# Genestat
# Statistics for genomic research

**Innovation area:** Health

**Key Innovator:** Marit Holden

**Partners:** UiO, NR, NTNU, OUS, PubGene

**Research staff scientists:** Eivind Hovig (principal investigator, OUS), Arnoldo Frigessi (principal investigator, UiO), Marit Holden (principal investigator, NR), Ingrid Glad (UiO), Ingunn Tvete (NR), Ørnulf Borgan (UiO), Heidi Lyng (OUS), Anne-Lise Børresen-Dale (OUS), Vessela Kristensen (OUS), Therese Sørlie (OUS), Egil Ferkingstad (NR), Bettina Kulle Andreassen (UiO), Knut Liestøl (UiO), Ole Christian Lingjærde (UiO), Magne Aldrin (NR), Lars Holden (NR), Fang Liu (Pubgene), Hiroko Solvang (OUS), Trevor Clancy (PubGene), Vegard Nygaard (UiO), Magne Thoresen (UiO) , Ole Andreas Andreassen (UiO), Simen Myhre (OUS)

**Ph.D students:** Halfdan Rydbeck (UiO), Marissa Leblanc (UiO), Linn C. Bergersen (UiO), Sveinung Gundersen (OUS), Øystein Sørensen (UiO), Kristoffer Hellton (UiO), Tonje Gulbrandsen Lien (UiO), Jonas Paulsen (UiO)

**Postdocs:** Geir Kjetil Sandve (OUS), Kukatharmini Tharmaratnam (UiO)

**International contacts and collaborators:** Sylvia Richardson (Imperial College, London), Clelia di Serio (Univ. Salute, Milano), Alessandro Ambrosi (Univ. Salute, Milano), Anestis Antoniadis (Grenoble), Jens Overgaard (Aarhus), Elja Arjas (Helsinki University, Finland) and many others

## Scope:

Produce statistical instruments to understand molecular mechanisms, on the basis of -omics data, contributing to biological discovery and thus combating diseases.

## Results in 2012 and plans:

***1. The Genomic Hyperbrowser – statistical methods***

*a) Statistical methods for genomic 3D data***:** In 2012 we started extending the statistical analysis to take the three-dimensional structure of DNA into account. Recently, genome-wide data on spatial co-localization of DNA is becoming available through the use of chromosome conformation capture together with high-throughput sequencing. The goal would be to create a web-based statistical system that allows generic analysis of genome annotation data in a three-dimensional setting, similar to the existing line-based statistical analysis of genomic data. The incorporation of three-dimensional localization allows extensions in several distinct directions: a) It can allow standard line-based analyses to be performed locally in bins corresponding to 3D co-localization; b) It can allow statistical analyses of questions that are directly related to three-dimensional structure, e.g. regarding unexpected 3D co-localization of related genomic features or unexpected correlation between values associated to co-localized genomic element; and c) It can allow comparative studies of the 3D organization of DNA across different tissue types and conditions. In 2012 we have submitted a first paper. The work on 3D co-localization will continue in 2013-2014. We hope that the extension, when implemented, will boost the external use of the Genomic Hyperbrowser (GH), which has not happened yet.

*b) Null models:* In the Genomic Hyperbrowser we use in most cases a preservation hierarchy in building the null hypothesis. The p-value depends critically on the choice of preservation and randomization of what is not preserved. There are similar situations in other applications outside genomics. In many tests, but not always, there is monotonicity between stochastic orderings of null hypothesis and p-values. This link between ordering of null models and p-values can be exploited to guide the use of the GH. We are writing a paper that gives sufficient condition for monotonicity. If we succeed to make a useful theory for the cases in the GH, we will be able to suggest how to guide the user in their investigations, for given computational time.

*c) Multitrack:* So far we have focused on hypothesis testing comparing two tracks. The number of possible relations explodes with more tracks. We started with the three track problem and the n-track problems for overlap between segment tracks. We have written a first draft paper on comparing several tracks. The extension to many tracks, and a way to perform in a more unsupervised fashion "discovery protocols", can be useful in personalized medicine.

***2. The Genomic Hyperbrowser – implementation and promotion***

A key aspect is the need for a user-friendly implementation of all the novelties. In 2012, the data handling side and the user interface of the system have been improved. Also, algorithms and methods have been added or improved, including some of the statistical methods mentioned above. In 2013 and 2014, we will continue improvements of the system, including implementing new methods that have been / will be developed for analyses of genomic 3D data. We need to have a better strategy on how to promote the GH. We see that there are many papers that could use the GH but do not do it. We should contact each of these groups, and make them aware, offer assistance etc., so that their next paper will use the GH.

***3. Predictive gene signature for response to post-mastectomy radiotherapy***

Radiotherapy (RT) is known to improve loco-regional control, disease free survival and to have a long-term improvement on overall survival in high-risk patients suffering from breast cancer. We have found a gene signature consisting of 7 genes, whose transcription interacts with the effect of postmastectomy radiotherapy (PMRT) to modify the hazard of local recurrence of the cancer. These genes have been identified by whole genome microarray using RNA from fresh frozen human samples. The signature has been technically replicated and verified in corresponding formalin fixed, paraffin embedded tissue. We have filed a patent with Invent2, and they are now leading the commercialization project. We have applied for FORNY funding from NFR and Helse Sør-Øst. This funding will be used to validate the signature in a new cohort of 1000 women and to do a second statistical analysis with a new method to confirm/extend the seven genes. This is planed for 2013. In 2014 we hope that the patent can be sold to a major pharmaceutical industry, while further validated towards the creation of a clinical chip. Regression with p>>n is the key method used. This project is in collaboration with OUS and the University of Aarhus.

***4. Statistical methodology for genomic data***

*a) p>>n regression:* This is the most common situation in genomics, and calls for variable selection for prediction. We have developed a new method that allows to discard a very large proportion of variables, before running the Lasso method for variable selection and learning. The paper is submitted. In 2013 and 2014, we wish to investigate the use of this idea in several new directions: survival models; monotone non-linear regression; network estimation; interaction detection. In particular this last point should be one of the first genome-wide studies of full interactions. This might lead again to interesting discoveries.

*b) Data integration:* On each patient, we can measure multiple -omics data (DNA, CNV, RNA, methylation, exposures etc.) and we want to develop a new Bayesian method to use all these data sources contemporaneously, in order to increase the power of detection of new signatures and reduce uncertainty. We have started by a state-of-the-art review of existing approaches, and will continue with the development of alternatives. Focus will be in breast cancer and nutritional factors in cancer. This project will continue in 2013 and 2014.

*c) Network estimation:* This is a very important problem in all molecular biology. We wish to achieve state-of-the-art competence first, so that we can apply the methods to the many genomics data that require it. We want to investigate the possibility to develop new methods that allow exact calculations of posteriors for Wishart priors on precisions matrices. The way we have suggested to legalize a non-positive-definite matrix for DNB should be considered also in this context.

Besides what is mentioned above, we have for many years helped researchers in biology and medicine with experimental design and statistical analysis of their genomic data. This support will continue in 2013 and 2014. Particular attention will be given to mental disorders.

## Papers:


Ambrosi A, Glad IK, Pellin D, Cattoglio C, Mavilio F, Di Serio C, Frigessi A. Estimated Comparative Integration Hotspots Identify Different Behaviors of Retroviral Gene Transfer Vectors. PLoS Comput Biol, 2011; Vol. 7(12): e1002292.

Andrey A. Shabalin, Haakon Tjelmeland, Cheng Fan, Charles M. Perou, Andrew B. Nobel, Merging two gene-expression studies via cross-platform normalization. Bioinformatics, 2008; Vol. 24(9): 1154–1160.

Athanasiu L, Mattingsdal M, Kähler AK, Brown A, Gustafsson O, Agartz I, Giegling I, Muglia P, Cichon S, Rietschel M, Pietiläinen OP, Peltonen L, Bramon E, Collier D, Clair DS, Sigurdsson E, Petursson H, Rujescu D, Melle I, Steen VM, Djurovic S, Andreassen OA. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. Journal of Psychiatric Research, 2010; Vol. 44(12): 748-753.

Baumbusch, LO; Aarøe, J; Johansen, FE; Hicks, J; Sun, H; Bruhn, L; Gunderson, K; Naume, B; Kristensen, VN; Liestøl, K; Børresen-Dale, AL and Lingjærde, OC. Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. BMC Genomics 2008, 9:379.

Bergersen, Linn Cecilie; Glad, Ingrid Kristine; Lyng, Heidi. Weighted Lasso with Data Integration. Statistical Applications in Genetics and Molecular Biology, 2011; Vol. 10(1).

Boye K, Grotterød I, Aasheim HC, Hovig E, Maelandsmo GM: Activation of NF-kappaB by extracellular S100A4: analysis of signal transduction mechanisms and identification of target genes. Int J Cancer., 2008; Vol. 123(6): 1301-10.

Brown, AA; Jensen, J; Nikolova, YS; Djurovic, S; Agartz, I; Server, A; Ferrell, RE; Manuck, SB; Mattingsdal, M; Melle, I; Hariri, AR; Frigessi, A; Andreassen, OA. Genetic variants affecting the neural processing of human facial expressions: evidence using a genome-wide functional imaging approach. Translational Psychiatry, 2012; Vol. 2: e143, doi: 10.1038/tp.2012.67.

Bruhn, Sören; Barrenäs, Fredrik; Mobini, R; Andersson, B; Chavali, S.; Egan, B. S.; Hovig, Eivind; Sandve, Geir Kjetil; Langston, M. A.; Rogers, Gary; Wang, H.; Benson, Mikael. Increased expression of IRF4 and ETS1 in CD4(+) cells from patients with intermittent allergic rhinitis.. Allergy 2011; Vol. 67(1): 33-40.

Bøhn, SK; Myhrstad, MC; Thoresen, M; Holden, M; Karlsen, A; Tunheim, SH; Erlund, I; Svendsen, M; Seljeflot, I; Moskaug, JØ; Laake, P; Arnesen, H; Tonstad, S; Collins, AR; Drevon, CA; Blomhoff, R. Blood cell gene expression associated with cellular stress defence is modulated by antioxidant-rich food in a randomised controlled clinical trial of male smokers. BMC Medicine, 2010; Vol. 8:54.

Bøhn, SK; Russnes, KM; Sakhi, AK; Thoresen, M; Holden, M; Moskaug, JØ, Myhrstad, MC; Olstad, OK; Smeland, S; Blomhoff, R. Stress associated gene expression in blood cells is related to outcome in radiotherapy treated head and neck cancer patients. BMC Cancer, 2012; Vol. 12:426.

Bøvelstad HM, Nygård S, Borgan Ø: Survival prediction from clinico-genomic models - a comparative study. BMC Bioinformatics, 2009: Vol. 10:413.

Clancy, Trevor; Pedicini, Marco; Castiglione, Filippo; Santoni, Daniele; Nygaard, Vegard; Lavelle, Timothy J.; Benson, Michael D.; Hovig, Eivind. Immunological network signatures of cancer progression and survival. BMC Medical Genomics, 2011; Vol. 4:28.

Disanto, G; Sandve, GK; Berlanga-Taylor, AJ; Morahan, JM; Dobson, R; et al. Genomic Regions Associated with Multiple Sclerosis Are Active in B Cells. PLoS ONE, 2012; Vol. 7 (3): e32281.

Disanto, G; Sandve, GK; Berlanga-Taylor, AJ; Ragnedda, G; Morahan, JM; Watson, CT; Giovannoni, G; Ebers, GC; Ramagopalan, SV. Vitamin D receptor binding, chromatin states and association with multiple sclerosis. Human Molecular Genetics, 2012; Vol. 21 (16): 3575-3586.

Di Serio, C; Pellin, D; Ambrosi, A; Glad, I; Frigessi, A. Biostatistics Meets Bioinformatics in Integrating Information from Highdimensional Heterogeneous Genomic Data: Two Examples from Rare Genetic Diseases and Infectious Diseases. In: Computational Intelligence Methods for Bioinformatics and Biostatistics, Lecture Notes in Computer Science, 2012; Vol. 7548: 10-20.

Djurovic S, Kähler AK, Kulle B, Jönsson EG, Agartz I, Le Hellard S, Hall H, Jakobsen KD, Hansen T, Melle I, Werge T, Steen VM, Andreassen OA. A possible association between schizophrenia and GRIK3 polymorphisms in a multicenter sample of Scandinavian origin (SCOPE). Schizophr Res., 2009; Vol. 107(2-3): 242-8.

Ferkingstad, Egil; Frigessi, Arnoldo; Lyng, Heidi: Indirect genomic effects on survival from gene expression data. Genome Biology, 2008; Vol. 9:R58.

Ferkingstad, Egil; Frigessi, Arnoldo; Rue, Haavard; Thorleifsson, Gudmar; Kong, Augustine: Unsupervised empirical Bayesian multiple testing with external covariates. Annals of Applied Statistics, 2008; Vol. 2(2): 714-735.

Fjeldbo, CS; Misund, K; Günther, CC; Langaas, M; Steigedal, T; T Lægreid, A; Bruland, T. Functional studies on RNA-transfected cell microarrays. Methods in Molecular Biology, 2010; Vol. 629: 175-191.

Gundersen, Kalas, Abul, Frigessi, Hovig, Sandve, Identifying elemental genomic track types and representing them uniformly, BMC Bioinformatics, 2011; Vol. 12(1): 494-515.

Halle, CL; Andersen, EKF; Lando, M; Aarnes, EK; Hasvold, G; Holden, M; Syljuåsen, RG.; Sundfør, K; Kristensen, GSB; Holm, R; Malinen, E; Lyng, H. Hypoxia-induced gene expression in chemoradioresistant cervical cancer revealed by dynamic contrast-enhanced MRI. Cancer Research, 2012; Vol. 72 (20), 5285-5295.

Holden, M; Deng, S; Wojnowski, L; Kulle, B: GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. Bioinformatics, 2008; Vol. 24(23): 2784-2785.

Husberg C, Nygård S, Finsen AV, Damås JK, Frigessi A, Øie E, Waehre A, Gullestad L, Aukrust P, Yndestad A, Christensen G.: Cytokine expression profiling of the myocardium reveals a role for CX3CL1 (fractalkine) in heart failure. J Mol Cell Cardiol, 2008; Vol. 45(2): 261-269.

Jemtland R; Holden M; Reppe S, Olstad OK; Reinholt FP; Gautvik VT; Refvem H; Frigessi A; Houston B; Gautvik KM. Molecular disease map of bone characterizing the postmenopausal osteoporosis phenotype. Journal of Bone and Mineral Research, 2011; Vol. 26(8): 1793-1801.

Joshi, H; Nordgard, SH; Frigessi, A; Børresen-Dale, AL; Kristensen, VN. Overrepresentation of transcription factor families in the genesets underlying breast cancer subtypes. BMC Genomics, 2012; Vol. 13:199.

Jønsson V, Samuelsen SO, Tjønnfjord G, Johannesen T: Looking for CLL genes. Leukemia & Lymphoma, 2008; Vol. 49(1): 10-11.

Kähler AK, Djurovic S, Kulle B, Jönsson EG, Agartz I, Hall H, Opjordsmoen S, Jakobsen KD, Hansen T, Melle I, Werge T, Steen VM, Andreassen OA: Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. Am J Med Genet B Neuropsychiatr Genet., 2008; Vol. 147B(7): 1089-1100.

Kähler AK, Djurovic S, Rimol LM, Brown AA, Athanasiu L, Jönsson EG, Hansen T, Gústafsson O, Hall H, Giegling I, Muglia P, Cichon S, Rietschel M, Pietiläinen OP, Peltonen L, Bramon E, Collier D, St Clair D, Sigurdsson E, Petursson H, Rujescu D, Melle I, Werge T, Steen VM, Dale AM, Matthews RT, Agartz I, Andreassen OA. Candidate gene analysis of the human natural killer-1 carbohydrate pathway and perineuronal nets in schizophrenia: B3GAT2 is associated with disease risk and cortical surface area. Biol Psychiatry., 2011; Vol. 69(1): 90-96.

Kraus I, Driesch C, Vinokurova S, Hovig E, Schneider A, von Knebel Doeberitz M, Dürst M: The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer Res., 2008; Vol. 68(7): 2514-2522.

Kresse, SH; Rydbeck, H; Skårn, M; Namløs, HM; Barragan-Polania, AH; et al. Integrative Analysis Reveals Relationships of Genetic and Epigenetic Alterations in Osteosarcoma. PLoS ONE, 2012; Vol. 7 (11): e48262.

Kulle, B; Frigessi, A; Edvardsen, H; Kristensen, V; Wojnowski, L: Accounting for haplotype phase uncertainty in LD estimation, Genet Epidemiol. 2008 February; 32(2): 168-78. Erratum in: Genet Epidemiol., 2008; Vol. 32(6): 586-587.

Lando, M; Holden, M;Bergersen, LC; Svendsrud, DH; Stokke, T; Sundfør, K; Glad, IK; Kristensen, GB; Lyng, H: Gene Dosage, Expression, and Ontology Analysis Identifies Driver Genes in the Carcinogenesis and Chemoradioresistance of Cervical Cancer. PLoS Genetics, 2009; Vol. 5(11): e1000719.

Lawyer G, Ferkingstad E, Nesvåg R, Varnäs K, Agartz I. Local and covariate-modulated false discovery rates applied in neuroimaging. NeuroImage., 2009; Vol. 47(1): 213-219.

Lawyer, Glenn; Nesvag, Ragnar; Varnas, Katarina; Frigessi, Arnoldo; Agartz, Ingrid: Investigating possible subtypes of schizophrenia patients and controls based on brain cortical thickness. Psychiatry Res., 2008; Vol. 164(3): 254-264.

LeBlanc, M; Kulle, B; Sundet, KS; Agartz, I; Melle, I; Djurovic, S; Frigessi, A; Andreassen, OA. Genome-wide study identifies PTPRO and WDR72 and FOXQ1-SUMO1P1 interaction

associated with neurocognitive function. Journal of Psychiatric Research, 2012; Vol. 46(2): 271-278.

Lyng H, Lando M, Brøvig RS, Svendsrud DH, Johansen M, Galteland E, Brustugun OT, Meza-Zepeda LA, Myklebost O, Kristensen GB, Hovig E, Stokke T: GeneCount: genome-wide calculation of absolute tumor DNA copy numbers from array comparative genomic hybridization data. Genome Biol., 2008; Vol. 9(5): R86.

Moger, TA; Haugen, M; Yip, BHK; Gjessing, HK; Borgan, Ø. A hierarchical frailty model applied to two-generation melanoma data. Lifetime Data Analysis, 2011; Vol. 17(3): 445-460.

Myhre, S; Mohammed, H; Tramm, T; Alsner, J; Finak, G; Park, M; Overgaard, J; Børresen-Dale, AL; Frigessi, A; Sørlie, T. In Silico Ascription of Gene Expression Differences to Tumor and Stromal Cells in a Model to Study Impact on Breast Cancer Outcome. PLoS ONE, 2010; Vol. 5(11): e14002.

Nakken S, Rødland EA, Rognes T, Hovig E: Large-scale inference of the point mutational spectrum in human segmental duplications. BMC Genomics, 2009; Vol. 10:43.

Nordgard SH, Alnaes GI, Hihn B, Lingjaerde OC, Liestøl K, Tsalenko A, Sørlie T, Lønning PE, Børresen-Dale AL, Kristensen VN: Pathway based analysis of SNPs with relevance to 5-FU therapy: relation to intratumoral mRNA expression and survival. Int J Cancer., 2008; Vol. 123(3): 577-585.

Nygaard V, Hovig E: Methods for quantitation of gene expression. Front Biosci., 2009; Vol. 14: 552-569.

Nygaard, V; Liu, F; Holden, M; Kuo, WP.; Trimarchi, J; Ohno-Machado, L; Cepko, CL.; Frigessi, A; Glad, IK.; van de Wiel, M.; Hovig, E; Lyng, H. Validation of oligoarrays for quantitative exploration of the transcriptome. BMC Genomics, 2008; Vol. 9:258.

Nygård, S; Borgan, Ø; Lingjærde, OC and Størvold, HL: Partial least squares Cox regression for genome-wide data. Lifetime Data Analysis, 2008; Vol. 14(2): 179-195.

Otnaess MK, Djurovic S, Rimol LM, Kulle B, Kähler AK, Jönsson EG, Agartz I, Sundet K, Hall H, Timm S, Hansen T, Callicott JH, Melle I, Werge T, Andreassen OA. Evidence for a possible association of neurotrophin receptor (NTRK-3) gene polymorphisms with hippocampal function and schizophrenia. Neurobiol Dis., 2009; Vol. 34(3): 518-524.

Psychiatric GWAS Consortium Coordinating Committee, Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. Am J Psychiatry, 2009; Vol. 166(5): 540-556.

Reppe, S; Refvem, H; Gautvik, VT; Olstad, OK; Høvring, PI; Reinholt, FP; Holden, M; Frigessi, A; Jemtland and R and Gautvik, KM: Eight genes are highly associated with BMD variation in postmenopausal Caucasian women. Bone, 2010; Vol. 46(3): 604-612.

Rimol LM, Agartz I, Djurovic S, Brown AA, Roddey JC, Kähler AK, Mattingsdal M, Athanasiu L, Joyner AH, Schork NJ, Halgren E, Sundet K, Melle I, Dale AM, Andreassen OA. Sex-dependent association of common variants of microcephaly genes with brain structure. Proceedings of the National Academy of Science of the United States of America, 2010; Vol. 107(1): 384-388.

Sandve, Geir Kjetil; Ferkingstad, Egil; Nygård, Ståle. Sequential Monte Carlo multiple testing. Bioinformatics, 2011; Vol. 27(23): 3235-3241.

Sandve GK, Gundersen S, Rydbeck H, Glad I, Holden L, Holden M, Liestøl K, Clancy T, Ferkingstad E, Johansen M, Nygaard V, Tøstesen E, Frigessi A, Hovig E. Inferential genomics at the sequence level. Genome Biology, 2010; Vol. 11(12).

Sandve, GK; Gundersen, S; Rydbeck, H; Glad, IK; Holden, L; Holden, M; Liestøl, K; Clancy, T; Drablos, F; Ferkingstad, E; Johansen, M; Nygaard, V; Tøstesen, E; Frigessi, A; Hovig, E. The differential disease regulome, BMC Genomics, 2011; Vol. 12:353.

Scharpf RB, Tjelmeland H, Parmigiani G, Nobel AB: A Bayesian model for cross-study differential gene expression (with rejoinder). Journal of the American Statistical Association, 2009; Vol. 104(488): 1318-1323.

Shabalin AA, Tjelmeland H, Fan C, Perou CM, Nobel AB: Merging two gene-expression studies via cross-platform normalization. Bioinformatics, 2008; Vol. 24(9): 1154–1160.

Solvang H, Lingjerde O, Frigessi A, Borresen-Dale A, Kristensen V: Linear and non-linear dependencies between copy number aberrations and mRNA expression reveal distinct molecular pathways in breast cancer. BMC Bioinformatics, 2011; Vol. 12:197.

Tesli, Martin Steen; Athanasiu, Lavinia; Mattingsdal, Morten; Kähler, Anna Katarina; Gustafsson, Omar; Kulle, Bettina; Werge, Thomas; Hansen, Thomas; Mors, Ole; Mellerup, Erling; Koefoed, Pernille; Jönsson, Erik G.; Agartz, Ingrid; Melle, Ingrid; Morken, Gunnar; Djurovic, Srdjan; Andreassen, Ole Andreas. Association Analysis of PALB2 and BRCA2 in Bipolar Disorder and Schizophrenia in a Scandinavian Case–Control Sample. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2010; Vol. 153B: 1276-1280.

Tøstesen, E: A stitch in time: Efficient computation of genomic DNA melting bubbles. Algorithms for Molecular Biology, 2008; Vol. 3:10.

Watson, CT; Disanto, G; Sandve, GK; Breden, F; Giovannoni, G; et al. Age-Associated Hyper-Methylated Regions in the Human Brain Overlap with Bivalent Chromatin Domains. PLoS ONE, 2012; Vol. 7(9): e43840.

Yip, BHK, Moger, TA; Pawitan, Y. Genetic analysis of age-at-onset traits based on case-control family data. Statistics in Medicine, 2010; Vol. 29(30): 3258-3266.

Zhao X, Rødland EA, Sørlie T, Naume B, Langerød A, Frigessi, A, Kristensen, V, Børresen-Dale, AL, Lingjærde, OC. Combining Gene Signatures Improves Prediction of Breast Cancer Survival. PLoS ONE, 2011; Vol. 6(3): e17845.

## Patent:

Sørlie, T; Frigessi, A; Børresen-Dale, A-L; Myhre, S; Mohammed, H; Overgaard, J; Alsner, J; Tramm, T. A gene signature associated with efficacy of postmastectomy radiotherapy in breast cancer. Provisional Patent number: INVEN-32535/US-1/PRO. Filed March 6, 2012. The provisional patent application for the gene signature has been filed by Inven2 AS on behalf of Oslo University Hospital and University of Oslo (U.S. Provisional Patent Application Serial No. 61/607,316).