PROJECT DELIVERABLE

Project acronym: **REQUITE**	GA number: **601826**
Project title: **Validating predictive models and biomarkers of radiotherapy toxicity to reduce side-effects and improve quality-of-life in cancer survivors**	
Funding Scheme: **Collaborative Project (FP7-HEALTH-2013-INNOVATION-1)**	
Project start date: **01 October 2013**	Duration: **60 months**
Project's coordinator: **Prof Catharine West (University of Manchester, UK)**	

Deliverable no.: **4.3**	Title: **Report on the selection of SNPs and copy number variants for validation, the genotyping of samples and descriptive analysis on REQUITE website**	
Due date: **Month 60 (30 September 2018)**	Actual date: **Month 62 (23 November 2018)**	
Aim of the Deliverable: To report on the genotyping undertaken of blood samples collected in the observational study (WP2) and processed by the CIGMR biobank (WP3). Describe the SNPs and CNVs identified as being associated with side effects following radiotherapy and that were included to improve tumour specific prediction models. **Deliverable D4.3 has been partly achieved with full analyses completed after the end of the project**.		
Lead beneficiary for this deliverable: **Chris Talbot (B4)**		
Personnel involved: **Catharine West, Rebecca Elliott, Antony Payton, Debbie Payne (B1); Petra Seibold (B2); Adam Webb, Kerstie Johnson (B4); Alison Dunning, Craig Luccarini, Laura Fachal (B6), Tom Burr, Teresa Coughlan (B11)**		

Dissemination level:		
PU	Public	**X**
PP	Restricted to other programme participants (within the Commission)	
RE	Restricted to a group defined by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

1. Objectives

A core aim of the REQUITE project was to find and validate genetic predictors of radiotherapy side effects. To do this the project collected blood samples from each patient at the time they were recruited, before radiotherapy.

2. Sample Collection & Processing

Blood samples were stored locally and sent in batches under temperature-controlled conditions to a laboratory at the University of Manchester that specialises in banking biological specimens: the CIGMR Biobank. At CIGMR the blood tubes were stored in ultra-cold freezers (-80C) until patient recruitment had been completed. CIGMR then used a robotic system to extract DNA from the blood and make sure the DNA samples were of the same concentration.

The DNA samples were then dispatched to a laboratory at the University of Cambridge that specialises in genetic analysis. They carried out testing for 533,631 genetic markers. The markers tested were single nucleotide polymorphisms (SNPs), which are places in the DNA at which patients can have different letters of the DNA code. SNPs occur normally throughout a person's DNA. SNPs that are particularly relevant to cancer and radiotherapy were scored (genotyped) using a high-throughput, low-cost system called Illumina OncoArray-500K.

3. SNP Genotyping & Imputation

Overall, 4,442 REQUITE samples were genotyped, and passed through a validated multi-step quality control pipeline. The process filters out samples with poor quality DNA and SNPs with low call rates or abnormal results that are unexpected when working with high quality DNA. In genome wide association studies it is important to control for ethnicity, which can otherwise lead to false positive and negative associations. For that reason patients with >80% European ancestry, as defined by ancestry informative SNPs, were included in the main analyses. Patients of other ethnicities will also be studied and included in future analyses subject to consideration of statistical power.

After these quality control steps data were retained for 483,517 SNPs on 4,223 patients. Data were also integrated from additional lung cancer patients from the CONVERT trial collected via University of Manchester and also cohorts from University of Gent. The inclusion of these lung samples increased the cohort size such that raw genotype data are available for 4,634 patients and after the quality control steps data were retained for 4,304 patients of which there were:

- 1,948 breast cancer cases
- 1,728 prostate cancer cases
- 628 lung cancer cases

Following the quality control step, imputation was performed of the genotypes of SNPs and copy number variants not directly scored by the Oncoarray. Imputation is the process by which the most probable genotype of other SNPs in the genome is calculated based on the known (scored) genotypes. Imputation was carried out using IMPUTE2 software version 2.3.2 and 1000 Genomes project data as the reference panel. A total 21,465,139 polymorphisms were imputed of which 13,077,651 passed imputation quality (r^2 metric >0.3 threshold). Considering common polymorphisms minor (allele frequency >0.05); a total of 7,431,964 were imputed of which 7,282,498 passed the quality threshold.

In summary, after quality control, a total of 2.04 billion genotypes were directly scored, and 55.2 billion imputed genotypes were recorded. These data are now stored centrally and are available for genome wide association analyses once final checks on the two-year follow-up data are completed. Work will continue beyond the end of the project funding.

4. Timing of Radiotherapy Delivery

By project end a genetic analysis was completed of circadian rhythm effects on radiotherapy response. Using a different breast cancer cohort (LeND), colleagues at the University of Leicester have shown a difference in side effects occurring two years after treatment between patients given radiotherapy in the morning and those given radiotherapy in the afternoon. Genotyping data on the REQUITE breast cancer patients from Leicester validated this finding and were used to show that variation in the *PER3* and *NOCT* genes determined whether patients showed this 'time-of-treatment effect' or not. This was published in Clinical Oncology journal (Johnson et al, 2018). The work will now be replicated in all the REQUITE cohorts, to try and confirm the result and repeated in prostate and lung cancer patients. If the result is replicated it would open the way to predict genetically which patients would benefit from being treated at a particular time of day – a cheap and easy way to personalise radiotherapy delivery.