

** indicates genes that may have been used in initial ICOR MATLAB analysis.

FALVAC-1	** FALVAC-1 is a vaccine against Plasmodium Falciparum. This is a good benchmark for a synthetic gene as it is a real candidate vaccine that is recombinantly produced (Reference Paper).
PEA	** Pseudomonas aeruginosa exotoxin A (PEA) is an important pathogenic factor 1 . It retains high immunogenicity even after detoxification, enabling its use as vaccine adjuvants and vaccine carriers. This is a good benchmark as it can be produced for a low-cost at a large-scale in E. coli. Further, the (Reference Paper) finds that codon optimization enhances expression of PEA in E. coli -- thus, if the tool presented in this research can achieve similar/better results than the paper and/or other approaches, it will be considered an improvement.
msox	** Monomeric sarcosine oxidase (Msox) is an important diagnostic enzyme and can catalyze the oxidation of N-methyl with the FAD cofactor. (Study) found that the expression of recombinant SOX "has been extensively studied." Thus, the sox gene serves as a good benchmark because the extent to which codon optimization affects it has been studied.
soxB	** Sarcosine oxidase subunit B (soxB) is an important diagnostic enzyme and can catalyze the oxidation of N-methyl with the FAD cofactor. (Study) found that the expression of recombinant SOX "has been extensively studied." Thus, the sox gene serves as a good benchmark because the extent to which codon optimization affects it has been studied.
PRL-3/PTP4A3	** Protein tyrosine phosphatase 4A3 (PTP4A3/PRL-3) is a well-documented protein that can be produced recombinantly in E. coli. This protein has been identified as a potential target to treat some cancers.
hPDF	** Human peptide deformylase (hPDF) is a target for cancer therapeutics. However, its expression is not very efficient in E. coli. This serves as a good target for benchmarks as past studies have noted the valuable potential of codon optimization of this gene.
PF3D7	** Circumsporozoite protein is used as a surface antigen on the sporozoite of a malaria parasite. It has been studied in papers that attempt to quantify gene expression.
PA	** Polymerase acidic protein (PA) plays a role in viral RNA transcription and replication. It is from the Influenza A virus. It has been studied in (codon optimization papers).
BIRC5	BIRC5 (Baculoviral IAP repeat-containing 5) is a multitasking protein that has roles in preventing apoptosis and cell proliferation (PubMed: 9859993 , PubMed: 21364656 , PubMed: 20627126 , PubMed: 25778398 , PubMed: 28218735). In Reference Paper , BIRC5 expression levels were analyzed in E. Coli. Sequence optimization (generated by using a sliding

	window → narrowed down) was used for this human gene to be expressed in <i>E. coli</i> .
BRAF1	BRAF1 (v-raf murine sarcoma viral oncogene homolog B1) is a protein kinase that transduces mitogenic signals from the cell membrane to the nucleus. This kinase phosphorylates MAP2K1 which activates the MAP kinase signal pathway (PubMed: 21441910 , PubMed: 29433126). This gene is a gene of interest because it is frequently mutated and allows a cell to become a tumor cell. In Reference Paper , BRAF1 expression levels were analyzed in <i>E. Coli</i> .
CAV1	CAV1 (Caveolin 1, caveolae protein) is involved in promoting cell cycle progression. It's a gene of interest because it is a tumor suppressor candidate making it potentially useful in recombinant expression. In Reference Paper , CAV1 expression levels were analyzed in <i>E. Coli</i> .
CD80	Cd80 (Mus musculus CD80 antigen) is a protein-coding gene whose protein, once activated, induces T-cell proliferation and cytokine production. In Reference Paper , Cd80 expression levels were analyzed in <i>E. Coli</i> .
CDK1	CDK1 (Cyclin Dependent Kinase 1) codes for a protein which is essential for G1/S and G2/M phase transitions. It has been used in calculating prognostic value in human cancer (Reference Paper). Higher expression levels correlated with a more-advanced tumor. Another reference paper analyzing the importance of this gene looked at expression and structure through recombinant expression in <i>E. coli</i> (CDK1 structures reveal conserved and unique features of the essential cell cycle CDK).
CEBPZ	CEBPZ (CCAAT/enhancer binding protein zeta) is a human protein-encoding gene that plays a role in responding to environmental stimuli (heat). In Reference Paper , CEBPZ expression levels were analyzed in <i>E. Coli</i> which makes it a good benchmark gene to compare to.
CLN3	CLN3 (Ceroid-lipofuscinosis, neuronal 3) is a gene that encodes a protein involved in lysosomal function. This is an interesting gene of study because it can cause neurodegenerative diseases once mutated (Batten disease). In Reference Paper , CLN3 expression levels were analyzed in <i>E. Coli</i> which makes it a good benchmark gene to compare to.
CREB1	CREB1 (cAMP responsive element binding protein 1) mutations can increase risk for certain diseases. Its most common variant is actually from a tumor. It is from a family of transcription factors that is expressed in the brain. In Reference Paper , CREB1 expression levels were analyzed in <i>E. Coli</i> .
CSNK1A1	CSNK1A1 (Casein Kinase 1 Alpha 1) belongs to the protein kinase superfamily and is involved in phosphorylating numerous proteins involved in cellular functions. Recombinant human CSNK1a1 protein, fused to His-tag at N-terminus, was expressed in <i>E.coli</i> by CSNK1A1_1-337aa Human, His tag, E.coli (GWB-ATG0D4) .

FGFR4	FGFR4 (Fibroblast growth factor receptor 4) comes under the kinases category and the protein encoded by this gene is a tyrosine kinase and cell surface receptor for fibroblast growth factors. Diseases associated with FGFR4 include Prostate Cancer and Rhabdomyosarcoma . In Reference Paper , FGFR4 expression levels were analyzed in <i>E. Coli</i> .
GSK3B	GSK3B (Glycogen Synthase Kinase 3 Beta) encodes for a protein that is a negative regulator of glucose homeostasis. Mutations that lead to defects in this gene are associated with Parkinson disease and Alzheimer disease. This is a gene of interest because it is found to be an important enzyme in glycogen metabolism. In Reference Paper , GSK3B expression levels were analyzed in <i>E. Coli</i> .
JUN	JUN (Jun Proto-Oncogene, AP-1 Transcription Factor Subunit) can often be mutated to cause a cell to become a tumor cell. JUN is the “putative transforming gene” of avian sarcoma virus 17 (Reference). It was expressed as a “transcription factor” in this Reference Paper in <i>E. coli</i> ; its expression was measured in vivo.
KIF11	KIF11 (Kinesin family member 11) was expressed recombinantly in <i>E. coli</i> in this study . “This gene encodes a motor protein that belongs to the kinesin-like protein family. Members of this protein family are known to be involved in various kinds of spindle dynamics.” (KIF11 Gene (Protein Coding)).
LAMP1	Lysosomal-associated membrane protein 1 (LAMP1) is a protein coding gene that is associated with Chediak-Higashi Syndrome and Gaucher's Disease . This is a gene of interest because it may play a role in tumor cell metastasis. It was expressed as a “transcription factor” in this Reference Paper in <i>E. coli</i> ; its expression was measured in vivo.
LEMD3	LEM domain containing 3 (LEMD3) is a human gene that is of interest. Mutations in this gene have been associated with osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. It was expressed as a “membrane protein” in this Reference Paper in <i>E. coli</i> ; its expression was measured in vivo.
MAPK1	Mitogen-Activated Protein Kinase 1 (MAPK1) is a part of the MAP kinase signal transduction pathway. It is of special interest because in this Reference Paper , when expressed in <i>E. coli</i> , there was a significant difference in the protein yield for their wild-type and optimized genes: 24.3 and 11.5 mg/L respectively.
MAPKAPK5	Mitogen-activated protein kinase-activated protein kinase 5 (MAPKAPK5) is a gene that encodes for a tumor suppressor. It is activated by the MAPK1 kinase and also plays an important role in the MAP kinase signal transduction pathway. It is of special interest because in this Reference Paper , when expressed in <i>E. coli</i> , there was a significant difference in the protein yield for their wild-type and optimized genes: 14.9 and 80 mg/L respectively.

NGFR	Nerve growth factor receptor (NGFR) is a protein coding gene associated with important pathways within the cell. Diseases associated with NGFR include Prurigo Nodularis and Infiltrative Basal Cell Carcinoma . In Reference Paper , NGFR expression levels were analyzed in <i>E. Coli</i> .
NOC2L	Nucleolar complex associated 2 homolog (<i>S. cerevisiae</i>) (NOC2L) can control major aspects of transcriptional regulation. It is an RNA/Ribosomal protein. In Reference Paper , NOC2L expression levels were analyzed in <i>E. Coli</i> .
OPRM1	"This gene encodes one of at least three opioid receptors in humans; the mu opioid receptor (MOR). The MOR is the principal target of endogenous opioid peptides and opioid analgesic agents such as beta-endorphin and enkephalins. The MOR also has an important role in dependence to other drugs of abuse, such as nicotine, cocaine, and alcohol via its modulation of the dopamine system." (Study). In Reference Paper , OPRM1 expression levels were analyzed in <i>E. Coli</i> .
PDCD11	Programmed Cell Death 11 (PDCD11) is a useful binding protein that colocalizes with U3 RNA (MIM 180710) in the nucleolus and is required for rRNA maturation and generation. It is important because, as a plasma protein, it is within the nucleolus of the cell and is a ribosomal protein. In Reference Paper , PDCD11 expression levels were analyzed in <i>E. Coli</i> .