READNA Publishable summary Final Periodic Report #3

www.cng.fr/READNA

Overall Project Summary

DNA analysis methods are tools to advance on the goal of understanding the information encrypted in our genomes. The overall objective of READNA (**REvolutionary Approaches and Devices for Nucleic Acid analysis**) is to develop a toolbox of nucleic acid analysis methods that will enable effective and economic deployment for the good of society. READNA has managed to make substantial headway on several different fronts during the 54 months of the project. In several lines developments and findings have led to key publications and have contributed to the understanding and advancement of nucleic acid analysis methodology. Efforts of the READNA consortium are substantially contributing to the advancement on the 1000 $ genome. READNA has also worked on the dissemination of its knowhow and on uniting the research community involved in nucleic acid technology development and application.

The READNA project has been very productive and successful with over **100** published or accepted papers, **20%** thereof in top tier publications, more than **160** presentations at conferences and more than **20** patents or applications. Several of the developments have or will be reaching the market soon. There are several spin-off companies that have been created as a result of READNA activities.

Project Highlights at the end of reporting period #3 - 30/11/2012:

Scientific:

- Distinction of the different nucleosides with nanopores (WP4)
- Extensive evolution of DNA handling methodology in nanofluidic devices (WP3)
- Exhaustive evaluation of targeted enrichment methodology (WP1)
- Development of novel enrichment technology with Selectors (WP1)
- Development of facile sequencing methods (WP2)
- Systems for epigenetic analysis (WP5)
- Methods for *in situ* sequence analysis (WP5)

Commercial:

- Spin-off of q-Linea
- Spin-off of HaloGenomics (acquired by Agilent)
- Spin-off of Genotype2Phenotype (Kalim Mir)
- Presentation of first results of strand sequencing with nanopores (ONT at AGBT in Marco Island, February 2012)
- Launch of Cell-O-Matic EU Project

The READNA project has made great progress towards its overall objectives during the 2nd and 3rd reporting periods. The five R+D workpackages have followed the projected evolution at the outset of the project.

READNA has substantially pushed the envelope of nucleic acid technology development. The technology developed by READNA will be deployed for the accumulation of high-resolution biological information. This information will be of use in many applications, such as the early detection of disease onset, identification of disease susceptibility and it will support decisions on the best mode of treatment of disease. READNA has prepared a diverse toolbox to turn personalized medicine and ultimately personalized healthcare into reality. The methods developed will also find their application in other fields such as agriculture and biotechnology. READNA can form the basis for improved quality of life.

WP Publishable Summary

WP1 (Near Term Innovations: Ancillary elements for 2nd generation DNA sequencers) has provided a benchmark of targeted enrichment methods and used the experience gained for the evolution of the development and further implementation of MegaPlex PCR and Selectors. These methods have been standardized, applied for targeted studies of genomic variability and the verification of rare variants, and are ready for widespread deployment. An SME, HaloGenomics has been spun out with this objective and has in the meantime been acquired by Agilent. Standards and formats for data analysis from 2nd generation sequencing experiments have been a second focal point of this WP. Several of the partners are now well embedded in international efforts to this end. A component has been added to examine the ethical and societal impact of high-resolution DNA analysis methods.

WP2 (Near Term Innovations: Improvement and extension of existing methods) focuses on the completion and stabilisation of the DNA analysis methods that had reached a proof of principle at the outset of the project. The DASH (dynamic allele-specific hybridisation) technique for genotyping has been evolved towards sequencing in an array format. Datta arrays have been developed for individual picolitre reactions. With ribo-PCR MS the simplest re-sequencing protocol has been established that requires only genomic DNA, primers and a PCR mastermix. Post cycling treatment consists of a simple treatment with NaOH and desalting before mass spectrometric analysis. Dedicated software for the interpretation of results has been developed. In a second mass spectrometry implementation a very economic, high resolution HLA typing protocol for HLA-A, HLA-B and HLA-DRB1 based on the analysis of microhaplotypes has been developed and software for 4-digit HLA allele-calling implemented. A third task focussed on the development of a system for sample characterisation that uses DNA, RNA and protein profiles analyzed by mass spectrometry.

WP3 (Fluorescence-based Single Molecule Sequencing) has prepared the stage for development of fluorescence-based 3rd generation sequencing methodology and long-range DNA analysis. Different sequencing chemistries have been developed using chemical ligation methods and FRET-based donor-acceptor systems on DNA polymerases. The resolution of optical detection has been improved to go below the diffraction limit of light. Microfluidic devices will be developed together with methods for DNA handling that allow record size DNA molecules to be manipulated and enzymatic reactions such as restriction or hybridisation reactions to be carried out inside the nanochannels. During the third reporting period workpackage 3 has made further progress on two fronts: (1) Sequencing biochemistry & (2) Obtaining sequence information in an ultra-long range

context. We started by investigating a number of novel concepts for sequencing. It is clear that we reached a fork in the road, one direction being chemical ligation-based sequencing which could become a cost-effective, approach for when a large number of molecules need to be enumerated such as in digital transcriptomics. The second road is real-time sequencing and its application on long-stretched molecules. Here we have explored a number of methods that monitor fluorescence modulation via FRET and this is the approach that is most relevant to our goal of sequencing genomic DNA in a long-range context.

We have made massive strides in handling and imaging of ultra-long genomic DNA fragments and obtaining contextual sequence information listed below:

1) We have been able to extract long DNA from single cells and chromosomes and have fabricated and tested new device designs that enable ultra-long DNA molecules to be visualized and reagents to be exchanged over them.
2) We have demonstrated one means for extracting sequence information in a long range context by FRET from a DNA intercalator dye to an incorporated nucleotide.
3) We have demonstrated a second means for extracting sequence information by detecting the delayed occurrence of polymerase conformational change associated with incorporation, when one of the nucleotides is limiting.
4) We have also developed methods to passivate the interior walls of the chips so that enzyme reactions can be carried out effectively, without various reaction components sticking to channel walls, which we had initially found to be a roadblock.
5) These advances have enabled us to conduct enzymatic reactions on DNA molecules stretched in nanofluidic channels, which is a major achievement and pre-requisite for applying the sequencing biochemistries we or others have developed, on ultra-long DNA.
6) Moreover, we have developed effective means for mapping megabase lengths of DNA, determining their identity by comparing to *in silico* maps based on reference genomes.
7) Furthermore, we have been able to extract the mapped DNA, amplify it and conduct next generation sequencing and FISH on the product, thereby reconciling gross (FISH), intermediate (Single Molecule mapping), and fine (short read sequencing) levels of genome organisation.
8) We have developed very efficient methods of chemically ligating DNA strands in the presence of DNA templates. These methods have been used in a preliminary study to demonstrate the feasibility of DNA sequencing by ligation. We have shown that very short fluorescently labelled oligonucleotides can be ligated together in a sequence-specific manner.

Our work therefore has succeeded in our goal of integrating sequence with its genomic context.

WP4 (<u>Nanopore Sequencing</u>) has been developing the use of nanopore-based measurements for single molecule nucleic acid sequence analysis. In line with the proposed plan of work, detection techniques have been developed that allow for the discrimination of all four nucleotides and 5-Me-cytosine as individual bases after exonuclease cleavage, and when present in the nanopore as an intact strand of DNA. Methods of control of DNA translocation have also been developed, one of which includes the successful coupling of exonuclease enzymes to a nanopore. Protein nanopores have also been placed into suitably sized apertures in solid materials, providing a further method to produce highly robust devices for clinical application.

Good progress has been made in this workpackage to develop the initial tools required to fulfil DNA sequencing by nanopores. This work continues to bode well for delivering a transformational approach to sequencing, one in which rapid, long-read direct analysis is implemented in highly scalable electronic read-out devices, suitable for a clinical setting as well as high-throughput research facilities. Progress to complete the research activities contained within the project covers several important synergistic aspects of research. First, protein nanopores have been developed that can measure and discriminate the four bases of DNA in their monophosphate forms. As these are the product of exonuclease activity, this is an important step in the technical progress. Equally important is the localisation of a suitable exonuclease in close proximity to the nanopore, which has also been achieved. Second, key developments have been made towards the creation of large stable arrays of nanopores by the insertion of protein-based nanopores into similarly sized apertures fabricated in solid materials. Such devices have the potential to provide a highly robust solution for analysis chips in the clinic. Also, we successfully explored the use of monoatomically thin graphene monolayers as an alternative membrane material instead of Silicon Nitride. Further progress has been made on droplet-interface bilayers, which raises the possibility of single cell DNA sequencing. Third, work on measuring static strands of DNA held inside mutant nanopore proteins has demonstrated the ability not only to resolve and discriminate single nucleotides within the strand, but also that this discrimination is sufficient to identity all four bases, and the modified variants methylcytosine and hydroxymethlycytosine. With the growing importance of the detection of natural modifications to DNA, it remains a high priority to demonstrate the capability of nanopores to measure methylation, both from a perspective of mapping such changes over a very long range, but also at the single base level via sequencing. Fourth, and in a development beyond that originally envisioned for the workpackage, measurement of static RNA strands inside nanopore proteins has also been shown to provide similar degrees of resolution and discrimination compared to DNA, including RNA base modifications, moving a step closer to direct analysis of RNA using nanopore devices.

WP5 (<u>New Genotyping Challenges</u>) is concerned with the development of methods for nucleic acid analyses for which currently no good, economic methods exists. In addition the WP is devoted to solving genotyping and epi-genotyping challenges, where the performance of current technology is not satisfactory. Existing nucleic acids analysis methods have been adapted for DNA methylation analysis and copy number variation analysis which have to deliver very precise quantitative results. Methods have been developed to analyse mutant nucleic acid molecules under challenging conditions such as *in situ* in histological sections or in a high background of wildtype molecules. The technologies will serve for genome-wide-, targeted multiplexed-, and clinical analysis needs. During the third reporting period work has progressed very well and the overall original plan has been respected. The efforts towards *in situ* sequencing in cells and tissue have reached a proof of concept stage and have been submitted for publication. The work in the WP is well integrated and the SME Olink is actively collaborating with several partners in the consortium. Olink and UU have spun out the SME Q-linea (www.qlinea.com) that commercialises the digital RCA approach pursued in this WP, initially for biodefense applications. We have also progressed well towards the development of a multiplex typing procedure for rare somatic variants in circulation, and for targeted multiplex methylation profiling. A unique approach to localized mutation detection in histology sections has been developed and published and the SME Olink will now adapt this technology to allow detection of oncogenic mutations directly in formalin-fixed paraffin embedded tissues. The WP aims to take certain technologies and apply them to diagnostic applications (e.g. *in situ* method will be developed for KRAS testing in biopsies).

WP6 (Training and dissemination) is responsible for the diffusion of know-how and results generated by READNA and training consortium members, the wider research community and the general public. Information has been diffused through the READNA website, a brochure, scientific publications and conference presentations. In the third reporting period several meetings and workshops have been organised. READNA has organized a total of 4 large workshops and in addition to the 1st and 2nd READNA Symposia on Advanced Nucleic Acid Analysis Methods that were organized in Berlin and Oxford, two large workshops have been organized in Paris involving speakers from the READNA consortium, both of the workshops had target audiences of over 200. The symposia were ideal opportunities for consortium members to expose current research being conducted within the READNA project. The READNA consortium was involved in organizing a workshop entitled "**d**e **n**ovo **G**enome **As**sembly assessment **P**roject workshop - **dnGASP**" in Barcelona (**5th - 7th April 2011**). The workshop was a collaborative effort among researchers to compare and evaluate methods and strategies for de novo genome assembly using data from 2nd generation sequencing platforms. A third symposium on Advanced Nucleic Acid Analysis Methods was organised in Barcelona (**28th September 2012**) involving speakers from READNA involving a target audience of **115** participants from both science and industry. In addition to READNA sponsoring speakers at two international sessions (**Paphos 2009 & Santorini 2011**) READNA also sponsored speakers for a session at the ESF 4th Functional Genomics & Disease Conference in Dresden (**15th April 2010**). The project brochure was distributed during the symposia and conferences, another effective dissemination tool. The mobility award program has been successfully completed with all **10** awards being used to initiate new interactions between READNA partners. The READNA website has been expanded and updated to ensure that it reflects the actual project situation. Finally the number of publications acknowledging READNA has surpassed **100** and **20%** are in high impact journals such as Nature Methods and Nature Nanotechnology.

WP7: Management

In the third reporting period the management structure has been reinforced. While maintaining the primary objective that management is solid and reliable while holding the administrative burden on researchers to a minimum. Communication to partners has been primarily by email. The website has also been used to communicate consortium activities and consortium sensitive activities can be found in the password protected section of the website. The 2nd plenary meeting was advanced to month 24 to enable consortium members to follow what progress had been made during the second year of the project. The 3rd Plenary meeting was organized at month 36 and allowed further refocusing to be done. A final plenary meeting was held in Barcelona at month 52 to allow partners an opportunity to present key results generated during the project. In addition the final plenary meeting was used to prepare the consortium for the final reporting. The SAB have been integrated into the consortium and have provided guidance in refocusing certain aspects of the project at month 24 during the 2nd plenary meeting. The SAB participated during the final meeting and once again provided their views and how they feel the post READNA should be handled. In addition the READNA project organized a mid-term review at month 32 in which the project was reviewed and evaluated to ensure that the project was performing according to plan by an external reviewer. A further improvement has been the increased frequency in coordination committee meetings (by phone) that has ensured important issues relating to the project have been discussed and resolved in a timely manner and that momentum of the project was maintained to the end of READNA. Each of the work package partners performed well within the planned project objectives and established interactions between workpackages have been further strengthened during the final reporting period.