C channel that modulates...” | sterol / ABCG5
chloride / MOD-1 |
| 20 | 15 | synthase, dehydrogenase, reductase | "neuronal C synthase (G)"  
"C transporter (G) polymorphism"  
"porcine C reductase (G)"  
"C dehydrogenase (G)"  
"C synthase (G) gene" | glucose / GLUT4  
glutamate / VGLUT1  
potassium / Kv4  
nitric oxide / nNOS  
serotonin / 5-HTT  
thiol / GILT  
Aldehyde / Ald4p  
5-aminolevulinate / ALAS1 |
| 21 | 10 | transporters | "G, a C transporter, ..."  
"low-affinity C cotransporter (G)"  
"a C binding protein, G"  
"C transfer protein (G) polymorphism"  
"C binding protein (G)" | ribavirin / ENT1  
sodium glucose / SGLT2  
methyl_CpG / Mecp2  
Cholesteryl_estер / CETP  
fatty_acid / hFABP |
| 24 | 24 | sequence, factor, moiety | "complete C sequence of G"  
"G is C exchange factor"  
"binding of [chemical] to the C moiety of G"  
"structural analysis of the C finger of G"  
"C binding domains of G" | amino_acid / GSTM4  
guanine_nucleotide / Rab3GEP  
heme / cytochrome_P-450  
zinc / THAP1  
nucleotide / CFTR |
| 27 | 74 | phosphorylation / kinases | "expression of receptor C kinase, G"  
"G receptor C kinases"  
"[gene name] (G) C phosphorylation"  
"C phosphorylation pathway"  
"implication of G and C kinase in..." | tyrosine / Trk  
tyrosine / ErB  
tyrosine / GIT1  
serine / STAT3  
creatine / esterase_D |
| 29 | 13 | phosphorylation / phosphatases | "G induces C phosphorylation"  
"G C phosphatase"  
"C phosphorylation sites on G"  
"conserved C residues in G"  
"a critical C residue in G" | tyrosine / Oncostatin_M  
tyrosine / Shp2  
Serine / IRS2  
histidine / lipoxygenase  
lysine / apolipoprotein_B-100 |
| 30 | 20 | inhibition / activation (via phosphorylation?) | "[other chemical] inhibits C activation of G"  
"efficacy of G C kinase inhibitors"  
"surface of G C domain"  
"discovery of C G inhibitors"  
"G induces rapid C phosphorylation" | phenylephrine / phospholipase_A  
tyrosine / EGFR  
zinc / TFIIIB  
glycine_hydraziđe / CFTR  
tyrosine / Prolactin |
A.2 Chemical-disease relationships

Here we describe the major clusters in Figure 3 of the main paper (reproduced below). We use "C" to represent a chemical, and "D" to represent a disease.

By far the largest set of chemical-disease relationships (from clusters 8g, 8h and 9) are treatment relationships, in which a chemical, C, is described as a treatment or potential treatment for a disease, D. Similar to cluster 3 from Figure 2, these relationships are mostly described in a static context: we don’t know why C is a useful treatment for D, but it is described as such without further elaboration.

While we did not choose to separate clusters 8g, 8h and 9 into different themes, there are subtle differences among these three clusters. Cluster 8g mostly describes evaluation of efficacy; C is investigated as an experimental treatment for D, or patients are described as receiving C for D without indication of whether C is useful. Dependency paths in cluster 8h tend to go further, indicating that the treatment was efficacious for D. Finally, cluster 9, a small cluster with only 14 dependency paths, includes statements about using C to prevent or reduce D, which is slightly different than treating D. However, due to the substantial similarities among these three clusters (some variant of the phrase “treatment for” appears in all three), we labeled all of them with the same theme.

Cluster 6, which also involves the word “treat”, refers mainly to the evaluation of side effects in C-treated patients. Despite its proximity to clusters 8 and 9 in the dendrogram, it is semantically more related to clusters 15 and 16, which describe side effects. In these clusters, D is not a disease that C is used to treat, but a side effect or adverse event resulting from treatment with C.
Cluster 20, which is close in meaning to clusters 15 and 16, includes statements implicating C in the pathogenesis of D. Here C is most often an endogenous compound. Whereas in clusters 15 and 16 we tend to see situations where a drug is intentionally administered to a patient or animal, causing an adverse event, cluster 20 refers to cases where levels of C (most often in serum or tissue) are associated with the risk or progression of D. These levels may result from external supplementation or overproduction of an endogenous compound by the body.

Related to cluster 20 are clusters 18 and 19, which describe biomarkers. In these situations, C is not implicated in the pathogenesis of D, but is instead referred to as an indicator, or marker, of disease progression. There is considerable overlap with the patterns used in cluster 20, but again the shift in meaning is subtle - a substance can be an indicator of D without causing D.

Finally, several clusters relate closely to the concept of disease treatment, but rather than stating “C is a treatment for D”, they instead describe observations about what C is doing. Clusters 1, 9, 21, 24, and 28 all refer to situations where C prevents D, or reduces the risk of D (note that cluster 9 appears both in the “prevents” theme, Pr, and in the “treatment/therapy” theme, T, in Table 2 in the main text). In contrast, clusters 26 and 30 refer to cases where C alleviates D, or reduces its effect. The implication here is that C is being used after D has already occurred.
Table A2: Cluster descriptions for chemical (C) – disease (D) interactions, following the cluster numbers illustrated in Figure 3 in the main text.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Size</th>
<th>Theme</th>
<th>Selected Descriptive Patterns</th>
<th>Entity Pair with Pattern (C / D)</th>
</tr>
</thead>
</table>
| 1              | 13           | prevents, reduces incidence | “C and [other drug] reduce [adverse event] after D”  
“C decreased levels of [substance] after D”  
“D of patients treated with C”  
“[women, men] receiving C to prevent D”  
“intravenous C reduces the incidence of D” | Isoflurane / cerebral_ischemia  
estrone / brain_injury  
triptans / coronary_spasm  
nevirapine / HIV-1_vertical_transmission  
magnesium / arrhythmias |
| 2              | 20           | inhibits growth / proliferation | “C significantly inhibited the growth of D”  
“C inhibits proliferation of D cells”  
“C inhibited [event(s)] in D cells”  
“C inhibited D growth”  
“C inhibits D growth in vitro” | celastrol / osteosarcoma  
Darbepoetin / hepatic_cancer  
NVP / RCC  
sorafenib / tumor  
Zebularine / acute_myeloid_leukemia |
| 3              | 46           | induction of effects in cells, esp. resistance; chemotherapy | “[event] induced by C in D cells”  
“C therapy for D”  
“C resistance in D”  
“D resistant to both C and [other drug]”  
“chemotherapy agents like C in D treatment” | fenretinide / neuroblastoma  
cisplatin / thoracic_malignancies  
Tamoxifen / breast_cancer  
imatinib / GIST  
doxorubicin / hepatocellular_carcinoma |
| 6              | 15           | treatment evaluations (esp. safety) | “C was measured in patients with D”  
“we evaluated the effects of C on D”  
“C is indicated for D”  
“C administered before/after D reduced [event]”  
“treatment of D with C” | Glutamic_acid / ischemic_stroke  
diphenidol / chronic_constriction_injury  
Bicillin_C-R / streptococcal_infections  
nicardipine / coronary_artery_occlusion  
sulfasalazine / juvenile_spondyloarthropathies |
| 8g             | 125          | treatment of disease (esp. evaluation of efficacy) | “C therapy for the treatment of D”  
“patients who received C for treatment of D”  
“D patients were treated with C”  
“effectiveness of C in D”  
“comparison of C and [other drug] in D” | indomethacin / PDA  
tigecycline / Acinetobacter_infections  
DMSO / amyloid_A_amyloidosis  
warfarin / atrial_fibrillation  
timolol / angle-closure_glaucoma |
| 8h             | 80           | treatment of disease (indication of efficacy) | “C may be useful for the treatment of D”  
“evaluate the protective efficacy of C in D”  
“C is a promising treatment option for patients with D”  
“C is approved for the treatment of D”  
“C is commonly prescribed for D” | OPC-18790 / congestive_heart_failure  
FTY720 / cerebral_ischemia  
bosutinib / CML  
anidulafungin / intra-abdominal_abscesses  
Colchicine / gout |
| 9              | 14           | treatment of disease | “C may be used for the prevention of D”  
“in [children, patients] with D following C treatment” | melatonin / premature_aging  
MPH / ADHD |
<table>
<thead>
<tr>
<th>Page</th>
<th>Quantity</th>
<th>Type</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>37</td>
<td>side effects (association)</td>
<td>“D associated with C therapy”&lt;br&gt;“the use of C has been associated with D”&lt;br&gt;“C intake was associated with D”&lt;br&gt;“D occurred after C”</td>
<td>clozapine / tachycardia&lt;br&gt;moxalactam / thrombocytopenia&lt;br&gt;caffeine / shorter_nocturnal_sleep_duration&lt;br&gt;oxaliplatin / hypersensitivity_reaction&lt;br&gt;alfentanil / hypotension</td>
</tr>
<tr>
<td>16</td>
<td>67</td>
<td>side effects (causal implications) / studies inducing effect</td>
<td>“administration of C resulted in D”&lt;br&gt;“C induces D”&lt;br&gt;“D was induced by administration of C”&lt;br&gt;“D was/were induced by infusion of C”&lt;br&gt;“patient developed D after receiving C”</td>
<td>vincristine / thrombocytopenia&lt;br&gt;Taxol / myalgias&lt;br&gt;lidocaine / Hypotension&lt;br&gt;ouabain / Cardiac_arrhythmias&lt;br&gt;ceftaroline / eosinophilic_pneumonia</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
<td>potential biomarkers</td>
<td>“C levels of D patients were significantly [lower/higher]…”&lt;br&gt;“monitoring of C in D rats”&lt;br&gt;“reduced C in D subjects”&lt;br&gt;“significant elevations of C in D subjects”&lt;br&gt;“effect of C on [biomarker level / event] in D patients”</td>
<td>homocysteine / hyperthyroid&lt;br&gt;homocysteine / hypertensive&lt;br&gt;selenium / asthmatic&lt;br&gt;leucine / MSUD&lt;br&gt;clozapine / schizophrenic</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>potential biomarkers</td>
<td>“effect of C supplementation in D”&lt;br&gt;“we studied the effect of C on D”&lt;br&gt;“C was well tolerated in [patient group] with D”&lt;br&gt;“blood C concentrations in patients with D”&lt;br&gt;“examine the C status of our D patients”</td>
<td>vitamin_D3 / Autism_Spectrum_Disorder&lt;br&gt;rosiglitazone / angiogenesis&lt;br&gt;tolterodine / incontinence&lt;br&gt;vitamin_C / diabetes_mellitus&lt;br&gt;magnesium / chronic_ambulatory_peritoneal_dialysis</td>
</tr>
<tr>
<td>20</td>
<td>63</td>
<td>levels associated with disease risk / progression</td>
<td>“high C levels are associated with increased risk of D”&lt;br&gt;“C implicated in D”&lt;br&gt;“effect of D on serum C levels”&lt;br&gt;“patients with D and increased C concentrations”&lt;br&gt;“C has been implicated in the pathogenesis of D”&lt;br&gt;“C intake may be associated with [lower/higher] risk of D”&lt;br&gt;“C supplementation and incidence of D: …”</td>
<td>cholesterol / coronary_heart_disease&lt;br&gt;bisphosphonates / osteonecrosis&lt;br&gt;testosterone / prostate_cancer&lt;br&gt;triglyceride / unstable_angina&lt;br&gt;Serotonin / migraine&lt;br&gt;PUFA / colorectal_neoplasia&lt;br&gt;beta-carotene / cancer</td>
</tr>
<tr>
<td>21</td>
<td>13</td>
<td>changed incidence / risk</td>
<td>“C use was associated with [increased/decreased] risk of D”&lt;br&gt;“C reduce(s) the risk of D”&lt;br&gt;“C may reduce the incidence of D in…”&lt;br&gt;“C was associated with a [lower/higher] risk of D”&lt;br&gt;“relation of C to risk of D”</td>
<td>Warfarin / ICH&lt;br&gt;Bisphosphonates / osteoporotic_fractions&lt;br&gt;Eicosapentaenoic_acid / cardiovascular_disease&lt;br&gt;Preconception_O3 / GDM&lt;br&gt;cholesterol / coronary_heart_disease</td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td>inhibits, suppresses</td>
<td>“C inhibited [other event] in D”&lt;br&gt;“the D action of C”&lt;br&gt;“C suppresses D through [mechanism]”</td>
<td>Ki23057 / gastric_tumours&lt;br&gt;diltiazem / hypotensive&lt;br&gt;Evodiamine / hyperalgesia</td>
</tr>
</tbody>
</table>
|   |   | “influence of C on D development”  
|   |   | “C significantly suppressed D”  
|   |   | histamine / seizure  
|   |   | AS1069562 / allodynia  
| 26 | 48 | inhibited / blocked disease progression  
|   |   | “the effects of C on the progression of D”  
|   |   | “C may protect against D”  
|   |   | “C blocked D in organ culture”  
|   |   | “C antagonized [other drug-induced] D”  
|   |   | “C attenuates D in mice”  
|   |   | “C ameliorated D by [mechanism]”  
|   |   | “C alleviates D in [disease model]”  
|   |   | minocycline / encephalopathy  
|   |   | Eicosapentaenoic_acid / atherosclerotic_disease  
|   |   | phenethyl caffeiate / hyperplasia  
|   |   | procyclidine / seizures  
|   |   | Simvastatin / pulmonary_fibrosis  
|   |   | EGB / endothelial_dysfunction  
|   |   | Propentofylline / hypersensitivity  
| 28 | 17 | preventive effects evaluated  
|   |   | “examine the effects of C on D”  
|   |   | “study was carried out to evaluate the effect of C on D”  
|   |   | “investigated possible beneficial effects of C on D”  
|   |   | “to assess the effect of C on D”  
|   |   | “C effective for the prevention of D”  
|   |   | metformin / cytotoxicity  
|   |   | atorvastatin / inflammation  
|   |   | AdCbl / atopic_dermatitis  
|   |   | nebivolol / endothelial_dysfunction  
|   |   | dronedarone / atrial_fibrillation  
| 30 | 23 | reduced, abolished, prevented  
|   |   | “C prevents D”  
|   |   | “C, a [description], prevented D”  
|   |   | “C is beneficial in D”  
|   |   | “D was reduced by C”  
|   |   | “C was effective in reducing D”  
|   |   | Itraconazole / fungal_infections  
|   |   | AMD3100 / anxiety_behaviors  
|   |   | lithium / tauopathies  
|   |   | gabapentin / Pain  
|   |   | buspirone / overall_anxiety_symptoms  

A.3 Gene-disease relationships

Here we describe the major clusters in Figure 4 of the main paper (reproduced below). We use "G" to represent a gene, and "D" to represent a disease.

Clusters 2h, 4, 6, 8, and 9 contain relationships that are quite similar to cluster 20 in Figure 3. All of these clusters describe situations where a protein (or chemical, in Figure 2 cluster 20) is implicated in the pathogenesis of a disease. Clusters 4 and 6 refer simply to increased levels of G in D, whereas clusters 8 and 9 more directly implicate the protein in D pathogenesis. Cluster 29 reflects a slightly different theme in which the protein promotes disease progression, rather than disease onset. The two themes share some overlap but are subtly different; cluster 29 focuses on cancers, discussing proteins promoting cell invasion, proliferation, and progression.

Clusters 2j and 3 include therapeutic relationships, where G is described as a treatment or potential treatment of D. Cluster 3 mostly describes trials of G in the treatment of D. While there are a few statements that could perhaps imply efficacy, such as “G therapy for patients with D”, the treatment relationships here are not described with anywhere near the definiteness of clusters 8 and 9 in Figure 3.

Clusters 5 and 7 are similar to clusters 18 and 19 in Figure 3 in that they do not ascribe a pathogenic role to the protein (or chemical) but instead refer to it as a biomarker. Cluster 7 contains statements where a protein, G, is described as “a robust diagnostic biomarker for D”, or “an indicator of D”, without insinuating that it causes D. Cluster 5 is very closely tied to cluster 6, but cluster 6 contains a few statements with causal implications, such as “G is a mediator of D”.

In clusters 10 and 12, the protein, G, is described as a drug target or potential target for the treatment of the disease,
D. Often this description does not include the word “target”, but it is implied - the statement refers to the utility of G inhibitors in treating D, for example.

Some statements in clusters 10 and 12 refer to mutations in G that have an effect on D. It’s implied that disruptions in the activity of G can impact the course of D. Clusters 13 and 14 address the issue of mutations more directly, either by describing studies that investigate the role of G mutations in the progression of D (cluster 13) or by directly implicating mutations in G as causal risk factors in D (cluster 14).

While clusters 5 and 7 refer specifically to biomarkers, clusters 15, 17 and 30 refer to overexpression of proteins in disease, usually in patient serum. These proteins could represent potential biomarkers as well, although they are not described in that way.

Clusters 18, 19 and 21 focus on regulation, specifically cases where improper regulation of a gene is linked to disease. There is substantial overlap between these ideas and those of overexpression, biomarkers, etc. but again the focus is subtly different.

The last set of clusters, 22, 26 and 27, focus explicitly on polymorphisms that increase disease risk. The terms “polymorphism”, “mutation”, and “variant” are all present. Cluster 22 focuses almost exclusively on tumor suppressor genes, which, when mutated, can cause cancers. Note that in this case it is mutations in the gene (the DNA) that are increasing risk, rather than the level or activity of a protein. There is some semantic overlap with clusters 13 and 14.
Table A3: Cluster descriptions for gene (G) – disease (D) interactions, following the cluster numbers illustrated in Figure 4 in the main text.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Size</th>
<th>Theme</th>
<th>Selected Descriptive Patterns</th>
<th>Entity Pair with Pattern (G / D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2h</td>
<td>44</td>
<td>therapeutic effects, esp. drug sensitivity, resistance</td>
<td>“G and response to [drug] in patients with D”&lt;br&gt;“G resistance in patients with D”&lt;br&gt;“serum G levels are associated with D”&lt;br&gt;“G sensitivity in D”&lt;br&gt;“comparison of G and [other drug] for detection of D”</td>
<td>TCF7L2 / type_2_diabetes&lt;br&gt;insulin / systemic_lupus_erythematosus&lt;br&gt;leptin / hepatic_steatosis&lt;br&gt;insulin / hypertension&lt;br&gt;cardiac_troponin_1 / ischemic_myocardial_injury</td>
</tr>
<tr>
<td>2j</td>
<td>33</td>
<td>influences disease treatment (some adjuvant therapies)</td>
<td>“the use of G in the treatment of D”&lt;br&gt;“D in patients treated with G”&lt;br&gt;“effect of G on [event] in D patients”&lt;br&gt;“G therapy in patients with D”&lt;br&gt;“efficacy of G in D”</td>
<td>parathyroid_hormone / osteoporosis&lt;br&gt;interferon_alpha_2b / Acute_renal_failure&lt;br&gt;prolactin / systemic_lupus_erythematosus&lt;br&gt;Erythropoietin / chronic_renal_failure&lt;br&gt;S-1 / colorectal_cancer</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>protein causes change in disease status</td>
<td>“injected G induces D”&lt;br&gt;“G promotes D”&lt;br&gt;“regulation of [event] by G in D”&lt;br&gt;“G inhibits D”&lt;br&gt;“G exacerbates D”</td>
<td>IL-1 / anorexia&lt;br&gt;VEGF-D / metastasis&lt;br&gt;TDP-43 / frontotemporal_lobar_degeneration&lt;br&gt;High-mobility_group_box_1 / ulcer_healing&lt;br&gt;VDUP1 / bacteremic_shock</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>levels / expression in disease</td>
<td>“G levels in D patients”&lt;br&gt;“expression of G in D”&lt;br&gt;“increased G levels in patients with D”&lt;br&gt;[“regulation/function] of G system in D”&lt;br&gt;“G level in D”</td>
<td>Interleukin-6 / headache&lt;br&gt;SFRP4 / primary_serous_ovarian_tumours&lt;br&gt;thyroglubulin / nontoxic_goiter&lt;br&gt;interleukin-6 / stroke&lt;br&gt;C-reactive_protein / atopic_dermatitis</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>levels / expression in disease</td>
<td>“G levels in patients with D”&lt;br&gt;“G levels in D patients”&lt;br&gt;“effects of [drug] on G in D patients”&lt;br&gt;“serum G levels in D”&lt;br&gt;“expression of G in D”</td>
<td>interleukin-6 / glomerulonephritis&lt;br&gt;Interleukin-2 / multiple_sclerosis&lt;br&gt;insulin / hypertensive&lt;br&gt;E-selectin / Kawasaki_disease&lt;br&gt;E-cadherin / carcinomas</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>biomarkers, diagnostic</td>
<td>“G is a robust diagnostic biomarker for D”&lt;br&gt;“G is an independent predictor of D”</td>
<td>TLE1 / synovial_sarcomas&lt;br&gt;Proinsulin / coronary_heart_disease</td>
</tr>
<tr>
<td>Page</td>
<td>Number</td>
<td>Section</td>
<td>Relevant Text</td>
<td>Relevant Genes/Proteins</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>---------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>role in pathogenesis</td>
<td>association of G with [event] in patients with D&quot;</td>
<td>Plasma_hyaluronidase / atherosclerosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;effects of G on D&quot;</td>
<td>TGFbeta-1 / breast_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;role of G in the development of D&quot;</td>
<td>chromogranin-A / neuroendocrine_tumors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;role of G in the pathogenesis of D&quot;</td>
<td>SERPINA3 / preeclampsia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;a novel gene, G, is associated with D&quot;</td>
<td>FCGR2A / rheumatoid_arthritis</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>role in disease course / pathogenesis</td>
<td>clinical impact of circulating G in D&quot;</td>
<td>miR-18a / oesophageal_squamous_cell_carcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G attenuates D&quot;</td>
<td>Wnt5a / pulmonary_arteriolar remodeling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G predicts [event] in patients with D&quot;</td>
<td>LTBP2 / acute_dyspnoea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;evidence for role of G in D&quot;</td>
<td>BRCA1 / gastric_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G: the link between D and [other disease]&quot;</td>
<td>HMGB1 / diabetes_mellitus</td>
</tr>
<tr>
<td>10</td>
<td>32</td>
<td>inhibitors used as therapies</td>
<td>&quot;G inhibitors in D: …&quot;</td>
<td>ACE / aortic_stenosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;D with G mutation(s)&quot;</td>
<td>TARDBP / amyotrophic_lateral_sclerosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;response to G inhibitors in patients with D&quot;</td>
<td>EGFR / squamous_cell_carcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G testing and management of D&quot;</td>
<td>EGFR / NSCLC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G gene amplification in D&quot;</td>
<td>c-erbB-2 / nasopharyngeal_carcinoma</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>drug targets (esp. cancer)</td>
<td>&quot;G signaling in D cells&quot;</td>
<td>Akt / colon_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G inhibitors in the treatment of D&quot;</td>
<td>MEK1/2 / malignancies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G as a strategic target in D therapy&quot;</td>
<td>ErbB1 / breast_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G: an attractive target for D therapy&quot;</td>
<td>Angiopoietin-2 / tumor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;[drug]: a C inhibitor for the treatment of D&quot;</td>
<td>tumor_necrosis_factor_alpha / rheumatoid_arthritis</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>evaluation of role of mutations in disease</td>
<td>&quot;G mutations in D&quot;</td>
<td>KRAS / lung_adenocarcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;mutations in G in D&quot;</td>
<td>GUSB / mucopolysaccharidosis_VII</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;characterization of G expression in D&quot;</td>
<td>MUC1 / papillary_thyroid_carcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G mutations are associated with [event] in D&quot;</td>
<td>KRAS / colorectal_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;role of G in D development&quot;</td>
<td>RSK2 / osteosarcoma</td>
</tr>
<tr>
<td>14</td>
<td>91</td>
<td>causal mutations</td>
<td>&quot;mutation of G in a patient with D&quot;</td>
<td>STK11 / Peutz-Jeghers_syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G mutation is associated with D&quot;</td>
<td>MTHFR / arterial_stroke</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;novel mutation in G gene associated with D&quot;</td>
<td>MYH7 / distal_myopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;characterization of G mutations causing D&quot;</td>
<td>GALC / Krabbe_disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;mutations of the G gene in patients with D&quot;</td>
<td>COL1A2 / osteogenesis_imperfecta</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;D: a novel G mutation…”</td>
<td>CISD2 / Wolfram_syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;D: novel G mutations and…”</td>
<td>NPC1 / Niemann-Pick_type_C_disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;the recurrent mutation of G in C patients&quot;</td>
<td>BRCA1 / breast_cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;G mutations can cause D&quot;</td>
<td>HIBCH / Leigh-like_disease</td>
</tr>
</tbody>
</table>
| 15 | 13 | levels, concentrations, expression | “G levels in patients with D”  
“serum G concentrations in D”  
“G expression in D cell lines”  
“diagnostic value of G in D patients”  
“prognostic relevance of G in D” | renin / thoracic_neuroblastoma  
leptin / hyperinsulinemia  
TIMP-1 / prostate_tumor  
interleukin_17 / lung_cancer  
CCN3 / Ewing_sarcoma |
| 17 | 12 | levels, overexpression | “serum G concentrations in patients with D”  
“serum G level in patients with D”  
“G overexpression in D”  
“G is overexpressed in D”  
“G expression in D patients” | erythropoietin / anemia  
thyroglobulin / subacute_thyroiditis  
cyclin_D3 / follicular_thyroid_carcinoma  
FOXG1 / hepatoblastoma  
SPARC / pancreatic_cancer |
| 18 | 43 | expression, mutations correlated with disease | “presence of G gene mutation in D patients”  
“frequency of G mutations in D”  
“association of D with G mutations”  
“association of G expression with D”  
“correlation between G expression and [event] in D” | BRAF / melanoma  
PTEN / thyroid_cancer  
PDH / cerebral_dysgenesis  
FcRn / lung_abnormalities  
COX-2 / colon_cancer |
| 19 | 32 | gene expression, regulation | “down-regulation of G in D cells”  
“expression of G mRNA in D”  
“mRNA expression of G in patients with D”  
“D cells expressing G”  
“regulation of G expression in D cells” | E-cadherin / breast_cancer  
CerbB-2 / nasopharyngeal_carcinomas  
KCNO1 / long_QT Syndrome_type_1_and_2  
P-gp / acute_myeloid_leukemia  
CYP1A1 / medulloblastoma |
| 21 | 66 | gene expression in cell lines | “G expression in D”  
“G expression in patients with D”  
“analysis of G expression in D”  
“effects of G on D cells”  
“G expression in D cells” | c-mpl / hematologic_disorders  
trypsinogen-1 / ulcerative_colitis  
SLC34A2 / ovarian_tumors  
p53 / hepatocellular_carcinoma  
MMP2 / prostate_cancer |
| 22 | 28 | tumor suppressor genes | “G as a D suppressor”  
“G acts as a D suppressor”  
“the gene G is a functional D suppressor”  
“G, a novel D suppressor”  
“G: a mediator of D” | Caspase-2 / tumour  
ECRG4 / tumor  
GADD45G / tumor  
SynCAM / tumor  
P-glycoprotein / melanoma_invasion |
| 26 | 26 | polymorphism | “association of variants of G with D”  
“association of the G polymorphisms with D”  
“genetic polymorphisms at G are associated with D”  
“mutations in the G gene in patients with D”  
“G polymorphisms are associated with D” | factor_V_Leiden / thrombosis  
interleukin-18 / type_1_diabetes  
SIRT1 / carotid_atherosclerosis  
P-protein / encephalopathy  
Chromogranin_A / hypertensive_renal_disease |
| 27 | 28 | polymorphism | “association of G gene polymorphism with D”  
“polymorphism of G in D”  
“mutation of the G gene in D” | vascular_endothelial_growth_factor / colon_cancer  
angiotensin-converting_enzyme / sarcoidosis  
endothelin-3 / Waardenburg-Hirschsprung_disease |
A.4 Gene-gene relationships

Here we describe the major clusters in Figure 5 of the main paper (reproduced below). We use "G1" to represent the first gene, and "G2" to represent the second gene.

The cluster themes in Figure 5 were the most difficult to parse among all the dendrograms. The vast majority of protein-protein relationships reflect some kind of change in activity or expression in the second protein based on the action of the first protein. Many of the relationships are similar to chemical-gene relationships in that a protein binds to another protein (cluster 10), increases its expression (clusters 21 and 22), or affects its expression in some other way that is not stated (clusters 7 and 17). All of these themes also appear in Figure 2.

However, there are a few other themes that are specific to protein-protein interactions. One protein can enhance the response of another to some stimulus (cluster 13), or activate or stimulate another protein by itself (clusters 14 and 16). A protein can be produced by a cell population expressing another protein, as in the case of lymphocytes (i.e. proteins produced by CD4-bearing T-cells), which is reflected in clusters 1, 2 and 6.

Clusters 24, 25, 28 and 30 all reflect similar relationships involving regulation and pathways, but are subtly different. Cluster 24 explicitly refers to signaling, with both protein members forming part of the same signaling pathway. Cluster 25 is a cluster of patterns reflecting abbreviations, where the two proteins involved are literally identical or part of the same protein complex. Clusters 28 and 30 speak more specifically of regulation, but contain several patterns that also refer to co-membership in the same pathway. All of these concepts are related.
Table A4: Cluster descriptions for gene (G1) – gene (G2) (usually protein-protein) interactions, following the cluster numbers illustrated in Figure 5 in the main text.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Size</th>
<th>Theme</th>
<th>Selected Descriptive Patterns</th>
<th>Entity Pair with Pattern (G1 / G2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90</td>
<td>cell populations</td>
<td>“G1 induction of human G2”</td>
<td>C5a / interleukin_1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“increased induction of G2 in G1 lymphocytes”</td>
<td>CD8 / interferon-gamma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 + G2 T-cell population”</td>
<td>CD25 / Foxp3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G2 induction of G1”</td>
<td>NfkappaB / Interleukin-1beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 expression [on, in] G2 T-cells”</td>
<td>CD161 / CD8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“an enriched G1 + G2 T-cell subset”</td>
<td>CD4 / CD8beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1-dependent G2 activation”</td>
<td>ERK / CREB</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>cell populations, regulation</td>
<td>“regulation of G2 expression by G1”</td>
<td>SOX10 / MITF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 induces G2 gene transcription”</td>
<td>TNF-alpha / MUC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“regulation of G2 by G1”</td>
<td>RECK / matrix_metalloproteinase-9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G2 expression in the G1 + cells”</td>
<td>CD34 / Bcl-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 / G2 ratio”</td>
<td>CD39 / CD8</td>
</tr>
<tr>
<td>6</td>
<td>39</td>
<td>cell populations, protein production / gene expression</td>
<td>“G1 production by G2 + T cells”</td>
<td>IL-17A / CD146</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 producing G2 + T cells”</td>
<td>IL-10 / CD8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 signaling in G2 + T cells”</td>
<td>IFN-gamma / CD4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 expression on G2 + T cells”</td>
<td>CXCR3 / CD8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“the role of G1 in the function of G2 + T cells”</td>
<td>CD28 / CD25</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>inhibits / induces expression</td>
<td>“G1 induces G2 expression”</td>
<td>Fos / Neurotensin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 inhibits G2 expression”</td>
<td>IL-15 / IL-7Ra</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“effect of G2 on G1 production”</td>
<td>MMP-9 / calcitonin-gene-related_peptide</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 secretion in G2 cells”</td>
<td>cholecystokinin / STC-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 induced G2 production”</td>
<td>TNF-alpha / TARC</td>
</tr>
<tr>
<td>10</td>
<td>76</td>
<td>binding, regulation of activity</td>
<td>“G1 binds G2”</td>
<td>HJURP / CENP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G2 interaction with G1”</td>
<td>Bcl-xL / Clusterin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 is a receptor for G2”</td>
<td>CD96 / CD155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 binding to G2”</td>
<td>Haptoglobin / apolipoprotein_A-I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 mediates activation of G2”</td>
<td>Bcl10 / NF-kappaB</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>enhances response (esp. hormones)</td>
<td>“G1 enhances [event] via G2”</td>
<td>Glypican-4 / insulin_receptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“changes in the G1 response to G2”</td>
<td>prolactin / thyrotropin-releasing_hormone</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 and G2 responses to [event]”</td>
<td>Prolactin / TRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“G1 in G2 receptor signaling”</td>
<td>Fc_gama_RI / p72syk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“exaggerated G2 response of G1”</td>
<td>thyrotropin-releasing_hormone / prolactin</td>
</tr>
</tbody>
</table>
| 14 | 67 | activation, stimulation, signaling | “G2 activates [protein] via G1”  
“G1 stimulates G2”  
“G1 modulates G2 signaling”  
“G2 stimulates G1 expression”  
“G1 induces phosphorylation of G2” | fucosyltransferase_1 / Calreticulin  
Akt / SREBP1c  
Hsp27 / p53  
EGFR / MUC1  
Thrombopoietin / STAT5 |
| --- | --- | --- | --- | --- |
| 16 | 23 | activation, targeting | “function of G2 in G1 receptor activation”  
“G2 promotes [event] by targeting G1”  
“G1 phosphorylation by G2”  
“role of G1 in the activation of G2”  
“regulation of G1 expression by G2” | TNFR1 / Ubc13  
EPB41L3 / miRNA-223  
NuMA / CDK1  
PP4 / JNK-1  
FGF8 / androgen_receptor |
| 17 | 13 | affects production (mostly induces) | “G2 induces the production of G1”  
“[protein] stimulates G2 production via G1”  
“regulation of G2 production by G1”  
“downregulation of G2 by G1”  
“enhancement of G2 by G1” | IgG1 / IL-27  
ERK1/2 / granulocyte_colony-stimulating_factor  
IFN_gamma / IL-18  
mir-25 / mitochondrial_calcium_uniporter  
TNF-alpha / IFN-gamma |
| 21 | 28 | induces expression / production | “G2 induces G1 production”  
“G1 modulates G2 expression”  
“induction of G1 expression by G2”  
“G2 upregulates G1 expression”  
“G1 stimulates G2 secretion in [cell type] cells” | beta-defensin-2 / Tat  
Stat3 / heat_shock_27kDa_protein  
iNOS / IL-1beta  
p16INK4a / p33ING1b  
Angiotensin_II / endothelin-1 |
| 22 | 56 | induces release / production | “G1 induces G2 expression”  
“G2 stimulates G1 secretion”  
“G1 stimulates G2 release”  
“G2 stimulates G1 production”  
“effect of G2 on G1 secretion” | CXCL12 / connective_tissue_growth_factor  
atrial_natriuretic_peptide / Thrombin  
Bradykinin / tissue_plasminogen_activator  
MCP-1 / Angiotensin_II  
renin / neuropeptide_Y |
| 24 | 62 | signaling, receptor binding | “G2 signaling via G1”  
“G1 / G2 costimulatory interactions”  
“coactivator G1 in G2 transcriptional activation”  
“G2 G1 signaling”  
“the G2 G1 receptor”  
“binding of G2 to the G1 receptor” | SMO-1 / TGF-beta  
ICAM-1 / LFA-1  
CBP / p53  
TCF / beta-catenin  
TNF / p55  
interleukin-1 / interleukin-18 |
| 25 | 26 | same or related protein: abbreviations | “G1 (G2) inhibitor”  
“expression of G1 (G2) protein”  
“G2 (G1) activity”  
“G2 (G1) expression”  
“G1 / G2 complexes” | mammalian_target_of_rapamycin / mTOR  
pentraxin_3 / PTX3  
PON1 / paraoxonase-1  
AURKA / Aurora_kinase_A  
PAI-1 / vitronectin |
| 28 | 26 | regulation of expression / production | “the roles of G1 / G2 in [event]”  
“G2 (G1) expression” | MMP-2 / TIMP-2  
M-CSF / macrophage_colony-stimulating_factor |
|   |   | activity | “binding of G1 / G2 proteins”  
|   |   |   | “G2 regulates G1 activity”  
|   |   |   | “synergistic effect of G1 / G2”  
|   |   |   | NF-kappa_B / Rel  
|   |   |   | RhoA / Shp-2  
|   |   |   | IL-6 / BSF-2  
| 30 | 28 | regulation of expression / activity | “upregulation of G2 activity by G1”  
|   |   |   | “regulation of G1 expression by G2”  
|   |   |   | “G1 regulation of G2”  
|   |   |   | “G2 regulation by the G1 pathway”  
|   |   |   | “prognostic significance of G1, G2, …”  
|   |   |   | CD28 / interleukin-4  
|   |   |   | TNF-alpha / TGF-beta  
|   |   |   | miR-133b / Connective_Tissue_Growth_Factor  
|   |   |   | JNK / ATF2  
|   |   |   | bcl-2 / PCNA |