Inicio > Proyectos y resultados > FP7 > Breast Cancer Somatic Genetics Study

Contenido archivado el 2024-06-18

Breast Cancer Somatic Genetics Study

Informe

Información del proyecto

BASIS

Identificador del acuerdo de subvención: 242006

Sitio web del proyecto

Proyecto cerrado

Fecha de inicio
1 Julio 2010

Fecha de finalización
31 Diciembre 2014

Financiado con arreglo a
Specific Programme "Cooperation": Health

Coste total
€ 13 875 833,15

Aportación de la UE
€ 10 499 999,00

Coordinado por
GENOME RESEARCH LIMITED
United Kingdom

Este proyecto figura en...

REVISTA RESEARCH*EU

Mujeres en la ciencia; e investigación para mejorar la vida de las mujeres

N.º 20, MARZO 2013

Final Report Summary - BASIS (Breast Cancer Somatic Genetics Study)

Executive Summary:
The International Cancer Genome Consortium (ICGC) was launched with the goal of coordinating a large number of research projects with the common aim of elucidating the detailed genomic changes found in many forms of common cancers that contribute to the burden of disease in people throughout the world.

The primary goals of the ICGC are to generate comprehensive catalogues of genomic abnormalities (somatic mutations, abnormal expression of genes, epigenetic modifications) in tumours from 50 different cancer types and/or subtypes which are of clinical and societal importance across the globe and to make the data available to the entire research community as rapidly as possible, and with minimal restrictions, in order to accelerate research into the causes and control of cancer. The ICGC actively facilitates and encourages communication among the members and provides a forum for coordination with the objective of maximizing efficiency among the scientists working to understand, treat, and prevent these diseases.

In BASIS, we proposed to generate comprehensive catalogues of somatic mutations in a minimum of 400 ER+, HER2-breast cancers under this ICGC model by high coverage, shotgun genome sequencing of both tumour and normal DNA.

Eventually all classes of mutations were expected to be detected including base substitutions, insertions, deletions, copy number changes, translocations and other chromosomal rearrangements. These catalogues of mutations would afford us statistical power to identify cancer genes that are mutated at a frequency of greater than 3% in this class of breast cancer. They will also carry signatures of past mutagenic processes operative in the development of each cancer and thus provide insights into the aetiology of the disease.

Complementary catalogues of transcriptomic changes (mRNA and miRNA profiles) and epigenomic changes (genome-wide DNA methylation) were generated for as many of the 400 cancer samples as possible to produce a comprehensive biological examination of ER+ breast cancer. Integrated analyses of these datasets have now been completed and the results compared to parallel datasets from other classes of breast cancer as well as other types of cancer.

The results of the BASIS study, as outlined in this report, provide a near definitive overview of the somatic genetics of breast cancer, a landmark milestone in understanding the disease. The main mutated cancer genes are known and the long tail of infrequently mutated cancer genes has been extensively explored. The mutated cancer genes remaining to be discovered in breast cancer are mostly operative at 2% or lower frequency. Activated fusion genes and non-coding driver mutations (other than at splice sites) probably play a very minor role in breast cancer.

To a large extent, this compendium of mutated cancer genes defines the pathways and druggable nodes in breast cancer that can be pursued by the pharmaceutical industry in the future. The mutation signatures that are found in breast cancer are now clearly defined. We know (or suspect) the mutational processes underlying approximately half of these, and thus the BASIS project has set the target for the next series of studies to establish the causes of breast cancer by understanding the process underlying the unknown signatures.

All of the data from this project will be deposited in the ICGC Breast Cancer DCC and in the European-Genome Phenome Archive, and as such will be freely available to bonafide researchers in order that they may continue to mine this information for future therapeutic value.

Project Context and Objectives:
All cancers arise due to somatically acquired mutations in the genomes of those afflicted that alter the function of key cancer genes. Understanding the critical mutational events underlying the development of cancer is paramount for advancing prevention, early detection and effective treatment of the disease.

Breast cancer is the most common cause of cancer death among women. Extraordinary recent advances in sequencing now make it a realistic aim to sequence large numbers of breast cancer genomes to find somatic mutations on a massive scale. In Europe, there are extensive, well-annotated collections of breast cancer cases that collectively are unparalleled elsewhere in the world. The individuals and institutions that have made these contributions are members of the BASIS consortium. BASIS proposed to definitively characterise the somatic genetics of ER+ve, HER2-ve breast cancer through generation of comprehensive catalogues of somatic mutations in 400 cases by high coverage genome sequencing coupled with integrated transcriptomic and methylation analyses.

Breast cancer is the most common class of cancer diagnosed in women worldwide with more than one million cases diagnosed annually and is responsible for >400,000 deaths per year making it the leading cause of cancer deaths in women. Histological, immunohistochemical, mRNA expression and genomic analyses have indicated that breast cancer is a heterogeneous disease composed of multiple biological subtypes. We adopted a pragmatic working classification of breast cancer based on histology combined with immunohistochemistry for a limited biomarker set including estrogen receptor (ER), progesterone receptor (PR) and HER2. This classification is outlined below:

- ER-, HER2+ tumors
- ER+, HER2+ tumors
- ER+, HER2-, PR+ tumors; ductal-type
- ER+, HER2-, PR- tumors; ductal-type
- Invasive lobular carcinoma
- “Rare” histological special types

BASIS focused on ER+ve, HER2-ve ductal-type breast cancers (both PR+ and PR-) which constitute the largest subgroup in this classification scheme and account for ~40% breast cancer cases.

The genomes of all cancer cells carry somatically acquired changes and the combined effects of a subset of these alterations, known as “driver” mutations, subvert normal control of cell proliferation, differentiation and death, thereby conferring a selective growth advantage on a clonal population of cells that ultimately manifests as a symptomatic cancer. Driver mutations are, by definition, in cancer genes.

The identification of cancer genes and exploration of their function has been fundamental to our understanding of cancer biology. Furthermore, in the last decade, proteins encoded by mutated cancer genes have become targets for the development of novel therapies, some of which are now in routine clinical practice as first line treatments, for example Herceptin in HER2+ve breast cancers and imatinib in chronic myeloid leukaemia.

The remaining somatic mutations in cancer genomes are “passengers” that do not confer growth advantage on the cancer cell. Nevertheless, passenger mutations bear the signatures of the mutational processes that generated them, including exogenous exposures and defects in DNA repair, and therefore provide substantial insights into how cancers are initiated.

In the past, investigation of somatic genetic changes in cancer genomes has been conducted piecemeal using diverse technological approaches including cytogenetics, fluorescence in situ hybridisation (FISH), limited sequencing, genotyping and comparative genomic hybridisation each of which is designed to detect only a subset of the mutations present. Even in aggregate, however, these approaches report only a minority of somatic mutations, often at low resolution.

The advent of second generation sequencing platforms has provided the opportunity to generate complete sequence data for large numbers of cancer genomes. This strategy can reveal all classes of somatic mutation. Coupled with analyses of epigenetic (methylation) change, comprehensive catalogues of somatic mutation can be generated for individual cancers, including all drivers and all passengers. Transcriptomic characterisation of both coding and noncoding genes can then provide the primary biological and phenotypic manifestations of these changes.

This remarkable prospect is, however, sobered by the realisation that there are more than 100 classes of cancer and that large numbers of cases of each will be required in order to detect many mutated cancer genes. Prompted by the size of the technical challenge, the need to avoid duplication of effort and the requirement for principles of data sharing, quality standards and ethical processes, we and others recently formed the International Cancer Genome Consortium (ICGC). The aim of the ICGC, summarised in a document released in April 2008 that is available at http://www.icgc.org/icgc_document is to establish these principles and provide scientific coordination of studies.

In BASIS, we proposed to generate comprehensive catalogues of somatic mutations in 400 ER+, HER2- breast cancers under the ICGC model by high coverage, shotgun genome sequencing of both tumor and normal DNA. All classes of mutations were detected including base substitutions, insertions, deletions, copy number changes, translocations and other chromosomal rearrangements. These catalogues of mutations will afford us statistical power to identify cancer genes that are mutated at a frequency of greater than 3% in this class of breast cancer. They will also carry signatures of past mutagenic processes operative in the development of each cancer and thus provide insights into the aetiology of the disease.

Complementary catalogues of epigenomic changes (genome-wide DNA methylation) will be generated for the same 400 cancer samples together with mRNA and miRNA expression profiles. Integrated analyses of these data will be carried out and compared to parallel datasets from other classes of breast cancer as well as other types of cancer.

Data will be made rapidly available to all scientific researchers with minimal restrictions. The results of this exhaustive and comprehensive set of studies will have enormous impact on our understanding of the causes and biology of breast cancer and will lead to major advances in detection, prevention and treatment in one of the most common diseases and causes of death in the developed world.

Objectives

- Collection of samples of tumor and normal tissues from ER+, HER2- breast cancer cases with relevant clinical information and stored under appropriate ethical supervision. (WP1)

- Review each case histopathologically to deliver DNA and RNA from at least 400 ER+, HER2- cases of sufficient quality and quantity for genomic analyses. (WP1)

- Tissue microarrays will be produced. (WP1)

- DNA from normal and tumor material from the 400 ER+, HER2- cases will be subjected to high coverage shotgun genome sequencing using next generation sequencing technology. (WP2)

- Sequence data will be analysed to generate complete catalogues of somatic mutations for each cancer sample, including base substitutions, small insertions and deletions, rearrangements, and copy number changes. (WP2)

- Statistical analysis of the combined set of somatic mutational catalogues from the 400 ER+, HER2- breast cancers will be conducted to identify new cancer genes and characterise patterns of somatic mutation. (WP2)

- Copy number changes will be independently evaluated in the 400 ER+, HER- breast cancers by hybridisation to high density genotyping arrays. (WP3)

- Genome-wide DNA methylation analyses will be conducted on the 400 ER+, HER- breast cancer cases using MethylCap technology followed by second generation sequencing. (WP4)

- Digital mRNA expression profiles on the 400 ER+, HER- breast cancers will be generated using mRNA sequencing employing second generation technologies. (WP5)

- MicroRNA expression profiles on the 400 ER+, HER- breast cancers will be generated by array analysis. (WP6)

- Data from all analyses will be organised for distribution and display and will be analysed and integrated to identify pathways and patterns of genomic change. (WP7)

- Data will be disseminated through a dedicated BASIS website, through the ICGC federated database system, through the Catalogue of Somatic Mutations in Cancer (COSMIC) database, through Array Express and GEO, through peer reviewed publications and presentations at symposia.

Project Results:
All cancers are abnormally proliferating clones of cells that arise due to somatic mutations. To achieve a comprehensive overview of the driver mutations that confer clonal proliferative advantage and the mutational processes generating somatic mutations in breast cancer, we sequenced the genomes of 402 oestrogen receptor positive and HER2 negative (0 or 1+ by immunohistochemistry) breast cancers and their matched constitutional normal genome using Illumina next generation sequencing machines through the sequencing pipelines of the Wellcome Trust Sanger Institute. We also produced copy number analyses, epigenomic profiles and interrogated the transcriptomes of a subset of the tumour samples. The subsequent data was integrated and subjected to full and robust analyses to inform our understanding of the genetic factors involved in the development of breast cancer.

WP1: Sample Curation
At the beginning of the Breast Cancer Somatic Genetics Study, a pathology working group was formed which was constituted of a group of ten dedicated and experienced breast pathologists. This group interacted intensively throughout the course of the project via joint two-day meetings and teleconferences.

All the cases proposed for inclusion in the project were jointly reviewed by two pathologists (neither of which had submitted the cases that they reviewed to the project) in order for them to discuss and reach a consensus for the scoring of microscopic tumour features using a double-headed microscope. During the BASIS project, 13 two-day pathology review sessions were held each attended by 4-6 pathologists.

In the initial period of the project, the pathology group developed optimal procedures for the isolation of DNA and RNA and also shared/discussed these procedures with the other participants in the project and also developed, then implemented, a standard scoring form for detailed histo-pathologic assessment of all breast carcinomas that were eligible to be analysed in the other work packages of the BASIS project.

In addition to scoring the histo-pathological features of each tumour, oestrogen receptor, progesterone receptor and HER2 staining was carried out and scored for each tumour by one pair of pathologists.

The cases submitted to the project for review came from a wide range of institutions both internal and external to BASIS. In total, the project received cases from 29 Institutions in 17 countries around the world. The breakdown of cases supplied by BASIS beneficiaries constituted approximately two thirds of the total number of cases in the larger ICGC Breast Cancer Working Group study – that is the study including ER+, triple negative and HER2+ cases.

All relevant pathology review data were collated and curated by the Research Coordinator and the Clinical Database Manager prior to entry into an excel spreadsheet. This data was then uploaded into a secure oracle database housed at the Wellcome Trust Sanger Institute.

This database also held the available clinical data relevant to the patient. Samples for inclusion in the project were selected from this database. Clinical data options were aligned to those collected by the ICGC DCC and the dictionary of responses was fixed as much as possible to ensure concordant interpretation of data from each institution. A formal guidance document was produced and issued to each supplying institute to assist the provision of clinical data.

To collect an optimal series of tumours for the BASIS project, over 2,500 breast carcinomas were reviewed, from which just over 400 ER+/HER2- tumours were selected for genome sequencing and subsequent analyses within the other work packages of the BASIS project.

We selected tumours with the highest percentages of tumour cells in the frozen specimen and selected breast carcinomas with specific histologic features, including rare types of breast carcinomas. Nonetheless, during the genomic analyses carried out in work packages 2 and 3, it became apparent that not all cases were of high cellularity. The deviation, we suspect, is a result of failure by contributing centres to provide nucleic acids from sections adjacent to the frozen which was reviewed and/or due to over estimation of tumour cellularity during manual review. Higher levels of stroma and normal contamination reduced the tumour content in some cases. Overall, the pathologists estimate of cellularity was higher than the actual cellularity (as derived from the genomic analyses) in 86% of cases and lower than the actual cellularity in the remaining 13% of cases.

Conversely, where adjacent tissue was supplied as the source of normal, a frozen section of that adjacent tissue was provided to confirm the absence of neoplastic cells. Again, contamination of the normal samples with tumour was observed during genomic analyses and this led to a small number of cases being lost from the project as the variants could not be called with sufficient confidence due to this contamination.

Tissue microarrays have been produced where material was available. A subset of the selected cases had been anonymised in line with the consent and ethical approvals within the supplying institute and, as such, no specific clinical or histological materials could be accessed for this group. A smaller group of samples could not be processed outside of the source country due to local ethical governance issues. However, for the majority of cases a paraffin block was available and TMA’s containing cores from 280 breast carcinomas have been produced at the Academic Medical Center. Staining for ER and HER2 was carried out on the whole tumour section as part of the standard pathology review process instead of directly on the newly constructed TMA’s (which are available for future translational research studies).

Work Package 2: Somatic Mutation Analyses

In order to maximise our power to extract biological meaningful readouts, and to understand the differences between subtypes, the BASIS samples selected in WP1 were subjected to genome sequencing in work package 2. They were then combined with a further cohort of cases (funded by other partners within the overall ICGC breast cancer working group) to provide a total of 672 cases (as of December 2014) for detailed genomic interrogation.

By combining the datasets from the whole ICGC Breast Cancer Working Group to produce a much larger cohort of cases, in fact, the largest cohort of whole-genome sequenced cancers of the same tissue-type, the Consortium has obtained an overarching view of the molecular pathology of Breast Cancer and are currently working towards a high impact publication in the near future.

Following quality control and initial analyses, this larger analyses cohort was composed of the following:
- 402 ER+, HER2- breast cancers (this project)
- 138 triple negative breast cancers funded by the Wellcome Trust Sanger Institute
- 83 mixed subtype cases funded by The National Center for Cancer Genomics, in collaboration with the Asian Medical Center and Hanyang Unversity
- 50 HER2 positive funded by the Institut Nationale du Cancer

Genome sequence coverage was >30-fold in the cancers and >20-fold in non-neoplastic tissues. Following routine processing and sequence alignment, somatic substitutions, insertions/deletions, rearrangements and copy number changes in exonic, intronic and intergenic genomic regions were called using a complex integrated suite of algorithms developed at the Sanger Institute. Full catalogues of mutations including base substitution somatic mutations, small indels, rearrangements and copy number changes were detected in 560 samples that have thus far undergone extensive curation.

To identify driver mutations in cancer genes, we used standard probabilistic approaches to identify mutation clustering beyond that expected by chance in the protein-coding exons of each gene, implementing corrections for mutation burden per sample, regional variation in sequence composition and sequence motifs. Essentially all the cancer genes known to be mutated in breast, and other cancers, were identified through this approach. Five potential new breast cancer genes were identified. In general, there exists a relatively limited number of commonly mutated cancer genes (arbitrarily defined as mutated in >10% cases), but a long tail of cancer genes with infrequent driver mutations both in breast cancer and most other cancer types.

We computationally searched our data for rearrangements predicted to cause in-frame fusion genes. Many were found, but none thought to be oncogenic gene fusions per se and likely represented passenger gene fusion events that arise in highly breakable parts of the genome (e.g. fragile sites).

The contribution of somatic driver base substitutions in non-coding regions of the genome (other than exonic splice sites) to neoplastic transformation is poorly understood for breast and other cancers, in large part because few whole genome sequences have thus far been analysed. The many non-coding germline substitutions recently found by Genome Wide Association Studies to confer small-elevated risks of diverse diseases renders this a plausible hypothesis. Such somatic mutations could, in principle, be in regulatory regions of cancer genes in which there also exist conventional coding sequence driver mutations in other cases, or may exclusively affect regulatory regions, as appears to be the case for promoter mutations of TERT.

To explore these possibilities, we searched for recurrence of somatic base substitution mutations in noncoding regions, correcting for background error rates, mutational profiles and numbers of mutations in each genome, using a general sliding window approach and also by focusing upon annotated regulatory regions including promoters and enhancers. These results will be presented in the manuscript that is in preparation.

We previously reported a mathematical approach to extract mutational signatures from catalogues of somatic mutations from cancer genomes and to estimate the contribution of each signature to each genome. In breast cancer, previous analyses of 21 whole genomes and, subsequently, 100 whole genomes and 800 exomes, revealed five base substitution signatures referred to as signatures 1, 2, 3, 8 and 13. We have applied this approach to the this series of breast cancer whole genome sequences and identify the full spectrum of mutation signatures identified previously and several new mutational processes that are also active in breast cancer (manuscript in preparation).

In all, we are able to construct comprehensive genomic profiles for each of the 560 cancers demonstrating the full breadth of inter-tumour variation in breast cancer. These findings are currently being woven together with the data from the remaining work packages to produce a journal article that is will be submitted for peer review by Spring 2015.

Work Package 3: Copy Number Analyses

During the course of the project we have established a primary data analysis pipeline for Affymetrix SNP6 microarray data required to provide important copy number data to other BASIS work packages. This pipeline includes methods for generation of basic copy number (CN) and allelic ratios (B allele frequencies, BAF) estimates, quality control tools, analysis tools for estimating allele specific copy numbers (ASCAT) from which genomic alterations are called, and tools for higher-level genomic analysis of the basic data. The pipeline is built around public bioinformatics tools and extensive in-house generated software. Notably, basic quality control work performed by WP3 has helped to identify numerous cases not suited for incorporation in the final study cohort.

Data from BASIS cases were merged with two other breast cancer cohorts with whole genome sequencing data, as outlined in WP2 above. This pooling of samples resulted in a final cohort of 560 cases, of which 555 had copy number and ASCAT data of sufficient quality. This increase in sample size provided a considerable increase in statistical power for analyses performed by WP3, and new insights into breast cancer tumourigenesis.

Basic data from WP3 provides an important overview of the characteristics of the studied cohort in relation to existing knowledge of copy number alterations in breast cancer. The overall pattern across all chromosomes of copy number alterations and allele-specific imbalances, such as loss of heterozygosity (LOH) in general or in copy number neutral regions specifically, correlate with previous studies, providing confidence in the results and in the cohort composition.

As expected compared with the literature, gains of 1q, 8q, 16p, distal part of 17q and 20q, and losses at 3p, 5q, 8p, 11q, 16q, and 17p represent the most frequently altered regions in the analysed cohort. Although these regions represent the overall most altered regions in a mixed breast cancer cohort, it has repeatedly been demonstrated that the different molecular subtypes of breast cancer are associated with differences in the spectrum of copy number alterations, and moreover that breast cancer may be further subdivided into clinically relevant subgroups by usage of copy number alterations alone or in concert with e.g. gene expression profiling.

Copy number alterations have been shown to be important in breast cancer in many ways, including to essentially define an entire clinically relevant subgroup of the disease on its own (the HER2/ERBB2-amplified subgroup), and to be useful in molecular refinement of the disease. In order to define target (driver) genes in breast cancer we integrated the extensive mutational analyses performed by WP2, which identified/defined a set of 70-plus breast cancer drivers, with copy number data (amplification or homozygous deletion).

Notably, many of the most frequently altered oncogenes in breast cancer are driven more by copy number alterations than by somatic mutations. This finding is in contrast to other tumour diseases, such as lung cancer (high frequency of EGFR and KRAS mutations) and melanoma (BRAF, and NRAS mutations), underpinning the importance of copy number analysis in breast cancer, not only diagnostic purposes (ERBB2-amplification) but also for molecular refinement.

Work Package 4: Methylation Analyses
By the end of the project, we concluded the phased generation of DNAme profiles from ER+/HER- breast tumours to reach a final number of 400. We have generated two datatypes for DNAme: (I), Whole-genome bisulfite sequencing; and (II), Infinium 450k.

Whole-genome bisulfite sequencing (WGBS) provides base resolution. We have generated WGBS profiles for a subset of 30 BASIS cases at a coverage level of 20x – 30x. These 30 cases were based on an initial clustering of 450k profiles from which at least 2 cases from each identified subgroup was selected for WGBS, so that the WGBS samples are a balanced representation of the total dataset. We have used these WGBS profiles to identify genomic elements of interest.

These include LMRs (low methylated regions, mostly distant regulatory regions), UMRs (unmethylated regions, mostly promoters), PMDs (partially methylated domains, megabase-sized regions). Whereas we have used both individual CpGs and LMRs to deduce DNAme subgroups and their signatures, PMDs were used in a meta-analysis that integrated different data types from other beneficiaries and work packages.

Infinium 450k arrays provide information at a lower, yet still considerable density and represents the most common DNAme platform used by most large consortia. In total, we generated 394 QC-passed 450k profiles (98.5% of the planned number). Using a widely applied method such as 450k arrays has allowed us to incorporate data from other consortia (such as The Cancer Genome Atlas) into our own analyses, which significantly increased the power to detect subgroup-specific features.

We performed an integrative analysis combining data from epigenetic (WP4), genetic (WP2), and transcriptomic (WP5) layers. This analysis has been extended with clinical metadata from WP1. Our main focus for this analysis has been on PMDs (partially methylated domains), which represent megabase-sized domains with loss of methylation in a patient- and PMD-variable manner.

Our analysis has revealed an important link between epigenetic and specific genetic aberrations in breast cancer patients, a feature that differs among the major pathological subtype classifier.

Work Package 5: RNA Analyses

By the end of the project, we generated and sequenced RNA libraries for 335 BASIS samples. Sequencing reads were mapped to the reference genome and transformed into expression values (FPKM) and normalised, yielding gene expression data that could be analysed and utilized by other beneficiaries for downstream analysis.

The generated gene expression data were explored using publically available and in house developed algorithms to identify novel transcript species and novel luminal subtypes. The complete in-depth exploration of the transcriptome is ongoing. The association with omics data from the other WPs has been completed.

Work Package 6: miRNA Analyses

miRNA expression data have been obtained from 399 of the final selected set of 400 ER+, HER2- breast tumour samples included in the BASIS study. Only one sample did not have RNA available. The complete data set (both raw data and processed data) were submitted to the BASIS analysis group at the last data-freeze in October 2014. The human miRNA microarray release v.19.0 from Agilent was applied; consisting of an 8x60K slide format printed using Agilent’s 60-mer SurePrint technology.

These arrays include 2006 miRNAs and are considerably updated and expanded compared to the miRNA microarray versions that were used for initial samples in the first years of the project; a total number of 826 new miRNAs were added on v.19 compared to the previous v.16.

Data have been successfully obtained from all samples with RNA available (399 out of the 400 selected samples), using the same version of Agilent human miRNA microarrays (v.19) ensuring high quality data. The data quality is very good and there are no batch effects witnessed between the two different array batches used.

Unsupervised clustering of the top 500 most variable miRNA's in these 399 cases identified two major clusters with some sub-clusters. Novel associations to other molecular classes and clinical parameters (like GISTIC classes and grade) were observed.

Work Package 7: Data Analyses and Integration

Over the past four years, we collectively development extensive quality control of the BASIS data, from somatic variant calling through to methylation and copy number alteration, populated International Cancer Genome Consortium biomarts with first mock data and then pilot data and then provided the entire BioMart datasets.

We developed a secure alignment server to provide the ability to share raw data to a broad set of approved researchers in the future and developed a consortium wide data sharing approach, based on a structured FTP site and data freezes.

We developed an extensive joint analysis of the BASIS data sets. This included work on: the distribution of somatic single nucleotide variation, and their likely mutational signature source; the detailing of protein coding driver mutations across the different samples; exploration of genomic features around single nucleotide variation, in particular in the non-coding sequences; correlations of copy number alterations, methylation, single nucleotide somatic variation and RNA expression levels; comparison of clinical phenotypes with all the above work. The resulting analysis provides the first comprehensive analysis of full genomes of breast cancer.

Potential Impact:
All cancers are due to somatic mutations. The impact of BASIS is in providing a definitive picture of the somatic genetics of breast cancer encompassing the whole genome, including protein coding and non-coding regions.

The BASIS study has defined a set of 119 mutated cancer genes, some commonly active and some rare, that are operative in breast cancer. These define the nodes and pathways to understand the biology of the disease and to launch drug =discovery efforts. The remaining cancer genes in breast cancer are very rare (operative in <2% breast cancer cases). They will also serve as biomarkers in future for defining sets of patients that may/or may not be sensitive to particular treatments.

The BASIS study has also defined the set of mutational signatures, and thus mutational processes, that are operative in breast cancer. These are essentially the causes of breast cancer and the definitive repertoire that has been established will serve as the basis for further studies to establish, in greater detail, the aetiology of breast cancer by epidemiologists and cancer biologists.

Breast cancer is a common disease affecting young and old women. The results of the BASIS study definitively set the scene for the next round of drug discovery campaigns by the pharmaceutical industry and outline the repertoire of biomarkers that will be used in future for stratifying patients. This will ultimately result in further improvement of treatment for patients, better prevention and better deployment of resources for health services.

List of Websites:

www.basisproject.eu

The website is not uptodate as there are no staff available to update it since the grant finished.

The lead contact on the project is Prof Sir Michael Stratton FRS - mrs@sanger.ac.uk.

Última actualización: 24 Agosto 2015

Permalink: https://cordis.europa.eu/project/id/242006/reporting/es