# Expanded View Figures

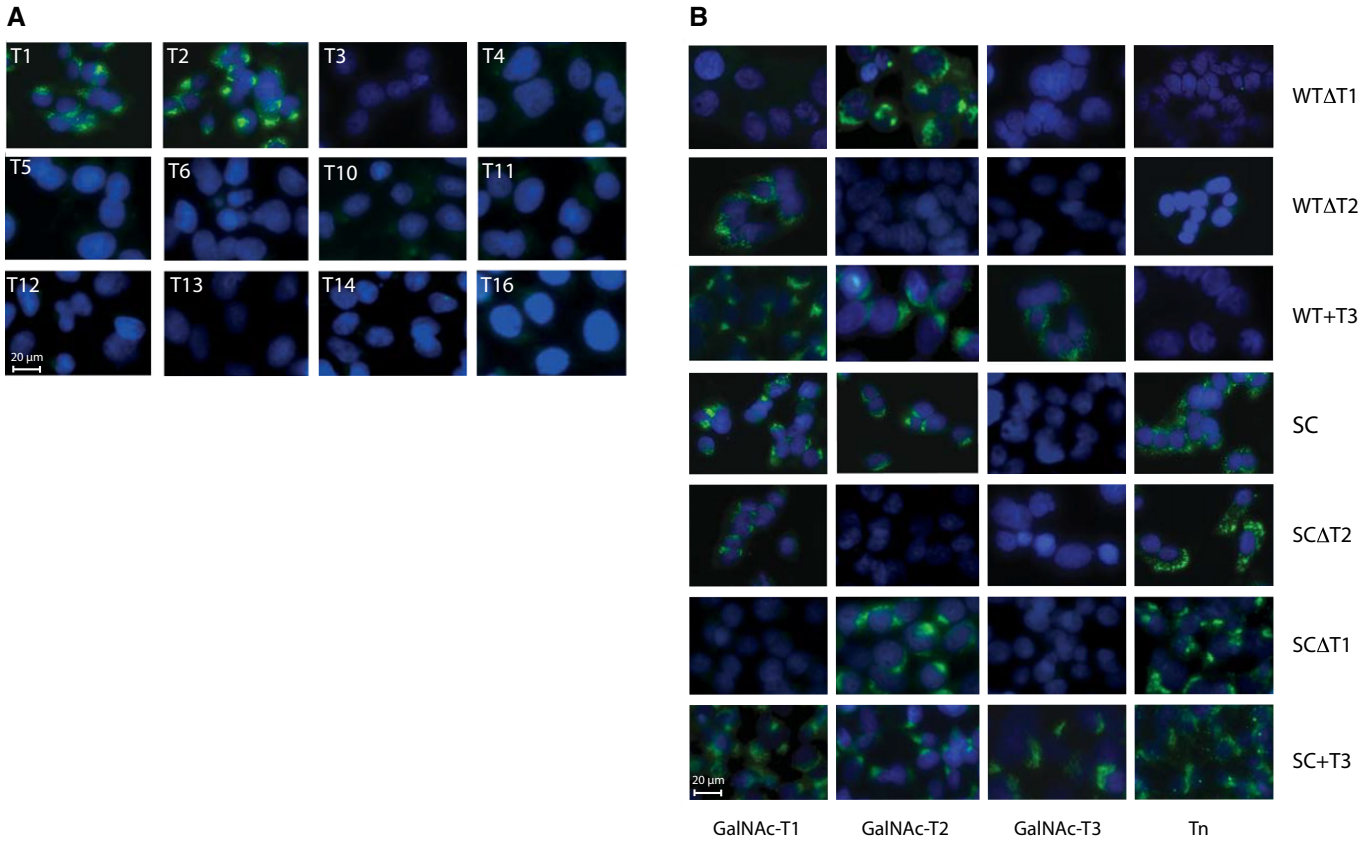



**Figure EV1. Characterization of GalNAc-T expression in HepG2$^{WT}$ and KO model cell lines.**

A ICC of HepG2$^{WT}$ cells with MAbs to a panel of human GalNAc-Ts. Scale bar, 20 μm.

B Characterization of generated isogenic HepG2 KO cell lines by ICC with MAbs to GalNAc-Ts and Tn (clone 5F4). The intensity of GalNAc-T3 labeling and subcellular localization was essentially identical to what we have previously demonstrated in a variety of human cancel cell lines expressing T3 [8]. Scale bar, 20 μm.

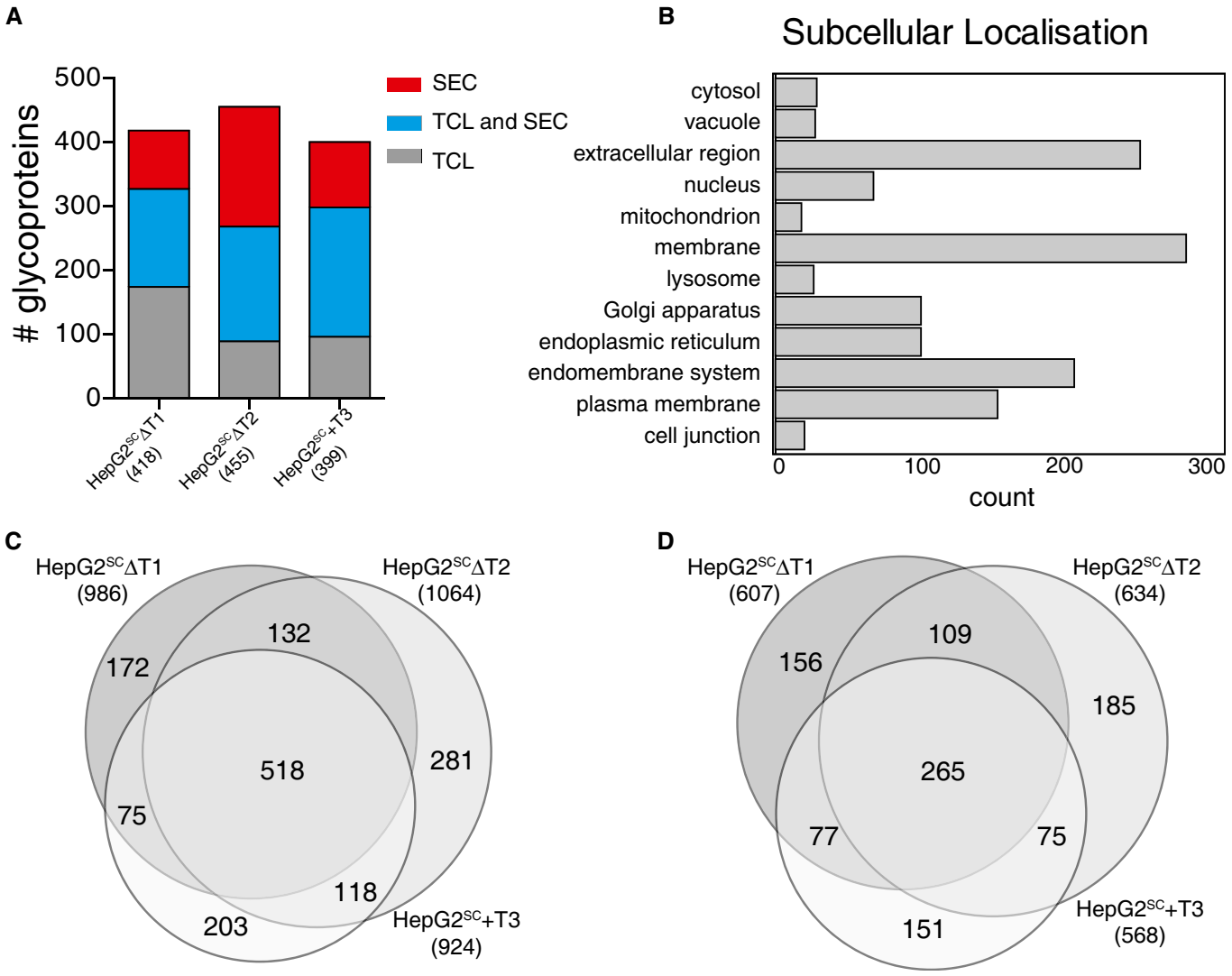



**Figure EV2. Differential O-glycoproteomes in HepG2$^{SC}$.**

A Column diagram showing the relative distribution of identified glycoproteins between TCL, SEC, and overlap between TCL and SEC in all samples HepG2$^{SC}$ΔT1, HepG2$^{SC}$ΔT2, and HepG2$^{SC}$+T3, illustrating that a comparable number of glycoproteins were identified in the three paired sample sets.

B Subcellular localization ontology of identified O-glycoproteins from HepG2$^{SC}$ showing an overrepresentation of extracellular, membrane bound, Golgi- and ER-resident proteins. Interestingly, we identified nuclear- and cytoplasmic-resident proteins, albeit at much lower frequencies. GalNAc and GlcNAc are exact isobars, and in this type of O-glycoproteomics studies, there is potential for including contaminating O-GlcNAc glycopeptides [9,10]. Upon closer examination, approximately 50% of these proteins had multiple GO annotations, including terms relating to the secretory pathway. However, manual inspection of the remaining proteins revealed that these had no annotations to the secretory pathway and were predicted not to have signal peptide sequences, suggesting that these did represent contaminating glycopeptides. These proteins were excluded from the statistical analysis and listed separately (Table EV2).

C, D Venn diagrams showing the distribution of all quantified peptides (C) or all quantified mono-glycosylated peptides (D) between HepG2$^{SC}$ΔT1, HepG2$^{SC}$ΔT2, and HepG2$^{SC}$+T3.

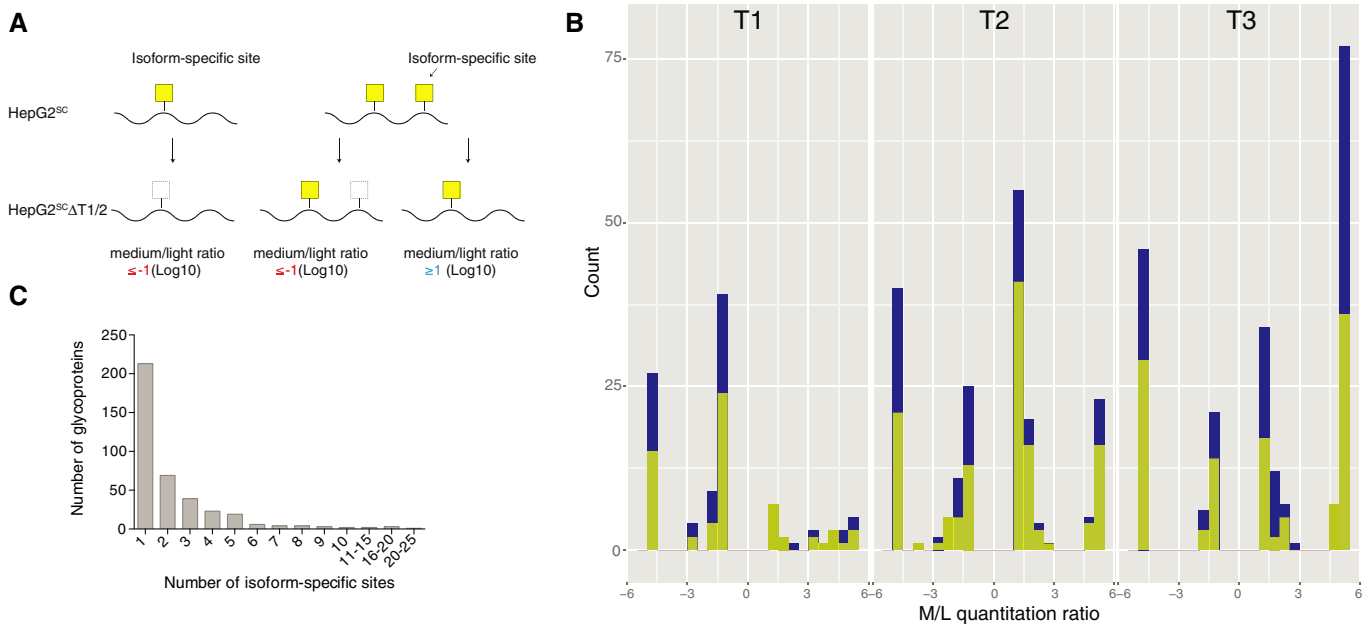



**Figure EV3. Considerations for isoform specificity.**

A Schematic depiction of how low and high M/L ratios and singlets may occur in both directions. Mono-glycosylated peptides with isoform-specific sites may be lost or downregulated (M/L ratio $\leq -1$ ($Log_{10}$)), and glycopeptides with two close adjacent O-glycosites, of which one is isoform-specific, may result in *de novo* appearance or upregulated (M/L ratio $\geq 1$ ($Log_{10}$)) of a mono-glycosylated peptide with the remaining non-isoform-specific site.

B Histogram showing the proportion of mono-glycosylated peptides that have potential glycosylation sites (sites previously identified in 12 human cell lines [2]) within the peptide sequence (green). Blue, total number of quantitated monoglycosylated peptides. A slightly higher proportion of the upregulated singlet glycopeptides for GalNAc-T1 (67% vs. 58%), GalNAc-T2 (72% vs. 51%), and the downregulated glycopeptides for GalNAc-T3 KI (60% vs. 41%), had previously been identified as glycopeptides with two or more O-glycosites, indicating that at least some of these sites may appear as the result of loss of (T1/T2) or gain of (T3) an isoform-specific glycosite.

C Distribution of isoform-specific sites per glycoprotein.

**Figure EV4. *In vitro* GalNAc-T enzyme assay validation of candidate isoform-specific O-glycosites.**

A 20-mer peptides designed to cover identified candidate isoform-specific O-glycosites as acceptor substrates for recombinant human GalNAc-T1, GalNAc-T2, and GalNAc-T3 enzymes (see Table EV3 for a summary of data). Total number of peptides tested for each isoform is shown in parenthesis. Since GalNAc-T3 is not endogenously expressed in HepG2 cells, we disregarded overlapping activity of this enzyme with a few substrates in qualifying isoform specificities of GalNAc-T1 and GalNAc-T2 in the HepG2 cellular context. Blue color marks isoform-specific peptides, dark gray marks peptides glycosylated by multiple isoforms, including the specific isoform (Tx), and light gray color the peptides glycosylated by multiple isoforms excluding the specific isoform (Tx). In testing GalNAc-T1 candidate isoform-specific sites, only 4/14 peptide designs served as substrates for any of the enzyme isoforms tested, and of these, 3/4 were found to be specific substrates for T1. For GalNAc-T2, 23/27 peptides were used by the enzymes tested and of these, 19/23 were specific for T2. For GalNAc-T3, 14/18 peptides served as substrates for the tested enzymes and of these, 9/14 were specifically glycosylated by T3.

B For comparative testing of the glycosites that were not predicted to be isoform-specific in the HepG2$^{SC}$ΔT1 analysis, we found that 2/35 peptides were weak substrates for T1, 14/35 was glycosylated by several isoforms, and 19/35 were not glycosylated by any isoform tested. For the isoform non-specific sites found in the HepG2$^{SC}$ΔT2, 0/38 peptides were specific for T2, 20/38 were glycosylated by several isoforms and 18/38 did not serve as substrates for any isoforms. For HepG2$^{SC}$ΔT3, 0/32 peptides were specific for T3, 16/32 were glycosylated by several isoforms and 16/32 were not glycosylated *in vitro*.

C Substantial efforts have been devoted to searching for consensus motifs for O-GalNAc glycosylation [11]. We therefore aligned the candidate isoform-specific glycosite sequences and analyzed these by frequency plots showing the relative frequency of amino acids ± 10 sites from GalNAc-T1, GalNAc-T2, or GalNAc-T3 isoform-specific glycosylation sites. For GalNAc-T1-specific sites, we demonstrated a preference for acidic amino acids in position +1 and −3 and for basic His in position +2 and −4. For GalNAc-T2-specific sites, we found a more clear preference (4- to 12-fold enrichment) for Pro in positions −1 and −3 and Trp in position −2. Interestingly, we also observed a diminishment of certain charged residues in position −3 to +1, as His and Asp residues are never found in these positions. For GalNAc-T3-specific sites, we found a subtle enrichment for basic and hydrophobic residues in positions −1, −2, +1, and +3. Interestingly, in general methionine, cysteine and the large bulky tryptophan residues are under-represented in sequences surrounding isoform-specific sites of all three isoforms. Overall, these results are well in line with previous *in vitro* glycosylation studies [11], and support the idea that isoform discriminating amino acid preference exists at least to some extent.

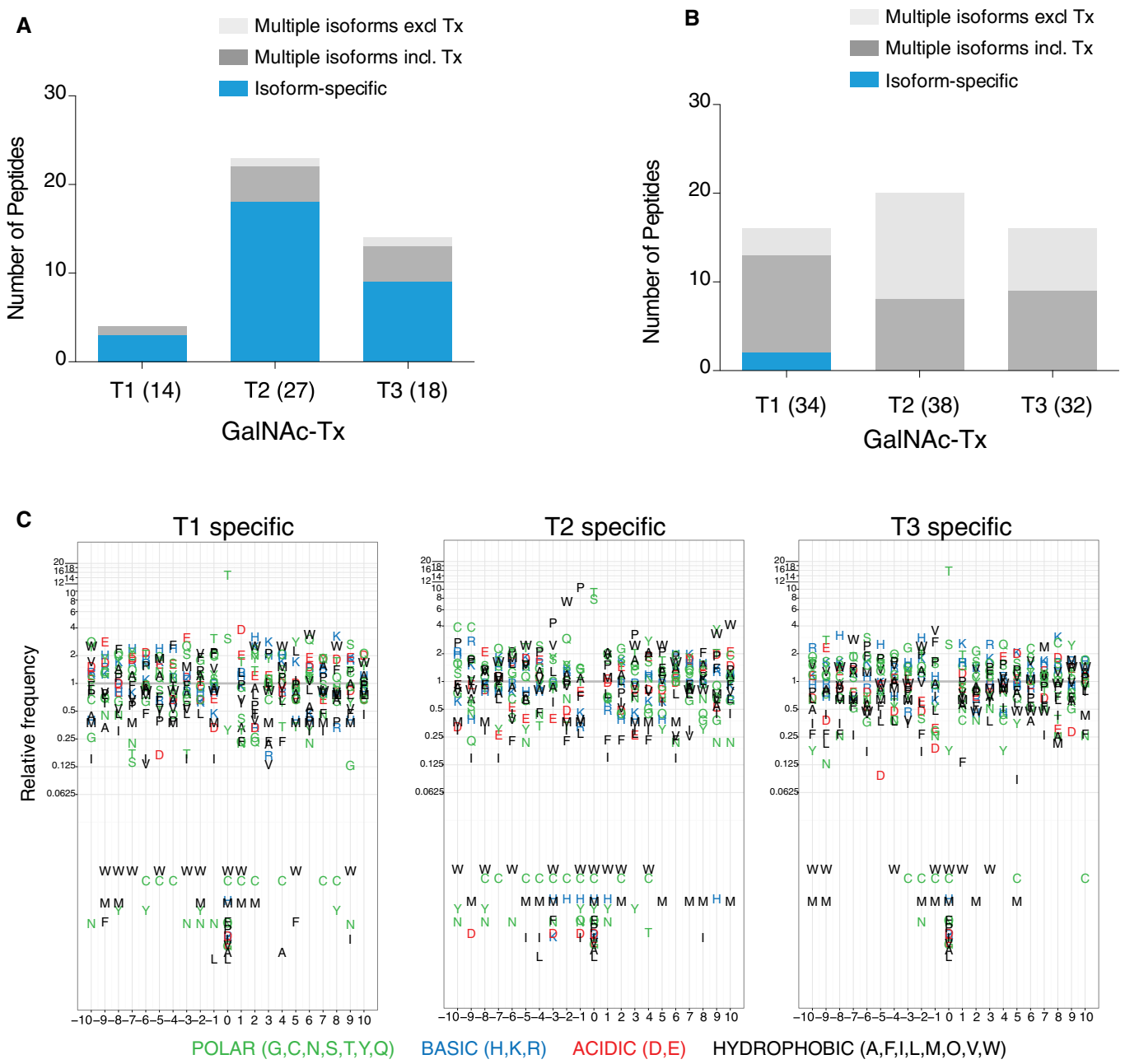



**Figure EV4.**

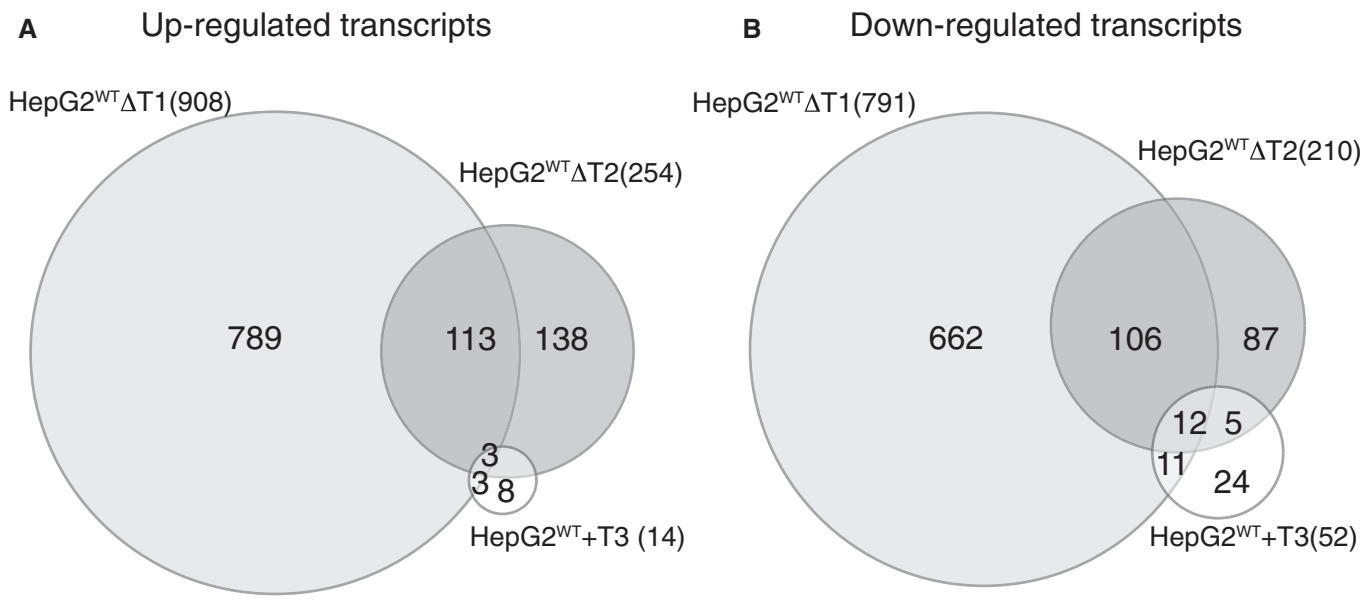
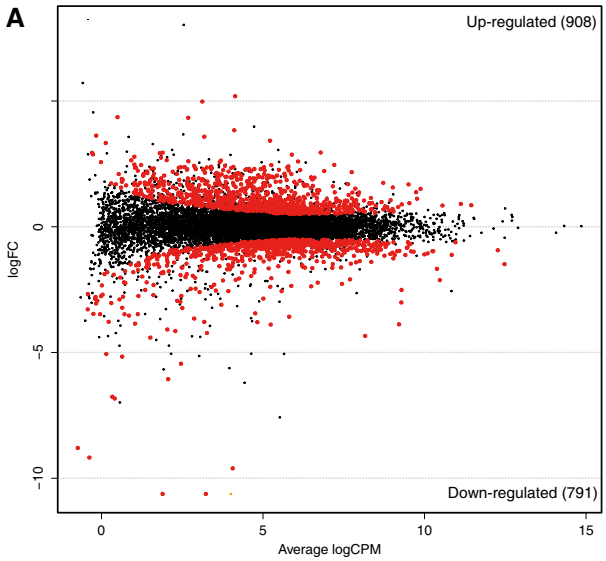



**Figure EV5. RNAseq transcriptome analysis of isogenic HepG2 cells.**
A, B Venn diagram showing distribution of up- (A) and downregulated (B) genes between HepG2$^{WT}$ΔT1, HepG2$^{WT}$ΔT2, and HepG2$^{WT}$+T3 compared to HepG2$^{WT}$.

**Figure EV6. Differential transcriptomics and Gene Ontology Enrichment analysis.**
A–C RNAseq expression data plotted for HepG2$^{WT}$/HepG2$^{WT}$ΔT1 (A), HepG2$^{WT}$/HepG2$^{WT}$ΔT2 (B), and HepG2$^{WT}$/HepG2$^{WT}$+T3 (C) with expression level as a function of fold change. Top 10 biological process ontology terms among up- or downregulated transcripts are shown in table format. The biological coefficients of variations between HepG2$^{WT}$ clones with the same gene KO/KI were 19, 25, and 29% for GalNAc-T1, GalNAc-T2, and GalNAc-T3, respectively (not shown). These variations are within the boundaries of what is expected when two individual clonal cell lines are compared. In contrast, the biological coefficients of variations between the two individual datasets for pools of HepG2$^{WT}$ cells were as expected much lower (2–3%).

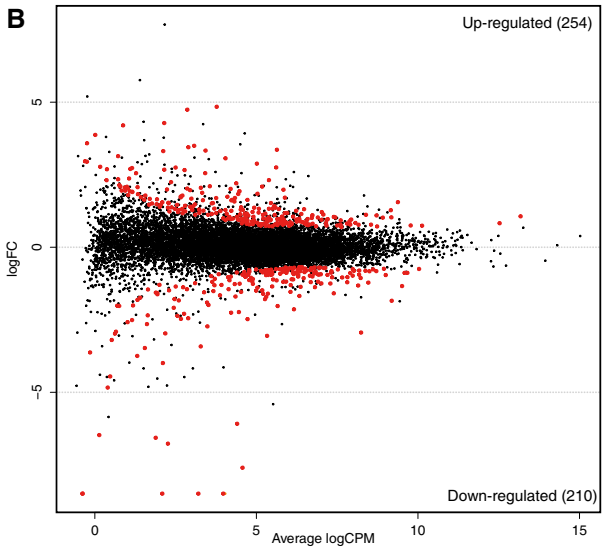



**Enriched terms among up-regulated genes**

| GO ID | p-value | OddsRatio | Exp. Count | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0048285 | 8.54E-24 | 4.10 | 26.60 | 88 | 537 | organelle fission |
| GO:0000278 | 2.00E-22 | 6.25 | 11.28 | 54 | 246 | mitotic cell cycle |
| GO:0051301 | 2.79E-20 | 3.55 | 29.90 | 88 | 612 | cell division |
| GO:0006260 | 1.18E-18 | 6.29 | 9.14 | 44 | 186 | DNA replication |
| GO:0044770 | 8.47E-17 | 4.707 | 13.37 | 51 | 274 | cell cycle phase transition |
| GO:0000722 | 1.13E-16 | 37.00 | 1.28 | 17 | 26 | telomere maintenance via recombination |
| GO:0032201 | 5.77E-16 | 39.13 | 1.18 | 16 | 24 | telomere maintenance via semi-conservative replication |
| GO:0006271 | 1.17E-12 | 25.02 | 1.22 | 14 | 25 | DNA strand elongation involved in DNA replication |
| GO:0006297 | 3.23E-12 | 28.15 | 1.08 | 13 | 22 | nucleotide-excision repair, DNA gap filling |
| GO:0032200 | 1.58E-11 | 8.24 | 3.51 | 21 | 71 | telomere organization |

**Enriched terms among down-regulated genes**

| GO ID | p-value | OddsRatio | Exp. Count | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0002253 | 1.17E-10 | 12.22 | 1.82 | 15 | 44 | activation of immune response |
| GO:0030449 | 1.32E-08 | 23.13 | 0.75 | 9 | 18 | regulation of complement activation |
| GO:0072376 | 2.54E-08 | 5.89 | 3.74 | 18 | 89 | protein activation cascade |
| GO:0006958 | 9.88E-07 | 7.10 | 2.14 | 12 | 51 | complement activation, classical pathway |
| GO:0009611 | 1.79E-06 | 1.82 | 50.15 | 84 | 1191 | response to wounding |
| GO:0002526 | 1.90E-06 | 5.45 | 3.08 | 14 | 74 | acute inflammatory response |
| GO:0002684 | 3.46E-06 | 2.18 | 23.90 | 48 | 572 | positive regulation of immune system process |
| GO:0019883 | 4.05E-06 | 22.94 | 0.505 | 6 | 12 | antigen processing and presentation of endogenous antigen |
| GO:0034976 | 5.64E-06 | 4.05 | 4.80 | 17 | 115 | response to endoplasmic reticulum stress |
| GO:0019885 | 1.43E-05 | 28.63 | 0.37 | 5 | 9 | antigen processing and presentation of endogenous peptide antigen via MHC class I |

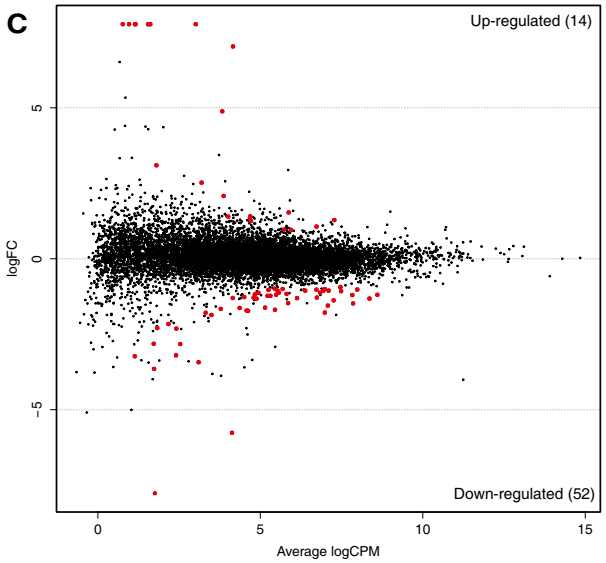



**Enriched terms among up-regulated genes**

| GOBPID | p-value | OddsRatio | ExpCount | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0015748 | 3.82E-04 | 6.93 | 0.95 | 6 | 73 | organophosphate ester transport |
| GO:0070328 | 4.55E-04 | 12.80 | 0.37 | 4 | 28 | triglyceride homeostasis |
| GO:0007009 | 4.87E-04 | 4.75 | 1.82 | 8 | 139 | plasma membrane organization |
| GO:0015866 | 5.06E-04 | 152.38 | 0.04 | 2 | 3 | ADP transport |
| GO:0071825 | 9.69E-04 | 10.24 | 0.44 | 4 | 34 | protein-lipid complex subunit organization |
| GO:0061024 | 1.22E-03 | 2.23 | 10.10 | 21 | 772 | membrane organization |
| GO:0006629 | 1.29E-03 | 1.97 | 15.99 | 29 | 1223 | lipid metabolic process |
| GO:0055088 | 1.90E-03 | 4.98 | 1.29 | 6 | 99 | lipid homeostasis |
| GO:1901264 | 2.15E-03 | 8.08 | 0.55 | 4 | 42 | carbohydrate derivative transport |
| GO:0072657 | 2.36E-03 | 2.91 | 4.00 | 11 | 306 | protein localization to membrane |

**Enriched terms among down-regulated genes**

| GOBPID | p-value | OddsRatio | ExpCount | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0009968 | 4.77E-05 | 2.70 | 9.84 | 24 | 847 | negative regulation of signal transduction |
| GO:0030334 | 1.33E-04 | 3.04 | 6.08 | 17 | 514 | regulation of cell migration |
| GO:0080134 | 1.52E-04 | 2.44 | 11.28 | 25 | 953 | regulation of response to stress |
| GO:0097190 | 2.29E-04 | 2.80 | 6.98 | 18 | 590 | apoptotic signaling pathway |
| GO:0030449 | 2.69E-04 | 14.82 | 0.32 | 4 | 27 | regulation of complement activation |
| GO:0023051 | 3.45E-04 | 2.11 | 17.08 | 32 | 1536 | regulation of signaling |
| GO:1900407 | 3.58E-04 | 13.63 | 0.34 | 4 | 29 | regulation of cellular response to oxidative stress |
| GO:0030855 | 6.01E-04 | 2.74 | 6.30 | 16 | 537 | epithelial cell differentiation |
| GO:0042127 | 6.32E-04 | 2.05 | 16.07 | 30 | 1358 | regulation of cell proliferation |
| GO:0072376 | 6.53E-04 | 6.20 | 1.05 | 6 | 89 | protein activation cascade |




**Enriched terms among up-regulated genes**

| GOBPID | p-value | OddsRatio | ExpCount | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0042697 | 6.85E-4 | Inf | 0.000685 | 1 | 1 | menopause |
| GO:0003190 | 1.36E-3 | 1604.6 | 0.001370 | 1 | 2 | atrioventricular valve formation |
| GO:2000820 | 1.36E-3 | 1604.6 | 0.001370 | 1 | 2 | negative regulation of transcription from RNA polymerase II promoter involved in smooth muscle cell differentiation |
| GO:0009450 | 1.36E-3 | 1604.6 | 0.001370 | 1 | 2 | gamma-aminobutyric acid catabolic process |
| GO:0036304 | 1.36E-3 | 1604.6 | 0.001370 | 1 | 2 | umbilical cord morphogenesis |
| GO:0010035 | 1.88E-3 | 15.59 | 0.260306 | 3 | 380 | response to inorganic substance |
| GO:0018242 | 2.05E-3 | 802.25 | 0.002055 | 1 | 3 | protein O-linked glycosylation via serine |
| GO:0043420 | 2.05E-3 | 802.25 | 0.002055 | 1 | 3 | anthranilate metabolic process |
| GO:0018243 | 2.73E-3 | 534.8 | 0.002740 | 1 | 4 | protein O-linked glycosylation via threonine |
| GO:0003199 | 2.73E-3 | 534.8 | 0.002740 | 1 | 4 | endocardial cushion to mesenchymal transition involved in heart valve formation |

**Enriched terms among down-regulated genes**

| GOBPID | p-value | OddsRatio | ExpCount | Count | Size | Term |
|---|---|---|---|---|---|---|
| GO:0050966 | 1.67E-4 | 145.52 | 0.02 | 2 | 7 | detection of mechanical stimulus involved in sensory perception of pain |
| GO:0046415 | 6.13E-4 | 66.12 | 0.04 | 2 | 13 | urate metabolic process |
| GO:0051593 | 6.13E-4 | 66.12 | 0.04 | 2 | 13 | response to folic acid |
| GO:0031295 | 1.18E-3 | 15.89 | 0.21 | 3 | 73 | T cell costimulation |
| GO:0007584 | 1.25E-3 | 9.38 | 0.47 | 4 | 165 | response to nutrient |
| GO:0009612 | 1.58E-3 | 8.77 | 0.50 | 4 | 176 | response to mechanical stimulus |
| GO:0060284 | 2.54E-3 | 4.22 | 1.89 | 7 | 660 | regulation of cell development |
| GO:0032904 | 2.86E-3 | Inf | 0.00 | 1 | 1 | negative regulation of nerve growth factor production |
| GO:0019343 | 2.86E-3 | Inf | 0.00 | 1 | 1 | cysteine biosynthetic process via cystathionine |
| GO:0033343 | 2.86E-3 | Inf | 0.00 | 1 | 1 | positive regulation of collagen binding |

**Figure EV6.**