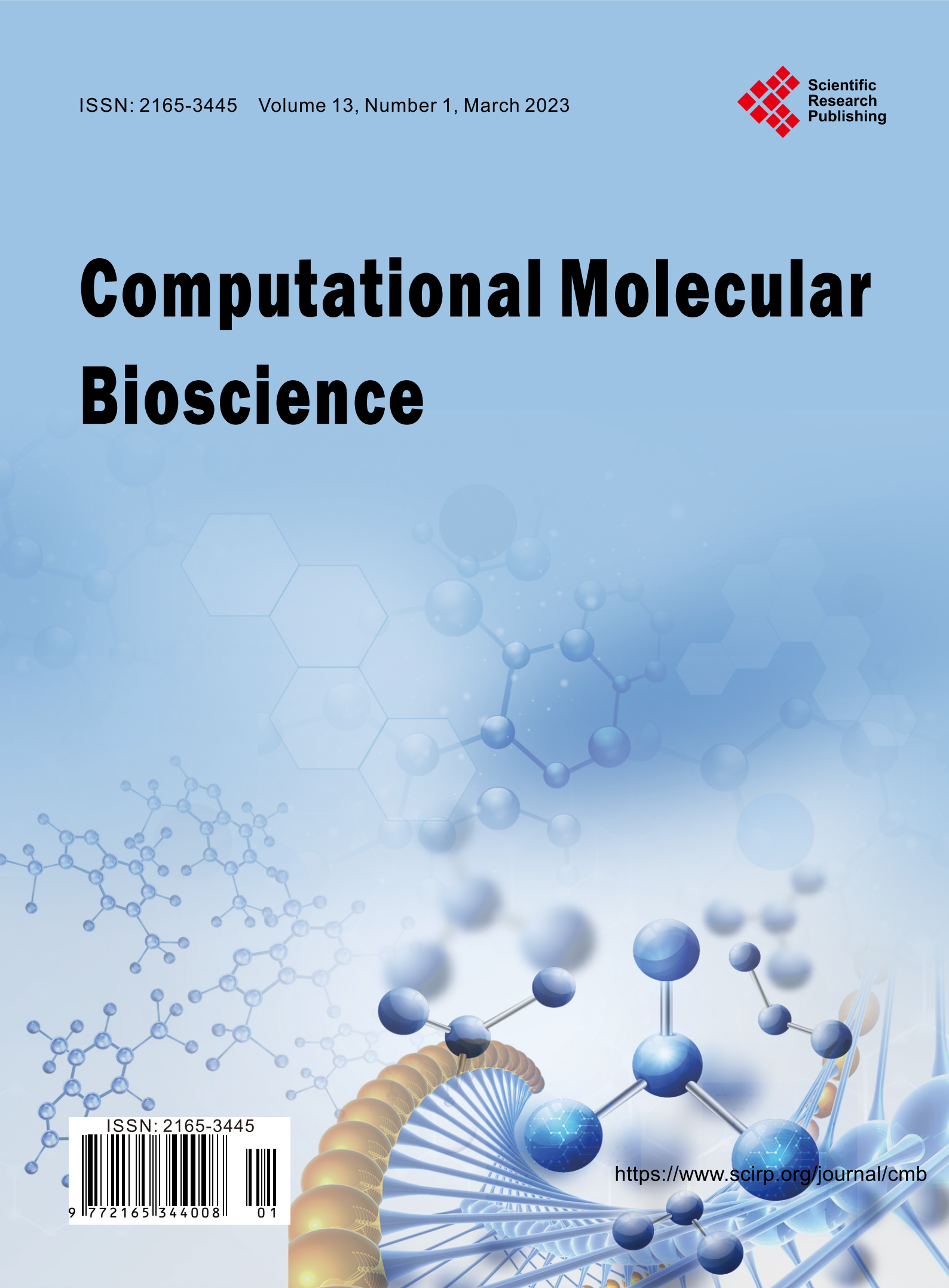
ISSN: 2165-3445 Volume 13, Number 1, March 2023

Scientific Research Publishing

Computational Molecular Bioscience



ISSN: 2165-3445

https://www.scirp.org/journal/cmb

Journal Editorial Board

ISSN: 2165-3445 (Print) ISSN: 2165-3453 (Online)

https://www.scirp.org/journal/cmb

Editorial Board

Computational Molecular Bioscience, 2023, 13, 1-20
https://www.scirp.org/journal/cmb
ISSN Online: 2165-3453
ISSN Print: 2165-3445

Table of Contents

Volume 13 Number 1 March 2023

Computational Molecular Bioscience (CMB)
Journal Information

SUBSCRIPTIONS

The *Computational Molecular Bioscience* (Online at Scientific Research Publishing, https://www.scirp.org/) is published quarterly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: cmb@scirp.org

Scientific Research Publishing

Computational Molecular Bioscience, 2023, 13, 1-20
https://www.scirp.org/journal/cmb
ISSN Online: 2165-3453
ISSN Print: 2165-3445

A Personalized Digital Code from Unique Genome Fingerprinting Pattern for Use in Identification and Application on Blockchain

Isaac Kise Lee[1,2]

[1]Eone Diagnomics Genome Center (EDGC), Incheon, South Korea
[2]NGENI Foundation, San Diego, CA, USA
Email: iklee@ngeni.org

How to cite this paper: Lee, I.K. (2023) A Personalized Digital Code from Unique Genome Fingerprinting Pattern for Use in Identification and Application on Blockchain. *Computational Molecular Bioscience*, **13**, 1-20.
https://doi.org/10.4236/cmb.2023.131001

Received: February 11, 2023
Accepted: March 28, 2023
Published: March 31, 2023

Abstract

With over 10 million points of genetic variation from person to person, every individual's genome is unique and provides a highly reliable form of identification. This is because the genetic code is specific to each individual and does not change over time. Genetic information has been used to identify individuals in a variety of contexts, such as criminal investigations, paternity tests, and medical research. In this study, each individual's genetic makeup has been formatted to create a secure, unique code that incorporates various elements, such as species, gender, and the genetic identification code itself. The combinations of markers required for this code have been derived from common single nucleotide polymorphisms (SNPs), points of variation found in the human genome. The final output is in the form of a 24 numerical code with each number having three possible combinations. The custom code can then be utilized to create various modes of identification on the decentralized blockchain network as well as personalized services and products that offer users a novel way to uniquely identify themselves in ways that were not possible before.

Keywords

Genomic Fingerprint, Digital Code, SNP's, Auxiliary Code, Marker Selection, Blockchain, Web3.0, Decentralized Identification (DID)

1. Introduction

There are approximately 8 billion people in the world and a large majority own computers, smartphones, and other internet-connected devices [1] [2]. According to a US survey in 2020, 87 percent of individuals have access to a computer

in their households, and about 5 billion active internet users in the world today [3]. This number is expected to grow exponentially with the recent revolution of IoT, Web3.0 and Metaverse applications [4]. Traditional forms of identity in use today have limited security and are often fragmented and inconvenient. A traditional username and password are commonly used to identify individuals, but password theft and phishing attacks raise serious concerns for many, and forgetting one's login information can be a significant inconvenience [5] [6]. Popular biometric identification methods such as facial recognition are increasingly common and at times can be more convenient than using a traditional username, but proper functionality involves ideal conditions, and using facial data for security purpose may raise concerns about a government or corporate overreach [7]. Blockchain technology-enabled identification methods could potentially allow for more secure management and storage of digital identities by providing unified, interoperable, and tamper-proof infrastructure with key benefits to enterprises, users, and IoT management systems [8] [9]. Having a unique, personalized identification code allows the individual to safely identify themselves without having to reveal other personal information to other parties with the added benefit of the code being truly unique to that individual.

Individual identity and personalization are integral to a functioning society and economy [10]. Having a safe, secure, and robust way to identify and personalize ourselves and our possessions is becoming increasingly crucial in today's digital society and global markets [11] [12]. At its most basic level, identity is a collection of claims about a person, place, or thing. For people, this usually consists of first and last name, date of birth, nationality, and some form of national identifier such as one's passport number, social security number (SSN), driving license, etc. [13] [14]. These data points are issued by centralized entities (governments) and are stored in centralized databases (central government servers) [15]. A digital identity arises organically from the use of personal information on the web and from shadow data created by the individual's actions online [16]. More robust identity and personalization management systems could be used to eradicate current identity issues such as inaccessibility, data insecurity, and fraudulent identities [8] [17] [18]. Security and identity are complex and ever-evolving issues for enterprise and government systems alike. Blockchain-based solutions could provide exceptional utility in solving issues common in current identity and digital systems [19] [20]. Blockchain technology allows users to create and manage digital identities through the combination of decentralized identifiers, identity management and embedded encryption. However, current blockchain-based identification methods rely on randomly generated digital code that is not unique to the individual [21] [22]. The genome fingerprint-based personalized digital code is unique to each individual and could potentially be used to verify ownership of identity even in the case when a private key or passcode is not available for the associated data.

The human genome holds roughly 3 billion base pairs [23] [24]. Each person's

genome sequence holds a unique combination formulated from random recombination from the parent's genome during fertilization [25] [26]. Half of the individual genome is inherited from the father and the other half from the mother. The genetic makeup is composed of four nucleotides (ACGT) and shares 99% similarities among all people [27]. Most genetic variations are in the form of single nucleotide polymorphisms (SNPs), which have been extensively studied by many acclaimed scientists after the completion of the Human Genome Project (HGP) in 2003 [28] [29]. SNPs are commonly obtained through several methods such as genome sequencing [30], genotyping arrays [31], or polymerase chain reaction (PCR) methods [2]. When using such methods, it is crucial to verify and validate SNP sites because validation success can vary based on SNPs chosen and different ethnic populations [32]. Most of these specific variations are rather rare and show an individual and population-specific occurrence, but some fraction of these variations arise very frequently in every population studied [33] [34]. These high allele frequencies (close to 50%) pan-ethnic common SNPs are considered to be natural results from evolution history and are not associated with any known phenotype or disease [35]. Most of these variations are bi-allelic, which means only two choices of nucleotides are observed at a given specific site [36] [37]. The patterns of these high allele frequencies bi-allelic SNP variations can be used to generate a personalized genetic code that can potentially identify and track individuals for use in forensics, such as the CODES method which uses fragment length polymorphisms [38]. SNPs have also been used to track animals such as U.S. beef cattle when suppliers require cattle identification for paternity analysis and breeding information [39]. We employ these bi-allelic, high frequency, pan-ethnic common SNP variations in the genome to prepare and create a unique genomic pattern akin to fingerprint authentication (Genome Fingerprinting). This information can provide accurate and powerful means to identify and distinguish individual people in a digital format by simply decoding the digital code without revealing other identifiable personal information. We employ a panel of selective, universal, high allele frequency bi-allele human SNP markers to efficiently generate a unique combination code that can be used in digital identification and personalization.

2. Methods

The proposed digital identification method comprises of steps that involve selecting genetic markers for identification, assembling information for the unique genetic identification, assigning digital codes for each of the features, creating the digital identification codes, decoding the digital identification code, registering the codes for blockchain, and implementing codes for various applications.

2.1. Digital Identification Format

In order to formulate a digital identification which holds all the necessary information from a wide selection of species of animals and plants with the possi-

bility of a rare event (identical twins or cloning), we propose the following digital identification format:

Species Code (2 - 4 digits), Gender Code (1 digit with Natural or Virtual birth separation) (1), Identification Codes (24 digits for numerical, and 12 digits for alphabetical), Auxiliary Code (1 or 2 digits).

2.2. Species Codes

Every animal, plant, bacteria, and virus on earth has genetic makeup (DNA and RNA) vital for growth, reproduction, and multiplication. Although the current study is primarily focused on using genetic code in identifying human individuals, the same concept of genetic based blockchain identification can be used in other species for identification, parental linkage, tracking, or management analysis in agricultural businesses. Therefore, species identification code should be included in the digital format. Below is an example of the proposed codes for some of the well-known species of interest (Table 1). The two-digit combination of the alphabetical code can cover up to 576 (24 × 24) different species in the code combination. Three (13,824 species) or four (331,776 species) digit species identification code combination may be implemented in future expansion.

2.3. Gender Codes

With respect to biological sex, humans are born as either male or female. However, gender classifications are simply not binary so other types (gender neutral, transgender, genetic abnormalities) should be considered in the gender code assignment to cover a wide spectrum of people. Furthermore, we are also exploring the possibility of a non-natural birth (Artificial or Virtual) individual such as cloning or digital creation in some applications, that could be differentiated from real (Natural) individuals to that of artificial or virtual individuals (Table 2).

Table 1. Examples of species code.

Species Codes	Scientific Name	Codes
Human	Homo sapiens	HS
Dog	Canis lupus	CL
Cat	Felis tus	FT
Babine	Bos taurus	BT
Swine	Sus scrofa	SS
Horse	Equus caballus	EC
Sheep	Ovis aries	OA
Tuna	Thunnus albacares	TA
Salmon	Salmo salar	SS
Alien	Xenomorph x	XX

Table 2. Gender code with possible outcomes for both natural and virtual birth.

Gender Codes	Male	Female	Other
Natural	1	2	3
Artificial	5	6	7

2.4. 24-Digit Numerical Codes

Because of SNP's mostly bi-allelic nature, each position of the markers has three different possible combinations as summarized in the table below. We can assign a code "1" for the known reference allele, "2" for the alternative allele, and "3" for the mixture of two alleles called heterozygotes. However, there is a possibility of an incomplete or failed sequencing result at a particular site that can stem from either low confidence call or missing call for a certain SNP [40] [41]. The missing calling makes it difficult to assign a piece of correct allele information on the given sites. Therefore, the missing genotypes are labeled "NA" in the sequencing output file and assigned a "0" in the code output. Typically, the standard whole genome sequencing coves roughly 95% of the entire genome region when it is sequenced in 30X read depth (standard Whole Genome Sequencing QC criteria), meaning that there is always a chance that one or more SNPs may not produce a satisfactory result to assign an appropriate code [42] [43]. In order to establish a criteria cutline and a measure of quality control, the minimum data point needed to assign a reliable identity code will be designated to 21 data points. Based on the 24 SNPs needed for human identification code generation used, the encoding system can accommodate and produce a valid result even when up to three SNP data points are missing (Table 3).

2.5. 12-Digit Alphabetical Codes

In order to accommodate the limitation of the numerical code while shortening it simultaneously, we converted the 24-digit numerical code into a 12-digit alphabetical code by combining two numerical codes into one alphabetical code (Table 4).

The numerical code has only four (0 - 3) options in each digit sites, but alphabetical codes have 16 different options in each digit sites and gives more options for downstream applications.

The 12-digit alphabetical codes can be easily converted back to a 24-digit numeric codes and be able to decode original SNP sequence of each individual people.

2.6. Auxiliary Codes

One of the drawbacks for using genetic-based identification codes is its inability to separate and distinguish individuals with identical genetic makeup. Such criteria are found in identical twins, triplets, quadruplets, and human clones [44]. Therefore, an auxiliary identification code is needed when two individuals with the same genetic makeup are to be identified. The auxiliary code can also be used

Table 3. Possible outcomes of each genetic markers' digital code including three possible combinations plus no result calls.

SNP Compositions	Allele1/Allele1	Allele2/Allele2	Allele1/Allele2	No Call
A/C	AA	CC	AC or CA	NA
A/T	AA	TT	AT or TA	NA
A/G	AA	GG	AG or GA	NA
C/T	CC	TT	CT or TC	NA
C/G	CC	GG	CG or GC	NA
T/G	TT	GG	TG or GT	NA
CODE	**1**	**2**	**3**	**0**

Table 4. Conversion tables for two numeric codes to one alphabetical code. The alphabet order is based on occurrence rate of each combination with A being the most common and P being the least common in the general population.

Numeric Code Combination	Alphabetical Code Conversion	Occurrence	
33	A	6	Most Common
32	B	5	Most Common
31	C	5	Most Common
23	D	5	Most Common
13	E	5	Most Common
22	F	4	Common
21	G	4	Common
12	H	4	Common
11	I	4	Common
30	J	3	Less Common
03	K	3	Less Common
20	L	2	Less Common
10	M	2	Least Common
02	N	2	Least Common
01	O	2	Least Common
00	P	1	Unlikely

to separate real (natural birth) individuals from artificial (virtual birth) individuals that are created in virtual environments through games or other computer algorithms. The auxiliary code can be numeric or alphabetic, both of which can be either single digit or double-digit, depending on the potential number of genetically identical individuals created in a natural (identical twins) or artificial (cloning) environment. For example, we are assigning the number "1" for single born real individuals and "2" for twins with identical genetic makeup (Table 5).

Table 5. Auxiliary code for genetically identical individuals such as twins and clones.

Codes	Single borne	Identical Twin	Identical Triple	Identical Quadruple
Born	1	2	3	4
Clone	Alphabetical Combination			

Note that non-identical twins have different genetic makeup and therefore will produce different codes. The auxiliary code is also designed to take into account the very rare event (one out of 280 billion cases when 24 SNPs are being used in the panel) when two unrelated individuals have the same sequencing code by accident.

2.7. SNP Identity Marker Selection

It is well established that about 0.1% of the genome accounts for all the genetic variations that are found from person to person [45] [46]. This means that there are potentially around three million base pair differences in a particular genome between two non-related individuals [47]. All of these variations in the genome can be used as a fingerprint for patterns when differentiating from other individuals. Therefore, each individual, except identical twins, has a unique genomic fingerprint that solely belongs to that person [48]. Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation, and they are easy to detect with most genotyping and sequencing technology platforms [49] [50]. Extensive research in the human genome project has led to more than 20 million variations being studied in different ethnicity and nationality backgrounds [51] [52]. Most of these variations are quite rare and show individual and population-specific distribution patterns, but some of these fractions arise very frequently in every population studied [53] [54]. The majority of these variations are bi-allelic, which means only two choices of nucleotides are observed within a given specific SNP site. The patterns and fingerprints of these high-frequency bi-allelic SNPs have been used in other applications, most notably to identify individuals for forensic purposes [55] [56].

In order to select the most informative markers for the identification and personalization purposes, SNPs that show a high minor allele frequency (>0.45) from all population studies were selected from HapMap and 1000 genome sequencing database [57] [58]. For example, there are over 4 million SNPs that have been genotyped in the CEU cohort within HapMap, and 218,000 SNPs showed minor allelic frequencies of greater than 45% [59]. The higher the minor allelic frequencies are, the more informative the result will be based on marker combinations present in the genomes [60]. Therefore, fewer markers are required when higher frequency SNP markers are used for testing, analysis, and interpretation for the identification and personalization purpose [61].

2.8. Marker Selection Workflow

Markers are selected from published HapMap and 1000 Genome Sequencing

database using the criteria below:

- Identify marker sets with high minor allele frequency (e.g., >45%) within the study population
- Select only bi-allelic SNP markers with good Hardy-Weinberg distribution
- Avoid markers with adjacent known polymorphisms to minimize potential allele drop in sequencing
- Find markers that are well separated from each other
- Avoid markers in a region containing duplicated sequence motifs

Number of SNP markers for digital identification code generation

Because each bi-allele SNPs has three different allele combinations (Reference, Alternative, Heterozygote), each SNP analysis could generate three distinct identification codes [62]. The possible combination of two unrelated SNP markers could potentially generate 9 (3 × 3) different codes, and three distinct SNP markers would result in 27 (3 × 3 × 3) different combination codes. When considering a world population of 8.1 billion people in recent world population survey, a 21 SNPs composition (Over 10 billion possible combination codes) could potentially differentiate all the people in the world (Table 6). Therefore, we propose to use 24 SNP combinations for human identification to create sufficient buffer and accommodate for future population growth, as well as virtual creation of artificial individuals.

2.9. Candidate SNP Marker Selection for Digital Identification

The table below showcases the proposed 100 markers to be used, all of which have an average 50% minor allele frequency ranging from 0.498 to 0.508 in all populations tested by the 1000 genome sequencing project (Table 7). Therefore, each of these SNPs is expected to be highly present in all populations. The results from any subset of these markers provide enough statistical power to reliably identify and separate every individual from the most genome data available.

2.10. Digital Identification Panel Design

We further selected 24 SNPs from the 100 candidates SNPs for our digital identity panel that showed the least ethnicity differences in all populations tested

Table 6. Theoretical combinations codes from the given SNP marker analysis based on three different allele combinations (bi-allele SNP analysis).

Num of Alleles	Num of SNPs	Possible Combination codes
3	20	3,486,784,401
3	21	10,460,353,203
3	22	31,381,059,609
3	23	94,143,178,827
3	24	282,429,536,481
3	25	847,288,609,443
3	26	2,541,865,828,329

Table 7. An example of a 100 candidate SNP marker selection from a public HapMap and 1000 Genome Sequencing database for digital identification and personalization purpose. RSID: Reference SNP Cluster ID, #CHR: Chromosome Number, position: SNP position in the chromosome.

RSID	#CHR	position	REF	ALT	EAS	SAS	EUR	AMR	AFR	Average
rs884918	chr16	10,059,826	T	C	0.5903	0.5337	0.499	0.4957	0.4054	0.50482
rs4901619	chr14	56,367,814	A	G	0.5903	0.4427	0.4135	0.5865	0.4705	0.5007
rs12194277	chr6	1.55E+08	T	C	0.5833	0.4213	0.5159	0.5605	0.4138	0.49896
rs9976404	chr21	35,551,600	C	A	0.5823	0.4202	0.505	0.5937	0.4191	0.50406
rs312368	chr4	19,275,229	G	A	0.5605	0.4264	0.4115	0.598	0.5151	0.5023
rs3829794	chr1	1.79E+08	T	C	0.5347	0.591	0.504	0.4164	0.4773	0.50468
rs7551188	chr1	25,273,200	T	C	0.5248	0.5307	0.4732	0.5893	0.407	0.505
rs2570569	chr9	18,700,566	A	G	0.5248	0.5174	0.4672	0.5807	0.4032	0.49866
rs7332772	chr13	38,270,164	G	A	0.5248	0.4427	0.5139	0.4222	0.618	0.50432
rs9225	chr2	47,086,112	C	T	0.4861	0.453	0.5934	0.4323	0.5416	0.50128
rs6940098	chr6	2,826,677	A	G	0.4831	0.4744	0.4453	0.5	0.618	0.50416
rs7596062	chr2	5,855,300	T	G	0.4821	0.4448	0.4891	0.5893	0.5091	0.50288
rs7688462	chr4	1.09E+08	C	T	0.4722	0.59	0.503	0.4625	0.4856	0.50266
rs1504192	chr5	18,173,795	T	C	0.4692	0.5337	0.4622	0.428	0.6059	0.4998
rs6572347	chr14	22,736,111	T	C	0.4692	0.4652	0.4304	0.513	0.6256	0.50068
rs2826401	chr21	21,939,786	C	A	0.4663	0.5266	0.4394	0.4496	0.6399	0.50436
rs2065378	chr9	89,188,290	C	A	0.4663	0.4346	0.5974	0.5706	0.4221	0.4982
rs10925027	chr1	2.48E+08	T	C	0.4653	0.4867	0.5954	0.4424	0.5242	0.5028
rs2932770	chr5	68,513,290	A	G	0.4623	0.5818	0.4245	0.4683	0.5537	0.49812
rs7642134	chr3	86,916,882	A	G	0.4603	0.5828	0.5825	0.4496	0.4425	0.50354
rs4322086	chr9	85,787,718	G	A	0.4563	0.4939	0.5169	0.5793	0.4697	0.50322
rs9411345	chr9	1.35E+08	C	A	0.4544	0.5777	0.4563	0.4121	0.6241	0.50492
rs11156119	chr6	1.57E+08	C	T	0.4405	0.5133	0.5179	0.5922	0.441	0.50098
rs236528	chr17	68,215,590	T	C	0.4385	0.5879	0.5	0.4841	0.4947	0.50104
rs12616545	chr2	1.19E+08	T	C	0.4325	0.4611	0.5726	0.438	0.6044	0.50172
rs6565479	chr17	78,712,423	A	G	0.4296	0.453	0.5805	0.4207	0.6082	0.4984
rs2680197	chr9	1.1E+08	T	G	0.4286	0.4397	0.5616	0.4741	0.6172	0.50424
rs631287	chr9	1.28E+08	G	A	0.4256	0.5828	0.5398	0.5115	0.4592	0.50378
rs4512966	chr13	1.11E+08	T	C	0.4246	0.4836	0.4821	0.5058	0.6006	0.49934
rs873233	chr1	1.53E+08	A	G	0.4187	0.4857	0.4264	0.5648	0.6112	0.50136
rs2808105	chr10	32,110,907	C	T	0.4187	0.4632	0.495	0.5375	0.6021	0.5033
rs11745164	chr5	1.61E+08	T	C	0.4187	0.4601	0.5537	0.5821	0.503	0.50352
rs10854351	chr21	33,382,451	T	C	0.4157	0.4918	0.5954	0.4409	0.5734	0.50344
rs12642421	chr4	1.12E+08	G	A	0.5784	0.5368	0.4493	0.4308	0.5212	0.5033
rs10737505	chr1	1.65E+08	C	T	0.5744	0.4765	0.5477	0.5086	0.407	0.50284

Continued

rs9302806	chr16	5,793,163	G	A	0.5744	0.4765	0.4284	0.4712	0.5492	0.49994
rs7693572	chr4	60,143,686	C	T	0.5734	0.498	0.4433	0.4539	0.525	0.49872
rs4843257	chr16	87,713,034	G	A	0.5645	0.4274	0.4751	0.5533	0.4697	0.498
rs3114431	chr7	20,347,570	A	G	0.5526	0.5654	0.4374	0.4294	0.5197	0.5009
rs687782	chr11	95,766,185	C	T	0.5516	0.5184	0.499	0.4323	0.4902	0.4983
rs2453244	chr1	32,975,903	T	G	0.5516	0.502	0.4314	0.513	0.5053	0.50066
rs10237006	chr7	1.57E+08	A	G	0.5486	0.4448	0.4911	0.5101	0.5068	0.50028
rs4428160	chr3	14,253,587	A	G	0.5456	0.4622	0.5318	0.4986	0.4788	0.5034
rs130423	chr22	33,984,394	A	G	0.5456	0.4611	0.5	0.4582	0.5492	0.50282
rs12303240	chr12	89,832,423	C	T	0.5446	0.4274	0.499	0.4712	0.5628	0.501
rs10930407	chr2	1.71E+08	C	T	0.5437	0.4949	0.4394	0.5706	0.4546	0.50064
rs7231112	chr18	57,612,891	G	A	0.5407	0.4448	0.4861	0.4683	0.5703	0.50204
rs12049869	chr11	1.28E+08	A	G	0.5377	0.4499	0.4861	0.4798	0.5393	0.49856
rs6860296	chr5	91,798,227	A	G	0.5337	0.5	0.4374	0.4712	0.5764	0.50374
rs2420477	chr10	1.2E+08	T	C	0.5278	0.5583	0.5249	0.4135	0.4735	0.4996
rs2638398	chr12	20,015,242	A	C	0.5228	0.4867	0.5577	0.5159	0.4092	0.49846
rs11074843	chr16	27,090,311	T	C	0.5228	0.4213	0.5467	0.5562	0.4705	0.5035
rs273701	chr18	23,181,382	C	A	0.5218	0.4376	0.4632	0.5548	0.5431	0.5041
rs3761161	chr20	10,110,237	G	A	0.5188	0.5307	0.5527	0.4712	0.447	0.50408
rs2172645	chr2	53,647,729	A	C	0.5188	0.5225	0.4364	0.4337	0.5968	0.50164
rs1048408	chr3	1.51E+08	G	A	0.5139	0.5102	0.4642	0.5692	0.4561	0.50272
rs708563	chr17	43,340,936	G	A	0.5139	0.4642	0.5229	0.5735	0.4478	0.50446
rs7939948	chr11	36,311,014	G	T	0.5129	0.5266	0.4722	0.4337	0.5582	0.50072
rs567681	chr11	1.16E+08	A	C	0.5129	0.5061	0.4523	0.4568	0.5605	0.49772
rs4684400	chr3	3,657,425	G	T	0.5129	0.5041	0.5318	0.4813	0.4781	0.50164
rs303898	chr6	25,039,602	G	A	0.5119	0.5593	0.505	0.428	0.5204	0.50492
rs10813796	chr9	32,359,091	A	G	0.5109	0.5061	0.5447	0.5274	0.4115	0.50012
rs17616971	chr15	36,260,238	G	A	0.5079	0.5123	0.5378	0.4222	0.5159	0.49922
rs7309088	chr12	1.04E+08	C	T	0.5079	0.4642	0.5368	0.4582	0.5356	0.50054
rs1884770	chr20	19,753,074	C	A	0.506	0.4816	0.502	0.5519	0.447	0.4977
rs2834323	chr21	35,359,935	A	G	0.505	0.5245	0.4235	0.5231	0.5182	0.49886
rs10169352	chr2	66,357,905	T	C	0.503	0.501	0.4583	0.5231	0.5287	0.50282
rs11037366	chr11	43,244,906	A	G	0.5	0.4632	0.5447	0.4107	0.5749	0.4987
rs2510245	chr9	1.37E+08	G	A	0.498	0.5327	0.5586	0.4971	0.4092	0.49912
rs12919933	chr16	3,137,261	G	A	0.497	0.545	0.498	0.4553	0.528	0.50466
rs1891110	chr10	1.25E+08	G	A	0.4931	0.4898	0.5785	0.4597	0.4834	0.5009
rs2924466	chr5	8,077,509	A	G	0.4891	0.4581	0.502	0.562	0.4879	0.49982

Continued

rs10955282	chr8	1.03E+08	G	A	0.4841	0.4836	0.4573	0.4712	0.5961	0.49846
rs3737669	chr1	9,164,133	G	A	0.4831	0.5225	0.5358	0.5476	0.4153	0.50086
rs252217	chr5	14,118,565	G	T	0.4831	0.4427	0.5775	0.5375	0.4735	0.50286
rs255888	chr5	1.11E+08	C	T	0.4792	0.4796	0.5517	0.5562	0.4561	0.50456
rs9290003	chr3	99,906,993	T	C	0.4722	0.544	0.4314	0.4798	0.5734	0.50016
rs1474016	chr10	1.25E+08	A	C	0.4722	0.4683	0.5507	0.4914	0.5356	0.50364
rs1573612	chr12	12,047,429	T	C	0.4722	0.4417	0.4503	0.5778	0.5772	0.50384
rs2736043	chr8	15,713,330	G	A	0.4683	0.5307	0.494	0.4683	0.5537	0.503
rs4831495	chr8	13,726,384	C	T	0.4663	0.501	0.5646	0.549	0.4175	0.49968
rs7203781	chr16	25,759,028	T	G	0.4563	0.4877	0.5716	0.5605	0.4123	0.49768
rs4594899	chr5	1,132,216	A	G	0.4544	0.5368	0.4374	0.5576	0.5098	0.4992
rs9345612	chr6	66,083,780	G	A	0.4534	0.5153	0.5437	0.5058	0.472	0.49804
rs168080	chr5	1.77E+08	A	G	0.4534	0.4693	0.5159	0.4914	0.5877	0.50354
rs7815790	chr8	98,902,730	A	G	0.4504	0.454	0.5775	0.5591	0.4493	0.49806
rs11672485	chr19	53,833,712	T	C	0.4454	0.4663	0.5417	0.4697	0.5915	0.50292
rs1876648	chr5	1.21E+08	G	A	0.4425	0.5511	0.5527	0.4236	0.5469	0.50336
rs2022499	chr20	36,966,531	G	T	0.4425	0.4898	0.5398	0.5288	0.5159	0.50336
rs4772827	chr13	86,963,707	T	G	0.4425	0.4898	0.4891	0.5086	0.5877	0.50354
rs897297	chr10	1.26E+08	A	G	0.4405	0.5143	0.5586	0.5793	0.4115	0.50084
rs2970627	chr8	36,884,712	T	C	0.4315	0.5327	0.5696	0.4395	0.5424	0.50314
rs4240927	chr1	2.22E+08	C	T	0.4276	0.5583	0.5189	0.428	0.5598	0.49852
rs4682576	chr3	1.1E+08	C	T	0.4276	0.5215	0.5577	0.5346	0.4644	0.50116
rs6664070	chr1	1.17E+08	A	C	0.4256	0.4734	0.5229	0.5519	0.5265	0.50006
rs10171464	chr2	1.39E+08	C	T	0.4256	0.4489	0.5338	0.5476	0.5537	0.50192
rs6889876	chr5	27,804,605	C	T	0.4236	0.4346	0.5716	0.5245	0.5386	0.49858
rs460697	chr5	95,018,939	A	G	0.4206	0.546	0.5348	0.4625	0.5287	0.49852
rs2323661	chr17	15,300,309	A	C	0.4127	0.5225	0.4801	0.5144	0.5628	0.4985
rs6955473	chr7	90,866,589	G	A	0.4117	0.4928	0.5795	0.5317	0.4841	0.49996

(Table 8). The chosen SNPs show minor allele frequency between 0.43 - 0.58 in every population tested, allowing for a wide variety of combinations for every individual across all populations. The fixed universal marker set can be used in all individuals in the same species for the digital identification panel.

2.11. Personal Digital Identity Code Assignment from 24 SNPs Panel Collection

Table 9 below is used to generate a personalized digital code for two real individuals (Person 1 male, and Person 2 female) using selected 24 SNP markers in the panel.

Table 8. 24SNP marker selection for the development of a digital identification panel. RSID: Reference SNP Cluster ID, #CHR: Chromosome Number, position; SNP position in the chromosome, REF: Reference Allele, ALT: Alternative Allele, EAS: East Asian, SAS: South.

RSID	#CHR	position	REF	ALT	EAS	SAS	EUR	AMR	AFR	Average
rs10169352	chr2	66,357,905	T	C	0.503	0.501	0.4583	0.5231	0.5287	0.50282
rs1048408	chr3	1.51E+08	G	A	0.5139	0.5102	0.4642	0.5692	0.4561	0.50272
rs4428160	chr3	14,253,587	A	G	0.5456	0.4622	0.5318	0.4986	0.4788	0.5034
rs4684400	chr3	3,657,425	G	T	0.5129	0.5041	0.5318	0.4813	0.4781	0.50164
rs255888	chr5	1.11E+08	C	T	0.4792	0.4796	0.5517	0.5562	0.4561	0.50456
rs252217	chr5	14,118,565	G	T	0.4831	0.4427	0.5775	0.5375	0.4735	0.50286
rs2924466	chr5	8,077,509	A	G	0.4891	0.4581	0.502	0.562	0.4879	0.49982
rs9345612	chr6	66,083,780	G	A	0.4534	0.5153	0.5437	0.5058	0.472	0.49804
rs10237006	chr7	1.57E+08	A	G	0.5486	0.4448	0.4911	0.5101	0.5068	0.50028
rs2736043	chr8	15,713,330	G	A	0.4683	0.5307	0.494	0.4683	0.5537	0.503
rs1474016	chr10	1.25E+08	A	C	0.4722	0.4683	0.5507	0.4914	0.5356	0.50364
rs12049869	chr11	1.28E+08	A	G	0.5377	0.4499	0.4861	0.4798	0.5393	0.49856
rs567681	chr11	1.16E+08	A	C	0.5129	0.5061	0.4523	0.4568	0.5605	0.49772
rs687782	chr11	95,766,185	C	T	0.5516	0.5184	0.499	0.4323	0.4902	0.4983
rs7939948	chr11	36,311,014	G	T	0.5129	0.5266	0.4722	0.4337	0.5582	0.50072
rs7309088	chr12	1.04E+08	C	T	0.5079	0.4642	0.5368	0.4582	0.5356	0.50054
rs1573612	chr12	12,047,429	T	C	0.4722	0.4417	0.4503	0.5778	0.5772	0.50384
rs12919933	chr16	3,137,261	G	A	0.497	0.545	0.498	0.4553	0.528	0.50466
rs708563	chr17	43,340,936	G	A	0.5139	0.4642	0.5229	0.5735	0.4478	0.50446
rs7231112	chr18	57,612,891	G	A	0.5407	0.4448	0.4861	0.4683	0.5703	0.50204
rs2022499	chr20	36,966,531	G	T	0.4425	0.4898	0.5398	0.5288	0.5159	0.50336
rs1884770	chr20	19,753,074	C	A	0.506	0.4816	0.502	0.5519	0.447	0.4977
rs3761161	chr20	10,110,237	G	A	0.5188	0.5307	0.5527	0.4712	0.447	0.50408
rs130423	chr22	33,984,394	A	G	0.5456	0.4611	0.5	0.4582	0.5492	0.50282
Average Frequency					0.44 - 0.55	0.44 - 0.55	0.45 - 0.58	0.43 - 0.58	0.45 - 0.57	0.5 - 0.55

3. Results

- Individual 1:

  Numerical Codes: 233331123133313113232333

  Alphabetical Codes: DACHCACCEDDA

- Individual 2:

  Numerical Codes 133223333321112302211231

  Alphabetical Codes: EBDAAGIDNGHC

Table 9. Digital identification code assignment from the SNP panel data.

Markers	RSID	#CHR	position	REF	ALT	Person 1 Genotype	Code 1	Person 2 Genotype	Code 2
1	rs10169352	chr2	66,357,905	T	C	CC	2	TT	1
2	rs1048408	chr3	1.51E+08	G	A	AG	3	AG	3
3	rs4428160	chr3	14,253,587	A	G	AG	3	AG	3
4	rs4684400	chr3	3,657,425	G	T	GT	3	TT	2
5	rs255888	chr5	1.11E+08	C	T	CT	3	TT	2
6	rs252217	chr5	14,118,565	G	T	GG	1	GT	3
7	rs2924466	chr5	8,077,509	A	G	AA	1	AG	3
8	rs9345612	chr6	66,083,780	G	A	AA	2	AG	3
9	rs10237006	chr7	1.57E+08	A	G	AG	3	AG	3
10	rs2736043	chr8	15,713,330	G	A	GG	1	AG	3
11	rs1474016	chr10	1.25E+08	A	C	AC	3	CC	2
12	rs12049869	chr11	1.28E+08	A	G	AG	3	AA	1
13	rs567681	chr11	1.16E+08	A	C	AC	3	AA	1
14	rs687782	chr11	95,766,185	C	T	CC	1	CC	1
15	rs7939948	chr11	36,311,014	G	T	GT	3	TT	2
16	rs7309088	chr12	1.04E+08	C	T	CC	1	CT	3
17	rs1573612	chr12	12,047,429	T	C	TT	1	CT	3
18	rs12919933	chr16	3,137,261	G	A	AG	3	AA	2
19	rs708563	chr17	43,340,936	G	A	AA	2	AA	2
20	rs7231112	chr18	57,612,891	G	A	AG	3	GG	1
21	rs2022499	chr20	36,966,531	G	T	TT	2	GG	1
22	rs1884770	chr20	19,753,074	C	A	AC	3	AA	2
23	rs3761161	chr20	10,110,237	G	A	AG	3	AG	3
24	rs130423	chr22	33,984,394	A	G	AG	3	AA	1

Generation of the Personalized Digital ID Code

Format: Species Code (2), Gender Code (1), Identification Codes (24), Auxiliary Code (1). Below are the two people used to generate the digital identification code using the format and composition of this paper. Individual 1 is a male human and single born. Individual 2 is a female human and single born. A 2D barcode can be created from the following code and can be linked to an HTML file or web page to link with additional information (Figure 1 and Figure 2).

- Individual 1: Human, Male, Single born

  HS,1,233331123133313113232333,1

  HS,1,DACHCACCEDDA1

- Individual 2: Human, Female, Single born

  HS,2,133223333321112302211231,1

  HS,2,EBDAAGIDNGHC,1

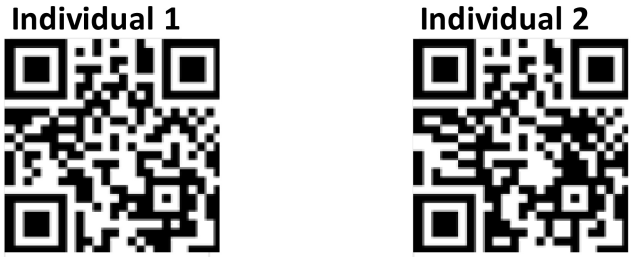


Figure 1. Generated 2D QR Code from the numerical genetic identity code from each individual using the UTF-8 encoding.

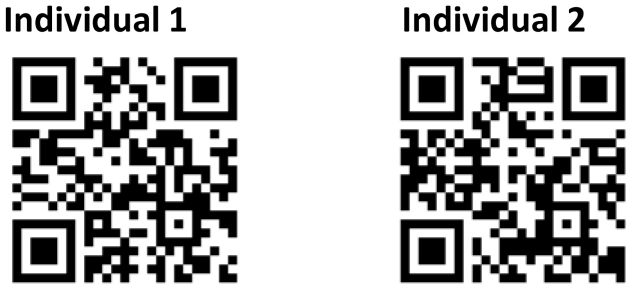


Figure 2. Generated 2D QR Code from the alphabetical genetic identity code from each individual using the UTF-8 encoding.

4. Discussion

The era of the Internet and the continual rise of social media platforms has shown the true value of secure identification and personalization. A large majority of consumers now expect companies to deliver personalized products and services, while at the same time, ensuring maximum privacy and control over their data [63].

In this paper, a detailed process of generating a personalized digital code from an individual's genome data was proposed. Through a process called genome fingerprinting, this code can be used as part of a decentralized identification (DID) system [64]. The genetic information obtained from genome fingerprinting can then be securely stored on blockchain, creating a genome based decentralized identifier (gDID) [65] [66]. This allows for secure authentication and verification of identity based on the individual's genomic information.

A genome-based code can be useful in a variety of ways in current Internet infrastructure as well as a decentralized one [67]. For example, genome data can be used to personalize virtual experiences by having custom avatars and characters. A real-world example that similarly captures this concept is CryptoKitties released in 2017 [68] [69]. CrytoKitties is a decentralized platform and digital collectible game built on the Ethereum blockchain. The game has a community of players that breed and raise virtual cats. Although the platform did not use actual genome data, each CryptoKitty, to a certain extent, had its own "virtual" genome, making each virtual cat unique and having its own set of traits and attributes that determine its appearance and abilities. CryptoKitties retained its

popularity through 2019 where the platform accumulated over 1.5 million users and over $40 million worth of transactions. The game's popularity has shown that personalized virtual experiences using blockchain technology has immense potential and using real genome data to augment such platforms can bring new, unreached levels of identity and personalization [69]. But perhaps a more universal use-case of utilizing a genome-based code is enhanced authentication and identity on Blockchain [70]. Genetic data is unique to an individual and cannot be changed or altered, inherently making it a highly secure form of identification [71]. Genome data can be used for biometric authentication in addition to finger or iris scanning [72]. A study has shown that a large majority of internet consumers frequently report password problems and frustrations, supporting the idea that novel forms of biometric authentication provide a more secure alternative to traditional passwords and usernames [73] [74]. Many users are also hesitant of putting their genome data out into the Internet so a decentralized storage solution like Blockchain can prevent data breaches that can compromise traditional centralized databases [75].

However, it is important to note that while genome fingerprinting provides a unique identifier, there are privacy and ethical concerns associated with the storage and use of genetic information, particularly in the context of decentralized identification [65] [76]. Careful consideration must be given to the security, privacy, and ethical implications of using genomic information in this manner, and appropriate safeguards must be put in place to ensure that the information is protected and used responsibly. In addition, the lack of clear regulations of DIDs at the moment can create uncertainty for businesses and individuals who want to use the technology [77]. This can cause further implications with scalability, as the average user might struggle with widespread adoption especially when the technology cannot keep up with the demand.

But with almost every mass technological adoption, society tends to appreciate technological revolutions only in hindsight. DIDs carry with it numerous advantages in comparison to existing forms of identification. First, DIDs enable individuals to have a greater degree of control over their personal data due to their reduced reliance on centralized intermediaries [78] [79]. Second, DIDs are based on open standards, allowing them to be used in a wide range of platforms and streamlining password authentication processes [80]. Such benefits are paramount in modern society because it improves the overall security and privacy of digital interactions, a key concern for many new technology innovations.

5. Conclusion

With the internet and social networking becoming an integral part of people's daily lives, secure identification and personalization are of utmost importance to many consumers and businesses. In this paper, we presented a novel way of securely identifying an individual online by using universal, high allele frequency bi-allele human SNP markers to efficiently generate a unique combination code.

We also presented real-world use cases for applying a genome-based DID in the industry as well as implications and technical challenges. When generating the code, it was important to select genetic markers that were common in all populations, so that any individual, no matter their background, is able to generate an identifier unique only to that person. Novel forms of identification and personalization based on genome data can open up new opportunities for personalization and offer better security and a better user experience for Internet users in the present as well as those in the decentralized future.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Madakam, S., *et al.* (2015) Internet of Things (IoT): A Literature Review. *Journal of Computer and Communications*, **3**, 164-173. https://doi.org/10.4236/jcc.2015.35021

[2] Gaudet, M., *et al.* (2009) Allele-Specific PCR in SNP Genotyping. In: Komar, A., Ed., *Single Nucleotide Polymorphisms. Methods in Molecular Biology*, Humana Press, Totowa, NJ, 415-424. https://doi.org/10.1007/978-1-60327-411-1_26

[3] Su, W., *et al.* (2020) Do Men Become Addicted to Internet Gaming and Women to Social Media? A Meta-Analysis Examining Gender-Related Differences in Specific Internet Addiction. *Computers in Human Behavior*, **113**, Article ID: 106480. https://doi.org/10.1016/j.chb.2020.106480

[4] Mystakidis, S. (2022) Metaverse. *Encyclopedia*, **2**, 486-497. https://doi.org/10.3390/encyclopedia2010031

[5] Ives, B., Walsh, K.R. and Schneider, H. (2004) The Domino Effect of Password Reuse. *Communications of the ACM*, **47**, 75-78. https://doi.org/10.1145/975817.975820

[6] Alabdan, R. (2020) Phishing Attacks Survey: Types, Vectors, and Technical Approaches. *Future Internet*, **12**, Article 168. https://doi.org/10.3390/fi12100168

[7] Wright, E. (2018) The Future of Facial Recognition Is Not Fully Known: Developing Privacy and Security Regulatory Mechanisms for Facial Recognition in the Retail Sector. *The Fordham Intellectual Property, Media and Entertainment Law Journal*, **29**, Article 611.

[8] Chen, J., Lv, Z. and Song, H. (2019) Design of Personnel Big Data Management System Based on Blockchain. *Future Generation Computer Systems*, **101**, 1122-1129. https://doi.org/10.1016/j.future.2019.07.037

[9] Chowdhury, M.U., *et al.* (2021) Blockchain Application in Banking System. *Journal of Software Engineering and Applications*, **14**, 298-311. https://doi.org/10.4236/jsea.2021.147018

[10] Allison, A., *et al.* (2005) Digital Identity Matters. *Journal of the American Society for Information Science and Technology*, **56**, 364-372. https://doi.org/10.1002/asi.20112

[11] Argento, L., *et al.* (2020) ID-Service: A Blockchain-Based Platform to Support Digital-Identity-Aware Service Accountability. *Applied Sciences*, **11**, Article 165. https://doi.org/10.3390/app11010165

[12] Sule, M.J., Zennaro, M. and Thomas, G. (2021) Cybersecurity through the Lens of

Digital Identity and Data Protection: Issues and Trends. *Technology in Society*, **67**, Article ID: 101734. https://doi.org/10.1016/j.techsoc.2021.101734

[13] Buckingham, D. (2008) Introducing Identity. MacArthur Foundation Digital Media and Learning Initiative.

[14] Schwartz, P.M. and Solove, D.J. (2011) The PII Problem: Privacy and a New Concept of Personally Identifiable Information. *New York University Law Review*, **86**, 1814.

[15] Tsai, C.H. (2015) The Application of a Personal Identification Database and Risk Management Mechanism. *International Journal of Social Sciences and Education Research*, **1**, 1009-1016. https://doi.org/10.24289/ijsser.279112

[16] Regan, P.M. (2002) Privacy as a Common Good in the Digital World. *Information, Communication & Society*, **5**, 382-405. https://doi.org/10.1080/13691180210159328

[17] Buccafurri, F., *et al.* (2018) Integrating Digital Identity and Blockchain. On the Move to Meaningful Internet Systems. OTM 2018 Conferences. *Confederated International Conferences: CoopIS, C&TC, and ODBASE* 2018, Valletta, 22-26 October 2018, 568-585.

[18] Lim, S.Y., *et al.* (2018) Blockchain Technology the Identity Management and Authentication Service Disruptor: A Survey. *International Journal on Advanced Science, Engineering and Information Technology*, **8**, 1735-1745. https://doi.org/10.18517/ijaseit.8.4-2.6838

[19] Stratopoulos, T.C., Wang, V.X. and Ye, J. (2020) Blockchain Technology Adoption. Use of Corporate Disclosures to Identify the Stage of Blockchain Adoption. *Accounting Horizons*, **36**, 197-220. https://doi.org/10.2308/HORIZONS-19-101

[20] Ning, X., Ramirez, R. and Khuntia, J. (2021) Blockchain-Enabled Government Efficiency and Impartiality: Using Blockchain for Targeted Poverty Alleviation in a City in China. *Information Technology for Development*, **27**, 599-616. https://doi.org/10.1080/02681102.2021.1925619

[21] Clavin, J., *et al.* (2020) Blockchains for Government: Use Cases and Challenges. *Digital Government: Research and Practice*, **1**, Article 22. https://doi.org/10.1145/3427097

[22] Zarrin, J., *et al.* (2021) Blockchain for Decentralization of Internet: Prospects, Trends, and Challenges. *Cluster Computing*, **24**, 2841-2866. https://doi.org/10.1007/s10586-021-03301-8

[23] Mills, R.E., *et al.* (2011) Mapping Copy Number Variation by Population-Scale Genome Sequencing. *Nature*, **470**, 59-65. https://doi.org/10.1038/nature09708

[24] Eichler, E.E. (2019) Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine*, **381**, 64-74. https://doi.org/10.1056/NEJMra1809315

[25] Chaisson, M.J., Wilson, R.K. and Eichler, E.E. (2015) Genetic Variation and the de Novo Assembly of Human Genomes. *Nature Reviews Genetics*, **16**, 627-640. https://doi.org/10.1038/nrg3933

[26] Knight, J.C. (2010) Understanding Human Genetic Variation in the Era of High-Throughput Sequencing. *EMBO Reports*, **11**, 650-652. https://doi.org/10.1038/embor.2010.126

[27] Ellegren, H. and Galtier, N. (2016) Determinants of Genetic Diversity. *Nature Reviews Genetics*, **17**, 422-433. https://doi.org/10.1038/nrg.2016.58

[28] Collins, F.S., *et al.* (1998) New Goals for the US Human Genome Project: 1998-2003. *Science*, **282**, 682-689. https://doi.org/10.1126/science.282.5389.682

[29] Ikegawa, S. (2012) A Short History of the Genome-Wide Association Study: Where

We Were and Where We Are Going. *Genomics & Informatics*, **10**, 220-225. https://doi.org/10.5808/GI.2012.10.4.220

[30] Gudbjartsson, D.F., *et al.* (2015) Large-Scale Whole-Genome Sequencing of the Icelandic Population. *Nature Genetics*, **47**, 435-444. https://doi.org/10.1038/ng.3247

[31] Senthilvel, S., *et al.* (2019) Development and Validation of an SNP Genotyping Array and Construction of a High-Density Linkage Map in Castor. *Scientific Reports*, **9**, Article No. 3003. https://doi.org/10.1038/s41598-019-39967-9

[32] Williams, L.M., *et al.* (2010) SNP Identification, Verification, and Utility for Population Genetics in a Non-Model Genus. *BMC Genetics*, **11**, Article No. 32. https://doi.org/10.1186/1471-2156-11-32

[33] The International HapMap Consortium (2005) A Haplotype Map of the Human Genome. *Nature*, **437**, 1299-1320. https://doi.org/10.1038/nature04226

[34] Miller, M.P. and Kumar, S. (2001) Understanding Human Disease Mutations through the Use of Interspecific Genetic Variation. *Human Molecular Genetics*, **10**, 2319-2328. https://doi.org/10.1093/hmg/10.21.2319

[35] Hinds, D.A., *et al.* (2006) Common Deletions and SNPs Are in Linkage Disequilibrium in the Human Genome. *Nature Genetics*, **38**, 82-85. https://doi.org/10.1038/ng1695

[36] Kruglyak, L. (1997) The Use of a Genetic Map of Biallelic Markers in Linkage Studies. *Nature Genetics*, **17**, 21-24. https://doi.org/10.1038/ng0997-21

[37] Xiong, M. and Jin, L. (1999) Comparison of the Power and Accuracy of Biallelic and Microsatellite Markers in Population-Based Gene-Mapping Methods. *The American Journal of Human Genetics*, **64**, 629-640. https://doi.org/10.1086/302231

[38] Tarach, P. (2021) Application of Polymerase Chain Reaction-Restriction Fragment length Polymorphism (RFLP-PCR) in the Analysis of Single Nucleotide Polymorphisms (SNPs). *Acta Universitatis Lodziensis. Folia Biologica et Oecologica*, **17**, 48-53. https://doi.org/10.18778/1730-2366.16.14

[39] Heaton, M.P., *et al.* (2002) Selection and Use of SNP Markers for Animal Identification and Paternity Analysis in US Beef Cattle. *Mammalian Genome*, **13**, 272-281. https://doi.org/10.1007/s00335-001-2146-3

[40] Handsaker, R.E., *et al.* (2015) Large Multiallelic Copy Number Variations in Humans. *Nature Genetics*, **47**, 296-303. https://doi.org/10.1038/ng.3200

[41] Nielsen, R., *et al.* (2011) Genotype and SNP Calling from Next-Generation Sequencing Data. *Nature Reviews Genetics*, **12**, 443-451. https://doi.org/10.1038/nrg2986

[42] Rehm, H.L., *et al.* (2013) ACMG Clinical Laboratory Standards for Next-Generation Sequencing. *Genetics Medicine*, **15**, 733-747. https://doi.org/10.1038/gim.2013.92

[43] Logsdon, G.A., Vollger, M.R. and Eichler, E.E. (2020) Long-Read Human Genome Sequencing and Its Applications. *Nature Reviews Genetics*, **21**, 597-614. https://doi.org/10.1038/s41576-020-0236-x

[44] Beck, J.J., *et al.* (2021) Biology and Genetics of Dizygotic and Monozygotic Twinning. In: Khalil, A., Lewi, L. and Lopriore, E., Eds., *Twin and Higher-Order Pregnancies*, Springer, New York, 31-50. https://doi.org/10.1007/978-3-030-47652-6_3

[45] Ku, C.S., *et al.* (2010) The Discovery of Human Genetic Variations and Their Use as Disease Markers: Past, Present and Future. *Journal of Human Genetics*, **55**, 403-415. https://doi.org/10.1038/jhg.2010.55

[46] Miller, R.D., *et al.* (2005) High-Density Single-Nucleotide Polymorphism Maps of

the Human Genome. *Genomics*, **86**, 117-126.
https://doi.org/10.1016/j.ygeno.2005.04.012

[47] Pray, L. (2008) Eukaryotic Genome Complexity. *Nature Education*, **1**, Article 96.

[48] Jain, A.K., Prabhakar, S. and Pankanti, S. (2002) On the Similarity of Identical Twin Fingerprints. *Pattern Recognition*, **35**, 2653-2663.
https://doi.org/10.1016/S0031-3203(01)00218-7

[49] Syvänen, A.C. (2001) Accessing Genetic Variation: Genotyping Single Nucleotide Polymorphisms. *Nature Reviews Genetics*, **2**, 930-942.
https://doi.org/10.1038/35103535

[50] Shaw, G. (2013) Polymorphism and Single Nucleotide Polymorphisms (SNPs). *BJU International*, **112**, 664-665. https://doi.org/10.1111/bju.12298

[51] Taylor, J.G., *et al.* (2001) Using Genetic Variation to Study Human Disease. *Trends in Molecular Medicine*, **7**, 507-512. https://doi.org/10.1016/S1471-4914(01)02183-9

[52] Huang, T., Shu, Y. and Cai, Y.D. (2015) Genetic Differences among Ethnic Groups. *BMC Genomics*, **16**, Article No. 1093. https://doi.org/10.1186/s12864-015-2328-0

[53] Gibson, G. (2012) Rare and Common Variants: Twenty Arguments. *Nature Reviews Genetics*, **13**, 135-145. https://doi.org/10.1038/nrg3118

[54] Schork, N.J., *et al.* (2009) Common vs. Rare Allele Hypotheses for Complex Diseases. *Current Opinion in Genetics & Development*, **19**, 212-219.
https://doi.org/10.1016/j.gde.2009.04.010

[55] Cawood, A. (1989) DNA Fingerprinting. *Clinical Chemistry*, **35**, 1832-1837.
https://doi.org/10.1093/clinchem/35.9.1832

[56] Patzak, J., Vrba, L. and Matoušek, J. (2007) New STS Molecular Markers for Assessment of Genetic Diversity and DNA Fingerprinting in Hop (*Humulus lupulus* L.). *Genome*, **50**, 15-25. https://doi.org/10.1139/g06-128

[57] Lynch, M. (1988) Estimation of Relatedness by DNA Fingerprinting. *Molecular Biology and Evolution*, **5**, 584-599.

[58] Wilkinson, S., *et al.* (2011) Evaluation of Approaches for Identifying Population Informative Markers from High Density SNP Chips. *BMC Genomic Data*, **12**, Article No. 45. https://doi.org/10.1186/1471-2156-12-45

[59] Eberle, M.A., *et al.* (2007) Power to Detect Risk Alleles Using Genome-Wide Tag SNP Panels. *PLOS Genetics*, **3**, e170.
https://doi.org/10.1371/journal.pgen.0030170

[60] Linck, E. and Battey, C. (2019) Minor Allele Frequency Thresholds Strongly Affect Population Structure Inference with Genomic Data Sets. *Molecular Ecology Resources*, **19**, 639-647. https://doi.org/10.1111/1755-0998.12995

[61] The International SNP Map Working Group (2001) A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms. *Nature*, **409**, 928-933. https://doi.org/10.1038/35057149

[62] Mardis, E.R. (2008) Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, **9**, 387-402.
https://doi.org/10.1146/annurev.genom.9.081307.164359

[63] Bleier, A., Goldfarb, A. and Tucker, C. (2020) Consumer Privacy and the Future of Data-Based Innovation and Marketing. *International Journal of Research in Marketing*, **37**, 466-480. https://doi.org/10.1016/j.ijresmar.2020.03.006

[64] Kuperberg, M. (2019) Blockchain-Based Identity Management: A Survey from the Enterprise and Ecosystem Perspective. *IEEE Transactions on Engineering Management*, **67**, 1008-1027. https://doi.org/10.1109/TEM.2019.2926471

[65] Kuo, T.T., *et al.* (2020) iDASH Secure Genome Analysis Competition 2018: Blockchain Genomic Data Access Logging, Homomorphic Encryption on GWAS, and DNA Segment Searching. *BMC Medical Genomics*, **13**, Article No. 98. https://doi.org/10.1186/s12920-020-0715-0

[66] Alshamrani, S.S. and Basha, A.F. (2021) IoT Data Security with DNA-Genetic Algorithm Using Blockchain Technology. *International Journal of Computer Applications in Technology*, **65**, 150-159. https://doi.org/10.1504/IJCAT.2021.114988

[67] Jin, X.L., *et al.* (2019) Application of a Blockchain Platform to Manage and Secure Personal Genomic Data: A Case Study of LifeCODE.ai in China. *Journal of Medical Internet Research*, **21**, e13587. https://doi.org/10.2196/13587

[68] Serada, A., Sihvonen, T. and Harviainen, J.T. (2021) CryptoKitties and the New Ludic Economy: How Blockchain Introduces Value, Ownership, and Scarcity in Digital Gaming. *Games and Culture*, **16**, 457-480. https://doi.org/10.1177/1555412019898305

[69] Ducuing, C. (2019) How to Make Sure My Cryptokitties Are Here Forever? The Complementary Roles of Blockchain and the Law to Bring Trust. *European Journal of Risk Regulation*, **10**, 315-329. https://doi.org/10.1017/err.2019.39

[70] Alghazwi, M., *et al.* (2022) Blockchain for Genomics: A Systematic Literature Review. *Distributed Ledger Technologies*: *Research and Practice*, **1**, Article 11. https://doi.org/10.1145/3563044

[71] Oldoni, F., Kidd, K.K. and Podini, D. (2019) Microhaplotypes in Forensic Genetics. *Forensic Science International*: *Genetics*, **38**, 54-69. https://doi.org/10.1016/j.fsigen.2018.09.009

[72] Pungila, C. and Negru. V. (2019) Accelerating DNA Biometrics In Criminal Investigations through GPU-Based Pattern Matching. *International Joint Conference SOCO*'18-*CISIS*'18-*ICEUTE*'18, San Sebastián, 6-8 June 2018, Proceedings 13.

[73] Bakken, S. (2021) Biometrics Light the Way for Secure Financial Services. *Biometric Technology Today*, **2021**, 10-12. https://doi.org/10.1016/S0969-4765(21)00072-2

[74] Mahier, J., *et al.* (2009) Biometric Authentication. In: Mehdi Khosrow-Pour, D.B.A. Ed., *Encyclopedia of Information Science and Technology*, IGI Global, Pennsylvania, 346-354. https://doi.org/10.4018/978-1-60566-026-4.ch059

[75] Gürsoy, G., *et al.* (2022) Storing and Analyzing a Genome on a Blockchain. *Genome Biology*, **23**, Article No. 134. https://doi.org/10.1186/s13059-022-02699-7

[76] Oestreich, M., *et al.* (2021) Privacy Considerations for Sharing Genomics Data. *EXCLI Journal*, **20**, Article 1243. https://doi.org/10.4135/9781071859544

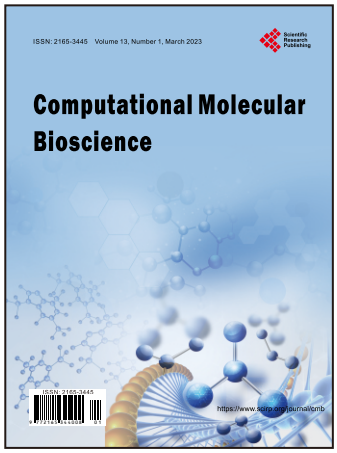
[77] Shabani, M. (2019) Blockchain-Based Platforms for Genomic Data Sharing: A De-Centralized Approach in Response to the Governance Problems? *Journal of the American Medical Informatics Association*, **26**, 76-80. https://doi.org/10.1093/jamia/ocy149

[78] Avellaneda, O., *et al.* (2019) Decentralized Identity: Where Did It Come from and Where Is It Going? *IEEE Communications Standards Magazine*, **3**, 10-13. https://doi.org/10.1109/MCOMSTD.2019.9031542

[79] Javed, I.T., *et al.* (2021) Health-ID: A Blockchain-Based Decentralized Identity Management for Remote Healthcare. *Healthcare*, **9**, Article 712. https://doi.org/10.3390/healthcare9060712

[80] Szalachowski, P. (2021) Password-Authenticated Decentralized Identities. *IEEE Transactions on Information Forensics and Security*, **16**, 4801-4810. https://doi.org/10.1109/TIFS.2021.3116429

Call for Papers



Computational Molecular Bioscience

ISSN: 2165-3445 (Print) ISSN: 2165-3453 (Online)
https://www.scirp.org/journal/cmb

Computational Molecular Bioscience **(*CMB*)** is an international journal dedicated to the latest advancement of Computational Molecular Bioscience. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of Computational Molecular Bioscience. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy. The journal publishes original papers including but not limited to the following fields:

- *Ab Initio* and Density Functional Calculations of Biomolecules
- Atomistic and Coarse Grained Molecular Dynamics
- Combined Computational and Experimental Studies of Biomolecular Interactions
- Combined Quantum Mechanical and Molecular Mechanical Methods (QM/MM)
- Computational Chemistry of Biomolecules, Ligands and Drugs
- Computational Drug Design: Structure-Based; Ligand-Based; Rational; *De Novo*
- Computational Modelling of Biomolecular Structures Interactions and Processes
- Computational Systems Biology and Chemistry
- Development and Applications of Monte Carlo Methods
- Development and Design of New Biological and Chemical Databases and Data Mining Techniques
- Development, Testing and Applications to Biomolecular Systems
- Enzymatic Reaction Mechanisms and Inhibition
- High Performance Computing in Molecular and Biomolecular Sciences
- Ligand Binding and Free Energy Calculations
- Modelling of Membrane Processes and Protein-Membrane Interactions
- Modelling Protein Structure, Conformational Dynamics an Interactions
- Molecular Mechanics, Force Field Development and Evaluation
- Molecular Visualizations and Data Analysis
- Multilevel Computational Simulations
- Nucleic Acids Structure, Dynamics and Interactions with Ligands
- Protein Folding
- Protein Ligand Docking New Algorithm, Codes and Applications
- Protein-Nucleic Acids Interactions
- Quantitative Structure-Activity Relationships (QSAR)
- Semiempirical Electronic Structure Calculations
- Structural Bioinformatics and Homology Modelling

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Website and E-Mail

https://www.scirp.org/journal/cmb **E-mail: cmb@scirp.org**

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: https://www.scirp.org

Subscription: sub@scirp.org

Advertisement: service@scirp.org