Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 9 Issue 6 June 2025

Research Article

QED Eukaryote Genetic Code and Principle of Information Flow and Biological Protein Synthesis

Rama Shankar Singh*

GEN-MED R&D, LLC, Orlando, FL 32817, USA

***Corresponding Author:** Rama Shankar Singh, GEN-MED R&D, LLC, Orlando, FL 32817, USA.

DOI: 10.31080/ASMS.2025.09.2097

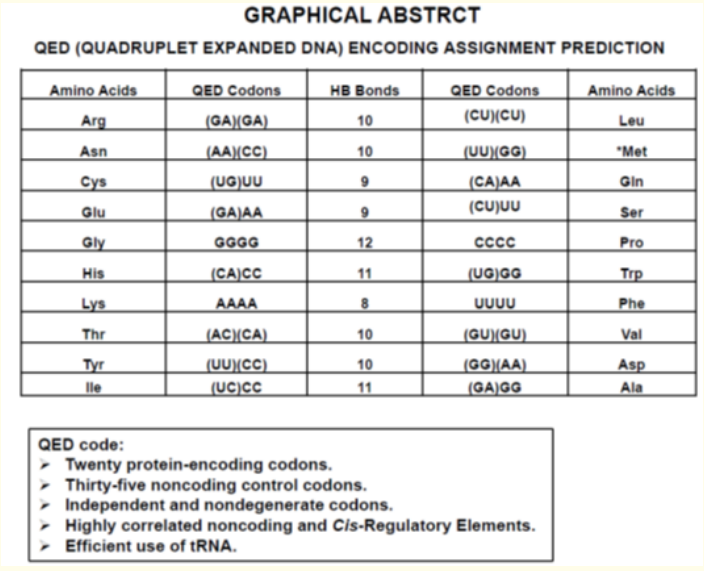
Received: March 12, 2025

Published: May 22, 2025

Abstract

The QED (Quadruplet expanded DNA) eukaryote genetic code model comprises all four DNA (T, A, C, and G) bases, the base position being independent and symmetric. In this model, the adjacent bases that naturally pair are designated noncoding. Based on these assumptions, the QED code consists of twenty nondegenerate independent protein-encoding and thirty-five nondegenerate noncoding codons, demonstrating a strong correlation with cis-regulatory elements. Despite variations in genetic pathways and information flow, the QED coding system applies to protein synthesis across eukaryotes, prokaryotes, and viruses. This broad applicability underscores fundamental principles of genetic information processing and protein biosynthesis. The search for gene-disease causality revealed the absence of a dedicated eukaryote genetic code for protein synthesis—a discovery that led to the development of the QED code. Since functional proteins are essential for cellular homeostasis, while dysfunctional proteins contribute to disease, addressing these dysfunctions is crucial for therapeutic advancements. The role of the QED code in understanding and correcting genetic dysfunctions at both the DNA and protein levels presents a transformative approach to disease treatment. This pioneering framework is expected to drive a paradigm shift in research and development, opening new avenues for treating monogenic rare diseases, multigenic cancers, and neurodegenerative disorders.

Keywords: Eukaryote; Quadruplet; Expanded; Genetic Coding; Nondegenerate; Prokaryote; Viruses; Central; Principle; Biology



Graphical Abstract

Introduction

The hypothesis of the Central Dogma of Biology [1,2] was to synthesize protein in Biology by flowing DNA information sequentially one way to protein. Subsequently, the triplet coding proposal [3] and its verification made protein synthesis a reality for prokaryotes but not viruses and eukaryotes. The broad application of QED code supports the principle of information flow and biological protein synthesis for eukaryotes, prokaryotes, and viruses.

The proposed QED (Quadruplet Expanded DNA) eukaryotic genetic code model is the first and only eukaryotic genetic code ever announced, introduced in December 2023 [4]. Its highly correlated noncoding QED codons are associated with cis-regulatory elements (Oct. 2024) [5]. The timing of this discovery coincides with my 50th anniversary of my studying DNA as the hereditary material, along with the four bases—T, C, A, and G—dating back to around 1974.

Motivation: Searching cure for incurable rare disease daughter suffered

The development of the QED genetic code is deeply personal. In 1968, while pursuing my Ph.D. in Electrical Engineering, my family faced an unimaginable challenge—our newborn child was diagnosed with a rare, incurable disease. The attending physicians suggested that the condition might be linked to our genes. At the time, the concept of genes was completely foreign to me, and the idea that they could be responsible for our child's illness was both bewildering and haunting. Genetic testing in 1968 was purely science fiction, yet this experience set me on a lifelong journey to uncover gene-disease causality relationships, ultimately leading to the development of the QED genetic code.

The relentless desire to understand gene-disease causality and seek potential cures for rare diseases fueled my work. Human eukaryotic cells carry hereditary DNA information and the molecular tools necessary to synthesize proteins, maintaining homeostasis. However, genetic variants, transcription errors, and splicing defects can lead to dysfunctional proteins, ultimately causing disease.

Scientific discoveries around the 1950s

The 1950s marked a significant period of scientific advancement, accelerating developments in various fields, including physical sciences and biology. Key breakthroughs during this time included Claude E. Shannon's theory of secure communication (1948) [6], the discovery of the transistor (1948) [7-9], and the elucidation of the DNA structure (1953) [10-12].

Shannon's theory of secure communication

Claude Shannon developed a mathematical framework for secure communication, ensuring reliable message transmission between two points (A and B) despite interference or interception. He introduced the concept of information measurement using bits, defining a unit of information as $\log_2 2 = 1$ bit. His work laid the foundation for modern encryption and coding systems, influencing fields such as cryptography, data compression, and digital communication.

Today, encryption strength is determined by the number of bits used in encoding. For example, modern 64-bit computing architectures support encryption methods ranging from 128 to 258 bits, meaning it would take 2^{128} to 2^{258} attempts to break the code. Additionally, advancements in communication technology have drastically improved data transmission speeds, with modern 5G networks offering download speeds of up to 20 Gbit/s and upload speeds of 10 Gbit/s, a testament to the impact of Shannon's work.

The discovery of the transistor

The invention of the transistor revolutionized modern technology, impacting audio, video, entertainment, and communication systems. As a three-terminal semiconductor device, the transistor can switch between ON (1) and OFF (0) states by applying voltage to its gate, forming the basis of digital circuits.

The development of monolithic planar transistor fabrication on silicon wafers enabled system-on-chip (SoC) integration. Manufacturing advancements have progressively reduced transistor sizes. For example, the A11 Bionic chip in the iPhone X (not an Apple endorsement) was fabricated using 10-nanometer (nm) technology—five times smaller than the 2-nm diameter of a DNA double helix—and contained approximately 4.5 billion transistors. Current U.S. semiconductor manufacturing aims to produce chips with 4-nm transistors, potentially allowing a single SoC to house around 12 billion transistors.

The central dogma of biology

Francis Crick was a visionary in molecular biology. After the discovery of DNA's structure, he hypothesized in 1958 [1,2] what would later be known as the Central Dogma of Biology, stating that genetic information flows in one direction: from DNA to messenger RNA (mRNA) to protein. At the time, no experimental data existed to support this hypothesis, but it would later become a fundamental principle of molecular biology.

In 1963, Crick further proposed the triplet coding hypothesis [3], suggesting that each amino acid is encoded by a set of three DNA bases. Protein synthesis occurs when three mRNA bases (codons) pair with corresponding transfer RNA (tRNA) anticodons in the ribosome. This process ensures error-free protein synthesis [13]. When perfect base-pairing does not occur, a phenomenon known as wobble base pairing [14] allows flexibility between the third codon base and the first anticodon base. By 1968, the full genetic code had been deciphered [15], identifying 64 triplet codons: 61 encoding amino acids, one serving as a START signal (AUG), and three functioning as STOP codons.

The nonoptimal nature of triplet coding

Despite its effectiveness, the triplet coding system is not mathematically optimal. According to Shannon's communication theory, the minimum number of bits required to encode 20 amino acids is $\log_2(20) = 4.32$ bits. However, the triplet codon system utilizes 6 bits ($2^6 = 64$ codons), making it redundant and degenerate, where multiple codons encode the same amino acid. Additionally, there are fewer than 20 distinct tRNAs, meaning that some tRNAs (iso-acceptor tRNAs) recognize multiple codons. The dual role of AUG as both a Met codon and a START signal remains unresolved in molecular biology.

The principle of information flow and protein synthesis

By 1958, protein synthesis was identified as the most critical biological process for prokaryotic organisms, as eukaryotes were not discovered until 1977. Subsequent advances in molecular biology expanded the understanding of protein synthesis across viruses, prokaryotes, and eukaryotes. The QED code later emerged as a broader theoretical framework accommodating these findings.

Biological information flow – left to right and protein synthesis

The flow of information in prokaryote, viruses and eukaryote is listed in Table 1.

Prokaryote	DNA	mRNA	Protein	Central Dogma	
	Gene Control	Operon	*Lac*	*Trp*	protein
Viruses	mRNA	cDNA	mRNA	protein	
Eukaryote	DNA	Transcription	Splicing	mRNA	Protein
	mRNA	DNA	Telomere	Telomerase	DNA repair
	DNA	DNA	Transposon		
	Protein	Protein	Disease	Future cure	Development
	Protein	DNA	Disease	Future cure	Development
	Virus	Trojan horse	Nano lipid	mRNA, DNA	Drug delivery
	Protein	Protein	DNA(Gene)	Circadian clock	Control

Table 1: Biological information flow – from left to right.

Prokaryote

In 1958, the prokaryote structure was the only known entity. The continuous distribution of mRNA bases and the resulting continuous protein synthesis allowed for the expression of only one gene—one protein. In the prokaryote, DNA and the resulting mRNA from transcription were located in the cytoplasm, where protein synthesis took place. The process of DNA to +mRNA to protein in the prokaryote is depicted in Figure 1, marking a significant milestone in our understanding of genetic processes.

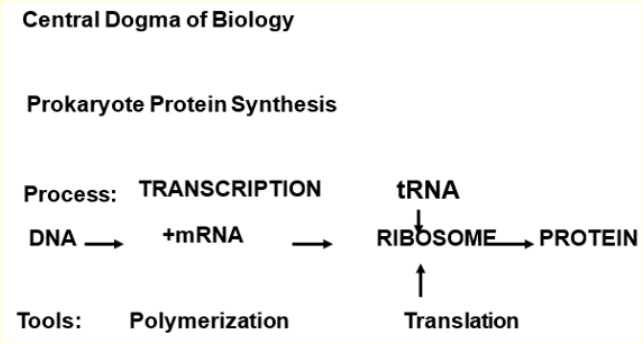


Figure 1

Viruses

The first violation of the central dogma of biology occurred in Viruses [16], where the sequence was viruses' +mRNA—RTP—cDNA—mRNA—protein. Due to a change in the flow of information in viruses, triplet coding lacked the tools and process to synthesize the protein. The virus's protein synthesis process and tools are discussed.

Viruses' mRNAs are single-stranded and double-stranded and have different polarities, requiring unique steps to synthesize the protein.

- + mRNA- Virus + mRNA is converted into cDNA (complementary DNA) using transcriptase (RT), then cDNA – mRNA- Protein flow is followed to synthesize the protein as sketched in Figure 2.
- - mRNA – First, -mRNA is converted into +mRNA using RDRNP (RNA-dependent RNA polymerase). Once +mRNA is available, the remaining protein synthesis process flow is the same as in Figure 2
- dsRNA—dsRNA has both—mRNA and + mRNA polarities. The first step is to separate—mRNA from +mRNA from dsRNA using RDRNP. Once +mRNA is available, the remaining protein synthesis flow is as in Figure 2.

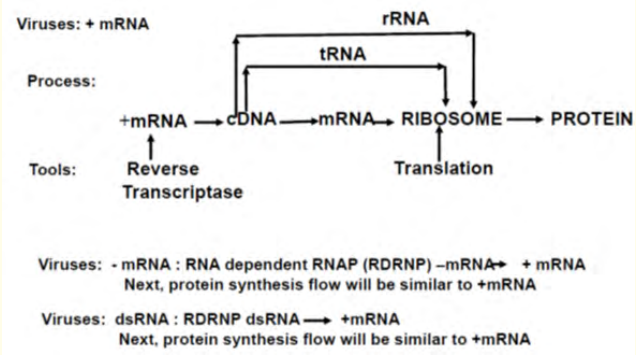


Figure 2: Viruses protein synthesis.

Prokaryote gene control

Triplet coding lacks any prokaryote gene control. The metabolite-induced Lactose digestion and Tryptophan synthesis each require three controlled enzyme syntheses. Since prokaryote protein synthesis is continuous, triplet coding cannot interrupt the process once it has started. The two well-known *lac* and *trp* operons [17-20] use a regulator, repressor, operator, and promotor to control gene polymerization for stopping and starting the enzyme synthesis.

Lac Operon

The lac operon default is negative. When the metabolite signal is absent (OFF), the regulator-generated repressor blocks the operator and promotor (OFF), blocking polymerization and protein synthesis, as sketched in Figure 3.

When the metabolite lactose signal is present, the regulator generates a repressor, which, with lactose, turns the operator ON, freeing the promotor to initiate the polymerization process for continuous synthesis of the three enzymes, as sketched in Figure 3.

Trp Operon

The *trp* operon default is positive, contrary to *the lac* operon. When the metabolite tryptophan signal is present, the regulator-generated repressor and metabolite keep the operator and promotor blocking (OFF) polymerization and tryptophan synthesis.

When the metabolite tryptophan signal is absent, the regulator-generated repressor with Tryptophan has no effect, the operator remains ON for polymerization, and three genes are continuously synthesized, as sketched in Figure 3.

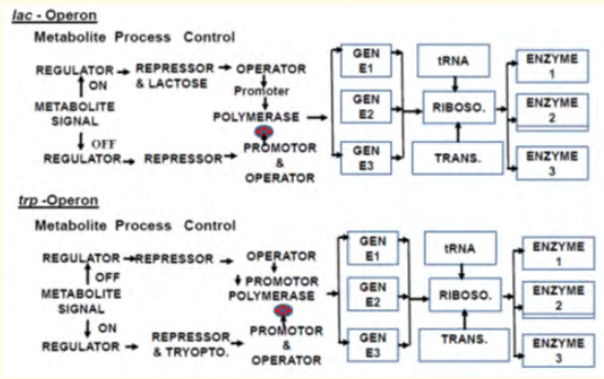


Figure 3: Prokaryote gene controls by lac and trp operons.

Eukaryote

In 1977, the discovery of the eukaryote's 'Split Gene' [21,22] revealed a complex genetic structure. DNA was found in the nucleus, and mRNAs were discretely distributed: < 2% encoding (exon) separated by > 98% noncoding (intron). The eukaryote transcription [23] process, starting with *Cis*-regulatory elements located upstream of the gene, introduced a new level of complexity. The process involved the start by TATA box, TFB protein, and controllers: promoter, activator, enhancer, and sensor to generate Pol-II followed by alternate splicing to get exon (mRNA) into the cytoplasm for protein synthesis. This complexity allows for the generation of multiple genes from one gene, making one gene-multiple proteins synthesis a possibility compared to prokaryote's one-gene- one-protein.

All eukaryote cells do not continually synthesize protein; they only do so when needed and triggered by Cell-Cell interaction. For protein synthesis, the specific Cell will generate a ligand, which will be detected and attached to the surface receptor of the destination cell to trigger the protein synthesis. The information flow sequence, process, and controls for protein synthesis are sketched in Figure 4. The information flows are.

DNA, Polymerization (Pol-I, Pol-II and Pol-III), Transcription, Splicing, mRNA, Protein.

Cis-regulatory elements and noncoding are unique features of the eukaryote, requiring new genetic coding. The proposed QED eukaryote genetic code [4,5] model has both protein-encoding and noncoding coding highly correlated with *Cis*-regulatory elements that meet the eukaryote requirements.

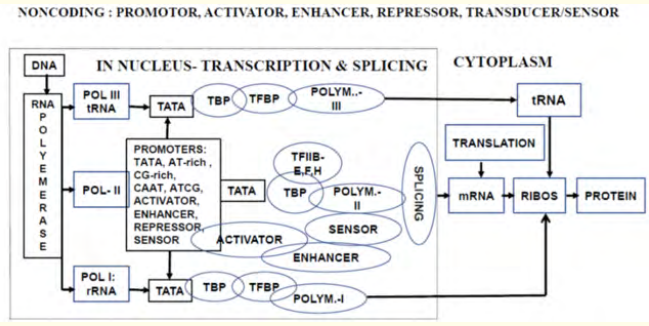


Figure 4: Eukaryote protein synthesis.

Cell's published leading papers, especially the Cell's fiftieth issue, have been a treasure trove in finding progress in molecular biology and genetics. Researchers have followed the central dogma of the biology of prokaryote protein synthesis and triplet genetic code for fifty years. Without using eukaryote genetic code, attempts have been made to fit *Cis*-regulatory elements [24,25] and noncoding [26,27] following prokaryote and triplet genetic code. Because of the structural differences among prokaryotes, viruses, and eukaryotes and the lack of triplet coding controls, the development of a new versatile QED eukaryote genetic code is not just a possibility, but an urgent necessity to synthesize proteins for all. The proposed QED eukaryote genetic [4,5] model meets the requirements of eukaryotes, prokaryotes, and viruses.

mRNA to DNA

Eukaryote DNA is packed in chromosomes, and Telomeres protect the ends. When Cells divide, DNA is separated and divided. The DNA polymerase opens the double-stranded, and each leading and lagging strand is duplicated. The leading strand is uninterrupted, but the lagging strand is duplicated with a series of missing repeating $(CCCAA)_n$ DNA bases at the end. The telomerase enzyme, which has an mRNA template, fills the missing bases at the DNA end with no gap. The use of mRNA to repair DNA end was recognized by awarding the 2009 Medicine and Physiology Nobel Prize to Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak for their work [28].

DNA to DNA

DNA to DNA is a typical jumping of genes from one chromosome to the other chromosome during mitosis of the Cell. While experimenting with maize in the thirties and forties', Barbar McClintock observed [29,30] colorful kernels in maize and termed it the movement of genes. Her experiment with maize was like Mendel's genetic experiment with peas in 1886. In 1983, the Nobel Prize in Medicine and Physiology was awarded to Barbara McClintock for discovering 'mobile genetic elements.' These elements, also known as transposons, are DNA sequences that can change their position within a genome, and they are now active in research on health and diseases, particularly in understanding genetic mutations and their role in disease development.

Protein to protein

Protein—Protein transfer is an anticipated approach for curing rare diseases by correcting dysfunctional proteins into functional proteins using the nondegenerate encoding QED eukaryotic genetic code. The degenerate triplet genetic code will have ambiguity when choosing a unique codon. The procedure is sketched in Figure 5.

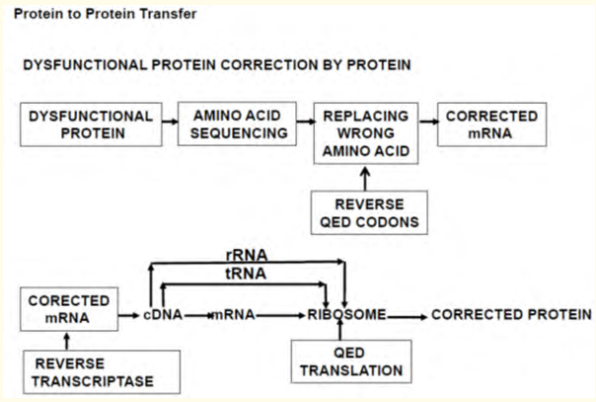


Figure 5

First, a dysfunctional protein is identified, and its amino acid is sequenced to find the wrong amino acid. Next, the incorrect amino acid is replaced by the correct one. The reverse QED codon generates the correct mRNA bases. Finally, reverse transcriptase generates cDNA, and QED codons translate into functional proteins.

Protein to DNA

The protein-DNA approach is like a hybrid gene therapy that corrects dysfunctional proteins at the DNA level. The mutated gene is identified, followed by DNA sequencing to identify the mutated bases. A corrected mRNA sequence is generated using CRISPR base-editing tools. The reverse transcriptase of corrected mRNA generates cDNA, followed by mRNA, and QED codons translate into functional proteins, as sketched in Figure 6.

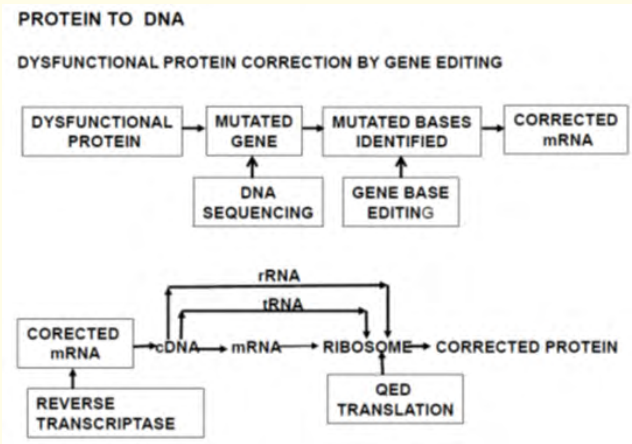


Figure 6

Virus in Lipid Nanoparticle: A Trojan Horse for Modified mRNA vaccine delivery

The best example is the COVID-19 vaccine developed using the nucleoside-modified Uridine (U) to pseudouridine mRNA procedure [31]. The modified mRNA is packed in a lipid nanoparticle [32] and delivered to the muscle cell. Moderna (mRNA-1273) used modified mRNA to manufacture COVID-19 and distributed it for vaccination, which saved millions of lives during the COVID-19 2020-2021 pandemic.

Protein-protein to Gene: Two proteins controlling Gene Circadian Clock

Eukaryote cells produce protein to maintain our body's homeostasis and control our daily day-night activities by the circadian clock. The circadian clock timing is disturbed when traveling between different time zones, causing jet lag. Since proteins produced in the Cell support all these activities, proteins produced in the Cell also control the circadian clock. Recently, using the Drosophila period (PER) gene, it was shown [33] that one protein by PER acts as negative feedback, blocking the PER mRNA production. The other two proteins control the transcription to turn the circadian clock ON for the production of mRNA. Here, the proteins produced by the cell control circadian clock activities.

QED eukaryote code and human diseases

Since the QED eukaryote code was developed to find cures for incurable rare diseases, How QED can aid in developing cures for rare diseases is illustrated using the three incurable rare diseases: Harlequin Ichthyosis (HI), Cystic Fibrosis, and Sickle cell disease.

Harlequin Ichthyosis (HI), Cystic Fibrosis, and Sickle Cell incurable monogenic rare diseases

The approach does not exhaust scientific discussion of the disease but points to the role of QED coding in overcoming the triplet genetic coding hurdle in finding the cure. More than 7000 rare diseases are listed on the NIH website, but there is no cure, only the management of the symptoms. So far, NIH has not yet even developed eukaryote genetic code like QED. Three examples are selected to illustrate the concept. HI and CFTR are related to epithelial, sickle to the blood cells. The genetic mutations that substitute the wrong amino acids create dysfunctional proteins, and the diseases are listed in Table 2.

DISEASES	Changes	mRNA bases	Amino Acid	Triplet Codon	QED Codon	Function	Gene	Ref.
Harlequin Ichthyosis	Normal		Thr	ACAN, N (T,C,A,G)	(AC)(CA)	Lipid Secretion	ABCA12	GeneTesting
OMIM 242500	Mutation	DupAA	Arg	CGN, AGA, AGG	(GA)(GA)	No lipid secretion		34
	Normal	G	Cys	UGU,UGC	(UG)UU	Lipid Secretion		35, 36
	Mutation	A	Tyr	UAU,UAC	(UU)(CC)	No lipid secretion		37, 38
Cystic Fibrosis	Changes	mRNA bases	Amino Acid	Triplet Codon	QED Codon	Function	Gene	Ref.
OMIM 602421	Normal	ATC, TTT	Ile, Phe	AUU, AUC, AUA	(UC)CC	Cl- ion secretion	ABCC7	39, 40, 41
	Mutation	AT-/--T= ATT	Ile, Phe-loss	AU-/--U, ~~(UUU)~~	UUUU	No Cl- ion secretion		42, 43
Sickle Cell	Changes	mRNA bases	Amino Acid	Triplet Codon	QED Codon	Function	Gene	Ref.
OMIM 603903	Normal	GAG	Glu	GAG, GAA	(GA)AA	Circular Blood cell	S(HbS)	44
	Mutation	GTG	Val	GUN(U,C,A,G)	(GU)(GU)	Sickle Blood Cell	(HbF)	45,46

Table 2: Monogenic (Mendelian) rare diseases sample.

HI - More than sixty incurable Ichthyosis skin diseases are listed on the Foundation for Ichthyosis and Related Skin Type website [34], and its status [35]. HI is an autosomal recessive life-threatening severe skin disease. The normal functioning gene ABCA12 [36-38] secretes lipids from the Cell. When a mutation occurs, protein ABCA12 fails to secrete the lipid accumulating in the epidermis, causing the disease. Gene testing provides the mutation and corresponding changes in the amino acids, as shown in Table 2.

DupAA mutation causes Thr to be replaced by Arg, and another mutation, G, replaced by A, causes Cys to be replaced by Tyr. The corresponding degenerate triplet codons from the codon table and QED nondegenerate encoding codons are listed in Table. 2. The dysfunctional protein will require replacing the wrong amino acid with the correct one to cure the disease. However, the hurdle has been selecting the proper codon from the degenerate triplet codon. Since QED codons are nondegenerate, it is easy to select unambiguously the required codons as listed in the Table. 2.

Cystic fibrosis (CFTR)

CFTR is an autosomal recessive disease due to a mutation in [39,40] ABCC7 gene [41], which fails to secrete the Chlorine Ion (Cl^-) from the Cell, causing life-threatening accumulation of mucus in the epithelial Cell of the lung. CFTR is located on 7 Chromosomes at band q31.2 (7q31.2). Mutation at codon 507 ATC (lie) and 508 TTT(Phe), the corresponding amino acid are in (), causes the disease. At 507, C is deleted, and the first two TTs at 508 leads to ATT (lie). Thus, Phe is lost, and the defect is listed as Phe 508 del CFTR. The corresponding triplet and QED codons are listed in Table 2. Thus, selecting a QED codon over a triplet is easy without confusion.

No cure has been found for CFTR so far. However, the appearance of molecular targeted therapies [42,43] is a ray of hope for its cure.

Sickle cell

Sickle cell disease (CD) is an autosomal recessive blood disease due to a monogenic mutation in oxygen-carrying Hemoglobin protein [44]. Hemoglobin has four subglobin proteins. Two globin proteins at birth (HbF) produce normal blood and turn it off when adults. The other two-adult globin (HbS) produce sickle-shaped blood cells and the source of disease. A single pair mutation from G A G (glutamine) to G T G (Val) replaces Glutamine with Val amino acid, resulting in a dysfunctional protein causing the disease. The corresponding triplet and QED codons are listed in Table 2.

No cure exists. Applying molecular targeted approaches to suppress HbS in the presence of HbF [45,46] may provide a path for gene therapy.

Sickle cell and gene therapy

Third International Summit on Human Genome Editing: Expanding Capabilities, Participation, and Access held on March 6–8, 2023, at the Francis Crick Institute in London, the UK Royal Society and Academy of Medical Sciences reported the gene therapy of CD. Victoria Gray reported treatment of her CD at the conference. Around 2022, it was achieved by implanting edited stem cells by CRISPR-cas9 in her bone marrow to change the gene HbS to control adult hemoglobin production. So far, no side effect has been reported.

The cost of the first CD gene therapy was relatively high. Nobody quoted the price, but it is estimated to be around $4 to $6 million. The following and other gene therapies may cost around $2 million. How many can afford it? The challenge is how to reduce the cost. New technology and innovative system processing may be needed, similar to reducing the genome's cost of sequencing.

Cancer

Multigenic cancer has a five-year life extension goal. Once cancer is detected, regular treatment follows, followed by surgery, radiation, and chemotherapy. When cancer metastasizes, no successful cure exists. Chemotherapy attacks all the cells of the body. No biological technique exists to deliver chemotherapy to the cancerous Cell, as GPS is used to reach a destination. Why has NCI not developed such a technique to cure cancers?

Neurodegenerative Alzheimer's, Huntington's, ALS, and Parkinson's

The gene-disease causality in neurodegenerative diseases is not well understood, and no cure exists for these diseases. The diagnosis is mainly based on pathophysiology observations. The approach here is not a scientific review but rather a study of the role of the QED code in resolving the occurrence of tandem repeat (TR) in Huntington's and other diseases. Most TRs are noncoding QED codons, which could aid in identifying the TR-disease relationship.

Human tandem repeats (TR) are a significant cause of several neurodegenerative diseases [47-49] located in different regions of the gene: the 5'-UTR, exon, intron, and 3'-UTR. TR in the exon area may synthesize proteins, but the intron is a noncoding area. Then, what is the mechanism for TRs to cause a disease?

TR (CAG)n and (GCX)n occur in exon regions. In triplet coding, CAG encodes glutamine but causes HD and several other diseases. GCX encodes Alanine, which causes neurological, muscular, and other developmental diseases.

HD - Gln is encoded by CAA and CAG. In QED code, CAA only encodes Gln, and CAG is noncoding. CAA encoding yields polyglutamine but not by CAG. When TR <30, no HD is observed, but it is active when TR >36 or higher.

ALS - Ala is encoded by GCX (X: any T, C, A, and G). Under QED, GCX is noncoding.

The QED noncoding codons CGX, GCX, CXG, XCG, and CGXX are synonymous and include most TR. For example, (CGG)n in 5-'URT, (GGGGCCC)n in an intron, and (CTG)n in 3'-UTR are all noncoding under QED code.

Since only dysfunctional proteins cause the disease, the question of how noncoding TRs are causing it is a good puzzle for Molecular Biologists and geneticists to solve. One possibility is that TR occurs at the exon/intron or intron/exon interfaces, interfering with the exon area and yielding dysfunctional proteins and disease.

Method

QED coding model development

The idea of QED code model development originated while reviewing the verification of the triplet code. In 1968, Robert W. Holley, H. Gobind Khorana, and Marshall N. Nirenberg (1968 Medicine Nobel Prize) verified 61 encodings, one START, and two STOPs triplet codons [15]. Holley established the tRNA structure. Khorana and Nirenberg synthesized polyribonucleotide. Khorana reported (Khorana's Nobel Prize lecture, pages -345-346) that poly-rAU containing the two bases in strictly alternating sequence elicits no response from the ribosomes in the cell-free system and are chain terminator.

The detailed synthesis was reported [50], (Table 6: Ribopolynucleotides: Di- and Tri-nucleotide Sequences) and listed in Table 3.

Poly r-AU	No response from the Ribosomes- a chain terminator		
	(UAA)	Also, a chain terminator	
Poly r- CG	No synthesis, expecting a similar response-noncoding		

Poly r-GUA	val (GUA)	ser (AGU)	(UAG) chain-terminator
Poly r-GAU	asp (GAU)	met (AUG)	(UGA) chain-terminator
	Codon Table	met(AUG)	(AUG) chain-terminator

Table 3: Ribopolynucleotides with repeating di- and tri- nucleotide sequences (Ref. 50, Table 6).

Consider the two STOP codons in Table 3: U(AG) and U(GA). The G position is independent and symmetric.

The central idea of the QED model is to expand adjacent di-ribopolynucleotides (AU) noncoding to quadruplet-ribonucleotides noncoding. The proposed QED (quadruplet expanded DNA) eukaryote model is based on the following assumptions:

1. All four DNA (A, T, C, and G) bases are involved; in mRNA, T is replaced by U
2. The base positions are independent; i.e., for any A and B, AB and BA are equivalent.
3. The base positions are symmetric; i.e., for any A and B, (AB) and (BA) are synonymous
4. The self-complementarity forming adjacent base pairs with any two adjacent NN (N any A, T, C, or G) bases, (AT) NN and (CG) NN, is noncoding.

Following the assumption (3), (AT)(NN)) and (NN)(AT)) are synonymous; likewise, (CG)(NN) and (NN)(CG) are synonymous.

The independent QED codons are generated using the property of a square symmetric matrix. A N x N square symmetric matrix with N rows and N columns has N x (N+1)/2 independent elements.

Four DNA bases (T, A, C, and G) arranged in a 4 x 4 square symmetric matrix will yield 4 x (4+1)/2 = 10 independent elements: two AU and CG are part of noncoding and eight parts of encoding. Next, arranging in a 10 x10 square symmetric matrix will yield 10 x (10+1)/2 = 55 independent elements. Out of fifty-five, the independent noncoding and encoding elements are estimated following QED assumptions.

- Noncoding (AU)NN are AU(4X4)- 16
- Noncoding (CG)NN are CG(4X4)- 16
- Noncoding (AU)(AU)) – 1
- Noncoding (CG)(CG) – 1
- Noncoding (AU)(CG) – 1

These thirty-five are independent noncoding, and the remaining (55-35=20) twenty are encoding.

The independent twenty encodings and thirty-five noncoding codon bases are generated using four DNA (T, C, A, and G) bases [4,5] following the QED assumptions and listed in Table 4.

	QED (Quadruplet expanded DNA) eukaryote codons				Hydrogen Bond
	Codons	Synonymous Isocodons, T(U)			
1	UUUU	UUUU			8
2	CCCC	CCCC			12
3	AAAA	AAAA			8
4	GGGG	GGGG			12
5	(AA)(CC)	(CC)(AA)			10
6	(UC)CC	(CU)CC	CC(UC)	CC(CU)	11
7	(UG)UU	(GU)UU	UU(UG)	UU(GU)	9
8	(UG)GG	(GU)GG	GG(UG)	GG(GU)	11
9	(CA)CC	(AC)CC	CC(CA)	CC(AC)	11
10	(UU)(GG)	(GG)(UU)			10
11	(AC)(CA)	(AC)(AC)	(CA)(CA)	(CA)(AC)	10
12	(GA)(GA)	(GA)(AG)	(AG)(GA)	(AG)(AG)	10
13	(GU)(GU)	(GU)(UG)	(UG)(UG)	(UG)(GU)	10
14	(GA)(GG	GG(GA)	GG(AG)	(AG)GG	11
15	(CA)AA	(AC)AA	AA(CA)	AA(AC)	9
16	UU(UC)	UU(CU)	(UC)UU	(CU)UU	9
17	(AG)AA	AA(GA)	AA(AG)	(GA)AA	9
18	(AA)(GG)	(GG)(AA)			10
19	(CU)(CU)	(CU)(UC)	(UC)(UC)	(UC)(CU)	10
20	(UU)(CC)	(CC)(UU)			

Table 4: QED twenty independent encoding codons, its synonymous and HB.

Similarly, the thirty-five independent noncoding bases were also generated [4,5].

Table 5 has thirty-five independents noncoding (italics font) and synonymous codons. QED noncoding was also highly correlated with *Cis*-Regulatory elements. The cis-regulatory elements base and corresponding correlated noncoding QED base are listed in the Table 6.

Number *	Noncoding	Noncoding Synonymous			H.B.	Cis- correlation
1	(TA)(TA)	(TA)(AT)	(AT)(TA)	(AT)(AT)	8	TATA -Trans. Start
2	(CG)(CG)	(CG)(GC)	(GC)(CG)	(GC)(GC)	12	(CG)(CG)-Intron
3	(AU)GG	GG(AU)	GG(UA)	(UA)GG	10	(AU)GG- START
5	(UG)(AG)	(GU)(AG)	(UG)(GA)	(GU)(AG)	10	(UG)(AG)-STOP
8	(UA)(GA)	(AG)UA)	(UA)(AG)	(GA)(AU)	9	(UA)(GA)-STOP
10	(UA)AA	AA(UA)	(AU)AA	AA(AU)	8	(UA)AA- STOP

6	(UG)AA	AA(UG)	(GU)AA	AA(GU)	9	AA (UG) Promotor
7	(TA)(GT)	(GT)(TA)	(TA)(TG)	(GT)(AT)	9	(TA)(TG) Promotor
11	(TA)(AC)	(AC)(TA)	(TA)(CA)	(AC)(AT)	9	CAAT Box, Promotor
15	TT(AC)	(AC)TT	(CA)TT	TT(CA)	9	TT(AC), Promotor
16	TT(AG)	(GA)TT	(AG)TT	TT(GA)	9	TT(AG), Promotor
22	AA(CT)	(CT)AA	(TC)AA	AA(TC)	9	AA(TC), Promotor
30	(CT)(TA)	(TC)(TA)	(CT)(AT)	(TC)(AT)	9	(TA)(CT), Promotor
12	(TT)(AA)	(AA)(TT)			8	TT AA, Promotor
14	TT(TA)	(TA)TT	(AT)TT	TT(AT)	8	TT(TA), Promotor
4	(UG)(AC)	(AC)(UG)	(UG)(CA)	(AC)(GU)	10	(UG)(CA), Promotor
9	(UA)(GC)	(UA)(CG)	(CG)(UA)	(CG)(AU)	10	(GC)(AU), Promotor
17	TT(CG)	(CG)TT	TT(GC)	(GC)TT	10	TT(CG), Promotor
18	CC(TA)	(TA)CC	(AT)CC	CC(AT)	10	(CC)(AT), Promotor
23	AA(CG)	(GC)AA	(CG)AA	AA(GC)	10	AA(CG), Promotor
28	(AC)(AG)	(AC)(GA)	(CA)(GA)	(CA)(AG)	10	(AC)(AG),Promotor
32	(CT)(AC)	(TC)(AC)	(CT)(CA)	(TC)(CA)	10	(TC)(AC), Promotor
33	(CT)(AG)	(TC)(AG)	(CT)(GA)	(TC)(GA)	10	(CT)(AG), Promotor
34	(CT)(TG)	(TC)(TG)	(CT)(GT)	(TC)(GT)	10	(CT)(AG),Promotor
19	CC(TG)	(TG)CC	(GT)CC	CC(GT)	11	CC(TG), Promotor
20	CC(AG)	(AG)CC	(GA)CC	CC(GA)	11	CC(AG), Promotor
24	GG(CT)	(CT)GG	(TC)GG	GG(TC)	11	GG(CT), Promotor
26	GG(AC)	(AC)GG	(CA)GG	GG(CA)	11	GG(AC), Promotor
27	(AC)(CG)	(CA)(CG)	(CA)(GC)	(AC)(GC)	11	(AC)(CG), Promotor
29	(AG)(CG)	(GA)(CG)	(AG)(GC)	(GA)(GC)	11	(AG)(CG), Promotor
31	(CT)(CG)	(TC)(CG)	(CT)(GC)	(TC)(GC)	11	(CG)(TC), Promotor
35	(GT)(CG)	(TG)(CG)	(GT)(GC)	(TG)(GC)	11	(CG)(TG), Promotor
13	(CC)(GG)	(GG)(CC)			12	(CC)(GG), Promotor
21	CC(CG)	(CG)CC	(GC)CC	CC(GC)	12	CC(CG), Promotor
25	GG(CG)	(CG)GG	(GC)GG	GG(GC)	12	GG(CG), Promotor

Table 5: The sequence number were rearranged while predicting their assignments.

Cis-regulatory elements and eukaryote cell

Cis-regulatory	Noncoding QED code	Table.... row #
TATA Box	(TA)(TA)	1
CAAT Box	(CA)(TA)	11
CG/GC	(CG)(CG)	2
YCAY	(TC)(AT)	30
(Y-T (U)Or C)	CC(AT)	18
	(TC)(AC)	32
UAGG	(UA)GG	3
UGCAUG	(GC)(AU)	9
UGCAUG	(UG)(CA)	4
AT-Rich	AT-Rich	7,14
GC-Rich	CG or GC- Rich	17,21,23,25
		27,29,31,35

Table 6: The correlated *cis*-regulatory elements and noncoding QED codons.

QED encoding assignment prediction – Table 7

The triplet and prosed eukaryote genetic theoretical QED code model is required to encode protein. Triplet applies mainly to prokaryotes because DNA is already in the cytoplasm. However, in eukaryotes, DNA is in the nucleus and requires transcription and splicing to get mRNA in the cytoplasm for protein synthesis, and triplet lacks such control. The differences in the protein synthesis process for Prokaryotes, Viruses, and Eukaryotes are shown in Figure 1, 2, and 4.

The QED's theoretical code was proposed in 2053 by extending laureate Khorana's observation, which needs verification. Triplet coding was proposed in 1963, and it took five years to verify in 1968. Likewise, the QED code verification may take as long as triplet coding or higher. Since QED-like triplet is anticipated to encode a protein, the QED code assignment prediction and to expedite its verification was made by following steps.

- First, QED coding assumptions were applied to the triplet encoding table. Under the fourth assumption, (AU)NN and (CG)NN, N being any T, C, A, and G, are noncoding. Consequently, any triplet (AU)N and (CG)N codons are noncoding and were crossed out from the triplet encoding table. Thus, only fifteen encoding triplet codons met the criteria [4] (Table 3).

- Second, Nirenberg showed [51,52] that poly-U, poly-A, and poly-C encode the amino acids Phe, Lys, and Pro, respectively, establishing a direct link among mRNAs, tRNAs, amino acids, codons, and anticodons in protein synthesis at ribosomes. Also, he showed that oligo chain lengths of 3 and 4: $(oU)_3$ and $(oU)_4$ showed nearly the same activities. Therefore, it is reasonable to predict that if triplet UUU can encode Phe, quadruplet UUUU could also encode Phe. Therefore, four QED code assignment projections were made: UUUU- Phe, AAAA-Lys, CCCC-Proo, and GGGG -Gly.

- Third, the Matching procedure was used for the twelve allowed encoding QED codons. Since two triplet bases could encode only sixteen amino acids, a third base was added, making it degenerate and allowing a dangling bond at the third base. Thus, the first two triplet bases were compared with the two QED bases for the matching procedure. The corresponding QED code assignment prediction was made if a match occurred, encoding the corresponding amino acid. The result is in Table 4(a) and (4b) [4]. An example of the assignment prediction for Arg was: Arg/R–AGA, AGG: If G is added to AGA and A is added to AGGA, then under QED assumptions 2 and 3, (AG)(GA) will represent both. Thus, in Table 4(a), QED (AG)(GA)-Arg/R assignment is predicted. Similar procedures were followed for the other eleven protein encoding codons.

- Fourth, the remaining five: Ala (GCN), Asp (GA/U, C), lle (AUN), Met (AUG), and Tyr (UA/U, C) have (GCN) and (AUN), where N is any T(U), C, A and G that correspond to noncoding QED code features and were crossed out. For these, additional conditions were imposed to predict the assignment: Maintain the Hydrogen Bond as needed for codon-anticodon pairing and not require wobble hypothesis; replacing (AU) by (AA) or (UU) will preserve quadruplet H.B; (CG) replacement by (CC) or (GG) will preserve sextuplet H.B. Extending the above Third Method's procedure, the assignment prediction was made.

Ala - the triplet code GCN, N being U, C, A, or G encodes Ala. Under QED, adjacent GCs are not allowed. Since C has a triple bond, replacing C with G, as GGN, will have the identical HB. Now replace GGN with GGA. Under QED, C, and U are not allowed, but G and A are allowed at the fourth position, making GGAG acceptable. Thus, (GG)(GA) will encode Ala.

Asp - the triplet GA(U/C) encodes Asp. This could be GA(UC), but U and C are not allowed in QED. However, A replacing U and G replacing C will maintain the Hydrogen Bonds. Thus, GA(AG) or synonymous (GG)(AA) meets the requirement and is assigned for Asp.

Tyr- the triplet UA(UC) encodes Tyr. Under QED, A and G are not allowed. A combination of (UU)(CC) meets the requirement encoding Tyr.

Ile - the triplet AUH (H being U or C or A) encodes Ile. Under QED, adjacent AUs are not allowed, but UUs or AAs are okay. Thus, UC (U or C) or (UC)(CC) will satisfy encoding Ile.

Amino Acids	QED Codons	HB Bonds	QED Codons	Amino Acids
Arg	(GA)(GA)	10	(CU)(CU)	Leu
Asn	(AA)(CC)	10	(UU)(GG)	*Met
Cys	(UG)UU	9	(CA)AA	Gln
Glu	(GA)AA	9	(CU)UU	Ser
Gly	GGGG	12	CCCC	Pro
His	(CA)CC	11	(UG)GG	Trp
Lys	AAAA	8	UUUU	Phe
Thr	(AC)(CA)	10	(GU)(GU)	Val
Tyr	(UU)(CC)	10	(GG)(AA)	Asp
lle	(UC)CC	11	(GA)GG	Ala

Table 7: QED (Quadruplet expanded DNA) EUKARYOTE ENCODING.

The QED encoding table has some unique and interesting properties. The encoding codon of one is the anticodon of the other: (GA)(GA) encodes Arg, and its anticodon (CU)(CU) encodes Leu. Similar attributes are noted for the other. Also, the anticodon of tRNA $^{(CU)(CU)}$ Leu will decode the (GA)(GA) Arg codon, maintaining proper HB bonds and assuring a secure protein synthesis. Consequently, efficient use of tRNA is anticipated.

Summary

The Quadruplet Expanded DNA (QED) eukaryote genetic code is the first and unique protein-encoding and noncoding controlling codons and has established a strong correlation between noncoding QED codons and Cis-regulatory elements, providing a unified encoding framework for eukaryotes, prokaryotes, and viruses. Expanding beyond the Central Dogma of Biology, which primarily applies to prokaryotic protein synthesis, QED coding supports a broader Principle of Information Flow and Biological Protein Synthesis across all life forms. QED coding offers a transformative approach to treating diseases by enabling the correction of dysfunctional proteins. This breakthrough is poised to drive a paradigm shift in biomedical research, paving the way for new therapies targeting monogenic rare diseases, multigenic cancers, and neurodegenerative disorders.

Data Availability

N/A.

Code Availability0

N/A.

Acknowledgements

I am thankful to Dr**.** Nawin Mishra, Distinguished Professor Emeritus of Genetics and Genomics, University of South Carolina, Columbia, SC 29208, for his encouragement and valuable comments regarding this work. My daughter, Usha Singh's rare, incurable disease was the catalyst to start the lifelong journey to find a gene-disease causality relationship. The outcome is the QED genetic coding for eukaryotic cells. This work would never have materialized without the unwavering full support of my family: my wife, Amrita Singh, and my children, Asha Singh and Om Prakash Singh (a.k.a. Tom Singh).

Author Contribution

Rama Shankar Singh - 100%.

Competing Interest

No competing interest.

Additional Information

Correspondence and requests for materials should be addressed to Rama Shankar Singh, email: gen-med @comcast.net.

Bibliography

1. FH C Crick. "On protein synthesis". *Symposia of the Society for Experimental Biology* 12 (1958): 138-163.

2. FH Crick. "Central Dogma of Molecular Biology". *Nature* 227 (1970): 561-563.

3. FH Crick. "On the genetic code". *Science* 139 (1963): 461-464.

4. Rama Shankar Singh. "Quadruplet Expanded DNA (QED) Genetic Code for Eukaryotic Cells". *Acta Scientific Medical Sciences* 7.12 (2023): 70-82.

5. Rama Shankar Singh. "Correlation between Eukaryotic Noncoding QED Genetic Codes and Cis-Regulatory Elements". *Acta Medical Sciences* 8.10 (2024): 89-96.

6. Claude E Shannon. "A Mathematical Theory of Communication". *The System Technical Journal* 27 (1948): 379-423, 623-656.

7. W H Brattain and J Bardeen. "Nature of the Forward current in germanium point contacts". *Physical Review* 74 (1948): 231.

8. J Bardeen. "Surface states and Rectification at a Metal Semi-Conductor contact". *Physical Review* 71 (1947): 717.

9. 1956 Physics Nobel Prize awarded jointly to William Bradford Shockley, John Bardeen and Walter Houser Brattain "for their researches on semiconductors and their discovery of the transistor effect".

10. JD Watson and FHC Crick. "Molecular structure of Nucleic Acids". *Nature* 171 (1953): 737-738.

11. JD Watson and FHC Crick. "The structure of DNA". *Cold Spring Harbor Symposia on Quantitative Biology* 18 (1953): 123-131.

12. 1962 Nobel Prize in Medicine and Physiology awarded to F. H. C. Crick, J. D. Watson and M. H. F. Wilkins "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material".

13. 2009 Chemistry Nobel Prize awarded to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath "for studies of the structure and function of the ribosome".

14. F H Crick. "Codon--anticodon pairing: the wobble hypothesis". *Journal of Molecular Biology* 19 (1966): 548-555.

15. 1968 The Nobel Prize in Physiology or Medicine was awarded jointly to Robert W. Holley, Har Gobind Khorana and Marshall W. Nirenberg "for their interpretation of the genetic code and its function in protein synthesis".

16. H M Temin and S Mizutani. "RNA-dependent DNA polymerase in virions of Rous sarcoma virus". *Nature* 226.5252 (1970): 1211-1213.

17. Jacob F and Monod J. "Genetic regulatory mechanisms in the synthesis of proteins". *Journal of Molecular Biology* 3 (1961): 318-356.

18. 1965 Nobel Prize in Physiology and Medicine awarded to François Jacob, André Lwoff and Jacques Monod "for their discoveries concerning genetic control of enzyme and virus synthesis".

19. M Lewis. "The lac repressor". *C. R. Biology* 328 (2005): 521-548.

20. Tom Maniatis. "From bacterial operons to gene therapy- 50th anniversary of cell". 187 (2024): 617-6421.

21. SM Berget., *et al.* "Spliced segments at the 5' terminus of adenovirus 2 late mRNA". *PNAS* 74 (1977): 3171-3175.

22. 1993 The Nobel Prize in physiology or medicine awarded to R.J. Roberts and P. A. Sharp "For their discoveries of split genes".

23. 2006 The Nobel Prize in Chemistry awarded to Roger D. Kornberg "for his studies of the molecular basis of eukaryotic transcription".

24. Carl G de Boer and Jussi Taipale. "Hold out the genome- a roadmap to solving the cis-regulatory code". *Nature* 625 (2024): 41-50.

25. Lingna Xu and Yuwen Liu. "Identification, Design, and Application of noncoding Cis-regulatory elements". *Biomolecules* 14 (2024): 945-966.

26. Poliseno Laura., *et al.* "Coding, or non-coding, that is the question". *Cell Research* 34 (2024): 609-629.

27. Ling-Ling Chen and V Narry Kim. "Small and long non-coding RNA_ Past, present and future". *Cell* 187 (2024): 451-6485.

28. The Nobel Prize in Physiology or Medicine 2009 to Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak. "how chromosomes are protected by telomeres and the enzyme telomerase".

29. Barbara McClintock. "THE ORIGIN AND BEHAVIOR OF MUTABLE LOCI IN MAIZE". *PNAS* 36 (1950): 345-356.

30. Barbara McClintock. The Nobel Prize in Physiology or Medicine 1983. "for her discovery of mobile genetic elements".

31. 2023 Nobel Prize in Medicine and Physiology awarded to Katalin Karikó and Drew Weissman "for their discoveries concerning nucleoside base modifications that enabled the development of effective mRNA vaccines against COVID-19".

32. Baden L R., *et al.* "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine". *The New England Journal of Medicine* 384 (2021): 403-416.

33. 2017 The Nobel Prize in Physiology or Medicine was awarded jointly to Jeffrey C. Hall, Michael Rosbash and Michael W. Young "for their discoveries of molecular mechanisms controlling the circadian rhythm".

34. FIRST - Foundation for Ichthyosis & Related Skin Types, Inc.

35. Joosten MD W., *et al.* "New development in the molecular treatment of Ichthyosis_ Reve. of literature". *Orphanet Journal of Rare Diseases* (2022): 17-269

36. Kelsell David P., *et al.* "Mutations in ABCA12 Underlie the Severe Congenital Skin Disease HI". *The American Journal of Human Genetics* 76 (2005): 794-803.

37. Anna C Thomas., *et al.* "ABCA12 Is the Major Harlequin Ichthyosis Gene". *Journal of Investigative Dermatology* 126 (2006): 2408-2413.

38. Akiyama Masashi., *et al.* "Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer". *Journal of Clinical Investigation* 115 (2005): 1777-1784.

39. Jean L Marx. "The Cystic Fibrosis Gene Is Found, News and Comment". *Science* 245 (1989): 923-925.

40. Hunt John F., *et al.* "Cystic Fibrosis Transmembrane Conductance Regulator (ABCC7) Structure". *Cold Spring Harbor Perspectives in Medicine* 3 (2013): a009514.

41. Ratjen Felix., *et al.* "Cystic fibrosis". *Nature Reviews Disease Primers* 1 (2015): 15010.

42. Francis S Collins. "Realizing the Dream of Molecularly Targeted Therapies for Cystic Fibrosis". *NEJM* 381 (2019): 1863.

43. Hartmut Grasemann and Felix Ratjen. "Cystic Fibrosis". *NEJM* 389 (2023): 1693.

44. Inusa Baba PD., *et al.* "Sickle Cell Disease—Genetics, Pathophysiology, Clinical Presentation and Treatment". *International Journal of Neonatal Screen* 5 (2019): 20.

45. George Q Daley. "Welcoming the Era of Gene Editing in Medicine". *The New England Journal of Medicine* 1-4 (2024).

46. Frangoul H., *et al.* "CRISPR-Cas9 Gene Editing for Sickle Cell disease and beta-Thalassemia". *NEJM* 384 (2021): 252-260.

47. Anthony J Hannan. "Tandem repeats mediating genetic plasticity in health and disease". *Nature Reviews Genetics* 19 (2018): 286-298.

48. Kume., *et al.* "CGG repeat expansion in LRP12 in amyotrophic lateral sclerosis". *AJHG* 110 (2023): 1086-1097.

49. Troast Brett., *et al.* "Genome-wide detection of tandem DNA repeats that are expanded in autism". *Nature* 586 (2020): 80-103.

50. AR Morgan., *et al.* "Studies on polynucleotides, LIX. Further codon assignments from amino Acid incorporations directed by ribopolynucleotides containing repeating trinucleotide sequences". *Proceedings of the National Academy of Sciences of the United States of America* 56 (1966): 1899-1906.

51. Nirenberg M and Leder P. "RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes". *Science* 145 (1964): 1399-1407.

52. Jones OW and Nirenberg M W. "Qualitative survey of RNA codewords". *Proceedings of the National Academy of Sciences of the United States of America* 48 (1962): 2115-2123.

53. Peabody D S. "Translation Initiation at Non-AUG Triplets in Mammalian Cells". *Journal of Biological Chemistry* 264 (1989): 5031-5035.