IMI1 Final Project Report Public Summary

Project Acronym: ONCOTRACK

Project Title: OncoTrack - Methods for systematic next generation oncology biomarker development

Grant Agreement: 115234
Project Duration: 01/01/2011 - 31/12/2016

1. Executive summary

1.1. Project rationale and overall objectives of the project

Cancer is a disease characterized by the 'hallmarks of cancer', the biological attributes that tumour cells acquire during pathogenesis of a cancer and distinguish them from normal cells[4]. A typical feature of all malignancies is genome instability, which generates the genetic variability that fosters acquisition and maintenance of the biological hallmarks. Colorectal carcinoma is typical of most solid tumours in that it is a clinically heterogeneous disease, with only a fraction of the patients responding to available therapies. Accurate diagnosis and in particular selection of the most appropriate therapy for the individual patient remains a major unmet need in oncology.

The scientific programme of the OncoTrack consortium was based on the premise that establishing effective precision medicine for the indication requires a better molecular understanding of the disease and the validation of robust biomarkers for patient stratification. Our goal was therefore to recruit a clinically homogeneous series of patients and to conduct an exhaustive analysis of their tumour tissues as a basis for biomarker discovery. We recruited patients with both primary cancers and metastatic lesions and compared the genomic sequences of the cancers with the patient's germline DNA. In addition, we analysed the transcriptome, methylome and proteome of selected samples.

In order to collect information on the response of the tumours to therapeutic intervention, we established both xenograft models in nude mice as well as organoids in three-dimensional cell cultures. These have been treated with a panel of standard therapeutic agents as well as with experimental agents targeting the major intracellular signalling pathways. We hypothesize that markers derived from this systematic approach to analysis of cancer biology can be used to guide the optimal therapeutic strategy at the level of the individual patient, and will also provide on-going prognostic guidance for the clinician. In parallel to the molecular analysis of the tumours, we are developing powerful new molecular assays designed to provide less invasive diagnostic procedures for the cancer patient, with emphasis on the use of patient derived models for assessment of drug response, histological analysis of patients' tumours, and quantitative analysis of plasma proteins and circulating free DNA

1.2. Overall deliverables of the project

The project was designed to advance our understanding of the fundamental biology of colon cancers,

but this innovative approach to the identification of biomarkers and the diagnostic technologies under development should be generally applicable to other solid cancers. This project therefore has the potential to impact the delivery of personalized therapy to colon cancer patients directly, but in the long term may benefit a broader spectrum of tumour patients.

The major objectives of the study were as follows:

1.) To identify and validate biomarkers that advance our understanding of tumour heterogeneity; using advanced metastatic colon cancers as an example

2.) To develop and apply innovative, sensitive, specific and accurate methods to detect biomarkers from limited quantities of tissue collected by less invasive methods, suitable for use in routine clinical

laboratories

3.) Develop and apply new sensitive, specific and accurate methods to detect and characterize circulating tumour cells and an understanding of their utility in assessment of prognostic, predictive and/or pharmacodynamic markers

4.) Develop and apply novel methods for identification and quantification of cancer stem cells and an understanding of their utility as prognostic, predictive and/or pharmacodynamic markers

5.) To develop and apply techniques for the evaluation of circulating tumour DNA, RNA or protein

6.) Establishment of uniform SOPs and analytical standards for accepted biomarkers that conform to

expectations of the regulatory authorities
7.) An education and training plan for dissemination of relevant skills among the collaborating sites

1.3. Summary of progress versus plan since last period

The work of the OncoTrack consortium in the final reporting period was focussed on the timely completion of milestones and deliverables scheduled for the final year of the project and dissemination of the achievements of the consortium. The reprioritization of the analytical programme and rescheduling of the outstanding Milestones and Deliverables was reported in the updated Description of Work, submitted with Amendment 9, dated September 28th 2016. With the exception of deliverables that could not be achieved for technical reasons (see Sections 2.2 and 2.3 of the report), the consortium was able to meet all scientific goals of the Description of Work.

WP1: Genome Sequencing and Proteomic Profiling

Sequencing and Data Analysis:

WP1 completed all sequence analyses for the so-called 'complete patients' for inclusion in the consortial publication. 'Complete Patients' are those for whom tumour tissue, normal tissue/peripheral blood and a xenograft (PDX) and/or 3D cell culture (organoids, PDO) are available. From these patients, whole exome (in some instances whole genome) sequencing, transcript sequencing and methylome analysis were conducted. In addition, confirmatory genome sequencing and transcriptome analysis was performed on the xenograft and cell culture models. Drug response data for the models was provided by WP2 (PDO) or WP6 (PDX).

In collaboration with WP8, a complete analysis of the genomic landscape of the patients' tumours was performed and correlation analysis to relate drug response of the biological models to their genomic and transcriptomic characteristics. Details of this work have been published and are described in the Summary Report for WP1 in Section 2.4 below.
Additionally, sequence data was generated for newly recruited patients and for fully characterized PDX and PDO models as they became available. Summaries of the sequencing libraries that were generated and the source materials are shown in Figure 1 below.

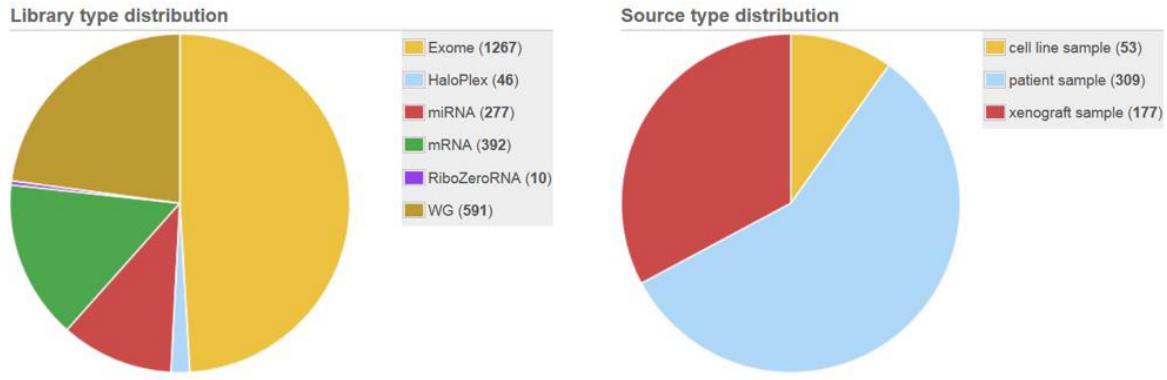

Figure 1: *Graphical summary of the sequencing libraries that were generated and the source materials, as deposited in the OncoTrack DataBase, February 25th 2016.*

Proteomic Profiling:

Our original intention was to perform data-driven proteome analyses to obtain an overview of proteins and protein modification states in the samples, using established analysis pipelines at the MPI-MG, based on a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Scientific), to be used primarily for the analysis of the tumours and xenografts and a similar pipeline at the Broad Institute (Cambridge, MA, USA), focusing primarily on the 3D cell culture (PDO)/cancer stem cell samples (financed through Bayer AG and coordinated through WP2). In particular, the work at the Broad

institute should focus on analysis of phospho-epitopes to provide information about the activation state of signal transduction pathways. Generation of sufficient protein from the PDO cultures, however, proved to be technically and financially unrealistic and it was therefore decided to perform the analyses at Broad with xenograft material. It was also decided to supplement the activities at the Broad with additional measurements of phosphopeptides in collaboration with the leading mass spectrometry service company Proteome Sciences Inc. (Cobham, Surrey, England). Proteome Sciences uses a proprietary data analysis system, SysQuantR, a phosphoproteomic workflow that in combination with Tandem Mass Tag labelling (TMTR) permits multiplexed analysis of pooled samples.

WP2: Cancer Stem Cells

Patient tumour-derived cell cultures were established as planned until the acquisition of primary tumours and metastases was terminated. Overall, the success rate of cell culture generation was approximately 60%. A total of 93 primary cell cultures was registered in the OncoTrack Database.

Cellular materials were distributed to partners upon request. As described in the deliverable D.2.07, cancer stem cells were characterized in the following ways: i) FACS sorting from heterogeneous cell populations, ii) transcriptome analysis of sorted cells, iii) identification of stem cell markers, iv) in vivo
testing of tumour-forming potential. Drug response data for the standard testing panel were generated and provided to WP1 and WP8 for inclusion in the correlation analyses.

WP3: Circulating Tumour Cells

With the established amplification protocol, we obtained an adequate yield of amp-cDNA. Due to the
ribosomal background in the samples, however, it was necessary to develop a ribosomal depletion method prior to sequencing the samples. None of the tested methods led to satisfactory results, leading us to delay sequencing of these valuable samples. The technologies for alternative approaches, using targeted analysis such as Panel-Seq or Exome Re-sequencing are still under investigation. Thus, we had to postpone the final sequence analysis of the CTC samples until a time point after the end of OncoTrack.

WP4: Plasma: DNA and Proteins

The objective of WP4 was to establish biomarkers, procedures and technologies for minimally invasive companion diagnostics and evaluation of drug responses, by developing an efficient workflow
for high-resolution analysis of biomarkers. In the final reporting period, the work on the outstanding deliverable D04.02, “Companion diagnostic assay platform for selected biomarkers” could be completed.
A versatile FRET-based biomarker quantification platform was developed for highly specific and sensitive multiplex detection of proteins – for which high-quality affinity binders are available – nucleic acids (DNA and RNA), and any biomarker that can cleave a specific peptide sequence. The assay performs in solution and in a single incubation step without requirement for any washing or separation steps. The platform can therefore be used for routine, high-throughput assays using very small samples volumes of patient blood, serum or tissue extracts.

WP5: Pathology Samples

The activities of WP05 progressed as planned and the work package deliverables were completed with minor deviations resulting from technical issues. Major progress was made in the use of primer extension assays (PEA) for assay of proteins in solution, correlation of protein biomarker levels with drug treatment (D05.06) and in the innovative *in situ* expression profiling

methodology[5], (D05.03 and D05.07). The latter two reports demonstrate that a description of the spatial distribution of transcriptional activity of relevant biomarkers in tissue sections is now possible and that this can be applied to analysis of tumour heterogeneity.

Using the example of the interaction between the HER1, HER2 and HER3 membrane receptors, FRET imaging studies demonstrated the qualification and quantification of protein-protein interactions (D05.01) The successful application of proximity ligation assays (PLA) to detection of biomarkers in cultured tumour cells was reported (D05.04). Attempts to apply the same methods to fresh tissue sections identified several technical problems that the partner UU has addressed using sources of funding other than OncoTrack. This caused a delay in transferring the specific assays developed to analysis of signalling pathways relevant for colon cancer.
The efforts have brought about changes that will facilitate analysis of signalling pathway activity relevant for diagnostics and for selection and monitoring of therapy in future projects.

WP6: Xenografts

For WP6 all milestones have been achieved and all deliverables completed with the end of the project. In the last 12 months the characterization of the PDX models has been completed, the data have been provided to the consortium and used for the correlation analyses of chemoresponsiveness
with the genetic profile of the tumour which have been performed by WP1 and WP8.
As during the project course it has become clear, that all colon tumours are derived from CSCs (now the generally accepted CSC model), all of the xenografts that were established and characterized for chemoresponsiveness by WP06 should be considered as CSC derived. Additional drug treatment of CSC derived xenografts (as planned for Deliverable D06.07) was therefore unnecessary and so was not carried out.

WP7: Sample Logistics

The recruitment was continued as planned and was terminated in June 2015. A total of 44 patients (36 colon cancer, 8 rectal cancer) were recruited for cohort 1 and 217 patients (127 colon cancer, 84 rectal cancer, 5 colon and rectal cancer) were recruited for cohort 2. Therefore the milestones for acquisition of 30 patients for cohort 1 and 60 patients for cohort 2 were completed. The accrual of a further 100 patients with primary tumour including 45 high risk patients (M07.02, M07.03) was only partially completed because of the more demanding inclusion criteria, which reduced the number of eligible patients.
Acquisition of metastatic and primary tumour material from 30 of the patients from cohort 2 was partially completed. Due to the low frequency of disease progression during the project time frame, only 15 patients were recruited.
Fresh or shock-frozen tissue, blood plasma and tissue sections for histological analysis were distributed to partners in the consortium according to the Description of Work. The collection of pertinent follow-up patient information and the subsequent completion of a clinical database was completed according to plan.

WP8: Data Integration, Statistical Analysis and Modelling

Thanks to the efforts from the legal team in resolving issues of legal constraint, the collaboration with the IMI eTRIKS consortium was reactivated and a working group is finishing the final loading of data to the tranSMART instance hosted by the Centre de Calcul IN2P3/CNRS in Lyon.
Mass Spectrometry data has now been integrated into the OncoTrack DB and database functions optimised, including the development of visualisation tools for these data types.

1.4. Significant achievements since last report

WP1: Genome Sequencing and Proteomic Profiling

The major activities in WP1 in the reporting period centred on analysis and interpretation of data. As described in the fifth Periodic Report (2015), the consortium has concentrated its efforts in the preceding year on achieving the data generation for all samples belonging to the so-called "frozen list" of tumours. This collection includes patients from both Cohort 1 (patients recruited with metastatic disease) and 2 (patients presenting at first diagnosis with a primary tumour) for whom a xenograft and/or a cell culture model is available. Following completion of all sequencing activities and testing of the standard drug panels in the xenograft and cell culture models, WP1 proceeded with a deep molecular characterization of the tumours and their derived models. With assistance
of WP2 and WP6 who provided drug response data, an analysis of the molecular profiles associated with drug responses was completed. This work, summarizing the efforts of the Oncotrack consortium as a group, has been recently published in Nature Communications (Schutte et al. February 2017, http://rdcu.be/o9U8). Further, two patents have been registered describing novel biomarkers of drug response, as a direct follow up of this work (EP 16 19 6277.4
and EP 16 19 6281.6).
Activities in the field of proteomic profiling by mass spectrometry started after genomic derived data generation, with a focus on analysis of patient tumor material and PDX tumors from vehicle treated animals. Measurements are still ongoing at the Broad Institute and final data analysis is expected to be completed in the first half of 2017.

WP2: Cancer Stem Cells

Tumour xenografts were established from FACS-sorted patient tumour-derived cell cultures and characterized at the level of marker expression and transcriptome. *In vitro* and *in vivo* preclinical models are easily interchangeable. Full molecular characterization was completed for 46 patient tumour-derived cell cultures, 19 tumours were modelled *in vitro* and *in vivo*.
A comparison of the chemo-responsiveness of patient derived organoids in culture and the corresponding xenograft demonstrated the potential of PDO models for discriminating the effects of drug treatment on CSCs and bulk tumour cells. (see D06.07).

Using FACS, we were able to identify ALDH-Positive cells as CSCs. Whole transcriptome analysis of $ALDH^{Positive}$ cells from PDO and xenografts demonstrated that $ALDH^{Positive}$ CSCs are enriched for stem cell associated developmental gene sets. Genes differentially expressed in $ALDH^{Positive}$ cells from both PDO models and xenograft models were selected for functional analysis by small interfering RNA (siRNA)-mediated gene knockdown. Genes of interest were functionally tested by serial passaging of siRNA-transduced cells in non-adherent culture to determine their effect on CSC self-renewal and survival . These experiments demonstrated that Early Growth Response 2 (EGR2) and the glucose transporter Solute Carrier Family 2 Member 12 (SLC2A12, or GLUT12) are required for CSC growth and survival in non-adherent cell culture (see M06.04 and D02.07).

WP4: Plasma: DNA and Protein

Companion diagnostic assay platform for selected biomarkers

Using the commercial fluorescence clinical immunoassay reader KRYPTOR compact plus as the assay detection system, we developed several multiplexed FRET-based assays against protein biomarkers (immunoassays), micro-RNA biomarkers (hybridization assays), and enzymes (kinetic peptide cleavage assays) (see D04.02 for details).

WP5: Pathology Samples

PEA based proteome analysis of xenografts (UU)

The multiplexed, primer extension assay PEA Proseek Oncology Panel II was used to perform proteome analysis on extracts from xenograft tissue. This assay is a high throughput, multiplex immunoassay enabling simultaneous analysis of 92 inflammation-related protein biomarkers. The tissue samples were categorized into two groups; cetuximab responders and cetuximab nonresponders,
based on the xenograft responses to cetuximab treatment. Mice in each group were treated either with standard-of-care compounds (Oxaliplatin, Irinotecan, 5-FU), compounds targeting EGFR (cetuximab, AZD8931, Afatinib), MEK inhibitor (AZD6244), compounds targeting VEGF (Avastin, Regorafenib, Nintedanib), or compounds targeting oncogenic pathways (mTOR FR, IGF-1/2mAB, AZ1, Volitinib). Drug treatments were performed for both analysis of both short-term (24 hr) and long-term (up to 4 weeks) effects.
Analysis of 177 samples after short-term drug treatment revealed that the expression levels for several proteins were significantly altered when we compared the cetuximab responder and non-responder groups (for details see D05.06). CXL17 (a chemoattractant for monocytes, macrophages and dendritic cells), ABL1 (a protein tyrosine kinase involving in cell division, adhesion, differentiation, and stress response) and TLR-3 (a member of the Toll-like receptor family that plays roles in activation of innate immunity) are examples of common proteins that were significantly altered by two or more drugs (individually treated) in PDX from cetuximab responders, but not from non-responders. These proteins commonly altered upon drug treatment in cetuximab responders, may have prognostic and/or therapeutic value.
An extended validation PDX cohort including short- and long- term drug treatments is currently being analysed with the above mentioned Oncology Protein Panel and an Inflammation Protein Panel. This work on xenograft-derived samples complements the earlier reports on plasma protein levels (D04.03) and the corresponding data from tissue analyses (D05.05) in OncoTrack patient clinical samples.

Characterization of heterogeneity within tumours using in situ expression profiling (SU)

Development of an *in situ* gene expression profiling assay that can be used to localize transcriptional activity of gene(s) of choice in cells and tissue sections has been reported previously (ref. 5). This method was applied to characterize the IGF-2 expression heterogeneity among multiple PDX models derived from independent samples obtained the same original patient tumour (150-ML-M). Figure 2 below demonstrates the results of analysis for IGF2 expression, together with control markers for tumour and stromal components in PDX tissues (for further details see D05.03).
This work was expanded to examine the sensitivity and specificity of the assay for a panel of 64 transcriptional markers *in situ* in sections of 135-ML-P human tumour tissue. The assay was validated using the cell lines A549 and U2-OS, for which RNA sequencing (RNAseq) data was available.
To investigate whether morphologically distinct features of the tumour 135-ML-P also displayed distinct molecular profiles, we compared the *in situ* sequencing counts in two regions of interest
(ROI) with clearly different morphology in the two consecutive sections 135-ML-P-TC-01-09 and 135-ML-P-TC-01-10 (Figure 3). The two ROIs differ markedly in their stromal components. ROI 1 has a lymphoid follicle, some perivascular tissue and some epithelial cells whereas ROI 2 consists mostly of fatty tissue surrounded by fibrous components. The correlation between the data from the two morphologically distinct areas were much lower within the tissue sections than

between the same areas in the two separate sections, proving that our assay can reveal tissue heterogeneity in molecular profiles.

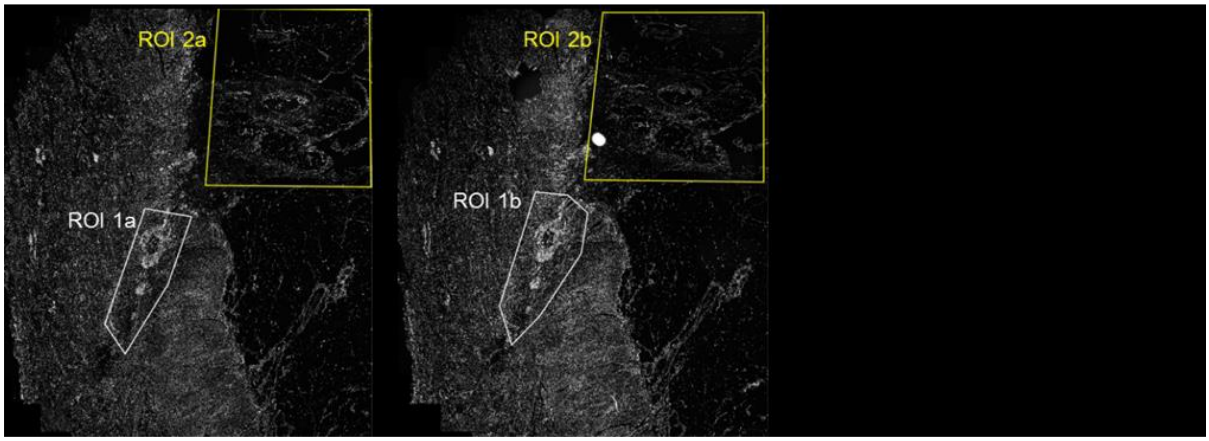

Figure 3. In situ sequencing data for our 64 plex assay was analyzed in two distinct areas (ROI 1 and 2) in the two consecutive sections a: 135-ML-P-TC-01-09 and b: 135-ML-P-TC-01-10.

Spatial distribution of the markers

To visualize the distribution of the genes across the tissue section 135-ML-P-TC-01-09, we generated density maps of in situ sequencing reads (Figure 4). The density maps clearly show that many markers have very different spatial distribution across the tissue section.

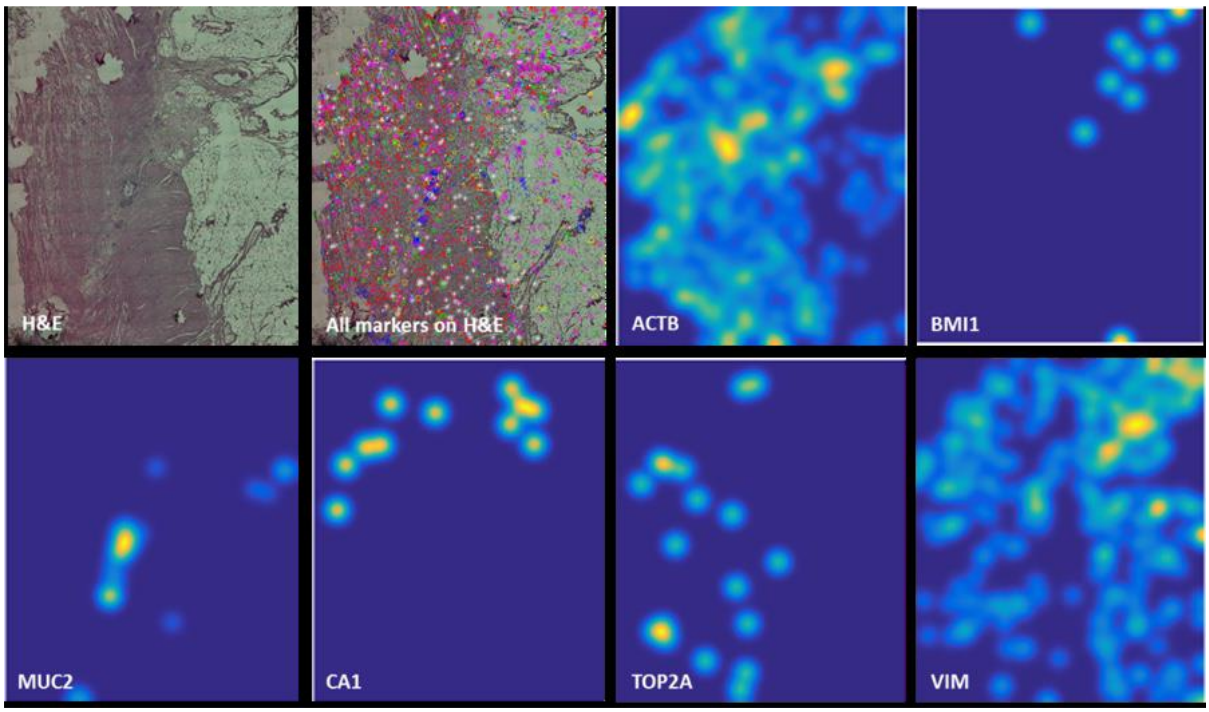

Figure 4. H&E staining and maps of situ seq reads across the tissue section 135-ML-P-TC-01-09. From upper left corner: H&E stained section without and with all in situ seq read on top. Density maps of ACTB in situ seq reads and some markers with distinct spatial distributions across the tissue (BMI, MUC2, CA1, TOP2A, and VIM).

Novel approaches for pathology assays (UPS)

The Partner Universite Paris-Sud in collaboration with Drs. Weitsman, Barrett and Ng (University College London) and WP6 obtained preliminary results from three patient-derived xenograft models (PDX) that show different base-line levels of HER1-HER2 and HER2-HER3 dimers (Fig. 5, black bars – control). After treatment of the PDX models with cetuximab, the dimerization of HER1-HER2 was not changed in a tumour that was highly sensitive to AZD8931, significantly increased in a moderately sensitive sample and disappeared in a resistant sample. Interestingly, the HER2-HER3 dimer was found at high levels in all samples and was significantly decreased after cetuximab treatment in sensitive and non-sensitive samples. These results suggest adaptive rewiring of HER family receptor signalling in response to cetuximab and require further investigation of HER family dimers in the rest of available samples to use this data in multiparametric modelling of the response to inhibition of HER1, HER2 and HER3 (for details see D05.01).

WP7: Sample Logistics

A total of 44 patients (36 colon cancer, 8 rectal cancer) was recruited for cohort 1 and 217 patients (127 colon cancer, 84 rectal cancer, 5 colon and rectal cancer) were recruited for cohort 2. The clinical database was completed.

WP8: Data Integration, Statistical Analysis, and Modelling

The OncoTrack DB continues to serve as a major platform for data exchange within the consortium and will be able to receive data until December 2017.
Implementation of MS data format and access / visualisation / download functions to extend the OncoTrack data integration platform.
Major progress in resolving legal issues for the collaboration with IMI-eTRIKS partners on tranSMART server hosting and data curation/loading. The tranSMART server hosted by IMI-eTRIKS at the Centre de Calcul IN2P3/CNRS in Lyon was set up with the latest eTRIKS-V3 platform in the 4th-quarter of 2016.
Major progress has been made in data analysis and correlation of molecular markers with drug response, resulting in the identification of predictive biomarkers, as reported in the consortium manuscript: Schutte, M. et al. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. Nat. Commun. 8, 14262 doi: 10.1038/ncomms14262 (2017).

1.5. Scientific and technical results/foregrounds of the project

WP1: Genome Sequencing and Proteomic Profiling

The large body of genome sequencing and proteome data represents a major part of the OncoTrack consortium foreground and one that is a valuable resource for future research projects in colon cancer. A significant part of this data has already been deposited in the European Genome-Phenome Archive (EGA) under accession number EGAS00001001752. The remaining data will be uploaded to the archive following completion of a data hosting agreement with EGA and ELIXIR. (See also Sections 4.4 and 5.2).

WP2: Cancer Stem Cells

PD3D models were heterogeneous and varied in frequency of CSCs, as shown by gene expression analysis and immunofluorescence staining. Heterogeneous gene expression patterns involved factors associated with self-renewal, proliferation, and differentiation. To enrich for cancer stem cells, a sub-fraction of ALDH-positive cells was sorted by FACS and subjected to further analysis for markers and tumour-forming potential. Genes associated with stem cell function were functionally tested by serial passaging of siRNA-transduced cells in non-adherent culture to determine their effect on CSC self-renewal and survival. These experiments

demonstrated that Early Growth Response 2 (EGR2) and the glucose transporter Solute Carrier Family 2 Member 12
(SLC2A12, or GLUT12) are required for CSC growth and survival in non-adherent cell culture.
WP2 established a repository of patient-derived 3D cell cultures (n>60). Cryo-stocks of cultures were deposited in the biobank at the Charite Universitatsmedizin Berlin. Forty-six cultures were fully characterized at the level of genomics and transcriptomics. These cultures represent an important resource for future research on the biology of cancer stem cells as well as for pre-clinical testing of novel drugs, drug combinations or drug repurposing. In personalized cancer treatment approaches, the cultures may serve as models for colorectal cancers exhibiting similar genetic make-up such as identical or functionally equivalent driver mutations and pathway activation patterns.

WP4: Plasma: DNA and Proteins

D04.02: Companion diagnostic assay platform for selected biomarkers
The performance of the multiplex homogeneous FRET immunoassays was demonstrated by detecting various biomarkers from a single serum sample. The selected biomarkers were neuron-specific enolase (NSE), carcinoembryonic antigen (CEA), the cytokeratin-19 fragment (Cyfra21-1), squamous cell carcinoma antigen (SCC), and the carbohydrate antigen (CA15.3). All biomarkers could be detected in a multiplexed manner from a single 50 µL serum sample at concentrations below the clinical cut-off values.
The homogeneous FRET-based assay was also adapted for detection of circulating nucleic acids in the companion diagnostic assay platform. We developed homogeneous FRET assays against micro-RNAs (miRNA) miRNA-20a, 20b, and 21 as model micro-RNAs and with significance for cancer diagnostic. The use of three different acceptor dyes allowed for simultaneous detection of the three different miRNAs. By applying a miRNA-specific padlock probe and an amplification step using rolling circle amplification, the assay sensitivity was improved by three orders of magnitude.

WP5: Pathology Samples

PEA based proteome analysis of xenografts (UU)
We found that a set of proteins is significantly altered between cetuximab responding and nonresponding
PDX samples, which underwent different drug treatments. An extended validating
PDX cohort including short- and long- term drug treatments is currently being analysed with the oncology protein panel and an Inflammation protein panel. We are further performing a comprehensive analysis on our data combined with patient's clinical information such as mutational status of genes, age, gender, and survival or death etc.

In addition, we implemented the recently developed proximity ligation assay coupled rolling circle amplification method (PLARCA) to assess 50 tissue lysate samples (25 tumour and 25 healthy tissue) from OncoTrack cohort and plasma samples to validate the technology for sensitive detection of proteins. PLARCA was able to detect proteins that were below the detection threshold for ELISA with p-value < 0.001.

Characterization of heterogeneity within tumours using in situ expression profiling (SU)

Our *in situ* expression profiling of IGF-2, stroma, and tumour markers in the 150-ML-M tumour samples. The analysis visualized the distribution of two distinct classes of tumour cells within the original patient derived tumour - one with low IGF-2 expression, and one with high. This provides
an explanation for the heterogeneity observed among the PDXs derived from this tumour.

WP6: Xenografts

WP6 established a collection of 106 patient-derived xenografts. Cryo-stocks of frozen tumour samples are maintained by Experimental Pharmacology & Oncology Berlin-Buch GmbH. Of these, 86 PDX have been fully characterized with regard to their response towards to a panel of 3 standard of care and 9 investigational drugs and 72 were fully characterized at the level of genomics and transcriptomics. These *in vivo* models represent an important resource for pre-clinical testing of novel drugs, drug combinations or drug repurposing and for correlation of drug response with molecular parameters as a basis for biomarker discovery.

WP8: Data Integration, Statistical Analysis, and Modelling

A consortium manuscript has recently been published (Jan 2017) in Nature Communications 'Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors', Schutte et al., 2017, Nature Communications.
The collaboration with the IMI eTRIKS consortium and technical setup of the TRANSMART platform has been summarized in a joint manuscript that is close to being finished. The manuscript will be submitted in early 2017.
OncoTrack-generated omics and response data were analysed through the ModCell™ system to generate predictions based on the classification of responder and non-responder from xenografts and 3D cell models from selected patients. For this, the ModCell™ system and the underlying PyBioS modelling software were applied. For the drug response modelling of the OncoTrack samples the following steps were performed: (i) The ModCell™ model was extended with the most relevant signalling pathways, (ii) drug response predictions were generated using Monte Carlo simulations, (iii) optimization of the ModCell™ model was conducted based on OncoTrack omics and response data, (iv) forward simulation was performed on a validation data set and (v) evaluation of validation data conducted.

1.6. Potential impact and main dissemination activities and exploitation of results

WP1: Genome Sequencing and Proteome Profiling

WP1 completed all sequence analyses for the so-called 'complete patients' that is, those for whom tumour tissue, normal tissue/peripheral blood and a xenograft (PDX) and/or 3D cell culture or organoids (PDO) are available. From these patients, whole exome (in some instances whole genome) sequencing, transcript sequencing and methylome analysis were conducted. In addition, confirmatory genome sequencing and transcriptome analysis was performed on the xenograft and cell culture models. Drug response data for the models was provided by WP2 (PDO) or WP6 (PDX). This collection of molecular data and biological models constitutes a unique resource for biomarker discovery and validation. At the same time, the models themselves offer an ideal platform for drug discovery and characterization
In collaboration with WP8, a complete analysis of the genomic landscape of the patients' tumours was performed and correlation analysis to relate drug response of the biological models to their genomic and transcriptomic characteristics. This work has yielded a novel classifier for response to the clinically important therapeutic agent cetuximab. The classifier could be validated using independent cohorts of xenograft models and also with molecular data from a clinical cohort.
We were also able to identify classifiers for response to other agents used clinically in the therapy of colon cancer, namely for 5-fluorouracil and for the tyrosine kinase inhibitors afatinib and sapitinib (AZD8931). The OncoTrack partner Charite Universitatsmedizin in Berlin is currently assembling data from a historical cohort for the validation of the 5-FU classifier. Validation of the TKIs will require collection of additional clinical response data - activities that are beyond the scope of OncoTrack.
Since it has been shown recently that proteomics data can outperform gene expression for

coexpression based gene function predictions (Wang J et al., 2017 Mol Cell Proteomics), we might expect additional biological insight and even better predictors of drug response after analysis of our proteomic profiling data. The combined efforts of all partners in OncoTrack have laid the groundwork
for novel predictive diagnostics that may benefit colon cancer patients in the near future.

WP2: Cancer Stem Cells

The EGR and SLC2A12 or GLUT12 genes have been reported as important regulators of early embryonic and placental development. We obtained novel insights suggesting that these genes are involved in the regulation of CSCs. This result adds new factors relevant for characterizing stem cell functions in CRC cell populations and potentially in other cancer entities.
Novel insights into CSC biology are an important contribution to understanding the contributions of CSCs to tumour biology and their role in development of resistance to established therapies. Advances in tumour biology, combined with the availability of well characterized models will contribute to the design and discovery of novel cancer therapies

WP4: Plasma: DNA and Proteins

D04.02: Companion diagnostic assay platform for selected biomarkers
WP4 successfully developed a versatile biomarker quantification platform that is able to detect selectively and sensitively multiple biomarkers in a single sample using a homogeneous assay format that does not require any washing or separation and only a single incubation step. The developed TRFRET
assays allowed for a fast and simple detection (with very low coefficients of variation) of circulating biomarkers at clinically relevant concentrations in small volume liquid samples (e.g., serum or plasma).
Technological advances such as reported here have the potential to increase the efficiency and reliability of routine clinical diagnostic procedures while the minimal sample requirements reduce inconvenience and discomfort for the patient.

WP5: Pathology Samples

PEA based proteome analysis of xenografts (UU)
Our proteomic analysis on PDX samples that underwent different drug treatments led to the identification of a set of proteins significantly altered between drug responding and non-responding PDX. These significantly altered proteins might serve as diagnostic or prognostic biomarkers, or potential therapeutic targets. Our finding may open new avenues for developing efficient methods for early diagnosis of cancer, its prognosis, and new therapeutic drugs.
Characterization of heterogeneity within tumours using *in situ* expression profiling (SU)
Our results showing the two distinct classes of tumour cells in the original patient tumour, highlights the importance of characterizing the heterogeneity of tumours in patients to provide an accurate molecular diagnosis. The study shows that our *in situ* expression profiling method may be a useful tool for such characterization and therefore as well as being a powerful new research tool might find applications in improved diagnostic procedures.

WP8: Data Integration, Statistical Analysis, and Modelling

This collaboration has explored and implemented the procedures and technologies in tackling the challenges in knowledge management of a complex multi-centre clinical project. It sets up a model in
cross-IMI efforts on building both an information/data integration platform and a collaboration platform. The experiences and results of this collaboration should benefit other IMI and EU/international projects on translational research in using appropriate IT systems to accelerate data

analysis, hypothesis generation and idea/outcome sharing.
The extension, improvement and validation of the ModCell™ platform within OncoTrack has also opened up a number of exploitation avenues, both commercially and research-driven. The improved platform is currently being deployed within a new SME-focused Horizon2020 project – CanPathPro – which aims to develop a commercial platform for the mechanistic modelling of cancer pathways. Further collaborative studies are in discussion with clinics, pharma and biotech companies for clinical validation of ModCell™ as well as application of the technology for commercially-orientated drug development projects.

1.7. Lessons learned and further opportunities for research

OncoTrack is a precompetitive research project that was created to tackle the general problem of identification and validation of clinically robust biomarkers in oncology. The consortium brought together for the first time scientists from pharmaceutical companies that previously had not participated in joint research projects, but who shared the common viewpoint that traditional approaches to biomarker development are fraught with shortcomings. Their experience in drug discovery, biomarker validation and the development of biological model systems could be shared with the academic and biotechnology groups, leading to synergistic effects in development of model systems and transfer of technologies among the partners.

- From the industry perspective, these processes worked best when research personnel (eg: postdocs) were dedicated to the project and worked within the company: This arrangement facilitates transfer of project output to the industry partners.
- Establishing and maintaining communication among Work Packages is a high priority, especially when close collaboration between or among geographically distant centres is necessary. The use of a webinar system and conscientious keeping of minutes should be encouraged.
- Regular cross-checks of progress and adherence to the Description of Work mitigate problems among interdependent Work Packages.
- It is essential to start sustainability planning early.
- Cooperation between consortia, especially putting legal agreements in place, is an administrative challenge. If possible, the legal framework should be agreed prior to the signature of the Project and Grant Agreements, to avoid loss of time later in the project.

A major part of the OncoTrack Foreground consists of the deep molecular characterization of the patient tissues and derived biological models. The extraordinary depth of information that the consortium has collected, combined with the use of the in vivo and in vitro models to quantify drug response provides a unique resource for research in tumour biology, biomarker discovery and for drug development. Partners in the consortium have already, before the conclusion of OncoTrack, initiated four new research co-operations including one H2020 FET Open project.
The joint efforts of the consortium have yielded a novel classifier for response to the clinically important therapeutic agent cetuximab. The classifier could be validated using independent cohorts of xenograft models and also with published historical data from a clinical cohort. After prospective validation in a new clinical trial, this classifier could potentially serve as a new companion diagnostic for cetuximab-based therapy. We were also able to identify classifiers for response to other agents used clinically in the therapy of colon cancer, namely for 5-fluorouracil and for the tyrosine kinase inhibitors afatinib and sapitinib (AZD8931). The OncoTrack partner Charite Universitatsmedizin in Berlin is currently assembling data from a historical cohort for the validation of the 5-FU classifier. Validation of the classifiers for the TKIs will require collection of additional clinical response data, activities that are beyond the scope of OncoTrack but could be the subject of a new research project