

BIOINFORMATICS AND STATISTICS

The Bioinformatics and Statistics group provides leadership on the collection and analysis of data for the research programs of the institute, by conducting research in bioinformatics and statistics and by performing state of the art analyses of a wide array of data types. Research topics include stratifying tumors into groups with distinct and homogeneous outcome and therapy response; the characterization of genes and pathways involved in tumorigenesis and understanding molecular regulatory mechanisms. A number of exemplary projects are presented below in more detail.

Systems Biology In collaboration with the Beijersbergen, Bernards and Jonkers groups we have established the Cancer Systems Biology Center. The aim of this center is to develop a strategy to tackle the complexity of molecular networks that govern breast tumorigenesis. This strategy is rooted in a modeling and experimental validation cycle spanning multiple levels of complexity including cell lines, mouse models and patients. As a start we are focusing on breast cancers for which no effective targeted therapies exist: ‘triple negatives’ and invasive lobular carcinomas. Since there are strong indications for the involvement of PI3K and MAPK signaling pathways in these subtypes, we are generating *in silico* models of therapy response by employing 1) normal and tumor cell lines as *in vitro* model systems; 2) quantification of functional activation of pathway components and associated cellular phenotypes; 3) computational modeling to create quantitative models of pathway behavior and resistance mechanisms; and 4) mouse tumor models as *in vivo* model systems. The identified models will be directly validated in proof-of-concept pre-operative trials. We hope that this strategy will yield improved diagnostic tools and tumor-specific treatments resulting in more tailored cancer therapy.

Oncogene discovery by direct association of insertion features with gene expression Insertional mutagenesis is a potent forward genetic screening technique to identify novel putative cancer genes. An important – yet unresolved – issue is to determine which genes are affected by the viral insertions retrieved from these screens. To address this, we developed RBM (rule-based mapping), an approach that automatically determines the most appropriate target genes for a set of integration sites based on a collection of literature derived rules. RBM exploits properties of individual insertions and genes, such as orientation and distance to the gene, and allows insertion mapping without prior knowledge of gene expression.

For both MuLV and SB, RBM produced superior association of insertions with gene expression compared to selecting the nearest gene as target. We also developed KC-RBM (kernel convolved RBM), a mapping approach that combines the power of the rules captured in RBM with data-driven common insertion site calling based on Gaussian kernel convolution. We demonstrate the superiority of RBM and KC-RM over existing approaches in recovering true positives from a list of independently, manually curated targets.

Extracting oncogenic pathways from insertional mutagenesis screens

We have developed combinatorial association logic networks (CALs): an approach to extract simple Boolean logic circuits which employ combinations of insertion loci to predict the expression pattern of downstream targets. In classical one-dimensional analyses, direct interactions between the insertion patterns and transcription levels across tumors are detected. However, when the insertion loci themselves interact, direct associations between the individual loci and transcript levels may become undetectable. Therefore, our method detects associations between transcript levels and the outputs of small Boolean logic networks that combine multiple genetic loci. The detection of logic networks requires solving a demanding optimization problem. By reformulating the objective function and applying a customized branch and bound algorithm, we obtain runtimes of up to four orders of magnitude faster than exhaustive search. We demonstrated our method on an insertional mutagenesis dataset, combining insertion data with transcriptional information from the same sample, finding known and novel associations between genes involved in Notch signaling.



Group leader Lodewyk Wessels

Lodewyk Wessels PhD Group leader
Michael Hauptmann PhD Academic staff
Marta Lopez-Yurda PhD Academic staff
Wilma Heemsbergen PhD Academic staff
Nicos Angelopoulos PhD Post-doc
Sander Canisius PhD Post-doc
Andreas Schlicker PhD Post-doc
Hayssam Soueidan PhD Post-doc
Jordi Vidal Rodriguez PhD Post-doc
Ewald Van Dyk MSc PhD student
Johann De Jong MSc PhD student
Christiaan Klijn MSc PhD student
Wouter Meuleman MSc PhD student
Jeroen De Ridder MSc PhD student
Jorma De Ronde MSc PhD student
Christine Staiger MSc PhD student
Bram Gerritsen Bioinformatician
Jelle Ten Hoeve Bioinformatician

Publications

De Ridder J, Gerrits A, Bot J, de Haan G, Reinders M, Wessels L. *Inferring combinatorial association logic networks in multimodal genome-wide screens.* *Bioinformatics.* 2010;26:i149-57

Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B. *Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.* *Mol Cell.* 2010;38:603-13

Klijn C, Bot J, Adams DJ, Reinders M, Wessels L, Jonkers J. *Identification of networks of co-occurring, tumor-related DNA copy number changes using a genome-wide scoring approach.* *PLoS Comput Biol.* 2010;6:e1000631

De Ronde JJ, Hannemann J, Halfwerk H, Mulder L, Straver ME, Vrancken Peeters MJ, Wesseling J, van de Vijver M, Wessels LF, Rodenhuis S. *Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response.* *Breast Cancer Res Treat.* 2010;119:119-26

Publications (continued)

De Ronde JJ, Klijn C, Velds A, Holstege H, Reinders MJ, Jonkers J, Wessels LF. *KC-SMARTR: An R package for detection of statistically significant aberrations in multi-experiment aCGH data. BMC Res Notes. 2010;3:298*

Horlings HM, Lai C, Nuyten DS, Halfwerk H, Kristel P, van Beers E, Jooisse SA, Klijn C, Nederlof PM, Reinders MJ, Wessels LF, van de Vijver MJ. *Integration of DNA copy number alterations and prognostic gene expression signatures in breast cancer patients. Clin Cancer Res. 2010;16:651-63*

Kool J, Uren AG, Martins CP, Sie D, de Ridder J, Turner G, van Uitert M, Matentzoglou K, Lagcher W, Krimpenfort P, Gadiot J, Pritchard C, Lenz J, Lund AH, Jonkers J, Rogers J, Adams DJ, Wessels L, Berns A, van Lohuizen M. *Insertional mutagenesis in mice deficient for p15Ink4b, p16Ink4a, p21Cip1, and p27Kip1 reveals cancer gene interactions and correlations with tumor phenotypes. Cancer Res. 2010;70:520-31*

Heemsbergen WD, Al-Mamgani A, Witte MG, Van Herk M, Pos FJ, and Lebesque JV. *Urinary obstruction in prostate cancer patients from the Dutch trial (68 Gy vs. 78 Gy): relationships with local dose, acute effects and baseline characteristics. Int J Radiat Oncol Biol Phys 2010;78:19-25*

He Y, Van 't Veer L, Lopez-Yurda MI, Van de Velde C, Marijnen C. *Do rectal cancer patients with PIK3CA mutations benefit from preoperative radiotherapy with regard to local recurrences? Clin Cancer Res, 2010;16: 6179*

Rasch CR, Hauptmann M, Schornagel J, Wijers O, Buter J, Gregor T, Wiggendaad R, De Boer JP, Ackerstaff AH, Kroger R, Hoebbers FJ, and Balm AJ. *Intra-arterial versus intravenous chemoradiation for advanced head and neck cancer: results of a randomized phase 3 trial. Cancer 2010;116:2159-2165*

Retel VP, Joore MA, Knauer M, Linn SC, Hauptmann M, and Van Harten WH. *Cost-effectiveness of the 70-gene signature versus Sankt Gallen guidelines and Adjuvant Online for early breast cancer. Eur J Cancer 2010;46:1382-1391*

Identification of networks of co-occurring oncogenic gains and losses

Collaborating oncogenic events can also be induced by copy number alterations. To detect such events in aCGH data, we developed a scoring framework to separate truly co-occurring aberrations from passenger mutations and dominant single signals present in the data. Analysis of high-resolution DNA copy number data from a panel of 95 hematological tumor cell lines correctly identified co-occurring recombinations at the T-cell receptor and immunoglobulin loci in T- and B-cell malignancies, respectively. In addition, our analysis revealed networks of co-occurring genomic losses and gains that are highly enriched for functional relationships. The co-occurring losses we find are independent of the canonical cancer genes within the network. Our findings suggest that large-scale, low-intensity copy number changes may be an important feature of cancer development or maintenance by affecting the gene dosages of a large interconnected network of functionally related genes. We have adapted this approach to be applicable to aCGH profiles derived from human tumors. This led to the discovery of co-occurrences uniquely associated with the triple negative subtype in breast cancer.

Evaluation of biomarkers predicting treatment response

Biomarkers predicting treatment response are useful for tailoring treatment. Before evaluation in prospective randomized trials, the first clinical evaluation often takes place in relatively small retrospective patient series or older trials. We investigated whether a classifier based on comparative genomic hybridization and initially developed to identify mutations in the BRCA1 gene of breast cancer patients might also detect loss-of-function of this gene due to causes other than mutation and could therefore predict sensitivity to double strand break inducing agents. The hypothesis was evaluated in 230 stage III breast cancer patients from an available clinical trial who had been randomized between adjuvant high-dose platinum-based and conventional anthracycline-based chemotherapy. We observed greater benefit from high-dose versus conventional chemotherapy among patients identified by the classifier based on a significant interaction between classifier and treatment. In a study of advanced non-small cell lung cancer, we demonstrated that epidermal growth factor (EGFR) ligands may predict response to EGFR-inhibitor treatment based on a clinical sample of 60 patients treated with EGFR-inhibitors and 60 matched control patients. Low concentrations of transforming growth factor-alpha and high concentrations of amphiregulin were associated with a better disease-specific survival among patients treated with EGFR-inhibitors versus control patients.

Long-term effects of therapeutic and diagnostic radiation exposure After the introduction of modern radiotherapy and chemotherapy, many cancers have become curable malignancies, e.g., Hodgkin lymphoma or testicular cancer. However, the life expectancy of survivors is compromised by the occurrence of late complications of treatment such as second malignancies and cardiovascular disease. In a cohort of about 500 long-term childhood cancer survivors treated with different potentially cardiotoxic therapies we showed that a high percentage of patients (27%) have an abnormal left ventricular shortening fraction, and that anthracycline dose, cardiac irradiation, and younger age at diagnosis were associated with subclinical cardiac dysfunction. In another project, we are evaluating the risk of cancers of the stomach and pancreas as second malignancies related to radiation (and chemo-) therapy received by survivors of testicular cancer. This will be performed in a multinational cohort with detailed assessment of radiation doses and the corresponding uncertainties. Although radiation doses from diagnostic procedures are usually much lower than therapeutic doses, some diagnostic imaging procedures may deliver doses high enough to present a non-negligible risk for sensitive individuals. Within a multi-national pilot study, we determined that it is feasible in the Netherlands to set up a nationwide cohort of approximately 100,000 children who have been exposed to ionizing radiation from diagnostic computed tomography (CT) scans, to estimate the radiation doses to different organs, and to evaluate their subsequent risk of childhood cancer, particularly leukemia. Funding has been obtained to conduct such studies in eight European countries on about one million children, including the Netherlands.