Molecular Oncology 2014

A DNA methylation-based definition of biologically distinct breast cancer subtypes

Olafur A. Stefanssona, Sebastian Morana, Antonio Gomeza, Sergi Sayolsa, Carlos Arribas-Jorbaa, Juan Sandovala, Holmfridur Hilmarsdottirb, Elinborg Olafsdottirc, Laufey Tryggvadottirc, Jon G. Jonassonc,d, Jorunn Eyfjordb, Manel Estellera,e,f,*

aCancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute, L'Hospitalet, Barcelona, Catalonia 08908, Spain

bThe Cancer Research Laboratory, Medical Faculty, University of Iceland, Reykjavik, Iceland

The Icelandic Cancer Registry, Reykjavik, Iceland

dDepartment of Pathology, Landspitali University Hospital, Reykjavik, Iceland

eDepartment of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain fInstitucio Catalana de Recerca i Estudis Avanc, ats (ICREA), Barcelona, Catalonia, Spain

Breast Cancer

- Most common cancer among women, ranks among the leading causes of cancer-related deaths.
- A heterogeneous disease with different subtypes showing distinct biological and clinical features.
- Prognosis of breast cancer patients has been improving over time with the development of subtype specific treatments
 - Tamoxifen for patients with hormone receptor-positive tumors
 - Trastuzumab (Herceptin) for patients displaying overexpression and amplification of the HER-2 oncogene

The importance of accurate subtype identification

- An important problem in breast cancer treatment is the definition of patient subsets that will require aggressive treatment options and close follow-up after treatment
- A major milestone on the way to this goal was the definition of five biologically and clinically meaningful breast cancer subtypes based on genome-wide expression analyses:
 - Luminal-A
 - Luminal-B
 - HER-2
 - Basal-like (Triple Negative: ER-, PR-, Her2-)
 - Normal-like

Trastuzumab (Herceptin) as an example for a subtype specific drug

- The HER2 pathway promotes cell growth and division when it is functioning normally; however when the HER2 receptors are overexpressed, cell growth accelerates beyond its normal limits.
- The HER receptors are proteins that are embedded in the cell membrane and communicate molecular signals from outside the cell (molecules called <u>EGFs</u>) to inside the cell, and turn genes on and off.

Trastuzumab (Herceptin) as an example for a subtype specific drug

- Trastuzumab is an antibody used for treating the HER2 subtype pateints, by binding the HER2 receptor and compensating for its overexpression.
- The original studies of trastuzumab showed that it improved overall survival in late-stage (metastatic) HER2positive breast cancer from 20.3 to 25.1 months
- Only 30% of the Her2 patients respond to it.
- Resistance to the treatment develops rapidly, in virtually all patients
- Trastuzumab costs about US\$70,000 for a full course of treatment
- It is possible to determine the "erbB2 status" of a tumor, which can be used to predict efficacy of treatment with trastuzumab











Epigenetics and breast cancer subtypes

- Molecular profiling of breast cancer subtypes have so far focused mainly on the expression level.
- Less is known about the contribution of epigenetic changes to the development of biologically distinct breast cancer subtypes
- Cancer related mutations often affect genes involved in regulating chromatin dynamics or the processing of epigenetic marks
- This highlights the importance the epigenome in cancer development and opens up new potentials for identifying patterns of potential relevance to patient prognosis and personalized medicine
- GOAL: Characterize the epigenetic profiles of breast cancer patients.

REMINDER

DNA Methylation and gene expression

- Methylation of CpG sites in the promoter of a gene may inhibit gene expression
- Most of the methylation differences between tissues, or between normal and cancer samples, occur a short distance from the CpG islands (at "CpG island shores") rather than in the islands themselves

DNA Methylation and cancer

- Methylation of CpG sites within the promoters of genes can lead to their silencing, a feature found in a number of human cancers (for example the silencing of tumor suppressor genes).
- In contrast, the hypomethylation of CpG sites has been associated with the over-expression of oncogenes within cancer cells

REMINDER

CpG sites and CpG islands

- CpG sites are regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide
- Cytosines in CpG dinucleotides can be methylated to form 5-methylcytosine.
- In mammals, methylating the cytosine within a gene can turn the gene off, a mechanism that is part of a larger field of science studying gene regulation that is called epigenetics.
- CpG islands (or CG islands) are regions with a high frequency of CpG sites

Infinium HumanMethylation450 BeadChip Kit

- Allows researchers to interrogate > 485,000 methylation sites per sample at single-nucleotide resolution
- Covers 99% of RefSeq genes, with an average of 17 CpG sites per gene region distributed across the promoter, 5'UTR, first exon, gene body, and 3'UTR.
- It covers 96% of CpG islands, with additional coverage in island shores and the regions flanking them.
- Methylation level of a CpG locus is estimated using beta values (B) which are the ratio of intensities between methylated and unmethylated alleles (rang: 0-1).



The context of the discussed paper in my project...

Clustering breast cancer samples by RNA-Seq data

RNA-Seq Breast Cancer Dataset



Clustering breast cancer samples by Methylation data





Meth27 bad out<mark>come samp</mark>le group.txt





Time



Unified Table of Significant Clinical Labels - Labels (pValue Threshold=1.00e-03)

Back to the paper's results

Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute, Spain

The PEBC Breast Cancer Methylation Dataset

- Infinium HumanMethylation450 BeadChips were used to measure DNA methylation on a genome-wide scale in a discovery cohort composed of
 - 40 tumors
 - 17 normal breast tissues
- Samples obtained by the Department of Pathology, University Hospital, Iceland.

The PEBC Breast Cancer Methylation Dataset

- Assigning PAM50 labels to the tumor samples (Gene expression arrays were not used):
 - Tissue microarrays (TMAs) were used to measure expression of subtype-specific markers, i.e., ER, PR, HER-2, Ki-67, EGFR and CK5/6, by immunohistochemistry (IHC).³
 - Tumors were assigned to breast cancer subtypes according to a validated classification scheme:
 - > Tumors positive for either ER or PR were classified as Luminal.
 - High levels of expression of Ki-67 (> 14%) in breast tumors with a Luminal phenotype were classified as Luminal-B (LumB), the remainder being classified as Luminal-A (LumA).
 - Tumors negative for both ER and PR while positive for HER2 (IHC score 3+) were classified as HER2 subtype.
 - Positivity for either CK5/6 or EGFR in tumors negative for both ER and PR were classified as Basal-like.

^{3.1} Genome-wide DNA methylation patterns and breast cancer subtypes

- 5000 differentially methylated CpGs were identified by comparing the 40 tumor samples to the 17 normal samples, included in the discovery cohort (SAMr pValue<0.05 + 10% mean difference threshold).</p>
- The 57 samples were then clustered over the 5000 differentially methylated CpGs using hierarchical clustering.
- Three clusters were called "significant" based on pvClust, with pValue<0.05 (AU > 95%), and compared to expression based subtypes.

pvClust

- An R package for hierarchical clustering with *p*-values
- For each cluster it calculates a p-values which indicates how strong the cluster is supported by data.
- pvclust provides two types of *p*-values: AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) value.
 AU *p*-value, which is computed by multiscale bootstrap resampling, is a better approximation to unbiased *p*-value than BP value computed by normal bootstrap resampling.



DNA methylation changes in breast tumors are non-random and define patterns correlated with clinically and biologically relevant subtypes. Cluster analysis of differentially methylated CpGs between breast cancers and normal breast tissue (the top 5000 most significant CpGs).



Genome-wide DNA methylation patterns and breast cancer subtypes

- Cluster 1 Mainly LumB
 - Enriched with Luminal-B tumors
 - Shows extensive DNA methylation of CpG islands, implying that they have acquired a methylator phenotype
 - "Interestingly, a few members of other subtypes also displayed this pattern (LumA, HER2)"
- Cluster 2 Mainly Basal
 - The DNA methylation patterns of most, but not all, Basal-like tumors were also distinctive in that the changes affected a different set of CpGs from those affected in most other tumors.

Genome-wide DNA methylation patterns and breast cancer subtypes

- In contrast, HER2 and Luminal-A (LumA) breast tumors were more heterogenous in terms of their DNA methylation patterns
- Only a small subset of the LumA breast tumors (4 / 12) showed evidence of a distinctive (i.e., statistically significant; AU > 95%) pattern of DNA methylation changes, emphasizing the biological heterogeneity within this subtype.

Genome-wide DNA methylation patterns and breast cancer subtypes

Conclusion:

DNA methylation changes in breast tumors are non-random and define patterns correlated with clinically and biologically relevant subtypes. Distinct epigenomic characteristics between breast cancers of the Luminal-B and basal-like subtypes

- Methylation signatures were derived for each subtype using a supervised test
 - Sub-type specific CpGs were identified using multi-class SAM + 10% difference between subtype means.
 - The test used the expression based subtype labels in order to detect differentially methylated CpGs.

Figure 1B The top 10 significant CpG's specifically characterizing each of the four "core" subtypes are shown, i.e. the LumA, LumB, HER2 and Basal-like subtypes

Top 10 significant CpG's for each breast cancer subtype



Validation of subtype specific differentially methylated CpGs in the TCGA cohort

- Identified subtype-specific CpG's were validated against an independent cohort (The Cancer Genome Atlas; TCGA) wherein breast cancer subtypes have been annotated for each tumor by the PAM50 assay using expression arrays.
- This analysis revealed consistent changes for LumB and Basal-like subtypes with 254 and 202 CpG's found, respectively, in both cohorts.
- In contrast, breast cancers of the LumA and HER2 subtypes showed very limited or no overlap at all.

Validation of subtype specific differentially methylated CpGs in the TCGA cohort



Figure 2A

Subtype-specific CpG methylation changes identified in relation to each of the four breast cancer subtypes (LumA, LumB, HER2 and Basal-like) were validated in an independent cohort obtained through the Cancer Genome Atlas.

The overlap, i.e. the number of CpG's consistently associated with each of the subtypes in both the TCGA and PEBC cohorts, is indicated by an arrow.

Normal-like

Normal breast tissue



Figure 2B The validated set of 254 LumB and 202 Basal-like specific CpG's shown in both cohorts

Analysis of functional DNA sequence elements

- The DNA methylation signatures for LumB and Basal-like tumors were analyzed in terms of functionally relevant DNA sequence elements.
- LumB signature predominantly involves CpG methylation of promoter sequences (54%, 137 of 254)
- Basal-like signature predominantly involved hypomethylation events occurring in <u>gene body</u> regions (26%, 53 of 202)

Figure 3

The validated DNA methylation signatures specific for LumB (254 CpG's) and B) Basal-like (202 CpG's) breast cancers differ significantly with respect to the sequence context in which CpG methylation changes tend to occur.

19

Promoter

11

Promoter

Promoter methylation events for the **LumB** subtype (χ^2 *P*-value = 0.002)

Gene body hypomethylation for the **Basal-like** subtype (χ^2 P-value = 0.02)

Promoters displaying methylation in association with either the Basal-like or LumB subtypes analyzed in terms of CpG islands, CpG shores or CpG poor promoter regions DNA methylation-based definition of breast cancer subtypes

- The following distinctive epigenomic features were observed:
 - CpG island **promoter methylation** in **LumB** tumors
 - Gene body hypomethylation in Basal-like tumors
- To establish how unique these "hallmark features" are to each of the two expression-based subtypes, a classifier was built.
 - Implemented as PAMr classifier which determined the "degree of similarity" (reflected in the cross-validation probabilities) for each tumor to the signatures of:
 - LumB-associated CpG island promoter methylation (consisting of 129 CpG's)
 - Basal-like associated gene body hypomethylation (consisting of 53 CpG's).

PAMr

Robert J. Tibshirani, Trevor J. Hastie, Balasubramanian Narasimhan, and Gilbert Chu

Sample classification from gene expression data, by the method of "nearest shrunken centroids"

Described at

Diagnosis of multiple cancer types by shrunken centroids of gene expression by Tibshirani, Hastie, Narasimhan and Chu (May 14, 2002).

The definition of DNA methylation-based subtypes in breast tumors.

- A) Cross-validated probability values derived from the pattern recognition algorithm (PAMr) indicating how robustly each tumor displays the validated signature of LumB-associated CpG island promoter methylation events (based on the validated catalogue of LumB-associated CpG island promoter methylation events, i.e. the 129 CpG's)
- B) The cross-validated probability values derived from PAMr indicating how robustly each tumor displays the Basal-like associated gene body hypomethylation signature (based on the 53 CpG's) shown for both the PEBC (left) and TCGA cohorts (right).

Defining novel epigenetics based subtypes: Epi-LumB and Epi-Basal

- The presence of LumB-linked methylome characteristics in an appreciable proportion of LumA and HER2 associated tumors provides the basis for defining a novel subtype hereafter referred to as Epi-LumB (the "Epi" prefix indicating its epigenetic nature).
- Given the unique methylome characteristics, i.e. gene body hypomethylation, we refer to the group of tumors that robustly display the gene body hypomethylation signature as the Epi-Basal subtype.

Defining the novel Epi-LumB and Epi-Basal subtypes based on CpG Signatures

- C) DNA methylation data over the validated catalogue of 129 "hallmark" CpG's characteristic of LumB tumors (i.e. those identified within CpG island promoters in association with the LumB subtype consistently in both the PEBC and TCGA cohorts) shown with respect to the novel Epi-LumB subtype.
- D) Similarly, the DNA methylation data over the validated catalogue of 53 "hallmark" CpG's characteristic of Basal-like tumors (i.e. gene body CpG's consistently associated with Basal-like tumors in both the PEBC and TCGA cohorts) are shown with respect to the novel Epi-Basal subtype.

Gene promoter methylation events affecting known cancer genes found in association with the DNA methylation-based subtypes

- In Epi-LumB samples, five genes were found to be
 - Differentially methylated on Epi-LumB samples
 - Exhibit significant inverse methylation-expression correlation on the TCGA dataset
 - Included in a list of 712 known tumor suppressor genes.
- The Epi-Basal subtype, in contrast, was not found to be associated with CpG methylation over the promoter region of known tumor suppressor genes.

Table 1

Epi-LumB specific CpG methylation events were found to affect a subset of previously known tumor suppressor genes.

The statistics shown describe the relation between CpG methylation and expression over Epi-LumBassociated TSG's in the TCGA cohort where data was available on both CpG methylation and expression (by RNA sequencing) for 731 tumors and 82 normal breast tissue samples.

TargetID	Gene symbol	R ²	Fold change expression (Unmethyla Methylated	e in P-value ated/ (adjusted))
cg14352983	L3MBTL4	0.264	2.667	4.96E-55
cg08336641	L3MBTL4	0.259	3.004	1.12E-53
cg14155416	L3MBTL4	0.255	2.621	7.90E-53
cg12924825	L3MBTL4	0.253	2.902	2.17E-52
cg18556788	L3MBTL4	0.245	2.652	2.02E-50
cg17688525	L3MBTL4	0.241	2.710	1.43E-49
cg03715143	ID4	0.238	2.850	8.89E-49
cg09232937	IRX1	0.191	6.137	2.08E-38
cg05724871	L3MBTL4	0.177	2.625	1.99E-35
cg14271531	ID4	0.147	2.676	5.32E-29
cg21167628	PTCH2	0.116	2.028	7.70E-23
cg20918243	RASSF10	0.109	9.006	2.33E-21
cg10530883	IRX1	0.106	5.251	8.22E-21

Distinct tumor evolutionary paths in association with Epi-LumB and Epi-Basal tumors

Patterns of DNA copy number changes associated with Epi-LumB and Epi-Basal tumors revealing divergent tumor evolutionary paths and candidate tumor suppressor genes

The clinical relevance of DNA methylation-based subtypes

- Developed locus-specific assays for analyzing a few selected markers that could serve as proxies for tumor classification.
- Pyrosequencing was used to analyze the selected proxy markers for Epi-LumB and Epi-Basal tumors in an <u>independent validation cohort of primary breast tumor</u> <u>samples from 310 patients</u>.
 - The Epi-LumB subtype was then assigned to tumors displaying methylation over the promoter region of two out of the three surrogate markers (TTBK1, ZNF132 and KCNA3).
 - The Epi-Basal subtype was then assigned to tumors negative for the Epi-LumB phenotype while positive for methylation of either TENC1 or ZNF671

Table 2

DNA methylation defined subtypes in an independent cohort validating the relation to the classification of breast cancers according to expression-based subtypes

	Basal- like	HER2	LumA	LumB	5NP [⊵]	Total
Epi-LumB	(8%) 2	(8%) 2	(29%) 7	(46%) 11	(8%) 2	(100%) 24
Epi-Basal	(40%) 10	(4%) 1	(24%) 6	(28%) 7	(4%) 1	(100%) 25
Other	0	0	(62%) 25	(35%) 14	(3%) 1	(100%) 40
						X ² = 31.0; P = 0.00014

Comparing the methylation –markers based subtypes to the expression based subtypes

Information on expression-based subtype classification was available in 89 of the tumors included in the validation cohort. b

5NP represents unclassified tumors due to negativity for the five phenotypic markers ER, PR, HER2, CK5/6 and EGFR.

а

The clinical relevance of DNA methylation-based subtypes

The Epi-LumB and Epi-Basal subtypes, defined according to the proxy-based classification system, were both found to be significantly associated with greater tumor size and poorly differentiated phenotypes. The clinical relevance of Epi-LumB and Epi-Basal tumors, defined according to selected proxy markers, analyzed with respect to parameters of clinical staging (tumor size and nodal metastasis status) and degree of differentiation (histological grade).

Tumorsize	T1a-c	T2 - T3	Total
Epi-LumB	(26%) 18	(74%) 52	(100%) 68
Epi-Basal	(30%) 17	(70%) 40	(100%) 56
Other	(52%) 46	(48%) 42	(100%) 88
			X ² = 13.7; P = 0.0010
Nodal metastases	Negative	Positive	Total
Epi-LumB	(35%) 23	(65%) 43	(100%) 64
Epi-Basal	(36%) 19	(64%) 34	(100%) 52
Other	(49%) 37	(51%) 38	(100%) 75
			X ² = 3.8; P = 0.15
Histological grading	+2/+1	+3	Total
Epi-LumB	(32%) 12	(67%) 25	(100%) 35
Epi-Basal	(32%) 11	(68%) 23	(100%) 33
Other	(78%) 47	(22%) 13	(100%) 54
			X ² = 27.6; P < 0.0001

The clinical relevance of DNA methylation-based subtypes

Importantly, the results revealed significantly shorter survival times for patients that develop Epi-LumB subtype breast tumors after adjustment for tumor size, the presence of lymph node metastases along with age and year at diagnosis (Hazards-ratio = 1.83; P = 0.035. Validation cohort (n = 210)

Patients with breast tumors classified as either Epi-LumB or Epi-Basal on the basis of proxy CpG methylation markers are associated with **reduced time to death due to breast cancer**.

A multivariate Cox's proportional hazards modeling of the survival data **wherein the Epi-LumB subtype was found to be an independent prognostic factor** after adjustment for tumor size, lymph node metastases and age- and year at diagnosis. Improved identification of highly aggressive breast cancers by combining methylation- and expression-based subtype definitions

- We find that breast cancer-specific survival times in LumB breast cancer patients do not differ depending on whether or not the tumors are positive for the Epi-LumB phenotype.
- Similarly, survival of patients with basal-like breast cancers does not differ depending on whether or not the tumors are positive for the Epi-Basal phenotype.
- These results indicate that the prognostic value associated with methylation- and expression-based breast cancer subtype definitions do not differ significantly - although we note that the number of patients in the independent cohort with available information on both definitions entails limited statistical power.

On ER+ tumors, Epi-LumB assignment contributes to the Cox model and is a marker for reduced survival (more than just expression based LumB)

• On ER- tumors, Epi-Basal assignment doesn't contribute to the Cox model when combined with the expression based definition.

Summary

- Methylation based signatures for two biologically distinct aggressive subtypes of breast cancer were identified.
- The signatures are characterized by differentially methylated genes and also by CpG context (Promoter Methylation for LumB subgroup and Gene Body hypomethylation for the Basal subgroup).

This suggests the existence of different methylation mechanisms active in biologically distinct cancer subtypes.

- Locus-specific assays were developed for selected proxy markers for identifying each of the methylation signatures.
- Clinical relevance and some prognostic value were demonstrated for the two methylation signatures.

Conclusions

Cancer subtype detection is difficult !

- High patient heterogeneity
- Dependency in clinical data collected over years by various organizations
- Challenging reproducibility and comparison to other published partitions of the patients.
- Defining subtypes assumes a clear partitioning of the data exist, when in reality samples are scattered continuously along many different axes.
- And still, every new factor that contributes to our ability to predict outcome or expose the underlying biological processes is a step in the right direction.