Publishable summary 2nd Periodic Report (01/06/2009 – 30/11/2010)

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 in the first 30 months. 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 made great progress towards its overall objectives during the 2nd reporting period up to month 30. The five R+D workpackages have followed the projected evolution at the outset of the project.
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 and are ready for widespread deployment. An SME, HaloGenomics has been spun out with this objective. 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. Starting from now a component is integrated in READNA 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. Several of the tasks in this WP have reached completion and efforts are being shifted to WP3, WP4 and WP5.

WP3 (**Fluorescence-based Single Molecule Sequencing**) has prepared the stage for development of fluorescence-based 3rd 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 have been 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.

WP4 (**Nanopore Sequencing**) 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. Progress has been delivered on track with the workpackage stated timelines.

WP5 (**New Genotyping Challenges**) is concerned with the development of methods for nucleic acid analyses for which currently no good, economic methods exists. On one hand *in situ* genotyping and sequencing methods were developed. This *in situ* genotyping procedure has successfully been developed while the *in situ* sequencing has provided very promising pilot results. These methods are being implemented for kRAS testing in biopsies. A mass spectrometry method for rare variant detection such as kRAS in circulation is advancing well with the proof-of-concept of a single-tube multiplex genotyping method that has been brought together with a highly specific allele-specific procedure.

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. So far READNA consortium members have published over 30 papers with support from READNA (Excel table of all READNA publications attached). 30% of these were in journals with a high impact factor for technological development such as Nature Methods and Nature Nanotechnology. READNA has organised 2 large public conferences, has contributed 3 sessions at international conferences and has organised 2 focussed invitation only workshops with around 30 participants. So far 7 staff exchanges between consortium members have been organised.

WP7 (**Management**) READNA is supported by the management workpackage that assures smooth operation.

The work of the READNA consortium has also led to the creation of two start-up companies (Q-linea and HaloGenomics).

READNA is pushing 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 for early detection of disease onset, identification of disease susceptibility and decisions on the best mode of treatment of disease. READNA is preparing the toolbox required to turn personalized medicine and ultimately personalized healthcare into reality. The methods developed will also find their application in other disciplines that rely on nucleic acid analysis, such as for the optimization of species for example in an agricultural setting. READNA can form the basis for improved quality of life.

WP1: Near Term Innovations: Ancillary elements for 2nd generation DNA sequencers

In the second reporting period, WP1 has extended the progress made in its first reporting period, towards evaluating and optimising systems for enriching genome sub-regions so that they provide input material for 2nd generation DNA sequencing. Specifically, we have now completed an empirical comparison of all existing methods, and thereby shown that: solution phase methods perform best; no hybridisation methods enrich to the degree achieved by amplification methods; and all methods suffer from considerable unevenness of target region recovery. These experiences not only help us define an optimum current method, but also guide us in creating our own effective and low-cost hybridisation based procedures. Regarding amplification based methods, we have made good progress in understanding why different genome regions recover with differing efficiency, and we have optimised the selector method to the point that it performs exceptionally well (now commercially available). An extension of the selector concept has even enabled single molecules to be amplified as circles to provide sequencing templates *in situ* - thus opening up exciting new possibilities for single cell analysis. Rare Mutation Enrichment methods are also being explored. Beyond this lab work, we have specified a core set of objective metrics for experiment reporting, and are looking into ethical dimensions of human genome sequencing by the kinds of methods being considered in the whole of READNA.

WP2: Near Term Innovations: Improvement and extension of existing methods

The objective of this WP was the advancement of methods that had passed their proof-of-concept at the outset of the project to a point where they could be deployed in studies or even a diagnostic device that could be generated based on this implementation. WP2 is made up of 5 tasks. 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 MS task focussed on the development of a system for sample characterisation that uses DNA, RNA and protein profiles analyzed by mass spectrometry. Four of the five tasks are completed while the fifth task will be completed in the remainder of READNA.

WP3: Fluorescence-based Single Molecule Sequencing

In this second periodic report Workpackage 3 has made progress on two fronts: (1) Sequencing biochemistry; (2) Towards sequencing in an ultra-long range context. We started by investigating a number of novel concepts for sequencing. It is clear that we have 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 RNA expression. The second road is real-time sequencing and its application on long-stretched molecules. Here one method reached a dead-end very quickly but a new method has arisen from its ashes. Progress on a second approach has been slow beyond the excellent proof of principle that was obtained in the first reporting period and additional resources are now available from a commercial partner.

Meanwhile, we have made massive strides in handling and imaging of ultra-long genomic DNA fragments. 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. This latter facility has enabled us to conduct enzymatic reactions on DNA molecules stretched in nanofluidic channels. Finally, substantial progress has been made on developing super-resolution and nanometric localization approaches that go beyond the capabilities of conventional diffraction-limited light microscopy.

WP4: Nanopore Sequencing

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 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 first 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.

WP5: New Genotyping Challenges

The work in this WP is devoted to solving genotyping and epi-genotyping challenges, where the performance of current technology is not satisfactory. Existing nucleic acids analysis methods will be adapted for DNA methylation analysis and copy number variation analysis which have to deliver very precise quantitative results. Methods are also being 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. The work is progressing very well and is overall following the original plan. The efforts towards *in situ* sequencing in cells and tissue have reached a proof-of-concept stage. 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 biodefence applications. We have also progressed well towards the development of a multiplex typing procedure for rare somatic variants in circulation. One patent has been filed based on work in the WP during the first 30 months as well as several papers submitted and published. For example, 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 work in this WP aims to take the technologies to diagnostic applications.

WP6: Training and dissemination

In the second reporting period several meetings and workshops have been organised. The 1st and 2nd READNA Symposia on Advanced Nucleic Acid Analysis Methods that were organized in Berlin and Oxford, respectively, had target audiences of 100 participants. The symposia were ideal opportunities for consortia members to expose current research being conducted within the READNA project. The READNA consortium was also responsible for organizing a workshop entitled "Tools for DNA sequencing and resequencing in diagnostics" in Berlin (14th of January 2010). The workshop regrouped three EU consortia READNA, TECHGENE and EuroGentest and had 25 participants. The workshop addressed how sequencing technologies can be further developed and ultimately integrated into a diagnostics laboratory, consisting of presentations from researchers working on the technologies and clinicians that are working with patients. READNA also sponsored speakers for a session at the ESF 4th Functional Genomics & Disease Conference in Dresden. The project brochure was distributed during this conference, another effective dissemination tool. The mobility award program is on schedule and a 7th award is currently in progress out of the 10 available in the program. The READNA website has been expanded and is regularly updated to reflect current progress. Finally the number of READNA acknowledged publications has surpassed 30 and 8 are in Nature journals.

WP7: Management

In the second 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 has been made during the second year of the project. The SAB have been integrated into the consortium and have provided guidance in refocusing certain aspects of the project at month 24. A further improvement has been the increased frequency in coordination committee meetings that has ensured important issues relating to the project have been discussed and resolved in a timely manner. Each of the work package partners are performing well within the planned project objectives and interactions between workpackages are now being nurtured.