**Title**

Optimization of Heterologous Glucoraphanin Production In Planta

**Permalink**

https://escholarship.org/uc/item/0nx5c3dv

**Journal**

ACS Synthetic Biology, 11(5)

**ISSN**

2161-5063

**Authors**

Barnum, Collin R
Endelman, Benjamin J
Ornelas, Izaiah J
et al.

**Publication Date**

2022-05-20

**DOI**

10.1021/acssynbio.2c00030

Peer reviewed

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

# Title: Optimization of heterologous glucoraphanin production *in planta*

**Authors**: Collin R Barnum[1,†], Benjamin J Endelman[1,†], Izaiah Ornelas[1], Roxanna Pignolet[1], and Patrick M Shih[1,2,3,4,5,*]

**Affiliations:**
[1] Department of Plant Biology, University of California, Davis, CA, USA.
[2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[3] Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA, USA.
[4] Department of Plant and Microbial Biology, University of California, Berkeley, CA
[5] Innovative Genomics Institute, University of California, Berkeley, CA

**\*** Correspondence to pmshih@berkeley.edu, University of California, Berkeley, 481 Koshland Hall, Berkeley, CA 94720

† Contributed equally to this work

**Abstract**:

Glucoraphanin is a plant specialized metabolite found in cruciferous vegetables that has long been a target for production in a heterologous host because it can subsequently be hydrolyzed to form the chemopreventive compound, sulforaphane, before and during consumption. However, previous studies have only been able to produce small amounts of glucoraphanin in heterologous plant and microbial systems compared to the levels found in glucoraphanin-producing plants, suggesting that there may be missing auxiliary genes that play a role in improving production *in planta*. In an effort to identify auxiliary genes required for high glucoraphanin production, we leveraged transient expression in *Nicotiana benthamiana* to screen a combination of previously uncharacterized coexpressed genes and rationally selected genes alongside the glucoraphanin biosynthetic pathway. This strategy alleviated metabolic bottlenecks which improved glucoraphanin production 4.74-fold. Our optimized glucoraphanin biosynthetic pathway provides a pathway amenable for high glucoraphanin production.

**Keywords** (6 max): Glucoraphanin, coexpression analysis, plant natural products, Transient expression, *Nicotiana benthamiana*

**Abbreviations**: GLS= glucosinolate, Met=methionine, 4MTOBA=4-methylthio-2-oxobutanoic acid, 2H2ESA= 2-hydroxy-2-(2-(methylthio)ethyl)succinic acid, 5MTOPA=5-methylthio-2-oxopentanoic acid, HM= homomethionine, 2H3PSA= 2-hydroxy-2-(3-(methylthio)propyl)succinic acid, 6MTOHA= 6-methylthio-2-oxohexanoic acid, DHM= dihomomethionine, 5MTPO= 5-(methylthio)pentanaloxime, 5MTPO-GSH= 5-(methylthio)pentanaloxime-Glutathione, H5MTPA= N-hydroxy-5-(methylthio) pentanimidothioic acid, H5MTPA= N-hydroxy-5-(methylthio)pentanimidothioic acid, DS-GE= desulfo-glucoerucin, GE= Glucoerucin, GR= Glucoraphanin, GIV= Glucoiberverin, GI= Glucoiberin, THM= trihomomethionine, 5MSOP= THM-glucosinolate, Leu= Leucine, HL= homoleucine, DHL= Dihomoleucine, HL-GLS= 2-methylpropyl-glucosinolate, DHL-GLS= 3-methylbutyl-glucosinolate, dCGS= allosterically insensitive cystathionine gamma-synthase, CGBP= core glucoraphanin biosynthetic pathway

## Introduction

Glucoraphanin is a methionine-derived glucosinolate that is highly abundant in cruciferous vegetables, such as broccoli[1], and serves as a chemical defense compound. To protect the plant from various pests, glucoraphanin is enzymatically converted to the highly reactive, bioactive isothiocyanate, sulforaphane[2]. Sulforaphane is not only beneficial as a plant defense compound, but also has been demonstrated to have a wide variety of beneficial health effects, with a strong role in the prevention of cancer[3–5]. Epidemiological evidence indicates that sulforaphane is effective in upregulating phase I and II detoxification enzymes, epigenetically regulating cancer genes, promoting apoptosis of cancer cells, and inducing cell cycle arrest[3–5]. While some sulforaphane-containing supplements are available, supplements are rarely clinically validated[6] and often prohibitively expensive. Therefore, improving the availability of glucoraphanin and sulforaphane through dietary consumption is of great interest, as it is a well-validated method of sulforaphane delivery[7].

Improving sulforaphane availability and consumption can be accomplished by increasing the number of foods that contain the stable precursor, glucoraphanin, through expanding the production of glucoraphanin to non-cruciferous vegetables. Mikkelsen et al.[8] successfully produced glucoraphanin in *Nicotiana benthamiana* through the expression of 11 genes identified in the model crucifer, *Arabidopsis thaliana*, although yields were far below what is normally produced in high glucoraphanin-producing *Brassica* species. Additionally, high amounts of leucine-derived glucosinolates were also produced during heterologous expression, which are minor components in *Arabidopsis thaliana*, indicating the pathway is not functioning as it does in its native host; therefore, we sought to identify genes and metabolic strategies to optimize the glucoraphanin biosynthetic pathway.

The core glucoraphanin biosynthetic pathway (CGBP) consists of 14 genes and is divided into three major sections: 1) chain elongation, 2) core structure formation, and 3) secondary modifications (Fig. 1). Glucoraphanin biosynthesis begins with the deamination of methionine to produce the α-keto acid, 4-methylthio-2-oxotetranoic acid (4MTOBA)[9]. 4MTOBA is transported into the chloroplast where two methyl groups are inserted into the carbon chain of methionine by the enzymes MAM1, IMPDH1, IPMI-LSU, and IPMI-SSU[10–14]. After two successive rounds of chain elongation, the elongated α-keto acid, 6-methylthio-2-oxohexanoic acid (6MTOHA), is transaminated to form an elongated form of methionine called dihomomethionine (DHM). DHM is then transported out of the chloroplast, through an unknown mechanism, where it undergoes seven enzymatic reactions that result in the formation of the glucosinolate, glucoerucin[8,15–20]. Glucoerucin is then S-oxygenated to form glucoraphanin[21].

The discovery of the glucoraphanin biosynthetic pathway prompted several studies to produce glucoraphanin or key pathway intermediates in microbial production platforms[22–24]; however, this approach experiences various shortcomings that limit efficacy and titers. While the minimal gene set for glucoraphanin production is known, heterologous expression in microbial systems fails to produce high glucoraphanin titers. Recent experiments from Yang et al.[24] highlighted this by attempting to improve glucoraphanin yields in *E. coli* through the coexpression of enzymes that enhance the

production of key methyl and sulfur donors. Despite their ability to increase donor molecule levels, there was no corresponding increase in glucoraphanin, indicating current enzymes are underperforming or additional auxiliary enzymes are needed. While screening candidate genes in *E. coli* is possible, some plant proteins experience issues with expression or activity in microbial hosts[23]. This could result in false negatives when screening candidate enzymes from plants for their ability to improve glucoraphanin yields. Production of indolylglucosinolates in yeast has also been examined; however, this also resulted in low titers, further suggesting some inherent biological hurdles that may exist in transferring the pathway out of plants [25].

There are numerous examples of glucosinolate pathway optimization *in planta*[10,16,26]. Studies utilizing *A. thaliana* knockouts and metabolic engineering efforts have elucidated enzymes with a preference for glucoraphanin production, and previous work has focused on optimizing the genes utilized in the chain elongation machinery[10,14]. This resulted in the identification of a set of enzymes able to increase the production of the key intermediate DHM by 30-fold compared to the previously recorded best[10]. Thus, pathway optimization *in planta* allows plant enzymes to be reliably screened for their ability to enhance glucoraphanin production.

While some improvements in the glucoraphanin biosynthetic pathway have been made, final glucoraphanin concentrations remain low, indicating a need for further pathway optimization. In this study, we sought to optimize the glucoraphanin biosynthetic pathway for high glucoraphanin production using *N. benthamiana* as a transient expression system. To accomplish this, we conducted a coexpression analysis with bait genes from the chain elongation machinery to identify candidate genes with putative roles in glucoraphanin biosynthesis. We also included rationally selected genes hypothesized to improve glucoraphanin production. Our screen resulted in an optimized set of genes including BCAT3, dCGS, IPMI2, and an acyltransferase with the CGBP to increase glucoraphanin yields by 4.74-fold compared to the expression of the CGBP alone.

## Results and Discussion

### Coexpression analysis identifies glucoraphanin-enhancing genes

Intermediate analysis of the CGBP revealed substantial buildups of intermediates in the chain elongation portion of the pathway compared to broccoli and *A. thaliana*, which are natural producers of glucoraphanin (Fig. S1). In hopes of identifying other genes involved in optimizing the chain elongation portion of the CGBP, we utilized the transcriptome database, ATTED-II[27] to conduct a cross-correlated coexpression analysis using two of the major chain elongation genes involved as bait, *MAM1* and *IPMI1* (Fig. 2A). Since these genes are largely specific to the production of short-chain aliphatic glucosinolates, we hypothesized that their use as bait genes would identify candidate genes that increase glucoraphanin production. From the coexpression analysis, we identified genes already known to be part of glucosinolate metabolism, validating this approach. Thirty-five genes were selected (Table S1) for transient coexpression with the entire, 14-gene, core glucoraphanin biosynthetic pathway in *N. benthamiana* (see Fig. 1, Table S2) via agroinfiltration.

Expression of the candidate genes (full list in Table S1) had varying effects on glucoraphanin production, ranging from drastic reductions to significant increases (Fig. 2b). For several candidate genes, the inhibitory effect on glucoraphanin production was expected as they are involved in competing pathways. For example, *MAM3* (CGBP7) is capable of carrying out six rounds of methionine chain elongation, which likely reduces the dihomomethionine for glucoraphanin production[28]. Twenty-eight of the genes did not show any significant change. This result is unsurprising since the production of glucosinolates in *Brassica* species is usually associated with pathogenesis or stress response, and these genes are likely involved in a general cellular stress response rather than directly in glucosinolate synthesis[29]. For example, *PMSR2* (CGBP14*)* is known for its involvement in reducing oxidative damage to proteins, protecting cells from general oxidative stress.

Though many genes did not increase glucoraphanin production, screening coexpressed genes yielded a gene that significantly improved glucoraphanin production. Expression

of CGBP13 improved glucoraphanin production by 1.48-fold. CGBP13 encodes a protein of unknown function, though sequence similarity places it in the broad family of HXXXD-type acyltransferases, herein referred to as acyltransferase. While they did not significantly improve glucoraphanin production, we also examined CGBP8 and CGBP24 as they were the second and third candidates from the coexpression screen. CGBP8 is an uncharacterized member of the cytochrome B5 family called *CYTB5-C;* however, when *CYTB5*-C was included in later experiments, it resulted in large variations in glucoraphanin levels, leading it to be excluded from further analysis (data not shown). The function of CGBP24 is also unknown, though sequence similarity places it in the α/ß-hydrolase superfamily, herein referred to as hydrolase. Expression of hydrolase results in an increase in homo-, dihomo-, and trihomomethionine, which may account for the minor increase in glucoraphanin at lower concentrations (Fig. S2).

**Rational selection of glucoraphanin-enhancing genes**

In addition to examining transcriptionally coregulated genes, we selected genes not pulled from the coexpression analysis whose function could alleviate hypothesized metabolic bottlenecks. Previous studies characterizing the cytosolic protein, BCAT4, found that it deaminates methionine (Fig. 1, compound 1) to form an α-keto acid (Fig. 1, compound 2)[8,10]. It was previously hypothesized that either BCAT4 or an endogenous enzyme in *N. benthamiana* is responsible for transaminating the elongated α-keto acid to form DHM (Fig. 1, compound 8)[8]. In an effort to optimize this step, we included the chloroplast-localized enzyme BCAT3 (CGBP37), which has previously been characterized to aminate methionine-derived α-keto acids that have undergone one or two rounds of chain elongation[30]. When expressed alongside the CGBP, CGBP37 increases glucoraphanin concentration the most out of all candidate genes, increasing glucoraphanin production by 1.75-fold relative to the CGBP (Fig. 2C).

Previous work in *E. coli* showed that DHM production could be enhanced through the addition of exogenous methionine[22]. We hypothesized that methionine levels are a limiting factor to glucoraphanin production as *in planta* methionine is generally in lower abundance. Previous studies in *Glycine max* and *N. benthamiana* utilized a feedback-

insensitive mutant of cystathionine γ-synthase (*dCGS*) to increase methionine levels[31,32]. Expression of CGBP36 improved glucoraphanin concentrations by 1.72-fold compared to the CGBP (Fig. 2C). Additionally, expression of *dCGS* with the chain elongation pathway (DHM36) increased methionine concentration by approximately 2-fold (Fig. 3). These two rationally selected genes, CGBP36 and CGBP37, provided substantially larger increases in glucoraphanin than any gene found through coexpression analysis.

Data on the total ion count of the putative CGBP intermediates appear to show a buildup of the chain elongation intermediates 2H2ESA and 2H3ESA, indicating a potential bottleneck (Fig. S1). Therefore, we sought candidate genes from the coexpression analysis that lowered the abundance of putative chain elongation intermediates. *IPMI2* (CGBP1) substantially decreases the abundance of putative chain elongation intermediates (Fig. 3) while having no statistically significant effect on glucoraphanin production (Fig. 2B, 4). *IPMI2* is a known gene involved in the methionine elongation pathway, though it is generally thought to be redundant to *IPMI1*[10,12]. *IMPI2* was included in later experiments to examine how a reduction in chain elongation intermediates may alter flux through the CGBP when expressed with additional genes.

**Stacked expression of top candidate genes improves glucoraphanin production**

To examine possible additive and synergistic effects on glucoraphanin expression, we co-expressed multiple genes in combination with the two highest producing candidate genes, *BCAT3* and *dCGS*. We chose to test the two highest genes from the coexpression analysis, acyltransferase and hydrolase. We also included *IPMI2*, which we hypothesized would increase flux through chain elongation due to its known role in methionine chain elongation. Gene combinations involving the expression of the CGBP, *BCAT3*, and genes found through coexpression analysis (Fig. 4A) resulted in significant increases relative to the expression of the CGBP and *BCAT3* alone (Fig. 4A) with CGBP41 having the greatest increase. Expression of the CGBP, *dCGS,* and genes found through coexpression analysis significantly increased glucoraphanin production

(Fig. 4B) relative to the expression of the CGBP and *dCGS* alone (Fig. 4B), with CGBP45 displaying the greatest increase.

When *BCAT3* and *dCGS* are expressed in tandem with the CGBP (CGBP50), there is a synergistic improvement in glucoraphanin concentration, compared to the expression of *dCGS* (CGBP36) or *BCAT3* (CGBP37) alone (Fig. 4C). Expression of *dCGS* alongside the chain elongation pathway and CGBP resulted in a buildup of methionine and various chain elongation intermediates, including the elongated α-keto acid (Fig. 3 and 5). Expression of *BCAT3* and *dCGS* with the CGBP (CGBP50) likely utilizes the buildup of the elongated α-keto acid to increase flux through the pathway. This is evident from a decrease in methionine, an increase in core structure biosynthesis intermediates, and an increase in glucoraphanin concentration (Fig. 5). While CGBP50 showed a significantly higher increase in glucoraphanin concentration compared to the CGBP, it was not significantly higher compared to the highest producing experiments, CGBP41 and CGBP45 (Fig. 4C). However, the expression of *IPMI2,* acyltransferase, *BCAT3* and *dCGS* with the CGBP (CGBP54) resulted in a 4.74-fold improvement in glucoraphanin production relative to the CGBP, which was significant relative to CGBP41 and CGBP45 (Fig. 4C). This could be from an apparent reduction in the level of specific chain elongation intermediates (compounds 2H2ESA, 2H3PSA, 6MTOHA) that resulted in further increases in DHM levels, causing a subsequent increase in core structure biosynthesis intermediates and glucoraphanin (Fig. 5). CGBP54 also significantly increased glucoraphanin production compared to CGBP41 and CGBP45 (Fig. 4C). When comparing CGBP45 and CGBP54, CGBP54 displays a slight decrease in some chain elongation intermediates and a buildup of core structure biosynthesis intermediates.

**Intermediate analysis elucidates potential bottlenecks in CGBP**

From our analysis of pathway intermediates, there is a preponderance of compounds that accumulate either to an equivalent or a greater extent than glucoraphanin. Some examples of this are the alternative elongated versions of methionine, homomethionine, and trihomomethionine, as well as their glucosinolate products. This could be explained

by the evidence that *MAM1* is known to proceed through the first two cycles of chain elongation, allowing it to make the precursors for homomethionine and dihomomethionine[13,14]. Additionally, the glucosinolates from these molecules could be produced since the core structure enzymes are functional on all observed lengths of elongated methionine. There is a substantial increase in the levels of glucoerucin as the glucoraphanin concentration increases, indicating that $FMO_{gs-ox1}$ is unable to convert the available pool of glucoerucin to glucoraphanin. A second $FMO_{gs-ox}$ gene could enhance glucoraphanin production by improving the rate of S-oxygenation or a more efficient form of this enzyme could be isolated or engineered, given this product is prevalent in both broccoli and *Arabidopsis* (Fig. S1).

Expression of the chain elongation pathway alone resulted in a build-up of the post-*MAM1* products, 2H2ESA, and 2H3PSA. Coexpression of *IPMI2* (DHM1) partially alleviated this bottleneck resulting in an increase in DHM (Fig. 3). Interestingly, when *IPMI2* (CGBP1) was expressed alongside the CGBP there was no evident increase in product formation during the initial screening (Fig. 2B). However, when *IPMI2* was expressed with other candidate genes and the CGBP there were synergistic effects resulting in higher glucoraphanin yields (Fig. 4).

When specific gene combinations are expressed, multiple bottlenecks appear in the second half of the core structure biosynthesis. For example, in all experiments with the addition of *BCAT3* and/or *dCGS*, the level of the post-*GSTF11* product, 5MTPO-GSH (Fig. 1, compound 11), was elevated compared to the core pathway alone, highlighting a new metabolic bottleneck. While *GGP1* was previously found to be necessary for the detectable production of glucoraphanin[8], it was originally identified through coexpression analysis of genes used in the production of benzyl glucosinolates[16]. Although sufficient for benzyl glucosinolate production, it is possible that *GGP1* is suboptimal for glucoraphanin production. While it is unclear why these buildups exist, they provide potential targets for further metabolic engineering and pathway optimization.

**Identification of additional compounds produced by the CGBP**

Expression of the CGBP in *Nicotiana benthamiana* produces a variety of additional glucosinolates. Previous studies have found an abundance of leucine- and/or isoleucine derived glucosinolates when expressing the CGBP in *N. benthamiana*[8]. Our study has confirmed the presence of these compounds when in *N. benthamiana* leaves expressing the CGBP (See Table S3 for retention times and MS/MS fragmentation). In addition, several gene combinations produced higher amounts of homoleucine (HL) and 2-methylpropyl-glucosinolate (HL-GLS) compared to expression of the CGBP alone, especially gene combinations expressing *BCAT3* (Fig. 5). This is in accordance with a previous *in vitro* characterization of recombinant *Arabidopsis thaliana* BCAT3, which showed high activity against the α-keto acids of leucine, isoleucine, valine, and, to a lesser degree, methionine[30]. *IPMI2* (CGBP1) and acyltransferase (CGBP13) also resulted in slight increases in HL and DHL, suggesting *IPMI2* and acyltransferase enhance general chain elongation but have no effect on methionine specificity. Little change in the abundance of dihomoleucine (DHL) or 3-methylbutyl-glucosinolate (DHL-GLS) was observed in any gene combinations, suggesting a preference of some pathway enzymes for HL-GLS precursors.

Changes in the abundance of several methionine-derived glucosinolates were also observed. Trihomomethionine and one of its glucosinolate derivatives, 5MSOP, displayed lower abundance in several gene combinations expressing *BCAT3* and *dCGS*, which aligns with the previously described preference of *BCAT3* for the α-keto acids that form homomethionine (HM) and DHM over the α-keto acid that forms THM[30]. HM and two of its glucosinolate derivatives, glucoiberin (GI) and glucoiberverin (GIV) showed substantial increases, especially in gene combinations expressing *BCAT3* and *dCGS* (Fig. 5). For example, CGBP50 increased GI concentration by approximately 6-fold compared to the expression of the CGBP alone (Fig. 5). This is likely due to the increase in methionine observed with the expression of *dCGS* and the additional transamination activity of *BCAT3*. While the production of these off-target compounds likely limits the amount of available methionine for use in the production of glucoraphanin, their presence could provide other benefits. For example, several studies have suggested the isothiocyanate derived from GI, could have beneficial effects similar to sulforaphane[33–35]. While the focus of the pathway optimization in this

study was glucoraphanin, the observed increase in GI production could be useful in future studies and plant engineering efforts.

**Conclusion**

The production of glucoraphanin in a non-cruciferous crop has the potential to enhance nutrition; however, high glucoraphanin production in heterologous systems has been previously unachievable. This work has elucidated genes that improve the yield of glucoraphanin by 4.74-fold compared to the previously established glucoraphanin biosynthetic pathway expressed in *N. benthamiana*, corresponding to a concentration of 2.05±0.32 µmol/g DW. Together, our results display an improved glucoraphanin biosynthetic pathway suitable for high glucoraphanin production in a heterologous plant and may lay the foundation for future stable plant transformants with high yields.

**Methods**

**Plant material**

*Nicotiana benthamiana* was grown in 3.5 inch square pots in a controlled environment facility under a 12/12 day/night cycle (12 hours light, 12 hours dark) at ~100 µmol photons $m^{-2}sec^{-1}$. Daytime temperatures were 26°C, and night temperatures were 25°C. Relative humidity was between 60 - 75%. Plants used in this study were 4 weeks old.

**Identification of coexpressed candidate genes**

The chain elongation portion of the CGBP has been hypothesized to be rate-limiting. For this reason, MAM1 and IPMI1 were used as bait genes in a coexpression analysis conducted with ATTED-II[27] to identify genes in *A. thaliana* that improve glucoraphanin yields [10,28]. The two sets of genes were initially trimmed by removing most of the genes that are known members of the CGBP, then cross-correlated to identify genes that are specifically present in both lists with a strong mutual rank score. Experiments involving the expression of the CGBP with individual candidate genes or combinations of candidate genes were assigned numerical codes. Expression of the 35 candidate genes identified through coexpression analysis with the CGBP were assigned CGBP1 –

CGBP35. Expression of the two rationally selected candidate genes with the CGBP were assigned CGBP36 and CGBP37. The experiments involving the expression of multiple candidate genes with the CGBP were assigned CGBP38 – CGBP56.

**Cloning**

*A. thaliana* Col-0 leaf RNA was isolated using the E.Z.N.A. Plant Kit (OMEGA). cDNA was generated from leaf RNA using the Superscript First Strand Synthesis kit (Invitrogen). Candidate genes were PCR amplified (see Table S1 for primers). Amplified candidate genes were then cloned into the binary vector PMS057 using Golden Gate assembly[36], Gibson assembly, or standard digestion and ligation assembly. 2-4 µL of the assembly reactions were transformed into DH5α chemically competent *E. coli* cells via heat shock as previously described[37]. Colonies were selected on LB agar plates containing 50 µg/mL kanamycin and sequence-verified using Sanger sequencing (McLab).

Sequence verified plasmids were used to transform *Agrobacterium tumefaciens* str. GV3101 by electroporation[38]. Competent cells were then plated on LB agar plates containing 50 µg/mL rifampicin, 10 µg/mL gentamicin, 50 µg/mL kanamycin.

**Infiltrations**

Overnight cultures of *A. tumefaciens* str. GV3101 were grown in LB to an OD600 between 0.8 and 1.2. Cultures were centrifuged at 4000xG for 10 min and the supernatant was removed. Bacterial pellets were resuspended in infiltration media (10 mM MgCl$_2$, 10 mM MES, 500 µM acetosyringone, pH: 5.6). Following an hour incubation, *Agrobacterium* strains containing the chain elongation pathway or CGBP and the candidate genes were mixed in various combinations to a final OD600 of 0.5. *A. tumefaciens* strains were normalized to the level of the highest number of strains used in an experiment. For experiments that had less than the highest number of strains, an additional *A. tumefaciens* strain harboring the unrelated gene (dsRed) was added to reach a final OD600 of 0.5. An *A. tumefaciens* strain harboring the p19 silencing suppressor was used in all experiments at the same concentration as other strains. *A.*

*tumefaciens* suspensions were syringe infiltrated into the abaxial side of the seventh leaf of 4-week old *N. benthamiana* in biological triplicate.

**Extractions**

*N. benthamiana* leaves were harvested 5 days post-infiltration. For experiments using *Arabidopsis thaliana* Col-0, plants were 7 weeks old when leaves were harvested. Broccoli was procured at a local market. Major veins of *N. benthamiana* were removed from the leaf tissue, and the tissue was frozen in liquid nitrogen before lyophilization. Lyophilized leaf tissue was bead beaten using a single steel bead at 20 Hz for 10 min. Following bead beating, 10 µL of extraction solution (80% MeOH, 20% $H_2O$, v/v) containing an internal standard (CUDA, Cayman Chemicals, ≥95% purity) at 5 ppm was added for every milligram of leaf tissue. Leaf tissue and extraction solvent were then bead beaten at 10 Hz for 20 min. Leaf tissue was then centrifuged at 10,000xG for 10 min, and the supernatant was transferred to a clean tube. The supernatant was frozen with liquid nitrogen and centrifuged at 10,000xG for 10 min. The supernatant was then transferred to a 96-well filter plate (0.2 µM; PVDF membrane; Corning) and centrifuged at 1500xG for 5 min into a clean 96-well autosampler plate, which was sealed before analysis by LC-MS/MS.

**LC-MS/MS analysis**

Liquid chromatography was performed using a Thermo Scientific Vanquish UHPLC with an Acquity UPLC BEH C18 column (2.1 x 100 mm, 1.7 µm particle size) using water with 0.1% formic acid (v/v) as eluent A and acetonitrile with 0.1% formic acid (v/v) as eluent B. Liquid chromatography analysis was carried out with the following elution profile at a flow rate of 0.450 mL/min: -1 to 0 min, 97% A; 0 to 5.5 min, 97% to 50% A;

5.5 to 6 min, 50% to 2% A; 6 to 12 min, 2% A; 12 to 13 min 2% to 97% A; 13 to 15 min, 97% A. The column preheater and compartment were set to 30°C.

The mass spectrometer (Thermo Scientific Q Exactive) equipped with an electrospray ionization source was run in negative and positive ionization modes. Negative ionization (Spray voltage, 2.50 |kV|; capillary temperature, 300°C; aux gas heater, 350°C; sheath gas flow rate, 45; aux gas flow rate, 10; sweep gas flow rate, 3) was used for the identification of compounds 3 and 6 (Fig. 1) and all glucosinolates. Positive ionization (Spray voltage, 3.50 |kV|; capillary temperature, 300°C; aux gas heater, 350°C; sheath gas flow rate, 45; aux gas flow rate, 10; sweep gas flow rate, 3) was used for the identification of Met, Leu, both of their elongation products as well as compounds 2, 7, 9, 11, 12 and 13. MS/MS analysis was completed using stepped normalized collision energy of 25, 35, 50.

**Quantification of glucoraphanin and pathway intermediates**

Quan Browser (Thermo Fisher) was used for quantitative analysis of the glucoraphanin concentrations. Purchased glucoraphanin standard (Extrasynthese, ≥98% purity) was used to build a concentration curve containing seven concentrations ranging from 0.1625 µmol/g DW to 10 µmol/g DW. All standards were prepared in triplicate in wildtype *N. benthamiana* tissue and processed using the extraction protocols as above to account for product loss during processing. Each standard was normalized to the CUDA internal standard.

Python scripts were developed, using pyOpenMS module, to extract MS peak values associated with the actual mass for all the putative intermediate compounds, as determined by the FreeStyle software (Thermo Scientific). An additional control parameter of retention time was added to the script to ensure all data used is from the putative intermediate compound in question. As most of these compounds did not have a standard for purchase and comparison, these values were used in a semi-quantitative

analysis based on the relative change in the compound compared to the core pathway alone. To achieve this, all values were initially normalized to the internal standard CUDA, followed by a secondary normalization relative to the CGBP. All CGBP pathway intermediates measured were not detected or detected at extremely low abundance in negative controls. Additionally, publicly available MS/MS fragmentation data was used to aid in compound identification when available.

## Acknowledgements

## Competing interests

The authors have no competing interests to declare.

## Funding

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Figures**

**Figure 1: Schematic of core glucoraphanin biosynthetic pathway (CGBP).** Pathway is broken up between the three major processes chain elongation, core structure biosynthesis and secondary modification. The numbers indicate the specific intermediate compound at each step in the pathway (greater details for each compound in Table S3). For compounds 3, 4, and 5, n=2. For compounds 6-13, n=3.
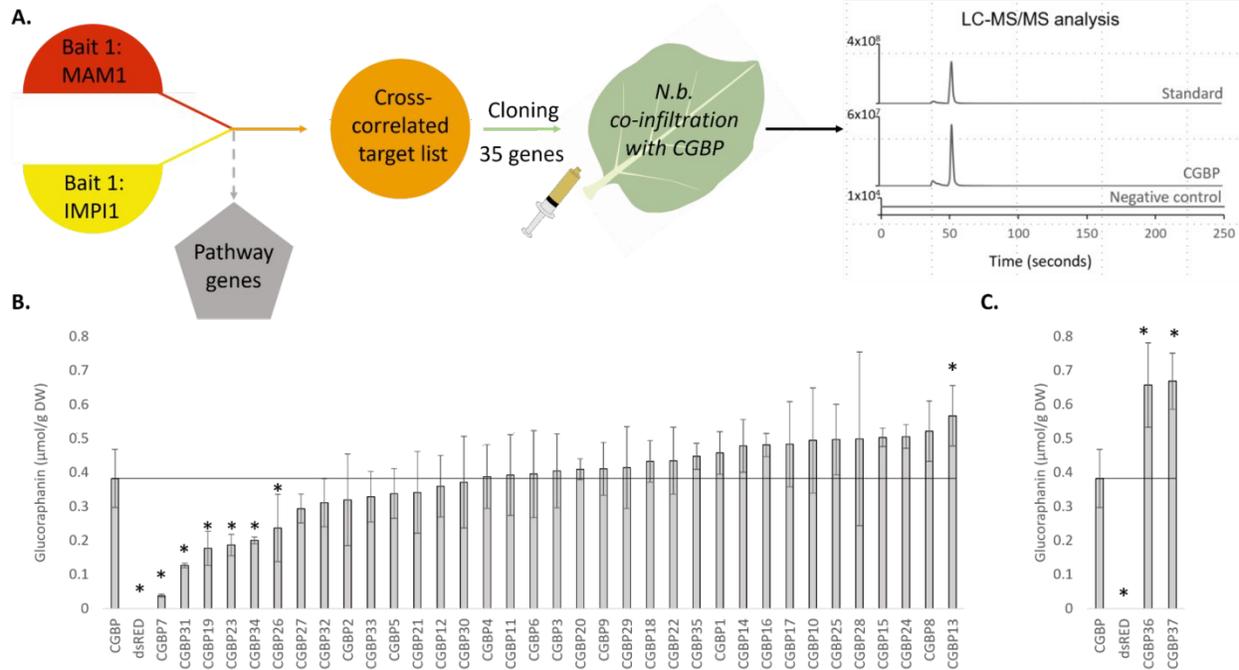
**Figure 2: Coexpression and rationally selected genes alter glucoraphanin concentration.** A. Schematic for coexpression analysis and metabolomics testing (full list identified genes in Table S3). B. Glucoraphanin levels produced by tandem expression of CGBP and genes identified through coexpression analysis. C. Glucoraphanin levels produced by tandem expression of CGBP and rationally selected genes. Black line represents the mean of CGBP. N=3. Statistical analysis by student t-test; *=p-value≤0.05 relative to CGBP.
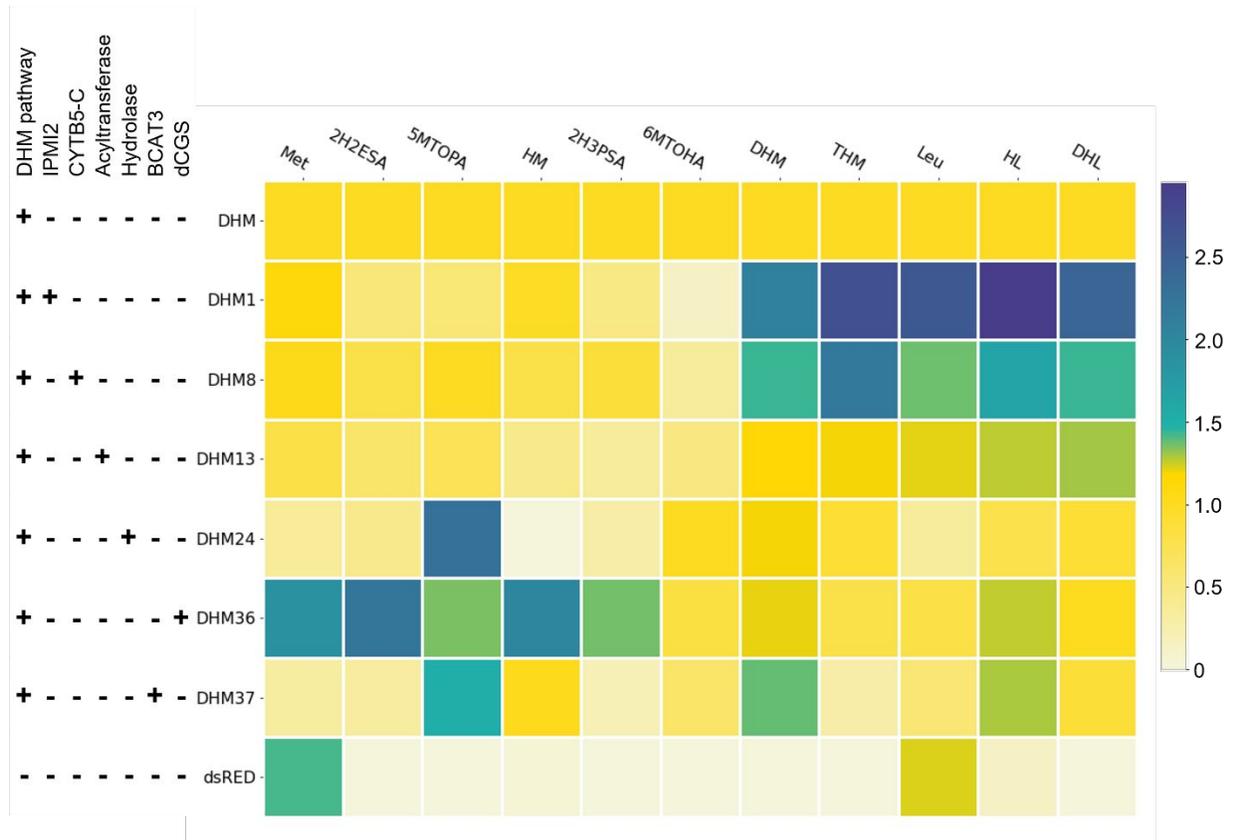
**Figure 3: Relative abundance of dihomomethionine (DHM) and pathway intermediates following expression of chain elongation genes in combination with enhancer genes**. The DHM pathway was expressed alone or with individual genes as indicated. All values were normalized to those observed when the DHM pathway was expressed alone (top row).

A.



| | dsRED | CGBP | CGBP01 | CGBP13 | CGBP24 | CGBP37 | CGBP38 | CGBP39 | CGBP40 | CGBP41 | CGBP42 | CGBP43 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Core pathway | - | + | + | + | + | + | + | + | + | + | + | + |
| IPMI2 | - | - | + | - | - | - | + | - | - | + | + | - |
| Acyltransferase | - | - | - | + | - | - | - | + | - | + | - | + |
| Hydrolase | - | - | - | - | + | - | - | - | + | - | + | + |
| BCAT3 | - | - | - | - | - | + | + | + | + | + | + | + |

B.



| | dsRED | CGBP | CGBP36 | CGBP44 | CGBP45 | CGBP46 | CGBP47 | CGBP48 | CGBP49 |
|---|---|---|---|---|---|---|---|---|---|
| Core pathway | - | + | + | + | + | + | + | + | + |
| IPMI2 | - | - | - | + | - | - | + | + | - |
| Acyltransferase | - | - | - | - | + | - | + | - | + |
| Hydrolase | - | - | - | - | - | + | - | + | + |
| dCGS | - | - | + | + | + | + | + | + | + |

C.



| | dsRED | CGBP | CGBP36 | CGBP37 | CGBP41 | CGBP45 | CGBP50 | CGBP51 | CGBP52 | CGBP53 | CGBP54 | CGBP55 | CGBP56 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Core pathway | - | + | + | + | + | + | + | + | + | + | + | + | + |
| IPMI2 | - | - | - | - | + | - | - | + | - | - | + | + | - |
| Acyltransferase | - | - | - | - | + | + | - | - | + | - | + | - | + |
| Hydrolase | - | - | - | - | - | - | - | - | - | + | - | + | + |
| BCAT3 | - | - | + | - | + | - | + | + | + | + | + | + | + |
| dCGS | - | + | - | - | - | + | + | + | + | + | + | + | + |

1
2
3
4
5
6
7
8
9
10
11
12

**Figure 4: Coexpression of multiple candidate genes shows synergistic enhancement of glucoraphanin concentration.** Quantified Glucoraphanin for core with addition of A. BCAT3 B. dCGS C. BCAT3 and dCGS each with three of the minor enhancer genes as indicated in the table below each graph. N=3. Statistical analysis by student t-test ‡= p-value≤0.05 relative to CGBP; *= p-value≤0.05 relative to CGBP37; ^=p-value≤0.05 relative to CGBP36; #= p-value≤0.05 relative to CGBP41; $= p-value≤0.05 relative to CGBP45. CGBP= core glucoraphanin biosynthetic pathway.
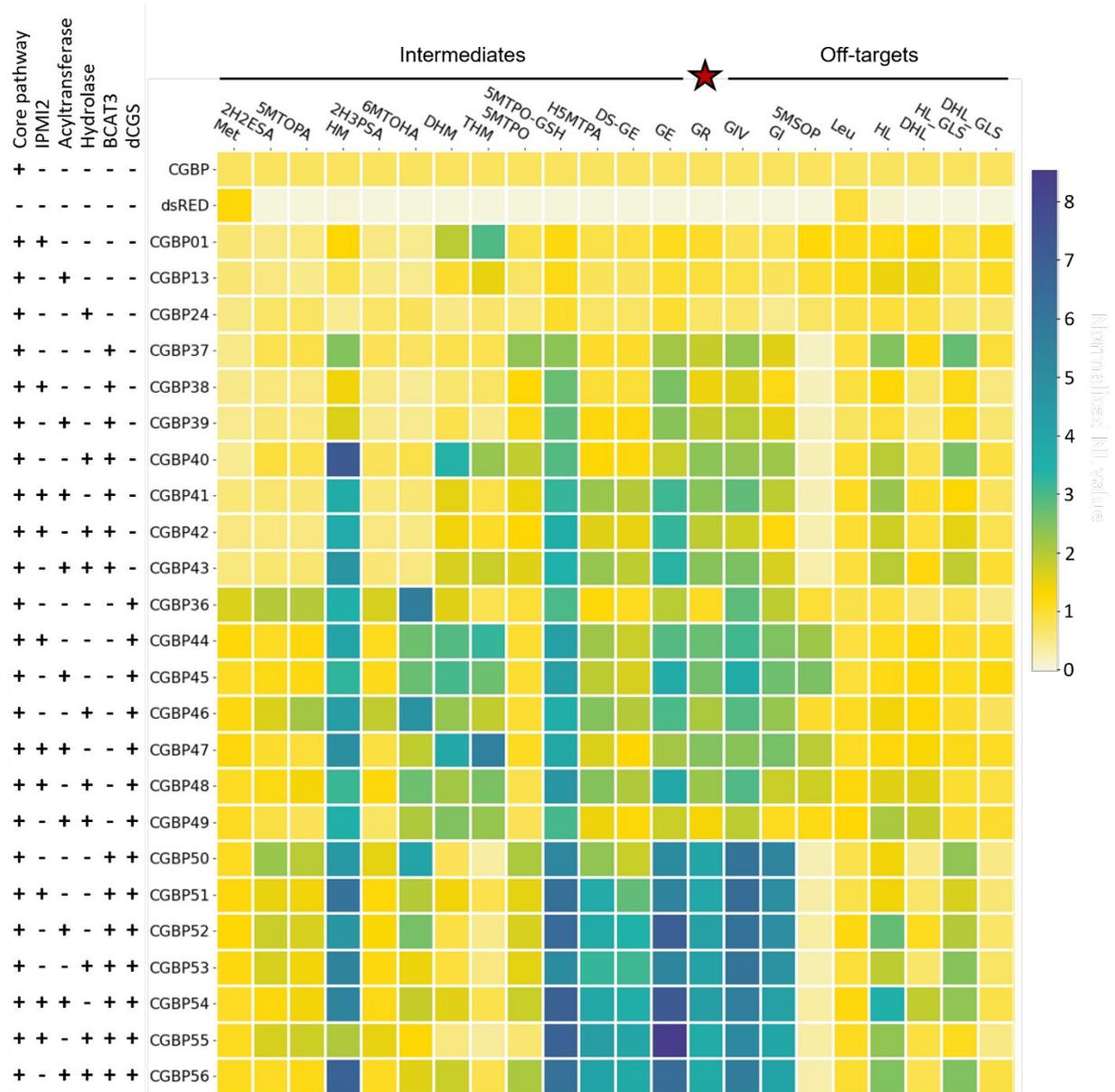


**Figure 5: LC-MS/MS analysis of putative pathway intermediates and off-target products generated by coexpression of select candidate genes.** All values were normalized to those observed when the CGBP was expressed alone (top row). N=3.Met=methionine, 2H2ESA= 2-hydroxy-2-(2-(methylthio)ethyl)succinic acid, 5MTOPA=5-methylthio-2-oxopentanoic acid, HM= homomethionine, 2H3PSA=  2-hydroxy-2-(3-(methylthio)propyl) succinic acid, 6MTOHA= 6-methylthio-2-oxohexanoic acid,

1
2
3 DHM= dihomomethionine, 5MTPO= 5-(methylthio) pentanaloxime, 5MTPO-GSH= 5-(methylthio)
4 pentanaloxime-Glutathione, H5MTPA= N-hydroxy-5- (methylthio)pentanimidothioic acid, DS-GE= desulfo-
5 glucoerucin, GE= Glucoerucin, GR= Glucoraphanin, GIV= Glucoiberverin, GI= Glucoiberin, THM=
6 trihomomethionine, 5MSOP= THM-glucosinolate, Leu= Leucine, HL= homoleucine, DHL= Dihomoleucine,
7 HL-GLS= 2-methylpropyl-glucosinolate, DHL-GLS= 3-methylbutyl-glucosinolate. Red star denotes
8 glucoraphanin.
9
10
11
12
13
14
15



**Supplemental figure 1: Relative glucoraphanin intermediate and off-target concentrations**.
Comparison done with *Arabidopsis thaliana Col-0*, store bought broccoli sprouts (BS) and core
glucoraphanin biosynthetic pathway (CGBP) expressed in *Nicotiana benthamiana*. Met=methionine,
2H2ESA= 2-hydroxy-2-(2-(methylthio)ethyl)succinic acid, 5MTOPA=5-methylthio-2-oxopentanoic acid,
HM= homomethionine, 2H3PSA= 2-hydroxy-2- (3-(methylthio)propyl) succinic acid, 6MTOHA= 6-
methylthio-2-oxohexanoic acid, DHM= dihomomethionine, 5MTPO= 5-(methylthio)pentanaloxime,
5MTPO-GSH= 5-(methylthio)pentanaloxime- Glutathione, H5MTPA= N-hydroxy-5-
(methylthio)pentanimidothioic acid, DS-GE= desulfo-glucoerucin, GE= Glucoerucin, GR= Glucoraphanin,
GIV= Glucoiberverin, GI= Glucoiberin, THM= trihomomethionine, 5MSOP= THM-glucosinolate, Leu=
Leucine, HL= homoleucine, DHL= Dihomoleucine, HL-GLS= 2-methylpropyl-glucosinolate, DHL-GLS= 3-
methylbutyl-glucosinolate.

**Supplemental figure 2: LC-MS/MS putative intermediate analysis of CGBP in combination with individual target genes of interest from coexpression analysis** (See Table S3 for gene accession numbers). Met=methionine, 2H2ESA= 2-hydroxy-2-(2-(methylthio)ethyl)succinic acid, 5MTOPA=5-methylthio-2-oxopentanoic acid, HM= homomethionine, 2H3PSA= 2-hydroxy-2-(3-(methylthio)propyl) succinic acid, 6MTOHA= 6-methylthio-2-oxohexanoic acid, DHM= dihomomethionine, 5MTPO= 5-(methylthio) pentanaloxime, 5MTPO-GSH= 5-(methylthio) pentanaloxime-Glutathione, H5MTPA= N-hydroxy-5- (methylthio)pentanimidothioic acid, DS-GE= desulfo- glucoerucin, GE= Glucoerucin, GR= Glucoraphanin, GIV= Glucoiberverin, GI= Glucoiberin, THM= trihomomethionine, 5MSOP= THM-glucosinolate, Leu= Leucine, HL= homoleucine, DHL= Dihomoleucine, HL-GLS= 2-methylpropyl-glucosinolate, DHL-GLS= 3-methylbutyl-glucosinolate. Red star denotes glucoraphanin.

## Tables

| Supplemental Table 1: Coexpression Analysis Genes, Primers and Mutual Rank Scores | | | | | |
|---|---|---|---|---|---|
| At # | Accession # | Gene | Name | Sequence (5' ---> 3') | Avg MR Score |
| 1 | At2g43100 | IPMI2 | Forward | atatagatcgaGGTCTCaAATGGCGTATTCTCTTCCTACATTTCC | 2.7 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAAGCTAATGATGGAATCATTCCCATC | |
| 2 | At5g10180 | SULTR2;1 | Forward | atatagatcgaGGTCTCaAATGAAAGAGAGAGATTCAGAGAGTTTTG | 19.65 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAAACTTTTAATCCAAAGCAAGCATCAAGA | |
| 3 | At1g78370 | GSTU20 | Forward | atatagatcgaGGTCTCaAATGGCGAACCTACCGATTCTTTTG | 7.4 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCAGAGATTGTTCTTCCTATACTCAGC | |
| 4 | At1g65860 | FMO GS-0X1 | Forward | atatagatcgaGAAGACtaAATGGCACCAACTCAAAACACAATCTG | 7.65 |
| | | | Reverse | atatagatcgaGAAGACtaAAGCTCATGATTCGAGGAAATAAGAAGGATG | |
| 5 | At1g21440 | Carboxylase | Forward | atatagatcgaGGTCTCaAATGTCGATGTTAATGGCGGCCAA | 8.5 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTATTTTGTTTCCCCTAGAGCGTTTCT | |
| 6 | At3g22740 | HMT3 | Forward | atatagatcgaCGTCTCaAATGGGATCTTTCGTGAAAGAAGAAACG | 15.7 |
| | | | Reverse | atatagatcgaCGTCTCaAAGCCTATTGCCCGAATTTGGGTTTTGATG | |
| 7 | At5g23020 | MAM3 | Forward | atatagatcgaGGTCTCaAATGGCTTCGTTACTTCTCACATCG | 11.35 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTATACAACAGCGGAAATCTGAGGG | |
| 8 | At2g46650 | CYTB5-C | Forward | atatagatcgaGGTCTCaAATGGCGAATCTAATTTCGTTTCACGAT | 13.4 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCCTACTTGTTGTTGTAGAATCTGAGAGC | |
| 9 | At4g30110 | HMAM2 | Forward | atatagatcgaGGTCTCaAATGGCGTCGAAGAAGATGACCAA | 37.35 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCCTATTCAATCACAATCTCTTTCAAGGTTC | |
| 10 | At5g44720 | Sulfurase | Forward | atatagatcgaGGTCTCaAATGGAGGAAGGTCTAAAGATTCAATCT | 28.55 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAAACAGCTGCTTCAGCTCTGG | |
| 11 | At4g14680 | ASP3 | Forward | atatagatcgaGGTCTCaAATGGCTTCCATGTCCACCGTC | 35.65 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAAACCGGAATCTTTTTCCGGAAGTTT | |
| 12 | At5g04950 | NAS1 | Forward | atatagatcgaGGTCTCaAATGGCTTGCCAAAACAATCTCGTTG | 21 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTACTCGATGGCACTAAACTCCTC | |
| 13 | At5g67150 | Transferase | Forward | atatagatcgaGGTCTCaAATGGCAGATGAAGTAGTAGTGATCTC | 22.2 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTATACAACACATACATGCTTCAAAAACTCT | |
| 14 | At5g07460 | PMSR2 | Forward | atatagatcgaGGTCTCaAATGGATTCTTCTCTGAAAACTCAGGAA | 27.9 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAGCCATAGCAGCGGATAGGG | |

| | | | | | |
|---|---|---|---|---|---|
| 15 | At3g22890 | APS1 | Forward | atatagatcgaGGTCTCaAATGGCTTCAATGGCTGCCGTC | 74.4 |
| | | | Reverse | atatagatcgaGGTCTCaAAAGCTTACACCGGAACCACTTCTGGTA | |
| 16 | At1g78490 | CYP708A3 | Forward | atatagatcgaGGTCTCaAATGAGCTCCATATGGAACGTTGC | 156.3 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCACTTGGTAGGAGACTGAGAGA | |
| 17 | At5g04590 | SIR | Forward | atatagatcgaGGTCTCaAATGTCATCGACGTTTCGAGCTCC | 51.85 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCATTGAGAAACTCCTTTGTATGTATCTATC | |
| 18 | At1g62800 | ASP4 | Forward | atatagatcgaGGTCTCaAATGAATTCCATCTTGTCAAGCGTCC | 32.35 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAGGCGATGCGAGTAACAACAG | |
| 19 | At3g01120 | MT01 | Forward | atatagatcgaGGTCTCaAATGGCCGTCTCATCATTCCAGTG | 34.8 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCAGATGGCTTCGAGAGCTTGAA | |
| 20 | At1g68600 | Transporter | Forward | atatagatcgaCGTCTCaAATGGGAGGTAAAATGGGATCAGTAC | 114 |
| | | | Reverse | atatagatcgaCGTCTCaAAGCTCAAACCTTAGGAATCTGATCAACAGC | |
| 21 | At3g57050 | CBL | Forward | atatagatcgaGGTCTCaAATGACATCTTCTCTGTCACTTCACTC | 90.25 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCCTAGAGAGGGAAGGTTTTGAAGG | |
| 22 | At2g34490 | CYP710A2 | Forward | atatagatcgaCGTCTCaAATGGTTTTCTCAGTTTCCATTTTTGCC | 38.9 |
| | | | Reverse | atatagatcgaCGTCTCaAAGCTCAGAGGTTCGGATACGTTACGA | |
| 23 | At2g37460 | umamit12 | Forward | atatagatcgaCGTCTCaAATGGAGGAAGTAAAGAAGAGGGATTG | 42.55 |
| | | | Reverse | atatagatcgaCGTCTCaAAGCTTAGACTGTTTCTACAGCTGTTCTTCT | |
| 24 | At3g23570 | alpha/beta-Hydrolases | Forward | atatagatcgaGGTCTCaAATGTCAGGTCATCAGTGCACCG | 98.55 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCACTTGAGATAGTCGATGAGCC | |
| 25 | At4g00880 | SAUR-like auxin-responsive | Forward | atatagatcgaGGTCTCaAATGGGTAACGGAGACAAAGTCATG | 47.1 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCAAACCCTAAAACACCGGATGAG | |
| 26 | At4g03050 | AOP3 | Forward | GCTTCTgtatattctgcccaaattcgcgATGGGTTCATGCAGTCCTCAACT | Negative Control |
| | | | Reverse | aaagaaaatttaatgaaaccagagttaaTTATTTCCCAGCAGAGACGCCAC | |
| 27 | At5g14200 | IPMDH1 | Forward | GCTTCTgtatattctgcccaaattcgcgATGGCGGCGTTTTTGCAAACGAA | Subsitute for IPMDH3 |
| | | | Reverse | aaagaaaatttaatgaaaccagagttaaTTAAACAGTAGCTGGAACTTTGGATTC | |
| 28 | At4g21960 | PRXR1 | Forward | atatagatcgaGGTCTCaAATGGGAGGCAAAGGTGTGATGAT | 151.6 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCAATGGTTCTTGTTTGCGAGATTACA | |
| 29 | At3g63110 | IPT3 | Forward | atatagatcgaGGTCTCaAATGATCATGAAGATATCTATGGCTATGTG | 103.2 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCACGCCACTAGACACCGC | |
| 30 | At5g10170 | MIPS3 | Forward | atatagatcgaGGTCTCaAATGTTCATCGAAAGCTTCAAGGTTGAA | 261.6 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCACTTGTACTCGAGAATCATGTTGTT | |
| 31 | At3g44990 | XTR8 | Forward | atatagatcgaCGTCTCaAATGGCTTTGTCTCTTATCTTTCTAGCT | 235.8 |
| | | | Reverse | atatagatcgaCGTCTCaAAGCTTAACATTCTGGTGTTTGGGTATGGTC | |
| 32 | At1g16060 | WRI3 | Forward | atatagatcgaGGTCTCaAATGTTCATCGCCGTCGAAGTTTC | 276.95 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTTAGCAATCATTTAACTCGCTGTAGAAATC | |
| 33 | At1g11840 | GLX1 | Forward | GCTTCTgtatattctgcccaaattcgcgGATTGTGCCAACAATGACTGGATTAC | 243.3 |
| | | | Reverse | aaagaaaatttaatgaaaccagagttaaACAATCAAAATTGGTCCGCAAATTCCG | |
| 34 | At5g01500 | TAAC | Forward | atatagatcgaGGTCTCaAATGGGAGAAGAGAAGTCTCTGCT | 294.4 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCTCAGGTTTGTTCATCGATTGTGTTAGG | |
| 35 | At4g38740 | ROC1 | Forward | atatagatcgaGGTCTCaAATGGCGTTCCCTAAGGTATACTTC | 204.3 |
| | | | Reverse | atatagatcgaGGTCTCaAAGCCTAAGAGAGCTGACCACAATCGG | |

**Supplemental Table 2: Core glucoraphanin biosynthetic pathway genes and their reactions**

| Name | Abbreviation | Accession # | Reaction^ |
|---|---|---|---|
| Branched Chain Amino acid Transaminase 4 | BCAT4 | At3g19710 | 1->2 |
| Bile Acid Transporter 5 | BAT5 | At4g12030 | transport |
| Methylthioalkylmalate Synthase 1 | MAM1 | At5g23010 | 2->3; 5->6 |
| Isopropylmalate isomerase large subunit 1 | IPMI-LSU | At4g13430 | 3->4; 6->7 |
| Isopropylmalate isomerase small subunit 1 | IPMI-SSU | At3g58990 | 3->4; 6->7 |
| Isopropylmalate dehydrogenase 1 | IPMDH1 | At5g14200 | 4->5; 7->8 |
| Cytochrome P450 79F1 | Cyp79F1 | At1g16410 | 8->9 |
| Cytochrome P450 83A1 | CYP83A1 | At4g13770 | 9->10 |
| Glutathione S-Transferase F11 | GSTF11 | At3g03190 | 10->11 |
| Gamma-glutamyl peptidase 1 | GGP1 | At4g30530 | 11->12 |
| Superroot 1 | SUR1 | At2g20610 | 11->12 |
| UDP-glycosyl transferase 74C1 | UGT74C1 | At2g31790 | 12->13 |
| Sulfotransferase 17 | SOT17 | At1g18590 | 13->14 |
| Flavin-Monooxygenase Glucosinolate S-Oxygenase 1 | FMOGS-OX1 | At1g65860 | 14->15 |

^=numbers match compounds in figure 1 and supplemental table 2

**Supplemental table 3: MS ion information for intermediates and off targets**

| compound number^ | name | Formula | MS ion polarity | m/z | Major MS2 peaks | RT (sec) |
|---|---|---|---|---|---|---|
| 1 | Methionine | C5H11NO2S | [M+H]+ | 150.0583 | 56, 104, 133 | 53 |
| 2 | 4MTOBA | C5H8O3S | [M+H]+ | 151.0423 | NA | * |
| 3 | 2H2ESA | C7H12O5S | [M-H]- | 207.0333 | 99, 127.9, 147 | 146 |
| 4 | 5MTOPA | C7H11O5S | [M-H]- | 207.0333 | NA | * |
| 5 | HM | C6H13NOS2 | [M+H]+ | 164.074 | 70, 100, 136 | 167 |
| 6 | 2H3PSA | C8H14O5S | [M-H]- | 221.0489 | 101, 161 | 175 |
| 7 | 6MTOHA | C8H13O5S | [M-H]- | 221.0489 | NA | * |
| 8 | DHM | C7H15NO2S | [M+H]+ | 178.0896 | 84, 105, 161 | 210 |
| 9 | 5MTPO | C6H13NOS | [M+H]+ | 148.0791 | 81, 84 | 141 |
| 11 | 5MTPO-GSH | C16H28N4O7S2 | [M+H]+ | 453.1472 | 116, 243, 273 | 170 |
| 12 | H5MTPA | C6H13NOS2 | [M+H]+ | 180.0511 | 89, 117, 145 | 155 |
| 13 | DS-GE | C12H23NO6S2 | [M+H]+ | 342.104 | 70, 229 | 154 |
| 14 | glucorucin | C12H23NO9S3 | [M-H]- | 420.0462 | 74.99, 96.96, 160.84 | 125 |
| 15 | glucoraphanin | C12H23NO10S3 | [M-H]- | 436.0411 | 74.99, 95.95, 96.96, 178 | 53 |
| **Off targets** | | | | | | |
| | Glucoiberverin | C11H21NO9S3 | [M-H]- | 406.0306 | 74.99, 96.96, 157.86 | 99 |
| | Glucoiberin | C11H21NO10S3 | [M-H]- | 422.0255 | 74.99, 96.96, 146 | 53 |
| | THM | C8H17NO2S | [M+H]+ | 192.1053 | 98, 174, 192 | 249 |
| | leucine | C6H13NO2 | [M+H]+ | 132.1019 | 86, 132 | 68 |
| | HL | C7H15NO2 | [M+H]+ | 146.1176 | 57, 60, 73, 114 | 116 |
| | DHL | C8H17NO2 | [M+H]+ | 160.1332 | 57, 72, 73, 114 | 169 |
| | HL-GLS | C12H23NO9S2 | [M-H]- | 388.0741 | 75, 97, 166 | 138 |
| | DHL-GLS | C13H25NO9S2 | [M-H]- | 402.0898 | 75, 97, 160 | 180 |
| **ISTD** | | | | | | |
| | CUDA | C19H36N2O3 | [M-H]- | 339.2653 | 71, 177, 214 | 404 |
| | | | [M+H]+ | 341.2806 | 100, 198, 216 | 404 |

^=compound numbers from figure 1; *=Below threshold of 10^4 ions; RT= Retention time

## References

(1) Fahey, J. W.; Zalcmann, A. T.; Talalay, P. The Chemical Diversity and Distribution of Glucosinolates and Isothiocyanates among Plants. *Phytochemistry* **2001**, *56* (1), 5–51. https://doi.org/10.1016/s0031-9422(00)00316-2.

(2) Bones, A. M.; Rossiter, J. T. The Enzymic and Chemically Induced Decomposition of Glucosinolates. *Phytochemistry* **2006**, *67* (11), 1053–1067. https://doi.org/10.1016/j.phytochem.2006.02.024.

(3) Elbarbry, F.; Elrody, N. Potential Health Benefits of Sulforaphane: A Review of the Experimental, Clinical and Epidemiological Evidences and Underlying Mechanisms. *Journal of Medicinal Plants Research* **2011**, 12.

(4) Jiang, X.; Liu, Y.; Ma, L.; Ji, R.; Qu, Y.; Xin, Y.; Lv, G. Chemopreventive Activity of Sulforaphane. *Drug Des Devel Ther* **2018**, *12*, 2905–2913. https://doi.org/10.2147/DDDT.S100534.

(5) Juge, N.; Mithen, R. F.; Traka, M. Molecular Basis for Chemoprevention by Sulforaphane: A Comprehensive Review. *Cell. Mol. Life Sci.* **2007**, *64* (9), 1105. https://doi.org/10.1007/s00018-007-6484-5.

(6) Kapoor, A.; Sharfstein, J. M. Breaking the Gridlock: Regulation of Dietary Supplements in the United States. *Drug Testing and Analysis* **2016**, *8* (3–4), 424–430. https://doi.org/10.1002/dta.1892.

(7) Yagishita, Y.; Fahey, J. W.; Dinkova-Kostova, A. T.; Kensler, T. W. Broccoli or Sulforaphane: Is It the Source or Dose That Matters? *Molecules* **2019**, *24* (19), 3593. https://doi.org/10.3390/molecules24193593.

(8)  Mikkelsen, M. D.; Olsen, C. E.; Halkier, B. A. Production of the Cancer-Preventive Glucoraphanin in Tobacco. *Mol Plant* **2010**, *3* (4), 751–759. https://doi.org/10.1093/mp/ssq020.

(9)  Schuster, J.; Knill, T.; Reichelt, M.; Gershenzon, J.; Binder, S. Branched-Chain Aminotransferase4 Is Part of the Chain Elongation Pathway in the Biosynthesis of Methionine-Derived Glucosinolates in Arabidopsis. *Plant Cell* **2006**, *18* (10), 2664–2679. https://doi.org/10.1105/tpc.105.039339.

(10) Crocoll, C.; Mirza, N.; Reichelt, M.; Gershenzon, J.; Halkier, B. A. Optimization of Engineered Production of the Glucoraphanin Precursor Dihomomethionine in Nicotiana Benthamiana. *Front Bioeng Biotechnol* **2016**, *4*. https://doi.org/10.3389/fbioe.2016.00014.

(11) Gigolashvili, T.; Yatusevich, R.; Rollwitz, I.; Humphry, M.; Gershenzon, J.; Flügge, U.-I. The Plastidic Bile Acid Transporter 5 Is Required for the Biosynthesis of Methionine-Derived Glucosinolates in Arabidopsis Thaliana. *Plant Cell* **2009**, *21* (6), 1813–1829. https://doi.org/10.1105/tpc.109.066399.

(12) He, Y.; Chen, B.; Pang, Q.; Strul, J. M.; Chen, S. Functional Specification of Arabidopsis Isopropylmalate Isomerases in Glucosinolate and Leucine Biosynthesis. *Plant and Cell Physiology* **2010**, *51* (9), 1480–1487. https://doi.org/10.1093/pcp/pcq113.

(13) Kroymann, J.; Textor, S.; Tokuhisa, J. G.; Falk, K. L.; Bartram, S.; Gershenzon, J.; Mitchell-Olds, T. A Gene Controlling Variation in Arabidopsis Glucosinolate Composition Is Part of the Methionine Chain Elongation Pathway. *Plant Physiology* **2001**, *127* (3), 1077–1088. https://doi.org/10.1104/pp.010416.

(14) Textor, S.; Bartram, S.; Kroymann, J.; Falk, K. L.; Hick, A.; Pickett, J. A.; Gershenzon, J. Biosynthesis of Methionine-Derived Glucosinolates in Arabidopsis Thaliana: Recombinant Expression and Characterization of Methylthioalkylmalate Synthase, the Condensing Enzyme of the Chain-Elongation Cycle. *Planta* **2004**, *218* (6), 1026–1035. https://doi.org/10.1007/s00425-003-1184-3.

(15) Bak, S.; Feyereisen, R. The Involvement of Two P450 Enzymes, CYP83B1 and CYP83A1, in Auxin Homeostasis and Glucosinolate Biosynthesis. *Plant Physiology* **2001**, *127* (1), 108–118. https://doi.org/10.1104/pp.127.1.108.

(16) Geu-Flores, F.; Olsen, C. E.; Halkier, B. A. Towards Engineering Glucosinolates into Non-Cruciferous Plants. *Planta* **2009**, *229* (2), 261–270. https://doi.org/10.1007/s00425-008-0825-y.

(17) Grubb, C. D.; Zipp, B. J.; Ludwig-Müller, J.; Masuno, M. N.; Molinski, T. F.; Abel, S. Arabidopsis Glucosyltransferase UGT74B1 Functions in Glucosinolate Biosynthesis and Auxin Homeostasis. *The Plant Journal* **2004**, *40* (6), 893–908. https://doi.org/10.1111/j.1365-313X.2004.02261.x.

(18) Klein, M.; Reichelt, M.; Gershenzon, J.; Papenbrock, J. The Three Desulfoglucosinolate Sulfotransferase Proteins in Arabidopsis Have Different Substrate Specificities and Are Differentially Expressed. *The FEBS Journal* **2006**, *273* (1), 122–136. https://doi.org/10.1111/j.1742-4658.2005.05048.x.

(19) Mikkelsen, M. D.; Halkier, B. A. Metabolic Engineering of Valine- and Isoleucine-Derived Glucosinolates in Arabidopsis Expressing CYP79D2 from Cassava. *Plant Physiology* **2003**, *131* (2), 773–779. https://doi.org/10.1104/pp.013425.

(20) Piotrowski, M.; Schemenewitz, A.; Lopukhina, A.; Müller, A.; Janowitz, T.; Weiler,

E. W.; Oecking, C. Desulfoglucosinolate Sulfotransferases from Arabidopsis Thaliana Catalyze the Final Step in the Biosynthesis of the Glucosinolate Core Structure*. *Journal of Biological Chemistry* **2004**, *279* (49), 50717–50725. https://doi.org/10.1074/jbc.M407681200.

(21) Hansen, B. G.; Kliebenstein, D. J.; Halkier, B. A. Identification of a Flavin-Monooxygenase as the S-Oxygenating Enzyme in Aliphatic Glucosinolate Biosynthesis in Arabidopsis. *The Plant Journal* **2007**, *50* (5), 902–910. https://doi.org/10.1111/j.1365-313X.2007.03101.x.

(22) Mirza, N.; Crocoll, C.; Erik Olsen, C.; Ann Halkier, B. Engineering of Methionine Chain Elongation Part of Glucoraphanin Pathway in E. Coli. *Metabolic Engineering* **2016**, *35*, 31–37. https://doi.org/10.1016/j.ymben.2015.09.012.

(23) Yang, H.; Liu, F.; Li, Y.; Yu, B. Reconstructing Biosynthetic Pathway of the Plant-Derived Cancer Chemopreventive-Precursor Glucoraphanin in Escherichia Coli. *ACS Synth. Biol.* **2018**, *7* (1), 121–131. https://doi.org/10.1021/acssynbio.7b00256.

(24) Yang, H.; Qin, J.; Wang, X.; El-Shora, H. M.; Yu, B. Production of Plant-Derived Anticancer Precursor Glucoraphanin in Chromosomally Engineered Escherichia Coli. *Microbiological Research* **2020**, *238*, 126484. https://doi.org/10.1016/j.micres.2020.126484.

(25) Mikkelsen, M. D.; Buron, L. D.; Salomonsen, B.; Olsen, C. E.; Hansen, B. G.; Mortensen, U. H.; Halkier, B. A. Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast Expression Platform. *Metabolic Engineering* **2012**, *14* (2), 104–111. https://doi.org/10.1016/j.ymben.2012.01.006.

(26) Møldrup, M. E.; Geu-Flores, F.; Olsen, C. E.; Halkier, B. A. Modulation of Sulfur Metabolism Enables Efficient Glucosinolate Engineering. *BMC Biotechnology* **2011**, *11* (1), 12. https://doi.org/10.1186/1472-6750-11-12.

(27) Obayashi, T.; Aoki, Y.; Tadaka, S.; Kagaya, Y.; Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol* **2018**, *59* (1), e3–e3. https://doi.org/10.1093/pcp/pcx191.

(28) Textor, S.; Kraker, J.-W. de; Hause, B.; Gershenzon, J.; Tokuhisa, J. G. MAM3 Catalyzes the Formation of All Aliphatic Glucosinolate Chain Lengths in Arabidopsis. *Plant Physiology* **2007**, *144* (1), 60–71. https://doi.org/10.1104/pp.106.091579.

(29) Chhajed, S.; Mostafa, I.; He, Y.; Abou-Hashem, M.; El-Domiaty, M.; Chen, S. Glucosinolate Biosynthesis and the Glucosinolate–Myrosinase System in Plant Defense. *Agronomy* **2020**, *10* (11), 1786. https://doi.org/10.3390/agronomy10111786.

(30) Knill, T.; Schuster, J.; Reichelt, M.; Gershenzon, J.; Binder, S. Arabidopsis Branched-Chain Aminotransferase 3 Functions in Both Amino Acid and Glucosinolate Biosynthesis. *Plant Physiology* **2008**, *146* (3), 1028–1039. https://doi.org/10.1104/pp.107.111609.

(31) Matityahu, I.; Godo, I.; Hacham, Y.; Amir, R. Tobacco Seeds Expressing Feedback-Insensitive Cystathionine Gamma-Synthase Exhibit Elevated Content of Methionine and Altered Primary Metabolic Profile. *BMC Plant Biology* **2013**, *13*

(1), 206. https://doi.org/10.1186/1471-2229-13-206.

(32) Song, S.; Hou, W.; Godo, I.; Wu, C.; Yu, Y.; Matityahu, I.; Hacham, Y.; Sun, S.; Han, T.; Amir, R. Soybean Seeds Expressing Feedback-Insensitive Cystathionine γ-Synthase Exhibit a Higher Content of Methionine. *J Exp Bot* **2013**, *64* (7), 1917–1926. https://doi.org/10.1093/jxb/ert053.

(33) Chambers, K. F.; Bacon, J. R.; Kemsley, E. K.; Mills, R. D.; Ball, R. Y.; Mithen, R. F.; Traka, M. H. Gene Expression Profile of Primary Prostate Epithelial and Stromal Cells in Response to Sulforaphane or Iberin Exposure. *The Prostate* **2009**, *69* (13), 1411–1421. https://doi.org/10.1002/pros.20986.

(34) Mitsiogianni, M.; Trafalis, D. T.; Franco, R.; Zoumpourlis, V.; Pappa, A.; Panayiotidis, M. I. Sulforaphane and Iberin Are Potent Epigenetic Modulators of Histone Acetylation and Methylation in Malignant Melanoma. *Eur J Nutr* **2021**, *60* (1), 147–158. https://doi.org/10.1007/s00394-020-02227-y.

(35) Wang, W.; Wang, S.; Howie, A. F.; Beckett, G. J.; Mithen, R.; Bao, Y. Sulforaphane, Erucin, and Iberin Up-Regulate Thioredoxin Reductase 1 Expression in Human MCF-7 Cells. *J. Agric. Food Chem.* **2005**, *53* (5), 1417–1421. https://doi.org/10.1021/jf048153j.

(36) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **2009**, *4* (5), e5553. https://doi.org/10.1371/journal.pone.0005553.

(37) Froger, A.; Hall, J. E. Transformation of Plasmid DNA into E. Coli Using the Heat Shock Method. *JoVE* **2007**, No. 6, 253. https://doi.org/10.3791/253.

(38) Lin, J.J. Electrotransformation of Agrobacterium. In *Electroporation Protocols for Microorganisms*; Humana Press: New Jersey, **1995**; Vol. 47, pp 171–178. https://doi.org/10.1385/0-89603-310-4:171.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18