



- (51) International Patent Classification:
C12Q 1/68 (2006.01) *G06F 19/00* (2011.01)
- (21) International Application Number:
PCT/EP2014/075708
- (22) International Filing Date:
26 November 2014 (26.11.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1322400.1 18 December 2013 (18.12.2013) GB
- (71) Applicant: **GE HEALTHCARE UK LIMITED**
[GB/GB]; Amersham Place, Little Chalfont Buckinghamshire HP7 9NA (GB).
- (72) Inventors: **HORTON, Jeffrey, Kenneth**; GE Healthcare UK Limited, The Maynard Centre, Forest Farm Estate, Whitchurch, Cardiff South Glamorgan CF14 7YT (GB). **TATNELL, Peter, James**; GE Healthcare UK Limited, The Maynard Centre, Forest Farm Estate, Whitchurch, Cardiff South Glamorgan CF14 7YT (GB). **STONE, Robert**; GE Healthcare UK Limited, The Maynard Centre, Forest Farm Estate, Whitchurch, Cardiff South Glamorgan CF14 7YT (GB).
- (74) Agents: **CAVILL, Ross, David** et al.; GE Healthcare UK Limited, Pollards Wood, Nightingales Lane, Chalfont St Giles Buckinghamshire HP8 4SP (GB).

- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: OLIGONUCLEOTIDE DATA STORAGE ON SOLID SUPPORTS

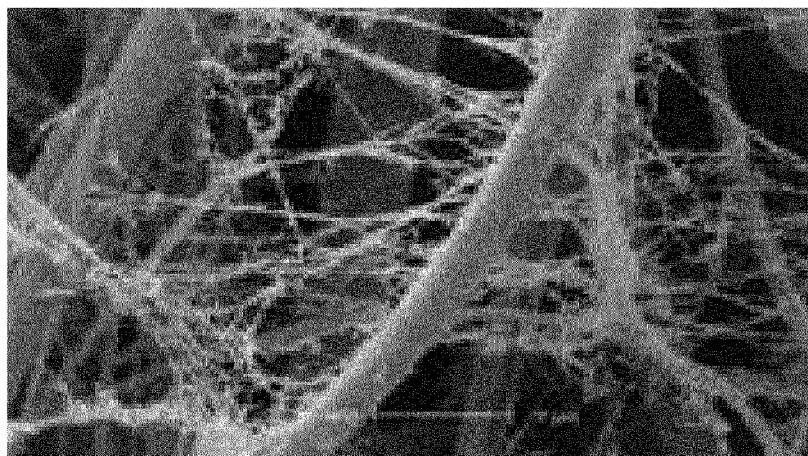


FIG. 6

(57) Abstract: A data storage medium is disclosed comprising a solid support matrix including an optional stabilising reagent or reagents in a dry form, for use as a support for artificially synthesised oligonucleotide sequences encoded with data. Preferably the matrix is fibrous (for example cellulose, or glass, fibres) formed into a support of sufficient strength to hold the oligonucleotide sequences. The stabilising reagents are preferably a combination of a weak base, and a chelating agent, optionally, uric acid or a urate salt, and optionally an anionic surfactant.



OLIGONUCLEOTIDE DATA STORAGE ON SOLID SUPPORTS

BACKGROUND OF THE INVENTION

5

Field of the Invention

The present invention relates to the storage and retrieval of data in the form of stabilised oligonucleotide nucleic acids.

10 Description of the Related Art

Publications referred to herein are incorporated by reference in their entirety.

The problem associated with storing the large volumes of data generated, is an ever-growing challenge. It is no longer feasible to store large amounts of data on a paper hard
15 copy. Digital technology provides a better solution but even the presently used forms of digital data storage have their limits. As increasing amounts of data, and more complex forms of information are created; storing these data using known digital media presents many difficulties. For example, computer hard drives and servers have limited capacity and require power and maintenance, while magnetic data storage media degrades. Very
20 long term storage is also problematic with known digital storage media, because, as well as the problems mentioned above, they can readily be affected by natural disasters and strong electromagnetic pulses for example from atomic weapon reactions or as occurs beyond Earth's atmosphere. There are also physical limits to the size of the equipment needed to store electronic data, meaning that it is becoming more and more problematic
25 to accommodate this equipment.

Church; Gao; and Kosuri (2012). "Next-Generation Digital Information Storage in DNA".
Science 337 (6102): 1628. [doi:10.1126/science.1226355] suggest that DNA can be
artificially synthesised in a form which encodes digital data. It is reported that DNA was
encoded with digital information that included an HTML draft of a 53400 word book, as
30 well as JPEG images and a java script algorithm. However, issues with errors were reported.

Further, Goldman; Bertone; Chen; Dessimoz; Leproust; Birney (2013). "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA". Nature 494 (7435): 77–80. [doi:10.1038/nature11875] report that DNA in the form of synthesised oligonucleotides could be prepared in a sequence which could then be read
5 later. Successful retrieval of that DNA data was reported also with no final errors, apparently because much of the DNA sequence data was overlapping, so intermediate errors could be identified and ignored. A sequential encoding technique was employed also to reduce errors. Since the synthesised oligonucleotides occupied very little volume, it was demonstrated that this technique could provide very high data density, and thereby
10 be an answer to holding ever-increasing volumes of data. In more detail, the authors were able to write and retrieve hours of high-definition video information using DNA sequences. The DNA was synthesized with defined sequence the order of which was indicative of the encoded digital information. In the study, 739 kilobytes of hard-disk information in the form of a MP3 of Martin Luther King's 'I have a dream' speech, a photo
15 of European Bioinformatics Institute in jpeg form, a text file of all of William Shakespeare's sonnets and a pdf of James Watson & Francis Crick's paper on the structure of DNA were taken and the digital information were all transformed into many overlapping complementary synthetically generated oligonucleotide DNA strands (~100 bps in length). The multiple DNA strands were subsequently sequenced, reassembled
20 and the digital information successfully recovered with 100% accuracy.

However, no thought was given by Church or Goldman et al to the long term storage of the DNA used to hold the data. Since DNA is in essence a string of monomers formed from nucleic acids, then like all organic matter, it can degrade, particularly in the
25 presence of enzymes, bacteria and moulds. It has been found that given ideal, cold, dry and dark conditions naturally occurring DNA can last for tens of thousands of years, for example as demonstrated by the DNA found within the remains of frozen woolly mammoths and the like. So, naturally it would be possible to keep the sequenced oligonucleotide data in cool, dry and dark conditions with a reasonable chance of
30 preserving it for long periods. The problem with that approach is that in most parts of the developed world, cool, dry and dark conditions require power and maintenance, which may not be available long term, for example due to natural disasters or wars. Moreover,

important data will require more certainty that the data will be preserved. So what is required is some form of physical storage arrangement which does not rely on power, and which is certain to keep the sequenced oligonucleotides in good condition, for subsequent decoding and reading of the data they contain. Further, it is postulated that
5 the relatively short sequenced oligonucleotides used by Goldman et al to hold data may be less stable than the relatively long DNA sequences occurring in nature, and so their successful long term storage may need more than the cool dry and dark conditions mentioned above.

10 Summary of the invention

The inventors of the present invention have realised that a solution to the above problem can be found in the use of a solid support matrix to which support can be applied oligonucleotide sequences in the form of artificially synthesised DNA encoded with data, preferably according to the description below. Preferably the solid support includes a
15 dried stabilising chemical. In particular they have found that a cellulose fibre paper and chemical coating on the paper (described in more detail below) sold under the trade name FTA (by Whatman Inc) provides a superior storage medium for oligonucleotides of about the length used by Goldman et al and longer.

20 Another issue is the monitoring of data integrity. The inventors propose herein a method and apparatus to facilitate visual monitoring of the DNA integrity and also a method and apparatus for monitoring DNA integrity by means of processing. These methods and apparatus will have applications beyond storage of data encoded in DNA and can be used for monitoring the integrity of naturally occurring DNA also.

25

According to a first aspect the invention provides a data storage medium comprising a solid support matrix for use as a dry storage support for oligonucleotide sequences encoded with data.

30 In an embodiment, said matrix includes a stabilising reagent or reagent mix in a dry form.

In an embodiment, the matrix is fibrous (for example cellulose, or glass, fibres) formed into a support of sufficient strength to hold the oligonucleotide sequences.

5 In an embodiment, the stabilising reagent or reagent mix is a reagent mix and comprises a combination of a weak base, and a chelating agent, optionally, uric acid or a urate salt, and optionally an anionic surfactant.

In an embodiment, the weak base is a Lewis base which has a pH of about 6 to 10, preferably about pH 8 to 9.5.

10

In an embodiment, the weak base is an organic and inorganic base, and if inorganic then optionally includes an alkali metal carbonate, bicarbonate, phosphate or borate (e.g. sodium, lithium, or potassium carbonate), and if organic then optionally includes, tris-hydroxymethyl amino methane (Tris), ethanolamine, tri-ethanolamine and glycine and
15 alkaline salts of organic acids (e.g. trisodium citrate).

In an embodiment, the weak base is Tris present either as a free base or as a salt, for example, a carbonate salt.

20 In an embodiment, the chelating agent binds multivalent metal ions with a comparable or better affinity than ethylene diamine tetraacetic acid (EDTA), and is preferably EDTA.

In an embodiment, the anionic surfactant includes a hydrocarbon moiety, aliphatic or aromatic, containing one or more anionic groups.

25

In an embodiment, the anionic surfactant is a detergent, for example sodium dodecyl sulphate (SDS) and/or sodium lauryl sarcosinate (SLS).

In an embodiment, the stabilising reagent comprises a chaotropic substance such as a
30 chaotropic salt, for example guanidinium thiocyanate.

In an embodiment, the storage medium includes a visual delineation around an area holding DNA data for example a printed boundary showing where the DNA is stored, or an indicating dye such as chlorophenol red or phenol red, which discolour, thereby showing where the DNA is stored.

5

According to a second aspect the invention provides a method of storing data, said method comprising the following steps: a) producing synthesised oligonucleotides having an nucleotide sequence representative of data to be stored; b) optionally replicating said oligonucleotides; and c) depositing said oligonucleotides, optionally replicated, onto a
10 storage medium according to the first aspect.

In an embodiment, at least steps b) and c) are performed in a liquid, and said liquid is allowed or caused to evaporate from the storage medium.

15 In an embodiment, step a) is performed according instructions published in the Goldman paper referred to herein.

Alternatively, step a) includes converting binary data into a quaternary code, and sequencing DNA nucleotides, according to the quaternary code to provide said
20 oligonucleotides.

In an embodiment, said conversion follows the n+1 rule described herein.

In an embodiment, wherein step a) includes incorporating a reporter into said sequence,
25 to provide an indication of data integrity, optionally said reporter providing an increase or decrease in detectability in response to a change in said integrity.

Preferably, said reporter is a fluorescent dye.

30 In an embodiment, said nucleotide sequence is at least 250 base pairs in length, and preferably at least about 1000 base pairs, for example at least about 963 base pairs.

In an embodiment, said replication is by means of a polymerase chain reaction (PCR), isothermal amplification, or rolling circle amplification.

In an embodiment, one or more sequences which aid replication are added to the end or
5 ends of the sequenced oligonucleotides.

According to a third aspect, the invention provides a method for recovering stored data from the storage medium of the first aspect when stored according to the method of the second aspect, said recovery method comprising the steps of:

- 10 i) extracting at least a portion of the replicated oligonucleotides from the storage medium;
- ii) optionally performing further replication of said oligonucleotides; and
- iii) performing an analysis of the oligonucleotides to determine their nucleotide sequence, and decoding said sequence to thereby to recover said stored data.

15 In an embodiment, said extracting step includes removing at least a portion of the storage medium, introducing into a vessel the at least a portion, and then performing said further replication step.

In an embodiment, said further replication step includes: adding a liquid to said at least a
20 portion.

In an embodiment, the analysis step includes determining the nucleic composition of the oligonucleotide by known techniques, for example using: an electropherogram; radiolabelling; single-molecule real-time sequencing (Pacific Biosciences, SMRT, Oxford
25 Nannpore); ion semiconductor (Ion Torrent [Life Technologies]); pyrosequencing (454 Life Sciences [Rouche]); sequencing by synthesis (Illumina); sequencing by ligation (SOLiD Systems[Applied Biosystems]); dideoxy chain termination (Sanger sequencing).

According to a third aspect the invention provides a method for determining the integrity
30 of DNA stored on a storage medium, for example a storage medium according to the first aspect, said method including the steps of: 1) performing an STR profile analysis on a portion of the DNA stored; or 2) using a reporter incorporated into the DNA sequence to

assess said integrity, for example where said reporter fluoresces, or stops fluorescing when said integrity is reduced.

The invention is further characterised by the claims.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a Oligonucleotide sequenced as deposited on an FTA card-
(sequence listing 1);

10 Figure 2 shows a portion of the sequence of Figure 1 recovered from an FTA card-
(sequence listing 2);

Figure 3 shows another portion of the sequence of Figure 1 recovered from an FTA card-
(sequence listing 3);

Figure 4 shows a comparison between recovery using FTA cards and plain paper;

15 Figure 5 shows the results of a DNA integrity experiment using STR profiling; and

Figure 6 shows a micrographic image of a data storage medium according to one
embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

20

In one embodiment digital information is encoded into sequenced DNA according to
directions given in Goldman et al 2013. This group used a number of common computer
file formats. In this case, five files comprised all 154 of Shakespeare's sonnets (ASCII
text), a classic scientific paper, a medium-resolution colour photograph of the European

25 Bioinformatics Institute (JPEG), a 26-second excerpt from Martin Luther King's 1963 'I
have a dream' speech (MP3 format). The data comprising each file were represented as
DNA information sequences (see Figure 1) known as base pairs (bp) formed from the
known constituents of DNA namely Adenine (A), Cytosine (C), Guanine(G) and Thymine
(T) . Each DNA sequence was split into overlapping segments, generating fourfold

30 redundancy, and alternate segments were converted to their reverse complement. These
measures reduce the probability of data loss. Each segment was augmented with
indexing information that facilitated the identification of the file from which it originated

and its location within that file. In all, the five files were represented by a total of 153,335 segments of DNA information, each comprising 117 nucleotides bps. Oligonucleotides were synthesised corresponding to the designed DNA segments, creating 1.2×10^7 copies of each oligonucleotide. Errors occur independently during synthesise (1 error per 5 500 bases) in the different copies of each string. Thus reducing the method's overall error rate. After resuspension, amplification and purification, a sample of the resulting library was sequenced to recover the information. The remainder was sub-aliquoted and re-lyophilized for long-term storage typically at minus 20° Celsius. Sequencing results yielded 79.6×10^6 read-pairs of 104 bases in length, from which the authors reconstructed 10 full-length (117-nt) oligonucleotides *in silico*. Full-length DNA sequences representing the original encoded files were then reconstructed digitally using software looking for the indexing information. The resulting DNA sequences could be fully decoded without intervention. Inspection confirmed that the original computer files had been reconstructed with 100% accuracy.

15

Goldman transposed digital binary computer information into ternary (base 3), requiring just three characters, and then encoded that information into the DNA. This allows the four nucleotide bases (A,C,G, and T) to represent 0,1, and 2 sequentially with 1 20 redundant base, and thus avoid repetitive use of bases which can cause data recovery errors due to present day sequencing machine shortcomings. For example, at one base A=0,C=1 and G=2, if the next base would be repeated then, for the next base C=0,G=1 and T=2, and so on. This ensures that a sequence of identical digits in the data is not represented by a sequence of identical bases in the DNA, helping to avoid reading mistakes.

25

Alternatively, a quaternary (base 4) system is proposed here, with or without the shifting base representation system mentioned immediately above. Conveniently, quaternary codes of 4 characters (A,C,G and T) provide a binary 8 bit byte (octet). ASCII code requires 7 bits, and a parity bit can be added to each character to provide error detection.

30

In a quaternary based system, the following encoding may be used:

English language characters	G E										
ASCII 7 bit code plus <i>parity bit</i>	10001111 10001010										
Quaternary code	2033 2022										
DNA sequence [where A=0,C=1,G=2 and T=3 always]	GATT GAGG										
DNA sequence [where n+1 is used, according to the following rule:]	GTCA GTAT										
<table border="0"> <tr> <td>Previous bases</td> <td>Next bases</td> </tr> <tr> <td>A=0,C=1,G=2 and T=3</td> <td>A=1,C=2,G=3 and T=0</td> </tr> <tr> <td>A=1,C=2,G=3 and T=0</td> <td>A=2,C=3,G=0 and T=1</td> </tr> <tr> <td>A=2,C=3,G=0 and T=1</td> <td>A=3,C=0,G=1 and T=2</td> </tr> <tr> <td>A=3,C=0,G=1 and T=2</td> <td>A=0,C=1,G=2 and T=3</td> </tr> </table>	Previous bases	Next bases	A=0,C=1,G=2 and T=3	A=1,C=2,G=3 and T=0	A=1,C=2,G=3 and T=0	A=2,C=3,G=0 and T=1	A=2,C=3,G=0 and T=1	A=3,C=0,G=1 and T=2	A=3,C=0,G=1 and T=2	A=0,C=1,G=2 and T=3	
Previous bases	Next bases										
A=0,C=1,G=2 and T=3	A=1,C=2,G=3 and T=0										
A=1,C=2,G=3 and T=0	A=2,C=3,G=0 and T=1										
A=2,C=3,G=0 and T=1	A=3,C=0,G=1 and T=2										
A=3,C=0,G=1 and T=2	A=0,C=1,G=2 and T=3										

The n+1 quaternary system used above increases data density over the ternary system used by Goldman, but still reduces the chances of strings of repeating bases which are difficult to read. The use of a quaternary system is not tied to the use of ASCII characters, and so any binary code can be transposed into base pairs according to the steps in the table above.

10 Recovering the data is possible by reversing the encoding shown in the table.

The DNA data storage medium of the present invention is based on a product known as FTA ® as supplied by GE Healthcare Inc, and variants of that product, for example FTA Elute, the compositions of which are given in more detail below, and which are herein referred to generally as FTA or FTA cards or FTA paper.

FTA includes stabilising reagent mix which comprises a combination of a weak base, and a chelating agent, optionally, uric acid or a urate salt, and optionally an anionic surfactant.

20

The "weak base" of the combination may be a Lewis base which has a pH of about 6 to 10, preferably about pH 8 to 9.5. One function of the weak base is to act as a buffer to maintain a composition pH of about 6 to 10, preferably about pH 8.0 to 9.5, for example, pH 8.6. Hence, a weak base suitable for the composition of the invention

5 may, in conjunction with other components of the composition, provide a composition pH of 6 to 10, preferably, about pH 8.0 to 9.5. Suitable weak bases according to the invention include organic and inorganic bases. Suitable inorganic weak bases include, for example, an alkali metal carbonate, bicarbonate, phosphate or borate (e.g., sodium, lithium, or potassium carbonate). Suitable organic weak bases include, for example, tris-

10 hydroxymethyl amino methane (Tris), ethanolamine, tri- ethanolamine and glycine and alkaline salts of organic acids (e.g. trisodium citrate). A preferred organic weak base is a weak monovalent organic base, for example, Tris. The Tris may be either a free base or a salt, for example, a carbonate salt.

15 A preferred chelating agent is a strong chelating agent. By "strong" chelating agent it is meant that the agent binds multivalent metal ions with a comparable or better affinity than ethylene diamine tetraacetic acid (EDTA). A preferred chelating agent according to the invention is EDTA.

20 Anionic surfactants are examples of surfactants which are useful in the present invention. A preferred anionic surfactant is a strong anionic detergent. As used herein, a "strong" anionic detergent includes a hydrocarbon moiety, aliphatic or aromatic, containing one or more anionic groups. Particularly preferred anionic detergents suitable for the invention include sodium dodecyl sulphate (SDS) and sodium lauryl

25 sarcosinate (SLS). In a preferred embodiment, the anionic detergent causes inactivation of most microorganisms which have protein or lipids in their outer membranes or capsids, for example, fungi, bacteria or viruses. This includes microorganisms which may be pathogenic to humans and are present in a biological sample. The detergent also functions to denature proteins such as DNAses thereby affording additional protection to

30 the applied DNA. The storage medium has a visual delineation (to the human eye or an automated device such as a camera and controller) around an area holding DNA data so that it is possible in future to view where the DNA was deposited. In this case the

delineation is a printed circle into which the DNA is deposited, but an indicating dye can also provide a delineation.

Alternatively, a product called FTA Ellute could be used as a storage medium, in which
5 case the stabilising reagent may comprise a single reagent, for example a chaotropic substance such as a chaotropic salt, for example guanidinium thiocyanate.

In order to demonstrate that a viable long term storage solution could be achieved a data storage medium in the form of a solid support matrix including a stabilising reagent or
10 reagents in a dry form, for use as a support for oligonucleotide sequences encoded with data, 2 experiments were performed:

Example 1

Sequencing of DNA fragments applied to FTA cards

15 Samples of plasmid DNA designated pUC19-p53 were spotted (stock concentration of 0.5 µg/ml in TE buffer) at a 1 ng mass onto FTA sample collection cards (GE Healthcare). These cards provide stabilising reagents according to the invention. This plasmid contains a 1038 base pairs of a gene named p53 mRNA co-corresponding to nucleotide sequence from 217 to 1255 of the p53 cDNA as described in GenBank accession number
20 BC003596 (IMAGE; 3544714, MGC; 646). This fragment was cloned into the SmaI site of the plasmid pUC19. The DNA sequence of the pUC-p53 DNA prior to application to FTA is shown in Figure 1. After application of the plasmid, the FTA cards were left to air dry for 60 min before being placed into multi-barrier pouches and stored at room temperature for at least a week (as recommended by the manufacturer).

25

DNA was extracted from the sample collection matrices using an Illustra™ tissue and cells genomic prep Mini Spin Kit (GE Healthcare) according to manufacturer's recommendations with modifications as outlined below.

1. The 3.0 mm marked disk containing the applied DNA sample was removed using a
30 Uni-Core punch and placed into a 1.5 ml sterile tube.
2. Sterile phosphate buffered saline (PBS) (1 ml) was added to each tube.

3. Tubes were centrifuged at 12 000 rpm for 2 min (All centrifugations were conducted in a Heraeus Biofuge™).
4. PBS was discarded and 50 µl of fresh sterile PBS was added to each tube.
5. The 3.0 mm punches were homogenized using a 20-gauge sterile needle.
- 5 6. Tubes were centrifuged at 5000 rpm for 10 sec.
7. DNA was isolated from the supernatant according to the manufacturer's instructions for the Illustra tissue and cells genomic prep Mini Spin Kit.

To generate DNA for sequencing two PCR products were generated using End-Point
10 PCR (see Figures 2 and 3 for details). This was carried out as follows: rTaq2 DNA Polymerase (GE Healthcare) was used together with the following primer sequences to target the p53 insert in the plasmid DNA sample extracted from the respective paper matrices.

- 15 The 963 base pair PCR product 1 (Figure 2) was generated using the following Primers (Sigma Genosys):

Primer 1 (M13 forward [-20]); 5'-GTAAAACGAACGGCCAGT-3'

Primer 2 (M13 reverse); 5'-CAGGAAACAGCTATGAC-3'

- 20 The 250 base pair PCR product 2 (Figure 3) was generated using the following Primers (Sigma Genosys):

Primer 3; (forward) 5'-GCGCACAGAGGAAGAGAATC-3'

Primer 4; (reverse) 5'-CCAAGGCCTCATTGAGCTCT-3'

- 25 A direct/punch-in PCR reaction was also performed, as an alternative procedure, in which a 1.2 mm diameter punch was used to excise a FTA disc from the card and this was placed directly into a PCR reaction supplemented with α -cyclodextrin as described in the UK patent application GB 1219137.5. The two PCR products were generated using the primers as described above.

30

PCR amplicons were generated using a PCR master mix (95 µl; 1 unit of enzyme, 20 pmol forward and reverse primers) and 5 µl sample (plasmid DNA extracted from the FTA

cards using the Illustra Kit) were added to appropriate wells of a 96-well plate and amplified accordingly (denature 94°C, 3 min. 30 cycles of 94°C, 30 sec; 55°C, 1 min; 72°C, 2 min; final soak at 72°C, 10 min). For direct amplification from the FTA card the 5 µl DNA sample was replaced with the appropriate volume of buffer and a FTA disc.

5

In order to check the integrity of the amplified DNA, 10 µl volume of each PCR was loaded on a 1% agarose/ TAE buffer gel (Tris, acetic acid and Ethylenediaminetetraacetic acid (EDTA)) along with DNA size standards and positive and negative (no template) PCR controls. To generate DNA sequencing reactions the PCR products and primers
 10 described above were used following the BigDye® Terminator v3.1 Cycle Sequencing Kit manufacturer’s recommendations.

To remove excess PCR primers, 2 µl of ExoproSTAR (GE Healthcare) was added to 5 µl of amplified DNA in a test tube. The test tube was sealed and spun briefly to ensure its
 15 contents were at the bottom of the tube. The tube was then incubated on a PTC200 thermocycler for 15 minutes at 37°C, followed by 15 minutes at 80°C and a hold at 4°C. The final volume of each reaction was made up to 10 µl by the addition of sterile water prior to the cycle sequencing reaction. Stock cycle sequencing mix was made up as follows:

20

Table 1. Sequencing mix

Constituent	1x	10x	15x
Big Dye Mix	4 µl	40 µl	60 µl
Primer(1.6pmol/ µl)	2 µl	20 µl	30 µl
5x Buffer	4 µl	40 µl	60 µl

Ten microliters of sequencing mix was added to each 10 µl PCR product from above. These samples were then amplified using both forward and reverse sequencing primers
 25 using the following protocol on a MJ Research PTC200 thermocycler: 96°C 30 seconds; 50°C 15 seconds; 60°C 4 minutes for 25 cycles followed by a 4°C hold on completion of the cycling. The amplified sequencing products were then processed using the Auto Seq G-50 columns: Columns were vortexed briefly, the caps loosened and the seals removed.

The columns were then placed into collection tubes. The tubes were spun for 1 minute at 2000g in an Eppendorf 5417C centrifuge and the flow through discarded. The columns were then transferred to fresh tubes and sequencing sample slowly applied to the top of the column. The tubes were spun for 1 minute at 2000 x g, the eluates were collected and
5 stored at 2-8°C in labelled 0.2ml PCR tubes prior to transfer to a sequencing facility for analysis.

Each product was sequenced six times. The consensus sequences of both PCR products (product 1 and 2) were compiled from all sequencing reactions performed. The
10 sequencing results of the amplified PCR pUC-p53 products are shown in Figures 2 and 3. Analysis of both consensus sequences indicates a complete homology with that described in the relevant regions of sequence described in Figure 1.

Figure 1 shows the sequence in pUC19 containing a fragment of the p53 cDNA as it was
15 deposited on the FTA support matrix which includes the stabilising reagents. Sequences in italics and bold refer to pUC19- and p53- specific sequences respectively. Sequences underline refers to PCR and DNA sequencing primers used in this study.

20 Figure 2 shows the recovered PCR product 1 sequence, and so demonstrates that recovery of DNA sequences of at least 963 base pairs which equates to 1900 bytes (about 2Kb) of digital information, is possible on FTA cards. Thus relatively long DNA sequences are stable in this storage format.

25 Figure 3 shows the recovered PCR product 2 sequence, and so demonstrates that recovery of DNA sequences of 250 base pairs which equates to 500 bytes (about 0.5Kb) of digital information, is possible on FTA cards.

Example 2

30 Quantitative PCR (qPCR) measurement of DNA recovered from FTA cards

qPCR was employed to quantify the amount of amplifiable DNA on each 1.2 mm diameter disks punched from FTA cards compared with a plain paper support (Whatman

product code: 31-ETF) control. The real-time PCR method targeted a 500 bp region of Lambda DNA previously applied to FTA cards and dried. The technique utilized the intercalating dye GelStar™ Nucleic Acid Stain (Lonza Group Ltd) to reveal the amplified product. This method provided the threshold cycle (Ct) values and mass amounts of
5 amplifiable DNA when read from the standard curve. The PCR master mix contained the following components:

- Fidelity™ DNA Polymerase (GE Healthcare);
- Lambda DNA (Promega Corp);
- GelStar Nucleic Acid Stain, used at 1:1000 dilution in a Tris/EDTA (TE) buffer; and
- 10 • Primers (Sigma Genosys):
 - Primer 1 forward 5'- GGTTATCGAAATCAGCCACAGCGCC-3'
 - Primer 2 reverse 5'- GATGAGTTCGTGTCCGTACAACACTGG-3'

A PCR master mix (20 µl; containing 2.5 U enzyme, 5 pmol forward and reverse primers,
15 and 1.5 µl GelStar Nucleic Acid Stain) and 5 µl sample or standard DNA dilution series, were added to wells of a 96-well plate and amplified on an Applied Biosystems 7900HT Fast Real-Time PCR System according to the specified cycling conditions (denature 95°C, 3 min. 40 cycles of 95°C, 30 sec; 55°C, 30 sec; 72°C, 1 min).

20 Data were generated using qPCR to reveal the amounts of amplifiable DNA that could be recovered from DNA-spotted FTA cards. The standard curve was used to accurately determine the amount of amplifiable DNA that could be recovered from FTA cards. Amounts of DNA recovered were compared with those derived from uncoated paper matrix 31-ETF (Figure 4).

25

Figure 4 shows DNA Recovery from FTA Cards as determined by quantitative PCR. DNA was applied to the matrices and recovered. The quantity of DNA recovered from FTA was compared with that recovered from uncoated paper. This graph demonstrates that
30 recovery of DNA data from FTA is superior to recovery from a support which has no stabilising reagents added.

To enhance the usefulness of the invention the inventors propose means for monitoring the integrity of the DNA for example visually on the solid support, or by minor processing of a portion of the storage medium (to leave the bulk of the replicated data on the storage medium).

5

In one embodiment, the data integrity monitoring technique relies on specific reporter groups attached or incorporated into the synthesized DNA either during or post synthesis. The reporters will generate an appropriate signal to indicate DNA integrity e.g. if the DNA remains intact or has become degraded. Such reporter groups could
10 potentially be based upon a range of molecules including florescent cye-dyes, quenchers and emitters, enzymatic reporters etc. If florescent dyes are used, then simple illumination of the storage medium will indicate the integrity of the data stored thereon. Alternatively the data integrity monitoring technique may be performed by removing a portion of the storage medium and performing short tandem repeat (STR) profiling
15 reaction that will interrogate 20 or so different STR loci generated as control DNA sequences within the synthesised DNA data. Such interrogation should report the correct profile, if the integrity of the DNA data is intact. Either system will potentially facilitate a relatively simple method in which the integrity of the DNA can be monitored during/after storage.

20

Example 3

Short Tandem Repeat Profiling as a data integrity check

In order to further demonstrate the integrity of the DNA applied to the FTA following elution, short tandem repeat (STR) profiling was carried out using punches taken from
25 dried blood spots on FTA Micro Cards. Dried blood spots for sample collection purposes were prepared as follows. Human whole blood, from 2 donors, (75 µl) was spotted on to FTA Micro Cards, allowed to dry at ambient temperature for 3 h, and stored in a desiccator before use. FTA cards were sampled within one week of spotting. FTA cards were punched at the centre of the sample spot in order to reduce variability.

30

The resultant punches were added directly to the STR profiling reagents and direct STR profiling was carried out on the punches using a PowerPlex 18D System (Promega,

Southampton, UK) over 26 amplification cycles, following the kit manufacturer's instructions. Thermal cycling conditions were as follows: 96°C for 2 min., then: 94°C for 10 sec, 60°C for 1 min for 26 cycles, then: 60°C for 20 min, followed by a 4°C rest.

- 5 The resulting PCR products were analysed on an ABI™ 3130xl Genetic Analyzer capillary electrophoresis system. The electrophoresis run cocktail mix was prepared using 1.0 µl of CC5 (Promega) internal lane standard plus 10.0 µl HiDi formamide per sample. Eleven microliters of run cocktail mix was added to 1.0 µl of sample. Instrument settings were as follows: injection time, 5 seconds; injection and run voltages, 3kV; run
10 time, 1,500 seconds with a G5 dye set. Results were evaluated with GeneMapper™ v3.2 software (Life Technologies, Paisley, UK).

Figure 5 shows an electropherogram prepared in GeneMapper software following DNA amplification from dried blood spots on FTA Micro Cards. Punches were prepared a 1.2
15 mm manual micro-punch. STR analysis was carried out using the PowerPlex 18D kits over 26 amplification cycles. The resulting PCR amplicons were analysed using an Applied Biosystems 3130xl Genetic Analyser and GeneMapper v3.2 software. The ordinate shows relative fluorescence units (RFU) ranging from 2000-8000, with 8000 RFU showing full scale deflection. The abscissa shows base pairs. The figures below the
20 peaks indicate the number of short tandem repeats associated with each allele. Periodical STR profiling reactions can be performed on the synthesised DNA applied to FTA cards and a quality measure such as peak heights etc. can be used to determine the quality and integrity of the applied synthetic DNA. The use of small 1.2 mm punches will potentially represent a preservation of DNA resource stored on FTA for long periods

25

It can be concluded that using a STR profiling is a viable way to determine the integrity of DNA data stored on FTA cards.

Where reporter groups are used, these reporters may provide for visual (manual or
30 automated) inspection of the solid support and the generation of a signal that indicates that the DNA remains intact to facilitate sequencing & hence data recovery. The following commercially-available florescent dyes that are known to interact with DNA and have

been used in fluorescence microscopy; PicoGreen, SYBR Green, Acridine Orange, Bisbenzimidazole trihydrochloride, HOE-33342 . Hoechst 33258, DAPI, Hoechst, Propidium Iodide etc. Visible dyes DNA binding dyes include; methylene blue, thionine etc Chen et al 2009 Nano Today, 4, 125--134 describes a FRET assay to monitor DNA stability.

- 5 Plasmid DNA was double-labeled with QD (525 nm emission) and specific nucleic acid dyes complexed with Cy5. The QD donor drives energy transfer stepwise through the intermediate nucleic acid dye to the final acceptor Cy5. Incorporating reagent into the DNA polymer could also be used.
- 10 The combination of data in the form of DNA on a storage medium according to the invention, coupled with a simple system to measure its integrity is a viable solution to archiving complex digital information.

Figure 6 shows FTA paper having relatively large fibres of cellulose material i.e. those thick elements running generally from top to bottom of the picture, intertwined with DNA strands, akin to a spider's web. This image illustrates the way that the DNA is held to the FTA in an intertwined way and thereby stabilised.

The experiments above show that FTA is a preferred data storage medium, although previous experiments have concluded that a solid matrix in the form of a cellulose fibre paper matrix, which has no stabilising reagent added, is capable of storing human DNA for many years- 27 years in the reported case. So, although desirable, it is not essential to include a stabilising reagent in or to the solid matrix.

25 Although fibrous solid matrix have been used in experiments- in this case a cellulose based paper product called FTA, it will be apparent to the skilled addressee that any material or surface which is man made and has a surface or subsurface having the ability to hold oligonucleotides in a similar manner to that shown in Figure 6 could be employed as a solid matrix support for this invention. For example, woven or non-woven fibrous materials, including man made, or naturally occurring polymer fibres, mineral fibre based materials such as glass fibre materials, surface treated solid materials for example, chemically or mechanically treated materials, including laser etched surfaces, all provided

30

with a surface micro roughness of sufficient roughness to hold DNA, porous materials such as porous polymers which have pores large enough to accommodate and hold the DNA stands, in a similar manner to that shown in Figure 6. In each case the surface or subsurface of the material accommodates DNA, but will yield the DNA when required, and thereby is also capable of accepting a stabilising reagent if used in dry form or in wet form for subsequent drying.

Recovering DNA from the storage medium has been carried out in the experiments described above, using known techniques which all involve liquid based replication of DNA by PCR, although other replication is possible, for example isothermal amplification, e.g. helicase dependant amplification, or for example rolling circle amplification. However, providing sufficient DNA can be recovered from the storage medium, that liquid based replication is not essential.

Additionally, removal of a portion of the storage medium to recover data is not essential. It is possible to elute a portion of the DNA data off the storage medium without removing a portion of the storage medium itself. In a further example DNA can be removed from the storage medium by abrading the storage medium which will itself hold multiple copies of the DNA data, and analysing directly the material abraded from the storage medium. The DNA data could even be analysed whilst remaining on the storage medium, for example by spectroscopic, or molecular analysis techniques.

Although embodiments only have been illustrated, it will be apparent to the skilled addressee that further modifications, variants, additions and omissions are possible within the scope and spirit of the invention defined herein.

Claims

1. A data storage medium comprising a solid support matrix for use as a dry storage
5 support for oligonucleotide sequences encoded with data.
2. A data storage medium as claimed in claim 1 wherein said matrix includes a stabilising reagent or reagent mix in a dry form.
- 10 3. A data storage medium as claimed in claims 1 or 2, wherein the matrix is fibrous (for example cellulose, or glass, fibres) formed into a support of sufficient strength to hold the oligonucleotide sequences.
4. A data storage medium as claimed in claims 2 or 3, wherein the stabilising reagent
15 or reagent mix is a reagent mix and comprises a combination of a weak base, and a chelating agent, optionally, uric acid or a urate salt, and optionally an anionic surfactant.
5. A data storage medium as claimed in claim 4, wherein the weak base is a Lewis base which has a pH of about 6 to 10, preferably about pH 8 to 9.5.
20
6. A data storage medium as claimed in claims 4 or 5, wherein the weak base is an organic and inorganic base, and if inorganic then optionally includes an alkali metal carbonate, bicarbonate, phosphate or borate (e.g. sodium, lithium, or potassium carbonate), and if organic then optionally includes, tris-hydroxymethyl amino methane
25 (Tris), ethanolamine, tri-ethanolamine and glycine and alkaline salts of organic acids (e.g. trisodium citrate).
7. A data storage medium as claimed in claims 4 or 5, wherein the weak base is Tris present either as a free base or as a salt, for example, a carbonate salt.
- 30 8. A data storage medium as claimed in any one of claims 4 to 7, wherein the chelating agent binds multivalent metal ions with a comparable or better affinity than ethylene diamine tetraacetic acid (EDTA), and is preferably EDTA.

9. A data storage medium as claimed in any one of claims 4 to 8, wherein the anionic surfactant includes a hydrocarbon moiety, aliphatic or aromatic, containing one or more anionic groups.

5

10. A data storage medium as claimed in claim 9, wherein the anionic surfactants is a detergent, for example sodium dodecyl sulphate (SDS) and/or sodium lauryl sarcosinate (SLS).

10 11. A data storage medium as claimed in claim 2, wherein the stabilising reagent comprises a chaotropic substance such as a chaotropic salt, for example guanidinium thiocyanate.

12. A data storage medium as claimed in any one of the preceding claims, wherein the
15 storage medium includes a visual delineation around an area holding DNA data for example a printed boundary showing where the DNA is stored, or an indicating dye such as chlorophenol red or phenol red, which discolour, thereby showing where the DNA is stored.

13. A method of storing data, said method comprising the following steps:

20

a) producing synthesised oligonucleotides having an nucleotide sequence representative of data to be stored;

b) optionally replicating said oligonucleotides;

c) depositing said oligonucleotides, optionally replicated, onto a storage medium
25 according to any one of claims 1 to 12.

14. A method of storing data as claimed in claim 13, wherein at least steps b) and c) are performed in a liquid, and said liquid is allowed or caused to evaporate from the storage medium.

30

15. A method of storing data as claimed in claim 13, wherein step a) is performed according instructions published in the Goldman paper referred to herein.

16. A method of storing data as claimed in claim 13, wherein step a) includes converting binary data into a quaternary code, and sequencing DNA nucleotides, according to the quaternary code to provide said oligonucleotides.

5

17. A method of storing data as claimed in claim 16, wherein said conversion follows the n+1 rule described herein.

18. A method of storing data as claimed in any one of claims 13 to 17, wherein step a) includes incorporating a reporter into said sequence, to provide an indication of data integrity, optionally said reporter providing an increase or decrease in detectability in response to a change in said integrity.

19. A method of storing data as claimed in claim 18, wherein said reporter is a fluorescent dye.

20. A method of storing data as claimed in any one of claims 13 to 19, wherein said nucleotide sequence is at least 250 base pairs in length, and preferably at least about 1000 base pairs, for example at least about 963 base pairs.

20

21. A method of storing data as claimed in any one of claims 13 to 20, wherein said replication is by means of a polymerase chain reaction (PCR), isothermal amplification, or rolling circle amplification.

22. A method of storing data as claimed in claim any one of claims 13 to 21, wherein one or more sequences which aid replication are added to the end or ends of the sequenced oligonucleotides.

23. A method for recovering stored data from the storage medium of any one of claims 1 to 10 when stored according to the method of any one of claims 13 to 22, said recovery method comprising the steps of:

- i) extracting at least a portion of the replicated oligonucleotides from the storage medium;
- ii) optionally performing further replication of said oligonucleotides; and
- iii) performing an analysis of the oligonucleotides to determine their nucleotide sequence, and decoding said sequence to thereby recover said stored data.

5

24. A method for recovering stored data as claimed in claim 23, wherein said extracting step includes removing at least a portion of the storage medium or the DNA stored thereon, and introducing into a vessel the at least a portion, or DNA, and then performing said further replication step.

10

25. A method as claimed in claim 24, wherein said further replication step includes: adding a liquid to said at least a portion.

26. A method for recovering stored data from the storage medium, as claimed in claim 15 23, wherein the analysis step includes determining the nucleic composition of the oligonucleotide by known techniques, for example using: an electropherogram; radiolabelling; single-molecule real-time sequencing (Pacific Biosciences, SMRT, Oxford Nanopore); ion semiconductor (Ion Torrent [Life Technologies]); pyrosequencing (454 Life Sciences [Rouche]); sequencing by synthesis (Illumina); sequencing by ligation 20 (SOLiD Systems[Applied Biosystems]); dideoxy chain termination (Sanger sequencing).

27. A method for determining the integrity of DNA stored on a storage medium according to any one of claims 1 to 12, said method including the steps of:

- 1) performing an STR profile analysis on a portion of the DNA stored according to any 25 one of claims 13 to 22; or
- 2) using a reporter incorporated into the DNA sequence to assess said integrity, for example where said reporter fluoresces, or stops fluorescing when said integrity is reduced.

1/4

GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA
 CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT
 CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT
 TGTGAGCGGA TAACAATTT ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTTGC
 ATGCCTGCAG GTCGACTCTA GAGGATCCCC TCCGTCTGGG CTTCTTGCAT TCTGGGACAG
 CCAAGTCTGT GACTTGCACG TACTCCCCTG CCCTCAACAA GATGTTTTGC CAACTGGCCA
 AGACCTGCCC TGTGCAGCTG TGGGTTGATT CCACACCCCG GCCCGGCACC CGCGTCCGCG
 CCATGGCCAT CTACAAGCAG TCACAGCACA TGACGGAGGT TGTGAGGCGC TGCCCCACC
 ATGAGCGCTG CTCAGATAGC GATGGTCTGG CCCCTCCTCA GCATCTTATC CGAGTGAAG
 GAAATTTGCG TGTGGAGTAT TTGGATGACA GAAACACTTT TCGACATAGT GTGGTGGTGC
 CCTATGAGCC GCCTGAGGTT GGCTCTGACT GTACCACCAT CCACTACAAC TACATGTGTA
 ACAGTTCCTG CATGGGCGGC ATGAACCGGA GGCCATCCT CACCATCATC ACACTGGAAG
 ACTCCAGTGG TAATCTACTG GGACGGAACA GCTTTGAGGT GCGTGTGTTGT GCCTGTCTCTG
 GGAGAGACCG GCGCACAGAG GAAGAGAATC TCCGCAAGAA AGGGGAGCCT CACCACGAGC
 TGCCCCCAGG GAGCACTAAG CGAGCACTGC CCAACAACAC CAGCTCCTCT CCCCAGCCAA
 AGAAGAAACC ACTGGATGGA GAATATTTCA CCCTTCAGAT CCGTGGGCGT GAGCGCTTCG
 AGATGTTCCG AGAGCTGAAT GAGGCCTTGG AACCTCAAGGA TGCCCAGGCT GGGAAGGAGC
 CAGGGGGGAG CAGGGCTCAC TCCAGCCACC TGAAGTCCAA AAAGGGTCAG TCTACCTCCC
GGGTACCGAG CTCGAATTCA CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCCTG
 GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
 AAGAGGCCCG CACCGATCGC CTTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCGCC
 TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTT ACACCGCATA TGGTGCACCTC
 TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC CCGACACCCG CCAACACCCG
 CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG
 TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGACGAA
 AGGGCCTCGT GATACGCCTA TTTTTATAGG TTAATGTCAT GATAATAATG GTTTCTTAGA
 CGTCAGGTGG CACTTTTCGG GGAAATGTGC GCGGAACCCC TATTTGTTTA TTTTTCTAAA
 TACATTCAA TATGTATCCG CTCATGAGAC AATAACCCTG ATAAATGCTT CAATAATATT
 GAAAAAGGAA GAGTATGAGT ATTCAACATT TCCGTGTCGC CCTTATTCCC TTTTTTGCGG
 CATTTTGCCCT TCCTGTTTTT GCTCACCCAG AAACGCTGGT GAAAGTAAAA GATGCTGAAG
 ATCAGTTGGG TGCACGAGTG GGTTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG
 AGAGTTTTCG CCCCAGAGAA CGTTTTCCAA TGATGAGCAC TTTTAAAGTT CTGCTATGTG
 GCGCGGTATT ATCCCGTATT GACGCCGGG AAGAGCAACT CGGTCGCCG ATACACTATT
 CTCAGAATGA CTTGGTTGAG TACTCACCAG TCACAGAAAA GCATCTTACG GATGGCATGA
 CAGTAAGAGA ATTATGCAGT GCTGCCATAA CCATGAGTGA TAACACTGCG GCCAACTTAC
 TTCTGACAAC GATCGGAGGA CCGAAGGAGC TAACCGCTTT TTTGCACAAC ATGGGGGATC
 ATGTAACCTG CTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA AACGACGAGC
 GTGACACCAC GATGCCTGTA GCAATGGCAA CAACGTTGCG CAAACTATTA ACTGGCGAAC
 TACTTACTCT AGCTTCCCGG CAACAATTAA TAGACTGGAT GGAGGCGGAT AAAGTTGCAG
 GACCACTTCT GCGCTCGGCC CTTCCGGCTG GCTGGTTTAT TGCTGATAAA TCTGGAGCCG
 GTGAGCGTGG GTCTCGCGGT ATCATTGCAG CACTGGGGCC AGATGGTAAG CCCTCCCGTA

FIG. 1

2/4

TCGTAGTTAT CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT AGACAGATCG
CTGAGATAGG TGCCTCACTG ATTAAGCATT GGTAAGTGTG AGACCAAGTT TACTCATATA
TACTTTAGAT TGATTTAAAA CTTCATTTTT AATTTAAAAAG GATCTAGGTG AAGATCCTTT
TTGATAATCT CATGACCAAA ATCCCTTAAC GTGAGTTTTT GTTCCACTGA GCGTCAGACC
CCGTAGAAAA GATCAAAGGA TCTTCTTGAG ATCCTTTTTT TCTGCGCGTA ATCTGCTGCT
TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGTTT GCCGGATCAA GAGCTACCAA
CTCTTTTTCC GAAGGTAACT GGCTTCAGCA GAGCGCAGAT ACCAAATACT GTTCTTCTAG
TGTAGCCGTA GTTAGGCCAC CACTTCAAGA ACTCTGTAGC ACCGCCTACA TACCTCGCTC
TGCTAATCCT GTTACCAGTG GCTGCTGCCA GTGGCGATAA GTCGTGTCTT ACCGGGTTGG
ACTCAAGACG ATAGTTACCG GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA
CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG CGTGAGCTAT
GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG GTATCCGGTA AGCGGCAGGG
TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGGAAA CGCCTGGTAT CTTTATAGTC
CTGTGCGGGT TCGCCACCTC TGACTTGAGC GTCGATTTTT GTGATGCTCG TCAGGGGGGC
GGAGCCTATG GAAAAACGCC AGCAACGCGG CCTTTTTACG GTTCTGGCC TTTTGCTGGC
CTTTTGCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG
CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC GAGTCAGTGA
GCGAGGAAGC GGAAGA

FIG. 1 (continued)

3/4

CAGGAAACAG CTATGACCAT GATTACGCCA AGCTTGCATG CCTGCAGGTC GACTCTAGAG
GATCCCTCC GTCTGGGCTT CTTGCATTCT GGGACAGCCA AGTCTGTGAC TTGCACGTAC
TCCCCTGCCC TCAACAAGAT GTTTTGCCAA CTGGCCAAGA CCTGCCCTGT GCAGCTGTGG
GTTGATTCCA CACCCCGCC CGGCACCCGC GTCCGCGCCA TGGCCATCTA CAAGCAGTCA
CAGCACATGA CGGAGGTTGT GAGGCGCTGC CCCACCATG AGCGCTGCTC AGATAGCGAT
GGTCTGGCCC CTCCTCAGCA TCTTATCCGA GTGGAAGGAA ATTTGCGTGT GGAGTATTTG
GATGACAGAA ACACTTTTCG ACATAGTGTG GTGGTGCCCT ATGAGCCGCC TGAGGTTGGC
TCTGACTGTA CCACCATCCA CTACAACCTAC ATGTGTAACA GTTCTGTCAT GGGCGGCATG
AACCGGAGGC CCATCCTCAC CATCATCACA CTGGAAGACT CCAGTGGTAA TCTACTGGGA
CGGAACAGCT TTGAGGTGCG TGTTTGTGCC TGTCCTGGGA GAGACCGGCG CACAGAGGAA
GAGAATCTCC GCAAGAAAGG GGAGCCTCAC CACGAGCTGC CCCCAGGGAG CACTAAGCGA
GCACTGCCCA ACAACACCAG CTCCTCTCCC CAGCCAAAGA AGAAACCACT GGATGGAGAA
TATTTACCCC TTCAGATCCG TGGGCGTGAG CGCTTCGAGA TGTTCCGAGA GCTGAATGAG
GCCTTGGAAC TCAAGGATGC CCAGGCTGGG AAGGAGCCAG GGGGGAGCAG GGCTCACTCC
AGCCACCTGA AGTCCAAAAA GGTCACTCT ACCTCCGGG TACCGAGCTC GAATTCACTG
GCCGTCGTTT TAC

FIG. 2

GCGCACAGAG GAAGAGAATC TCCGCAAGAA AGGGGAGCCT CACCACGAGC TGCCCCCAGG
GAGCACTAAG CGAGCACTGC CCAACAACAC CAGCTCCTCT CCCCAGCCAA AGAAGAAACC
ACTGGATGGA GAATATTTCA CCCTTCAGAT CCGTGGGCGT GAGCGCTTCG AGATGTTCCG
AGAGCTGAAT GAGGCCTTGG

FIG. 3

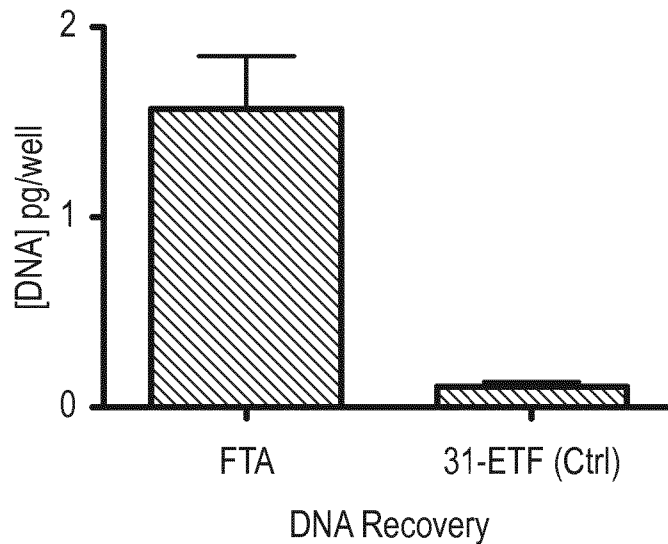


FIG. 4

4/4

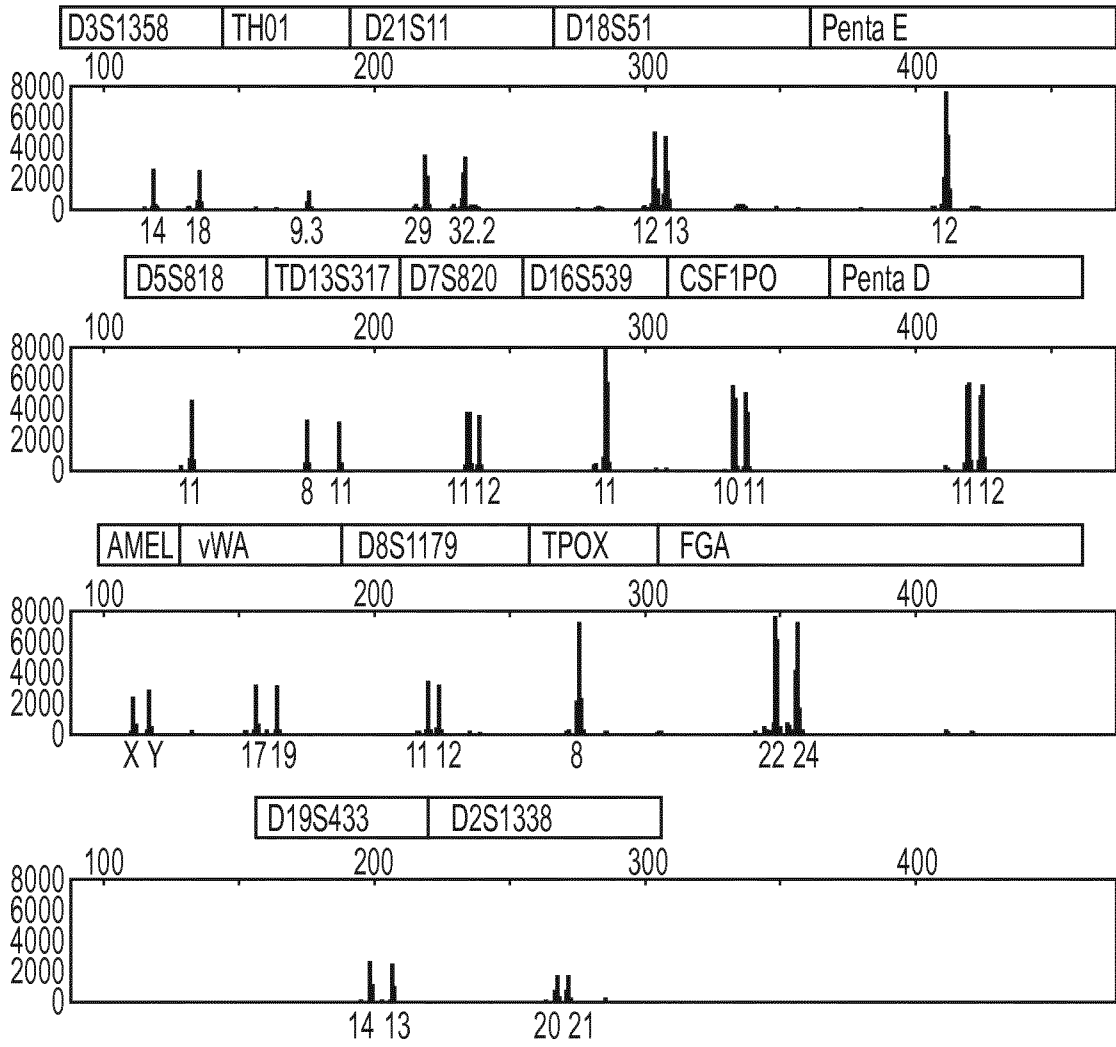


FIG. 5

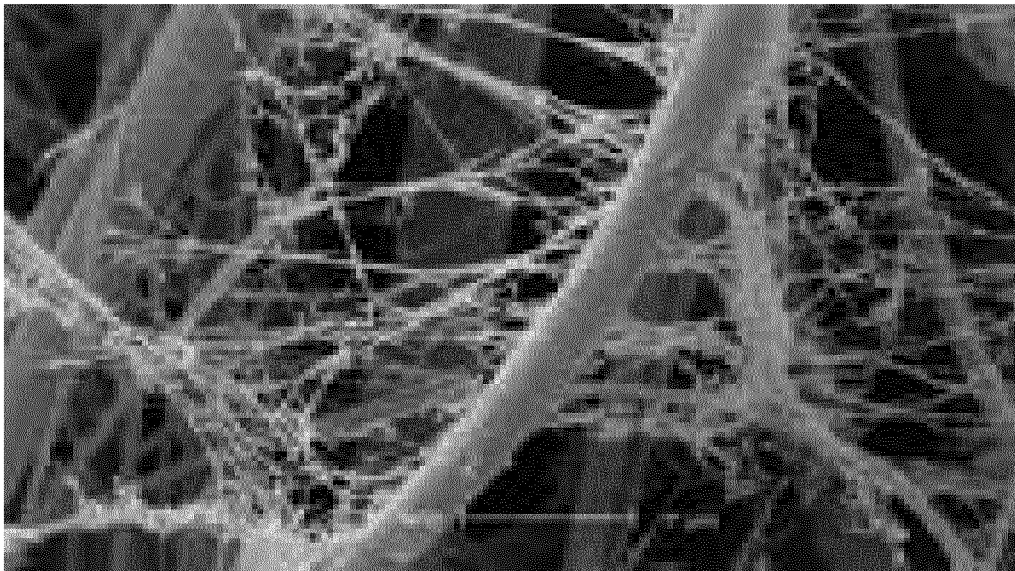


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/075708

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 G06F19/00
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
C12Q G06F
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96/39813 A1 (FLINDERS TECHNOLOGIES PTY LTD [AU]) 19 December 1996 (1996-12-19)	1-10
Y	in particular see claims 1-9, 16-17, 20, 22, 31-33, 39, 50 and pages 4-7, 9-10,12-13; the whole document	1-27
Y	WO 00/14505 A1 (FLINDERS TECHNOLOGIES PTY LTD [AU]; BURGOYNE LEIGH ALEXANDER [AU]) 16 March 2000 (2000-03-16) dry solid medium such as cellulose for storage of macromolecules such as DNA, see pages 7-11; the whole document	1-27
	----- -/--	

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>
---	---

Date of the actual completion of the international search 3 March 2015	Date of mailing of the international search report 12/03/2015
--	---

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Vix, Olivier
--	---

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/075708

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 90/03959 A1 (UNIV SOUTHERN AUSTRALIA [AU]) 19 April 1990 (1990-04-19) in particular see pages 3-6 and claims 1-21; the whole document	1-27

Y	WO 2013/178801 A2 (EUROP MOLECULAR BIOLOGY LAB [DE]) 5 December 2013 (2013-12-05) the whole document	1-27

Y	WO 03/025123 A2 (MOUNT SINAI SCHOOL OF MEDECINE [US]; BANCROFT CARTER [US]; CLELLAND CA) 27 March 2003 (2003-03-27) the whole document	1-27

Y	G. M. CHURCH ET AL: "Next-Generation Digital Information Storage in DNA", SCIENCE, vol. 337, no. 6102, 28 September 2012 (2012-09-28), pages 1628-1628, XP055082578, ISSN: 0036-8075, DOI: 10.1126/science.1226355 the whole document	1-27

Y	NICK GOLDMAN ET AL: "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA", NATURE, 1 January 2013 (2013-01-01), XP055050963, ISSN: 0028-0836, DOI: 10.1038/nature11875 the whole document	1-27

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/EP2014/075708

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9639813	A1	19-12-1996	AU 691114 B2 07-05-1998
		AU 5887696 A	30-12-1996
		CN 1192648 A	09-09-1998
		DE 69633239 D1	30-09-2004
		DE 69633239 T2	08-09-2005
		DE 96915910 T1	22-04-2004
		EP 0849992 A1	01-07-1998
		EP 1484111 A1	08-12-2004
		WO 9639813 A1	19-12-1996
WO 0014505	A1	16-03-2000	AU 5840099 A 27-03-2000
		CA 2352260 A1	16-03-2000
		EP 1125104 A1	22-08-2001
		US 2001039010 A1	08-11-2001
		WO 0014505 A1	16-03-2000
WO 9003959	A1	19-04-1990	CA 2000192 A1 05-04-1990
		WO 9003959 A1	19-04-1990
WO 2013178801	A2	05-12-2013	AU 2013269536 A1 18-12-2014
		CA 2874540 A1	05-12-2013
		KR 20150016572 A	12-02-2015
		WO 2013178801 A2	05-12-2013
WO 03025123	A2	27-03-2003	NONE