SCIENTIFIC REPORTS

OPEN

PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis

Shiv Kumar[1], Chuanjuan Tao[1], Minchen Chien[1], Brittney Hellner[1], Arvind Balijepalli[2], Joseph W. F. Robertson[2], Zengmin Li[1], James J. Russo[1], Joseph E. Reiner[3], John J. Kasianowicz[2] & Jingyue Ju[1]

[1]Center for Genome Technology & Biomolecular Engineering; Departments of Chemical Engineering and Pharmacology, Columbia University, New York, NY 10027, [2]Physical Measurement Laboratory, Semiconductor & Dimensional Metrology Division; and CMOS Reliability and Advanced Devices Group, National Institute of Standards and Technology, Gaithersburg, MD 20899-1070, [3]Department of Physics, Virginia Commonwealth University, Richmond, VA 23284.

We describe a novel single molecule nanopore-based sequencing by synthesis (Nano-SBS) strategy that can accurately distinguish four bases by detecting 4 different sized tags released from 5′-phosphate-modified nucleotides. The basic principle is as follows. As each nucleotide is incorporated into the growing DNA strand during the polymerase reaction, its tag is released and enters a nanopore in release order. This produces a unique ionic current blockade signature due to the tag's distinct chemical structure, thereby determining DNA sequence electronically at single molecule level with single base resolution. As proof of principle, we attached four different length PEG-coumarin tags to the terminal phosphate of 2′-deoxyguanosine-5′-tetraphosphate. We demonstrate efficient, accurate incorporation of the nucleotide analogs during the polymerase reaction, and excellent discrimination among the four tags based on nanopore ionic currents. This approach coupled with polymerase attached to the nanopores in an array format should yield a single-molecule electronic Nano-SBS platform.

D NA sequencing is a fundamental technology in the biological and medical sciences. Recently, several analytical methods have been developed to detect DNA or RNA at the single molecule level using chemical or physical microscopic technologies[1–3]. In particular, ion channels have been shown to detect individual DNA or RNA strands, leading to the promise of high-speed sequencing and analysis of DNA[4–13].

In 1996, Kasianowicz et al.[4] first demonstrated that the α-hemolysin (αHL) channel could be used to detect nucleic acids at the single molecule level. The αHL channel has a 1.5 nm-diameter limiting aperture[14–17], and its voltage-dependent gating can be controlled, such that the pore remains open indefinitely[17], making it an ideal candidate for nanopore-based detection and discrimination. Individual single-stranded polyanionic nucleic acids are driven through the pore by the applied electric field, and the polynucleotides cause well-defined, transient reductions in the pore conductance[4,8,10,12]. Because the residence time of the polynucleotide in the pore is proportional to the RNA or DNA contour length, it was suggested that a nanopore may be able to sequence DNA in a ticker-tape fashion if the current signatures of the four bases can be discriminated from each other.[4] Towards the goal of sequencing with nanopore[4,13,18], in another approach, an αHL channel with a covalently linked adaptor in the pore was used to identify unlabeled nucleoside-5′-monophosphates one by one following exonuclease cleavage[19]. However, a complete exonuclease-nanopore system based on this concept to sequence DNA has so far not been demonstrated.

Despite the ability of nanopores to detect and characterize some physical properties of DNA at the single molecule level, the more demanding goal of accurate base-to-base sequencing by passing a single stranded DNA through the nanopore has not yet been realized. Oxford Nanopore Technologies recently announced the ability to accomplish strand sequencing in a nanopore at 3-base resolution with an error rate of 4%[20]. Another group reported single base resolution strand sequencing with a nanopore, but had difficulty correctly determining homopolymer sequences[21].

The native αHL channel has an inherent ability for high-resolution molecular discrimination. For example, it can discriminate between aqueous H^+ and D^+ ions[17], and Robertson et al.[22] have recently demonstrated that the
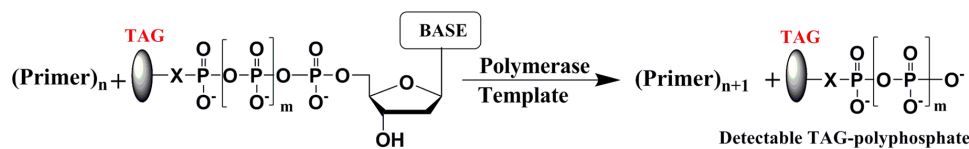
Figure 1 | Mechanism of primer extension and release of tag-polyphosphate for nanopore detection.

αHL channel can easily separate poly(ethylene glycol) (PEG) molecules at better than the single monomer level. In the latter study, a molecular mass or size spectrum estimated from the mean current caused by individual PEG molecules entering the pore easily resolves individual ethylene glycol repeat units. In addition, the mean residence time of the polymer in the pore increases with the PEG size[23]. Based on these previous investigations using nanopores to detect and distinguish molecules with different structures and the fact that DNA polymerase can recognize nucleotide analogs with extensive modification at the 5′-terminal phosphate group as efficient substrates[24-28], we propose a novel nanopore-based sequencing by synthesis (Nano-SBS) strategy that will accurately differentiate each of the four different sized tags attached to the 5′-phosphate of each nucleotide at the single molecule level for sequence determination. The basic principle of the Nano-SBS approach is described as follows. As each nucleotide is incorporated into the growing DNA strand during the polymerase reaction, its tag is released by phosphodiester bond formation (Fig. 1). The tags will enter a nanopore in the order of release, producing unique ionic current blockade signatures due to their distinct chemical structures, thereby determining DNA sequence electronically at single molecule level with single base resolution. We demonstrated that the 5′-terminal phosphate position of the nucleotide is unique in its ability to tolerate sizable modifications by large tags based on PEG molecules without affecting polymerase

recognition. This overcomes the inherent constraints imposed by the small differences among the 4 bases, a challenge which all other nanopore sequencing methods have faced for decades. Thus, the proposed Nano-SBS approach identifies individual bases by the detection and differentiation of the large tags released during the polymerase reaction instead of the small nucleotides themselves. The tags are large molecules that have slow diffusion rates, which greatly increase their chance of entering the nanopore and producing unique ionic current blockade signals. As proof-of-principle, we attached four different length coumarin-PEG tags to the terminal phosphate of 2′-deoxyguanosine-5′-tetraphosphate. We demonstrate efficient incorporation of the nucleotide analogs during the polymerase reaction, and better than baseline discrimination among the four tags based on their nanopore ionic current blockade signatures. This approach coupled with polymerase covalently attached to the nanopores in an array format should yield a single-molecule Nano-SBS platform.

Results

General principle of single molecule electronic DNA sequencing by synthesis using PEG-labeled nucleotides and nanopore detection. The single molecule electronic Nano-SBS system, which is shown schematically in Fig. 2, depicts the DNA polymerase bound in close proximity to the nanopore entrance. A template to be sequenced is added along with the primer. To this template-primer complex, four differently tagged nucleotides are added to the bulk aqueous phase. After polymerase catalyzed incorporation of the correct nucleotide, the tag-attached polyphosphate will be released and pass through the nanopore to generate a unique ionic current blockade signal, thereby identifying the added base electronically because the tags have distinct chemical structures. An example of four continuous nucleotide incorporation reactions with different tags for each base is shown in Supplementary Fig. S1. An array of nanopores, each with a covalently attached polymerase adjacent to the pore entrance, will allow single-molecule SBS.

This tag-based Nano-SBS system offers the following advantages over strand sequencing through nanopores: (1) it overcomes the inherent constraints imposed by the small differences among the 4 bases by instead using 4 large and distinct molecular tags, which are easily differentiated by a nanopore; and (2) there is no need to slow down the transit speed of the tag through the pore as long as the tag is detectable, because the polymerase extension and tag release rate is much slower than the tag interaction time with the pore. This would also eliminate phasing issues inherent to strand sequencing. Here, we describe the synthesis and efficient incorporation of a new class of nucleotide analogs with 5′-phosphate-attached tags. These tags consist of four different length PEGs and a coumarin moiety. We also demonstrate four distinct ionic current blockade patterns produced by these tags in an α-hemolysin channel at the single molecule level. This proof-of-principle study of the separate elements of the proposed Nano-SBS system demonstrates the feasibility of integrating them into a single molecule electronic SBS nanopore sequencer in the future.

Design, synthesis and characterization of PEG-labeled nucleotides. The four 5′-phosphate tagged 2′-deoxyguanosine-5′-tetraphosphates (Fig. 3) were synthesized according to the generalized synthetic scheme shown in Fig. 4. First, 2′-deoxyguanosine-5′-triphosphate (dGTP) was
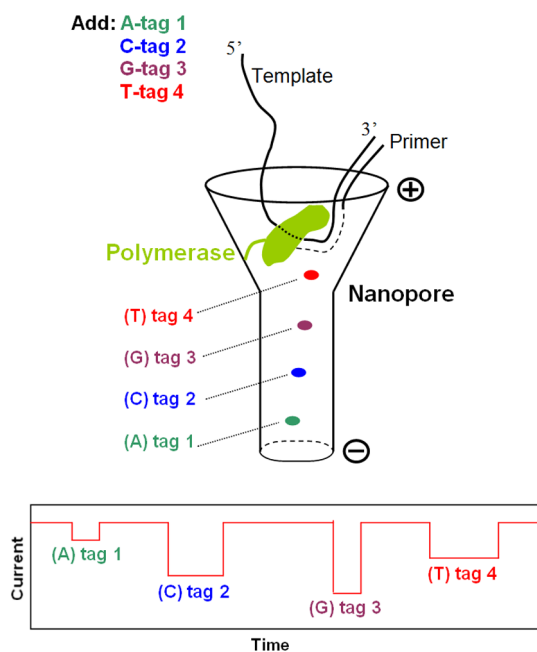

Figure 2 | Schematic of single molecule DNA sequencing by a nanopore with phosphate-tagged nucleotides. In the envisioned approach, each of the four nucleotides will carry a different tag. During SBS, these tags, attached via the 5′-phosphate of the nucleotide, will be released into the nanopore one at a time where they will produce unique current blockade signatures. For the purpose of illustration only, several tags are shown inside the pore in the order of their sequential release; in actuality only one tag will enter the pore at a time. A large array of such nanopores could lead to highly parallel, high throughput DNA sequencing.
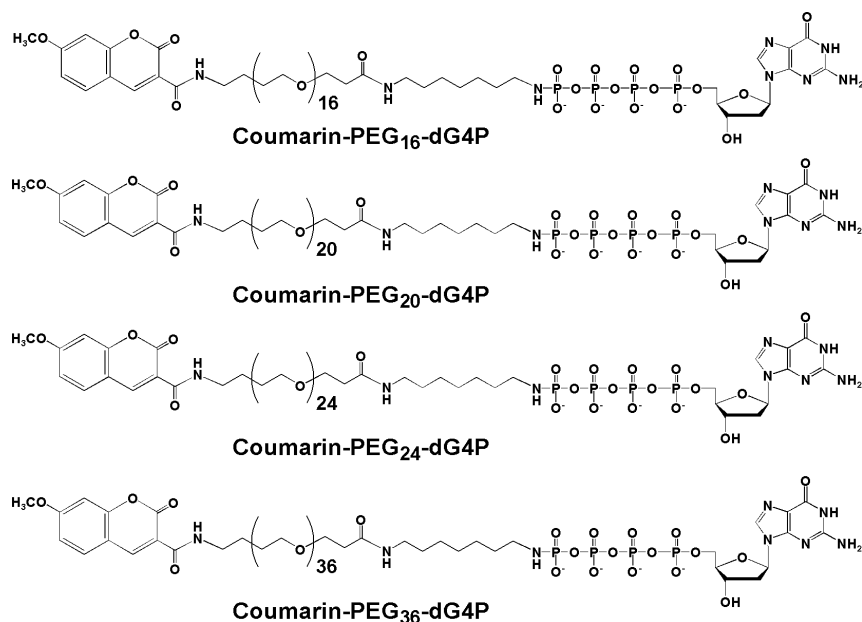
Figure 3 | Structures of four coumarin-PEG$_n$-dG4P nucleotide analogs.

converted to 2′-deoxyguanosine-5′-tetraphosphate (dG4P). Then, a diaminoheptane linker was added to the terminal phosphate of the tetraphosphate to produce dG4P-heptyl-NH$_2$ (Product A) in order to attach different length PEG tags. In a separate set of reactions, 6-methoxy-coumarin N-hydroxysuccinimidyl ester was reacted with one of four amino-PEG$_n$-COOH molecules with 16, 20, 24 or 36 ethylene glycol units, to produce coumarin-PEG$_n$-COOH molecules, which were subsequently converted to the corresponding NHS-esters (Product B). The coupling of dG4P-heptyl-NH$_2$ (Product A) with the coumarin-PEG$_n$-NHS esters (Product B) yields the four final nucleotide analogs, abbreviated coumarin-PEG$_n$-dG4Ps (Fig. 4, n = 16, 20, 24, 36). The coumarin moiety was used as a prototype modifier to further tune the size of the tag as well as to track the purification of intermediates and the final nucleotide analogs. Synthesis of the expected coumarin-PEG$_n$-dG4P molecules was confirmed by MALDI-TOF mass spectroscopy (Supplementary Fig. S2).

We next tested the coumarin-PEG$_n$-dG4P nucleotide analogs in polymerase extension reactions using the Therminator У DNA polymerase. A primer-loop-template was designed where the next complementary base was a C, enabling dGMP to be added to the DNA primer (Supplementary Fig. S3). Coumarin-PEG$_n$-triphosphate is released during the reaction (Supplementary Fig. S4). MALDI-TOF-MS confirmed that indeed each of the four coumarin-PEG$_n$-dG4P nucleotide analogs gave the correct extension product with 100% incorporation efficiency, as shown by the appearance of a single peak at ∼8,290 daltons in the mass spectra (Fig. 5). The absence of a primer peak at 7,966 daltons suggested that the reaction proceeded essentially to completion. An important feature of the Nano-SBS approach is that the extended DNA chains contain all natural nucleotides without any modifications, allowing SBS to continue over extensive lengths. All the extension products shown in Fig. 5 represented the incorporation of the coumarin-PEG$_n$-dG4P nucleotide analogs, with no products derived from potential residual dGTP or dG4P, since the molecules were purified twice in an HPLC system that separates these molecules effectively with a retention time difference of more than 10 min between the two groups of compounds. To further exclude this possibility, we treated the purified coumarin-PEG$_n$-dG4P nucleotide analogs with alkaline phosphatase, which would degrade any contaminating tri- or tetra-phosphate to the free nucleoside, and used the resulting HPLC-repurified coumarin-PEG$_n$-dG4P nucleotide analogs in extension reactions.

The tags released during incorporation of the coumarin-PEG$_n$-dG4P nucleotide analogs in polymerase reactions should be coumarin-PEG$_n$-triphosphates (coumarin-PEG$_n$-P$_3$). To reduce the complexity of the charge on the tags, we treated the released tags (coumarin-PEG$_n$-P$_3$) with alkaline phosphatase, yielding coumarin-PEG$_n$-NH$_2$ tags (Supplementary Fig. S4), and then analyzed these tags for their nanopore current blockade effects. In further developing the Nano-SBS system, we can pursue such treatment of the released coumarin-PEG$_n$-P$_3$ tags with alkaline phosphatase, which would be attached to the entrance of the nanopores downstream of the polymerase, to generate coumarin-PEG$_n$-NH$_2$ tags. Alternatively, we can optimize the conditions for using nanopores to directly detect the released charged coumarin-PEG$_n$-triphosphate tags. For the proof-of-principle studies reported here, in order to obtain large amounts of material for testing by MALDI-TOF MS and protein nanopores, we produced synthetic versions of the expected released tags (coumarin-PEG$_n$-NH$_2$) by acid hydrolysis of the four coumarin-PEG$_n$-dG4P nucleotide analogs to cleave the P-N bond between the polyphosphate and heptylamine moiety (Supplementary Fig. S4). The expected coumarin-PEG$_n$-NH$_2$ molecules were confirmed by MALDI-TOF-MS analysis, following HPLC purification (Fig. 6). MALDI-TOF-MS results indicated that the coumarin-PEG$_n$-NH$_2$ tags generated by acid hydrolysis were identical to the tags produced by alkaline phosphatase treatment of the released coumarin-PEG$_n$-triphosphate tags during the polymerase reaction.

Characterization of the current blocking effect of the tags in nanopores. To demonstrate the feasibility of our proposed electronic single molecule SBS approach, we measured a heterogeneous mixture of the four coumarin-PEG$_n$-NH$_2$ tags for their current blockade effects on a single αHL nanopore (Fig. 7). The top of Fig. 7 shows the profile of current blockade versus time. The lower left of Fig. 7 shows a representative subset of the time series data, indicating that inside the nanopore, PEG tags produce current blockades that are characteristic of their size. The relative frequency distribution of the histogram of blockade events ($<i>/<i_{open}>$) shows four well separated and distinct peaks for the four coumarin-PEG$_n$-NH$_2$ tags ($n = 36, 24, 20,$ and 16 from left to right respectively in Fig. 7, lower right). To highlight the wide separation of the peaks, and offer clear evidence that detection of a specific nucleotide might be accomplished by the unique blockade signal afforded by its released PEG
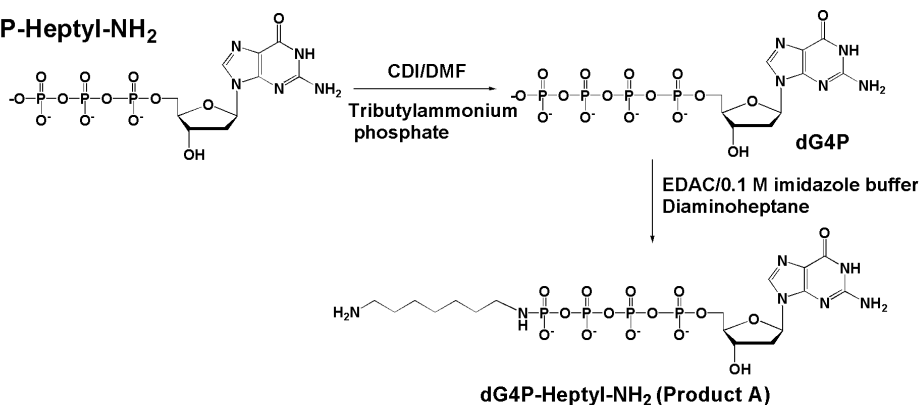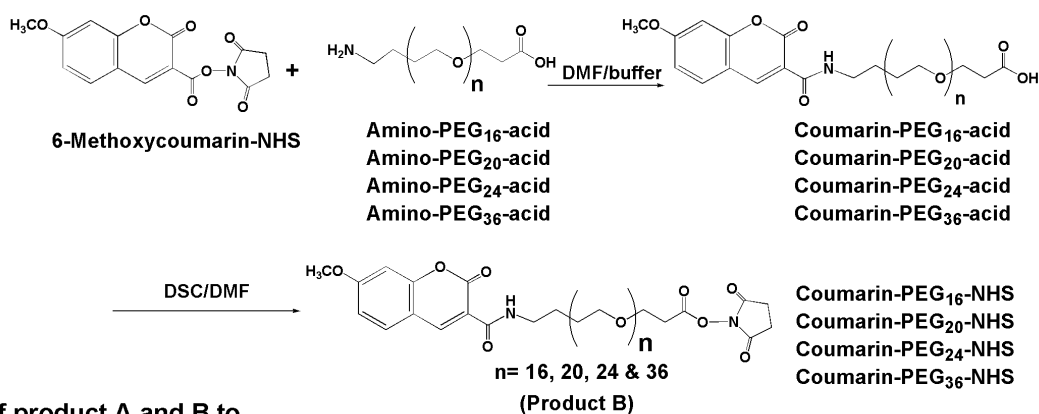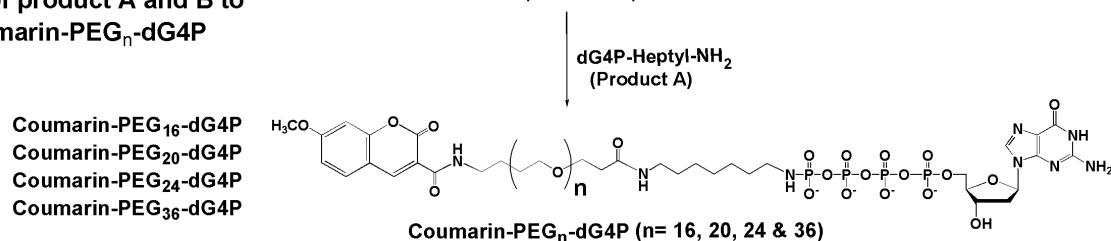
Figure 4 | Synthetic scheme for four coumarin-PEG$_n$-deoxyguanosine-5′-tetraphosphates (coumarin-PEG$_n$-dG4P, n = 16, 20, 24, 36).

tag, the peaks are fitted with single Gaussian functions and the corresponding 6 σ error distributions are shown (colored rectangles at top in Fig. 7, lower right). We also characterized separately each of the coumarin-PEG$_n$-NH$_2$ molecules with the pore (data not shown), which confirmed the identity of the different-sized PEG-related peaks shown in Fig. 7. These results suggest that a single base could be discriminated with accuracy better than 1 in 5x10^8 events, represented in Fig. 7, lower right, by using A, C, G and T designations, which would occur when four different nucleotides with four different length PEGs such as those tested here are used for DNA sequencing.

Discussion

As described above, a single αHL ion channel can separate single molecules based on their size, and easily resolves a mixture of PEGs to better than the size of a single monomer unit (i.e., < 44 g/mol)[16,18,22]. This high resolution arises from the interactions between the PEG polymer, the electrolytes (mobile cations), and amino acid side chains that line the αHL channel's lumen[16]. These interactions allow the pore to be used as a nanometer-scale sensor that is specific to the size, charge and chemical property of an analyte.

Here, such analysis is extended to PEGs with different chemical groups on either terminus. The single channel ionic current recording

in Fig. 7 (top and lower left) illustrates the blockades caused by the four different sized coumarin-PEG$_n$-NH$_2$ molecules, one at a time. As with unmodified PEG, each of the current blockades is unimodal (i.e., described well with Gaussian distributions and well-defined mean values).

To accurately discriminate between the four bases (A, C, G and T) for strand nanopore sequencing, one or more of the following strategies need to be adopted: (1) enhance and differentiate the strength of the detection signals; (2) develop an effective method to discern and process the electronic blockade signals generated; (3) control the translocation rate of nucleic acids through the pore, e.g., by slowing down DNA movement; and (4) design and make new and more effective synthetic nanopores. As we demonstrated here, the Nano-SBS approach has transformed the problem of resolving the 4 individual bases to that of discriminating among 4 large well-differentiated tags, which essentially solves the first three problems.

DNA sequencing by synthesis is the dominant platform for genomics research and personalized medicine[29-33]. Kumar et al. first reported the modification of nucleoside-5′-triphosphates, either by introducing more phosphate groups to produce tetra- and penta-phosphates and introducing fluorophores directly to the terminal phosphate or attaching a linker between the terminal phosphate and the fluorophore [24,25]. Tetra- and penta-phosphates were shown to be better DNA polymerase substrates, and fluorophore-labeled
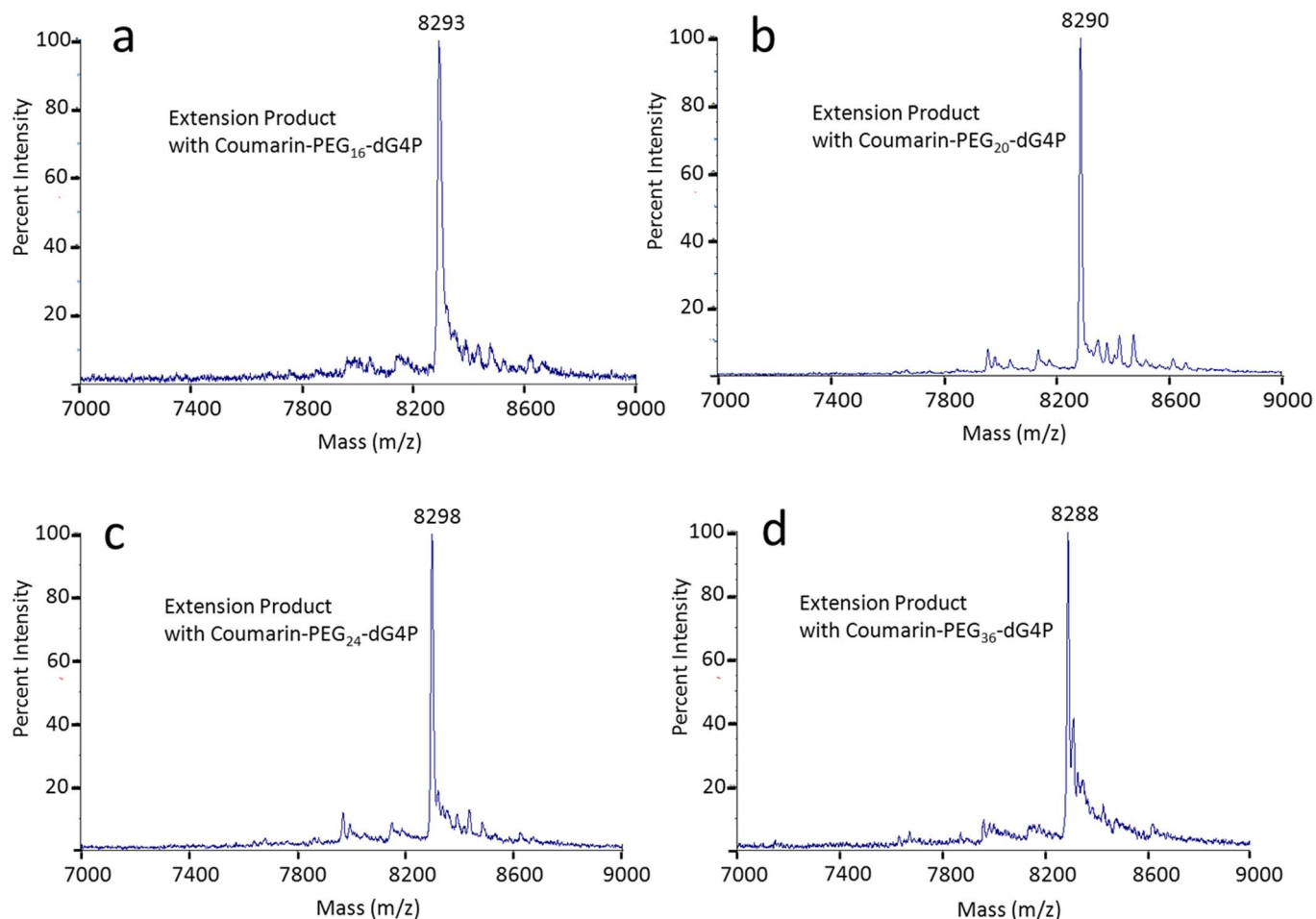
Figure 5 | MALDI-TOF MS measurement of the extension products obtained with the four coumarin-PEG$_n$-dG4P nucleotide analogs. A template-loop primer (Supplementary. Fig. S3), in which the template contained a C at the next position, was used along with one of the four coumarin-PEG$_n$-dG4P nucleotide analogs for the polymerase reaction. In each case, 2′-dGMP is incorporated into the growing DNA strand, yielding a single base primer extension product with 100% efficiency, shown as a single peak in each mass spectrum at ~ 8290 daltons.

phosphate nucleotides have been used widely for DNA sequencing[26–28,34]. Here, we have demonstrated a novel approach to enhance discrimination of the four nucleotides by modifying them at the terminal phosphate moiety with distinct large chemical tags for single molecule electronic SBS. The physical and chemical properties of the tag can be further adjusted to optimize the nanopore capture efficiency and measurement accuracy. For instance, the insertion of a positively charged linker consisting of four lysines or arginines between the polyphosphate and the PEG will produce precursors with a neutral charge and released tags with a net positive charge. Using the appropriate magnitude and sign of the potential[23], the released tags, but not nucleotide substrates, will be transported through the pore.

The coumarin moiety on the tagged nucleotides can be replaced with other molecules of larger size or different charge to further enhance nanopore discrimination. Clearly, it is important that every tag released in a polymerase reaction be maintained in the proper order for real-time single molecule Nano-SBS. Despite all these precautions, some unreacted nucleotide analogs might enter the pore. Thus, the ability to discriminate between cleaved tags and unreacted nucleotide analogs will be important; fortunately, these two groups of tags should be easily differentiated by a nanopore due to their significant size and charge differences. In addition, it has not escaped us that the tagged nucleotide Nano-SBS approach can be implemented in a straightforward way by adding the four nucleotides (A, C, G and T) labeled with identical tags on the 5′-phosphate in a stepwise fashion to reduce the overall complexity of the system, analogous

to pyrosequencing[30] and the Ion Torrent approach[33]. However, unlike those methods, the Nano-SBS approach has the advantage of single molecule sensitivity without the requirement for DNA amplification, and hence no issues with sequencing through homopolymeric regions, since tags released at each position of the homopolymer are detected discretely by the nanopore at single-molecule level.

The single molecule electronic Nano-SBS approach described here should be applicable to either protein nanopores (*e.g.*, αHL; *Mycobacterium smegmatis* porin A, MspA)[35], or solid-state nanopores[36–42]. These options will provide nanopores with different properties that are appropriate for detecting a library of tags. To implement this novel strategy for DNA sequencing, an array of nanopores[43] can be constructed on a planar surface to facilitate massively parallel DNA sequencing.

In conclusion, we have conducted proof-of-principle studies for a novel single molecule electronic Nano-SBS platform that will measure the tags released from the nucleotide substrates during the polymerase reaction, for sequence determination. In its full implementation in the future, it should be capable of long, accurate reads, and potentially offer very high throughput electronic single molecule DNA sequencing.

Methods

Synthesis of coumarin-PEG$_n$-dG4P nucleotide analogs. The synthesis of coumarin-PEG$_n$-dG4P involves three steps (A, B, C) as shown in **Fig. 4.** All of the nucleotide analogs were purified by reverse-phase HPLC on a 150 × 4.6 mm column (Supelco), mobile phase: A, 8.6 mM Et$_3$N/100 mM 1,1,1,3,3,3-hexafluoro-2-propanol in water
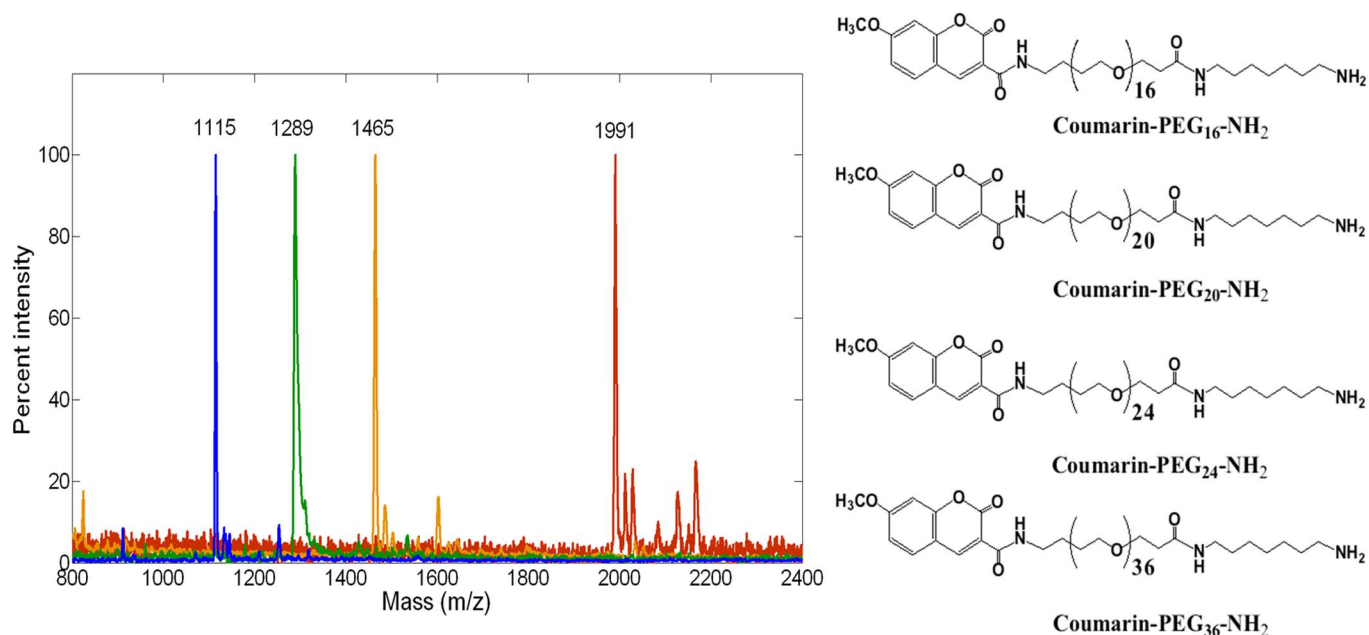
Figure 6 | Characterization of the released coumarin-PEG$_n$-NH$_2$ tags by MALDI-TOF MS. Coumarin-PEG$_n$-NH$_2$ tags generated by acid hydrolysis of coumarin-PEG$_{16}$-dG4P yielding coumarin-PEG$_{16}$-NH$_2$ (blue, m/z = 1115), coumarin-PEG$_{20}$-dG4P yielding coumarin-PEG$_{20}$-NH$_2$ (green, m/z = 1289), coumarin-PEG$_{24}$-dG4P yielding coumarin-PEG$_{24}$-NH$_2$ (orange, m/z = 1465) and coumarin-PEG$_{36}$-dG4P yielding coumarin-PEG$_{36}$-NH$_2$ (red, m/z = 1991), are identical to the tags produced by alkaline phosphatase treatment of the released coumarin-PEG$_n$-triphosphate tags during the polymerase reaction (Supplementary. Fig. S4), as shown by MALDI-TOF-MS analysis. A composite image of four separately obtained MS spectra is shown. The structures of the coumarin-PEG$_n$-NH$_2$ tags are shown to the right.

(pH 8.1); B, methanol. Elution was performed from 100% A isocratic over 10 min followed by a linear gradient of 0–50% B for 20 min and then 50% B isocratic over another 30 min.

Synthesis of 2′-deoxyguanosine-5′-tetraphosphate (dG4P). The synthesis of 2′-dG4P is carried out starting from 2′-dGTP. 300 μmoles of 2′-dGTP (triethylammonium salt) were converted to the tributylammonium salt by using 1.5 mmol (5 eq) of tributylamine in anhydrous pyridine (5 ml). The resulting solution was concentrated to dryness and co-evaporated twice with 5 ml of anhydrous DMF. The dGTP (tributylammonium salt) was dissolved in 5 ml

anhydrous DMF, and 1.5 mmol 1, 1-carbonyldiimidazole (CDI) was added. The reaction was stirred for 6 hr, after which 12 μl methanol was added and stirring continued for 30 min. To this solution, 1.5 mmol phosphoric acid (tributylammonium salt, in DMF) was added and the reaction mixture was stirred overnight at room temperature. The reaction mixture was diluted with water and purified on a Sephadex-A25 column using a 0.1 M to 1 M TEAB gradient (pH 7.5). The dG4P elutes at the end of the gradient. The appropriate fractions were combined and further purified by reverse-phase HPLC to yield 175 μmol of the pure tetraphosphate (dG4P).^{31}P-NMR: δ, −10.7 (d, 1P,α-P), −11.32 (d, 1P, δ-P), −23.23 (dd, 2P, β, γ-P); ESI-MS (-ve mode): Calc. 587.2; Found 585.9 (M-2).
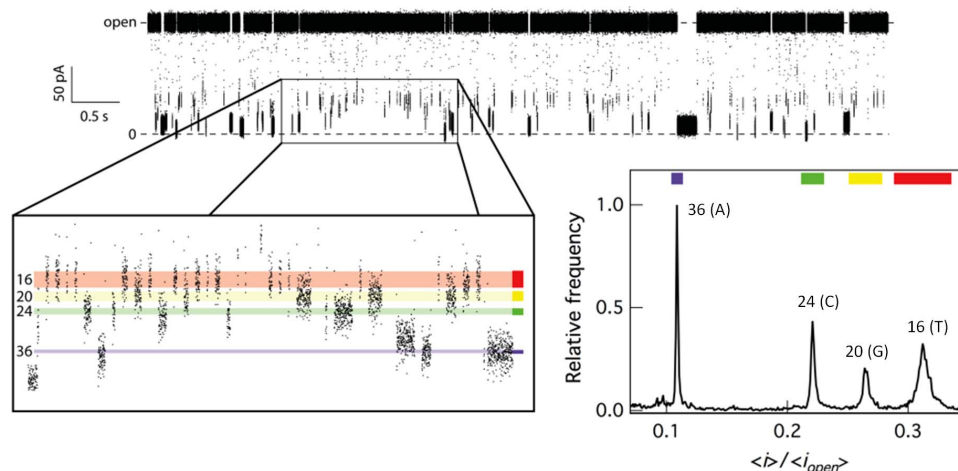

Figure 7 | Discrimination of the coumarin-PEG$_n$-NH$_2$ tags in protein nanopores at single molecule level. Four coumarin-PEG$_n$-NH$_2$ tags (n = 16, 20, 24 and 36), derived from the corresponding coumarin-PEG$_n$-dG4P nucleotide analogs, were pooled and diluted in 4 M KCl, 10 mM Tris, pH 7.2 for nanopore measurement. (**Top**) Profile of current blockade versus time. (**Lower Left**) A subset of the time series data indicates that when these PEG tags enter a single α-hemolysin ion channel, they cause current blockades that are characteristic of their size. (**Lower Right**) A histogram of the mean current blockade caused by 4 individual PEG tags (sizes 36, 24, 20 and 16 from left to right) shows baseline resolution with a 10 kHz measurement bandwidth. The colored bars at the top represent the 6 σ distribution of the data (assuming Gaussian distributions for each of four PEG tags that could represent each of the four DNA nucleotides), which suggests that a single base could be discriminated with an accuracy better than 1 in 5x10[8] events, represented in this figure by using A, C, G and T designations, which would occur when four different nucleotides with these four different length PEGs are used for DNA sequencing.

6

Synthesis of dG4P-heptyl-NH₂ (Product A). To 80 μmol dG4P in 2 ml water and 3.5 ml 0.1 M 1-methylimidazole-HCl (pH 6) were added 154 mg EDAC and 260 mg diaminoheptane. The pH of the resulting solution was adjusted to 6 with concentrated HCl and stirred at room temperature overnight. This solution was diluted with water and purified by Sephadex-A25 ion-exchange chromatography followed by reverse-phase HPLC to yield ~20 μmol dG4P-heptyl-NH₂ (Product A), which was characterized by ESI-MS (-ve mode): calc. 699.1; Found 698.1 (M-1).

Synthesis of coumarin-PEGₙ-NHS esters (Product B). Amino-PEGₙ-acids (1 eq) [Amino-d(PEG)₁₆, ₂₀, ₂₄, ₃₆-acids; Quanta Biodesign] were dissolved in 0.1 M sodium carbonate-sodium bicarbonate buffer (pH 8.6), followed by addition of coumarin-NHS (1 eq) in DMF, and the reaction mixture was stirred overnight. The coumarin-PEGₙ-acids obtained were purified by silica-gel chromatography using a CH_2Cl_2-MeOH (5–15%) mixture and the appropriate fractions were combined. These compounds were analyzed by MALDI-TOF MS (Supplementary Table 1).

Reaction of the coumarin-PEGₙ-acids with 1.5 eq. of disuccinimidyl carbonate (DSC) and 2 eq. of triethylamine in anhydrous DMF for 2 h yields the corresponding coumarin-PEGₙ-NHS esters (Product B). The coumarin-PEGₙ-NHS esters, which move slightly faster than the corresponding acids on silica-gel plates, were purified by silica-gel chromatography using a CH_2Cl_2-MeOH (5-15%) mixture and then used in the next step.

Synthesis of Coumarin-PEGₙ-dG4P. dG4P-heptyl-NH₂ (Product A) synthesized above was taken up in 0.1 M sodium carbonate-bicarbonate buffer (pH 8.6) and to this stirred solution was added 1 eq. of one of the coumarin-PEGₙ-NHS esters (Product B) in DMF. The resulting mixture was stirred overnight at room temperature and then purified on a silica-gel cartridge (15-25% MeOH in CH_2Cl_2 to remove unreacted coumarin-acid or -NHS ester and then 5:4:1 isopropanol/NH₄OH/ H₂O). The crude product was further purified twice by reverse-phase HPLC to provide pure coumarin-PEGₙ-dG4P: coumarin-PEG₁₆-dG4P (retention time, 31.7 min); coumarin-PEG₂₀-dG4P (retention time, 32.2 min); coumarin-PEG₂₄-dG4P (retention time, 33.0 min); coumarin-PEG₃₆-dG4P (retention time, 34.3 min). The structure of all the molecules was confirmed by MALDI-TOF MS (Supplementary Table 2).

DNA polymerase extension reactions using coumarin-PEGₙ-dG4P nucleotide analogs. Extension reactions were performed using a template-primer (5′-GATCGCGCCGCGCCTTGGCGCGGCGC-3′, M.W. 7966), in which the next complementary base on the template is a C, allowing extension by a single G (Supplementary Fig. S3). Each extension reaction was carried out in a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems) at 65°C for 25 minutes in 20 μl reactions consisting of 3 μM template-loop-primer, 1X Thermination γ buffer [50mM KCl, 20mM Tris-HCl, 5mM MgSO₄, 0.02% IGEPAL CA-630 (pH 9.2)], 2 units of Thermination γ DNA polymerase (New England Biolabs), and 15 μM of one of the coumarin-PEGₙ-dG4P nucleotide analogs. The DNA extension products were precipitated with ethanol, purified through C18 ZipTip columns (Millipore), and characterized by MALDI-TOF MS. As shown in Fig. 5 in the main text, four identical extension products (expected molecular weight 8,295) were obtained. These polymerase extension reactions for each coumarin-PEGₙ-dG4P were repeated and the released products (coumarin-PEGₙ-triphosphates, Supplementary Fig. S4) were treated with alkaline phosphatase (1U at 37°C for 15 min) to yield the coumarin-PEGₙ-NH₂ tags, which were extracted into dichloromethane and characterized by MALDI-TOF-MS.

Acid hydrolysis of coumarin-PEGₙ-dG4P to yield the coumarin-PEGₙ-NH₂ tags (Supplementary Fig. S4). Acetic acid was added to the coumarin-PEGₙ-dG4P nucleotide analogs to a final concentration of 10%, and the solution was vigorously shaken overnight to ensure the hydrolysis of the N-P bond between the δ phosphate and the heptylamine to yield the coumarin-PEGₙ-NH₂ tags. The solution was dried using a CentriVap and resuspended in an appropriate volume of water. A 1 μl aliquot was collected for MALDI-TOF MS characterization, and a second aliquot was measured at 260 nm and 350 nm using a NanoDrop ND-1000 spectrophotometer for further characterization of the resulting molecules. The resulting coumarin-PEGₙ-NH₂ tags were the expected size as measured by MALDI-TOF MS (see Fig. 6 and Supplementary Table 3).

Nanopore Measurements of the Coumarin-PEGₙ-NH₂ Tags. *Membrane and channel formation.* Single α-hemolysin (αHL) channels were inserted into solvent-free planar lipid bilayer membranes[44] fabricated across an ~ 80 μm diameter hole in a 25 μm thick Teflon partition separating two electrolyte solution wells as described previously[45]. 4 M KCl, 10 mM Tris titrated to pH 7.2 with citric acid was used throughout the experiment. Membranes were formed by first wetting the partition with 1 % v/v hexadecane/pentane. 10 mg/mL diphytanoyl phosphatidylcholine (DPhyPC) in pentane were spread at both air-electrolyte solution interfaces with the solution levels well below the hole in the Teflon partition. After 10 min, the solution levels were raised above the hole spontaneously to form a membrane. Approximately 0.5 μL of 0.5 mg/mL αHL was injected into the solution immediately adjacent to the membrane and the ionic current was observed until a single channel inserted into the membrane. The *cis* chamber contents were then exchanged with protein-free electrolyte solution to maintain a single channel for the duration of the experiment.

Coumarin-PEGₙ-NH₂ molecules (*n* = 16, 20, 24 and 36) were added to the *trans* side of the pore (defined as the β-barrel side of the channel) to a final concentration

between 0.4 μmol/L and 1 μmol/L of each component. Ionic current was recorded between two matched Ag/AgCl electrodes (3 M KCl) at a fixed potential (−40 mV) for approximately 15 min to achieve sufficient counting statistics. Data were recorded with a 4-pole Bessel filter at 10 kHz oversampled at 50 kHz.

Data analysis. Data were analyzed off-line with an in-house program written in LabVIEW (National Instruments) as described previously[23]. In brief, blockades were located with an event detector based on a simple threshold algorithm set at 5 σ of the current noise in the open state. When an event is detected, the points in the rise time and decay time were discarded (~ 60 μs and 20 μs, respectively). The mean blockade depth was calculated from the remaining points and the open channel current was calculated from the mean of 0.8 ms of open channel data separated 0.2 ms from the threshold. The data were reported as a ratio of the means ($<i>/<i_{open}>$) and the nanopore spectra were calculated as a histogram of these values.

1. Perkins, T. T., Quake, S. R., Smith, D. E. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* 264, 822–6 (1994).
2. Rief, M., Clausen-Schaumann, H. & Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat Struct Biol* 6, 346–9 (1999).
3. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271, 795–9 (1996).
4. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA* 93, 13770–3 (1996).
5. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys J* 77, 3227–33 (1999).
6. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. & Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc Natl Acad Sci USA* 97, 1079–84 (2000).
7. Henrickson, S. E., Misakian, M., Robertson, B & Kasianowicz, J. J. Driven DNA transport into an asymmetric nanometer-scale pore. *Phys Rev Lett* 85, 3057–60 (2000).
8. Vercoutere, W. *et al.* Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat Biotechnol* 19, 248–52 (2001).
9. Kasianowicz, J. J., Henrickson, S. E., Weetall, H. H. & Robertson, B. Simultaneous multianalyte detection with a nanometer-scale pore. *Anal Chem* 73, 2268–72 (2001).
10. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc Chem Res* 35, 817–25 (2002).
11. Kasianowicz, J. J. Nanometer-scale pores: potential applications for analyte detection and DNA characterization. *Dis Markers* 18, 185–191 (2003).
12. Kasianowicz, J. J. Nanopores: flossing with DNA. *Nat Mater* 3, 355–6 (2004).
13. Kasianowicz, J. J., Robertson, J. W., Chan, E. R., Reiner, J. E. & Stanford, V. M. Nanoscopic porous sensors. *Annu Rev Anal Chem* 1, 737–66 (2008).
14. Song, L. *et al.* Structure of Staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859–66 (1996).
15. Bezrukov, S. M., Vodyanoy, I., Brutyan, R. A. & Kasianowicz, J. J. Dynamics and free energy of polymers partitioning into a nanoscale pore. *Macromolecules* 29, 8517–22 (1996).
16. Krasilnikov, O. V. Sizing channels with neutral polymers. in *Structure and Dynamics of Confined Polymers,* ed. Kasianowicz, J.J., Kellemayer, M.S.Z. & Deamer, D.W., NATO Science Series, Kluwer Academic Publishers, The Netherlands 87, 97–116 (2002).
17. Kasianowicz, J. J. & Bezrukov, S. M. Protonation dynamics of the alpha-toxin ion channel from spectral analysis of pH-dependent current fluctuations. *Biophys J* 69, 94–105 (1995).
18. Kasianowicz, J. J. *et al.* Physics of DNA threading through a nanometer pore and applications to simultaneous multianalyte sensing. in *Structure and Dynamics of Confined Polymers,* ed. Kasianowicz, J.J., Kellemayer, M.S.Z. & Deamer, D.W., NATO Science Series, Kluwer Academic Publishers, The Netherlands 87, 141–64 (2002).
19. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4, 265–70 (2009).
20. Advances in Genome Biology and Technology (AGBT) Conference, Marco Island, FL, 2012 and Feb. 17, 2012 press release from Oxford Nanopore Technologies (http://www.nanoporetech.com/news/press-releases/view/39).
21. Manrao, E. A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* 30, 349–353 (2012).
22. Robertson, J. W. *et al.* Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc Natl Acad Sci USA* 104, 8207–11 (2007).
23. Reiner, J. E., Kasianowicz, J. J., Nablo, B. J. & Robertson, J. W. Theory for polymer analysis using nanopore-based single-molecule mass spectrometry. *Proc Natl Acad Sci USA* 107, 12080–5 (2010).
24. Kumar, S. *et al.* Terminal-Phosphate-Labeled Nucleotides with New Linkers. US Patent, 7,393,640 (2008).

25. Kumar, S. & Sood, A. Labeled Nucleoside Polyphosphates. US Patent, 7,041,812 (2006).
26. Kumar, S. *et al*. Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* **24**, 401–8 (2005).
27. Sood, A. *et al*. Terminal phosphate-labeled nucleotides with improved substrate properties for homogeneous nucleic acid assays. *J Am Chem Soc* **127**, 2394–5 (2005).
28. Eid, J. *et al*. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–8 (2009).
29. Ju, J., Li, Z., Edwards, J., Itagaki, Y. Massive parallel method for decoding DNA and RNA. US Patent, 6,664,079 (2003).
30. Margulies, M. *et al*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
31. Ju, J. *et al*. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **103**, 19635–19640 (2006).
32. Bentley, D. R. *et al*. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
33. Rothberg, J. *et al*. An integrated semiconductor device enabling non-optical genome sequencing, *Nature* **475**, 348–352 (2011).
34. Sims, P. A., Greenleaf, W. J., Duan, H. & Xie, X. S. Fluorogenic DNA sequencing in PDMS microreactors. *Nat Methods* **8**, 575–80 (2011).
35. Derrington, I. M. *et al*. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A* **107**, 16060–5 (2010).
36. Garaj, S. *et al*. Graphene as a subnanometre trans-electrode membrane. *Nature* **467**, 190–3 (2010).
37. Hall, A. R. *et al*. Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores. *Nat Nanotechnol* **5**, 874–7 (2010).
38. Merchant, C. A. *et al*. DNA translocation through graphene nanopores. *Nano Lett* **10**, 2915–21 (2010).
39. Schneider, G. F. *et al*. DNA translocation through graphene nanopores. *Nano Lett* **10**, 3163–7 (2010).
40. Storm, A. J. *et al*. Fast DNA translocation through a solid-state nanopore. *Nano Lett* **5**, 1193–1197 (2005).
41. Wanunu, M., Sutin, J., McNally, B., Chow, A. & Meller, A. DNA translocation governed by interactions with solid-state nanopores. *Biophys J* **95**, 4716–25 (2008).
42. Rosenstein, J. K., Wanunu, M., Merchant, C. A., Drndic, M. & Shepard, K. L. Integrated nanopore sensing platform with sub-microsecond temporal resolution. *Nat Methods* **9**, 487–492 (2012).
43. McNally, B. *et al*. Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett* **10**, 2237–44 (2010).
44. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc Natl Acad Sci U S A* **69**, 3561–3566 (1972).
45. Rodrigues, C. G. *et al*. Mechanism of KCl enhancement in detection of nonionic polymers by nanopore sensors. *Biophys J* **95**, 5186–5192 (2008).

Acknowledgments

Author contributions

JJ, CT, SK and ZL conceived the initial approach; SK synthesized the nucleotide analogs; MC, CT, BH and ZL characterized these molecules and carried out the SBS experiments; AB and JWFR performed and analyzed the nanopore experiments; JER and JWFR designed the data analysis algorithms used to process the data. JJ, SK, JJR, JJK and JWFR designed the experiments, assessed the results, and wrote the manuscript with MC.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kumar, S. *et al*. PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis. *Sci. Rep.* **2**, 684; DOI:10.1038/srep00684 (2012).

8