Cancer Informatics

Libertas Academica
FREEDOM TO RESEARCH

OPEN ACCESS
Full open access to this and thousands of other papers at http://www.la-press.com.

ORIGINAL RESEARCH

Significance Analysis of Microarrays (SAM) Offers Clues to Differences Between the Genomes of Adult Philadelphia Positive ALL and the Lymphoid Blast Transformation of CML

Colin Grace and Elisabeth P. Nacheva

University College London, London, UK. Corresponding author email: e.nacheva@ucl.ac.uk

Abstract: Philadelphia positive malignant disorders are a clinically divergent group of leukemias. These include chronic myeloid leukemia (CML) and de novo acute Philadelphia positive (Ph(+)) leukemia of both myeloid, and lymphoid origin. Recent whole genome screening of Ph(+)ALL in both children and adults identified an almost obligatory cryptic loss of *Ikaros*, required for the normal B cell maturation. Although similar losses were found in lymphoid blast crisis the genetic background of the transformation in CML is still poorly defined. We used Significance Analysis of Microarrays (SAM) to analyze comparative genomic hybridization (aCGH) data from 30 CML (10 each of chronic phase, myeloid and lymphoid blast stage), 10 Ph(+)ALL adult patients and 10 disease free controls and were able to: (a) discriminate between the genomes of lymphoid and myeloid blast cells and (b) identify differences in the genome profile of de novo Ph(+)ALL and lymphoid blast transformation of CML (BC/L). Furthermore we were able to distinguish a sub group of Ph(+) ALL characterized by gains in chromosome 9 and recurrent losses at several other genome sites offering genetic evidence for the clinical heterogeneity. The significance of these results is that they not only offer clues regarding the pathogenesis of Ph(+) disorders and highlight the potential clinical implications of a set of probes but also demonstrates what SAM can offer for the analysis of genome data.

Keywords: sam, significance analysis, arraycgh, ph+all, cml, lymphoid blast crisis, igh rearrangements, tarp, chr 9p, chr7p

Cancer Informatics 2012:11 173–183

doi: 10.4137/CIN.S9258

This article is available from http://www.la-press.com.

Introduction

Array CGH (aCGH) has shown itself to be a mature technology capable of detecting genomic gains and losses at a resolution of at least 250 base pairs. Clearly at this resolution there will be masses of data that will challenge the analyst, particularly in the light of the general variance of the genome from individual to individual—so called copy number variations (CNV) and knowledge that the function of a substantial part of the genome is still unknown hence referred to as 'orphan' or 'predicted'. It is also clear that genomic copy number aberrations (CNA) associated with diseased cells are likely to interfere with transcriptional pathways and affect gene function.

Significance Analysis of Microarrays (SAM) was developed by Tusher[1] as a straightforward way of comparing data sets using an internally generated false discovery rate (FDR) as the criterion for the identification of probes that significantly differ between 2 or more classes. It has been successfully used to analyse gene expression data and is now routinely applied.[2,3]

Philadelphia positive malignant disorders are a clinically divergent group of leukaemias with a unique identifying feature, the BCR/ABL1 fusion gene, usually resulting from the chromosome rearrangement t(9;22)(q34;q11) or its variants, that leads to constitutive expression of an aberrant tyrosine kinase. These include chronic myeloid leukaemia (CML) and de novo acute leukaemia of both myeloid Ph(+)AML and lymphoid origin Ph(+)ALL. The latter two disorders are clinically aggressive and therapy challenging even in the era of the powerful tyrosine kinase inhibitors. CML is a multistage progressive disorder which if untreated inevitably ends as fatal acute myeloid or lymphoid blast transformation. The latter, from which it has been reported to differ karyotypically, is usually clinically indistinguishable from Ph(+)ALL the most common type of ALL in adults.[4] Although non-random chromosome changes may accompany disease progression in CML, the genetic background of the malignant transformation from the benign chronic phase (CP) to acute leukaemia (blast crisis, BC) is poorly understood. Recent whole genome screening identified a spectrum of cryptic aberrations associated with disease progression[5,6] sometimes even present at the onset of CML.[7] Also similar investigations of Ph(+)ALL in both children and adults identified recurrent cryptic loss of *Ikaros*, required for the normal B cell maturation[8] in addition to the known deletions of the *p*16 *(CDKN2A)* gene.

These findings led us to look in CML and Ph(+) ALL for imbalances in DNA sequences significantly associated with the disease stage and lineage origin. We used array CGH data obtained from 40 anonymous bone marrow samples comprising 10 CML chronic phase, 10 CML lymphoid blast phase, 10 CML myeloid blast phase, 10 Ph(+)ALL from the UKALLXII(R) trial [9] and 10 peripheral blood samples from disease free individuals. Of the ALL samples 5 had t(9;22)(q34;q11) as a sole cytogenetic abnormality, one was Ph negative but BCR/ABL positive, 3 showed hyperdiploid karyotype (HEH) but none showed aberrations detectable by G banding in the short arm regions of chromosome 7 and 9 (Table 1). The presence of the *BCR*/*ABL1* fusion gene was confirmed in all samples by qPCR and/or FISH (D-FISH probe, Vysis, USA) as reported previously.[9] All Ph(+)ALL and 5 out 10 CML blast crisis samples had established B cell immunophenotype.[6,10]

Having identified genome regions of potential interest, ranked in order of significance, out of the thousands of array results, it is then a challenge to design further experiments to evaluate their contribution to the biology of the BCR/ABL positive disease.

Materials and Methods

Array CGH analysis was performed as described previously.[6] Briefly, Agilent (Wilmington, DE, USA) oligonucleotide arrays were hybridized following the manufacturer's protocol. 500 ng genomic test DNA was extracted from either peripheral blood or BM samples. Sex mismatched pooled DNA from peripheral blood mononuclear fraction of 6–8 disease free individuals (Promega, UK) was used as reference. Customized Agilent oligonucleotide arrays comprising 8 × 15 k probe sets per slide were designed from an analysis of active loci (hot spots) in the CML BC genome corresponding to pairs of probes that exceeded a 3SD threshold in 3 or more CML BC samples from a previous study.[5] Probes were selected to cover regions at ~1 k intervals except where the presence of repetitive sequences disallowed the inclusion of reliable probes.

The arrays were scanned, features extracted and the data analyzed using an Agilent scanner and

Table 1. Summary of chromosome and fish results for the Ph(+)ALL cases.

No	Case ID	Karyotype	BCR breakpoint	FISH * BCR/ABL1			FISH* *p16*	FISH** *Ikaros*	FISH** *PAX5*	FISH** *MLLT3*
				Fusion present	Extra fusion	Deletion der(9)				
1	296/165	46,XX,t(9;22)(q34;q11)[11]/ 46,idem,der(5)t(1;5)(q2;q2)[8]/ 46,XX[3]	m-BRC	Yes	No	No	Loss	Loss	ND	ND
2	297/139/267	46,XY,t(9;22)(q34;q11)[10]	m-BCR	Yes	No	No	Diploid	Diploid	ND	Diploid
3	298/140	46,XY [10]	m-BRC	Yes	No	Yes***	No	ND	ND	ND
4	299/136	57,XY,+X,+Y,+2,add(2)(p?),+4,+8, t(9;22)(q34;q11),+10,+13,+14,+1 5,+21,+der(22)t(9;22)[4]/ 57,idem,add(2)(p?),+del(2)(p11),+ 11,-15[3]/ 46,XY[5]	m-BRC	Yes	Yes	Yes***	Yes***	Yes***	No	ND
5	300/166	46,XY,t(9;22)(q34;q11)[10]	m-BRC	Yes	No	Yes***	No	Yes***	Yes***	ND
6	301/142	46,XY,t(9;22)(q34;q11.2)[9]/ 46,XY[1]	m-BRC	Yes	No	No	No	No	No	ND
7	302/170	46,XX,t(9;22;17)(q34;q11;q21)[1]/ 48,idem,+X,+14[7]/ 46,XX[4]	m-BRC	Yes	No	No	ND	ND	ND	ND
8	303/131	52,XY,+X,+4,+5,+7,+8, t(9;22)(q34;q11),+20[6]/ 46,XY[4]	m-BRC	Yes	No	No	ND	ND	ND	ND
9	304/138	46,XY,t(9;22)(q34;q11)[7]/ 46,XY[3]	m-BRC	Yes	No	Yes**	No	Yes**	ND	ND
10	305/130/269	46,XY,t(9;22)(q34;q11)[20]	m-BRC	Yes	No	No	ND	ND	ND	ND

Notes: *D-FISH using commercial probes (Vysis); **using customised dual colour/dual probes (BAC and/or fosmid); ***cryptic loss identified only by high resolution array CGH.
Abbreviations: m-BCR, minor breakpoint; M-BCR, major breakpoint; ND, not done.

Mathematica software (http://www.wolfram.com). In addition, all samples had been subjected to whole genome screening using 105 K Agilent oligonucleotide arrays as part of published study.[6]

The emergence of high throughput technology such as microarrays raises a fundamental statistical issue relating to testing hundreds of hypotheses thus rendering the standard *P* value meaningless.[11] The False Discovery Rate (FDR) concept is an alternative to the *P* value. Tusher[1] described such a method: Significance Analysis of Microarrays (SAM) and the implementation due to Chu et al has been incorporated by J Craig Venter Institute into their suite of 'MeV' routines.[12] SAM uses permutations of sample labels to estimate the FDR. We report the application of SAM for 5,000 permutations setting the median number of false significant probes to zero, for the supervised analysis of the myeloid and lymphoid blast crisis, chronic CML, Ph(+)ALL and control samples. Firstly, after removing data for the sex chromosomes, we constructed a table defining the log fluorescence ratio (FR) for each locus and assigned classes eg, Lymphoid blast phase CML (L) or Myeloid blast phase CML (M); Ph(+)ALL (ALL); Chronic phase CML (C); Control (Ctrl); Male (m) or Female (f). We chose a two class unpaired test and applied SAM to ask if there were any probes that were uniquely associated with either classification. All genome addresses are derived from build 35 (March 2006) of the Human Genome.

Results

Genomic difference between lymphoid and myeloid lineages

We applied SAM to seek correlations between genome imbalances and clinical presentation. We asked which probes were significantly involved in the discrimination between lymphoid and myeloid lineages using the classes of myeloid and lymphoid CML BC as a model. Altogether we identified 489 significant probes, the top 100 of which were restricted to the TCR, IKZF1 and IgH genomic regions. Figure 1 shows cluster analysis

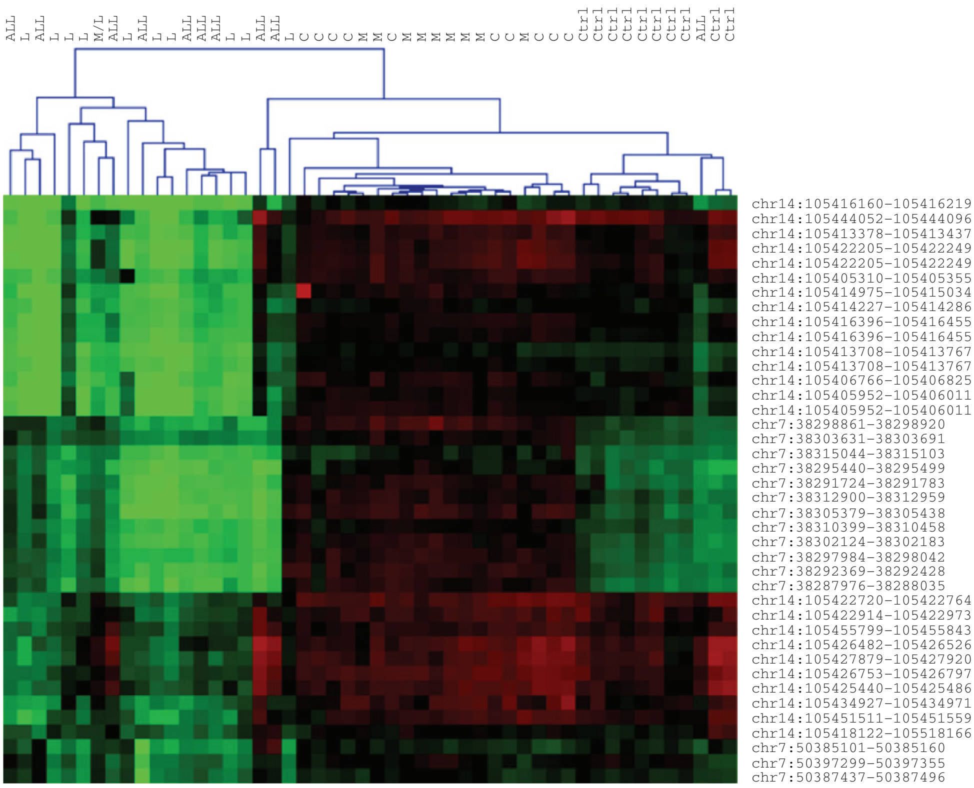

Figure 1. Top 40 most significant probes from a cluster analysis of 489, distinguishing lymphoid and myeloid BCR/ABL1 positive genomes.
Notes: SAM analysis listed a total of 489 significant probes that discriminate between lymphoid and myeloid BCR/ABL1 positive genomes. The top 20% of these probes were associated with the TCR @chr7:38,287,976-38,315,044 and the IgH region @chr14:105,405,310-105,518,122. Arrow points to homozygous deletions of the IgH probes (bright green) seen exclusively in Ph(+) samples with an early B cell lymphoid phenotype.

of the 40 most significant probes indicating losses occurring at genome address between 105,405,310 and 105,518,122 mbp in the IgH region, between 38,287,976 and 38,315,044 mbp in TCR and between 50,385,101 and 50,429,250 mbp in the sequences of IKZF1 (Table 2). Lymphoid samples including Ph(+)ALL clustered together displaying losses (Fig. 1 on the left), while the myeloid blast crisis and chronic CML samples formed a separate cluster with the control samples (Fig. 1 on the right). We noted that 3 samples sat at the myeloid/lymphoid borderline and that some of the control samples showed losses in the TCR region.

Table 2. The 40 most significant probes differentiating between lymphoid and myeloid lineages.

Probe	Chr	Target	Address mbp	Expected score	Observed score
A_18_P20213662	14	chr14:105414227–105414286	105.414227	–0.563993	–7.793448
A_18_P12135242	14	chr14:105413378–105413437	105.413378	–0.56458443	–7.7092276
A_16_P02991277	14	chr14:105422205–105422249	105.422205	–0.5605237	–7.3922744
A_16_P20175261	14	chr14:105426753–105426797	105.426753	–0.5587666	–7.2691813
A_18_P20215586	14	chr14:105422720–105422764	105.42272	–0.5601302	–7.137599
A_16_P02991277	14	chr14:105422205–105422249	105.422205	–0.56032795	–7.105713
A_16_P20175223	14	chr14:105416396–105416455	105.416396	–0.5630354	–6.8957753
A_16_P20175218	14	chr14:105414975–105415034	105.414975	–0.5638017	–6.803494
A_16_P40322315	14	chr14:105413708–105413767	105.413708	–0.5643853	–6.7806916
A_16_P20175223	14	chr14:105416396–105416455	105.416396	–0.56323034	–6.704411
A_16_P40322315	14	chr14:105413708–105413767	105.413708	–0.5641888	–6.6329618
A_18_P12136981	14	chr14:105422914–105422973	105.422914	–0.5599273	–6.6110125
A_16_P20175222	14	chr14:105416160–105416219	105.41616	–0.5634191	–6.426458
A_16_P01695449	7	chr7:38297984–38298042	38.297984	0.12489289	–6.2314205
A_16_P40322466	14	chr14:105455799–105455843	105.455799	–0.5500808	–6.107134
A_16_P17913603	7	chr7:38310399–38310458	38.310399	0.12737036	–6.08846
A_16_P37994547	7	chr7:38298861–38298920	38.298861	0.1250429	–6.059183
A_16_P02991291	14	chr14:105434927–105434971	105.434927	–0.5564316	–6.0575943
A_16_P40322313	14	chr14:105406766–105406825	105.406766	–0.56478184	–6.0261383
A_18_P12135286	14	chr14:105444052–105444096	105.444052	–0.5535429	–5.963885
A_16_P17913579	7	chr7:38302124–38302183	38.302124	0.1256673	–5.9104414
A_16_P37994526	7	chr7:38291724–38291783	38.291724	0.12381547	–5.8998003
A_18_P25472429	7	chr7:38303632–38303691	38.303632	0.12597458	–5.8726745
A_16_P02991281	14	chr14:105426482–105426526	105.426482	–0.55895203	–5.8314695
A_16_P02991264	14	chr14:105405952–105406011	105.405952	–0.5651721	–5.7746663
A_16_P17913610	7	chr7:38312900–38312959	38.3129	0.12768513	–5.768144
A_16_P02991284	14	chr14:105427876–105427920	105.427876	–0.5581815	–5.7465305
A_16_P40322308	14	chr14:105405310–105405355	105.40531	–0.56556815	–5.698316
A_16_P17913618	7	chr7:38315044–38315103	38.315044	0.12799555	–5.660891
A_16_P17941758	7	IKZF1	50.397299	0.1786682	–5.6329775
A_16_P17913559	7	chr7:38295440–38295499	38.29544	0.12458947	–5.617972
A_16_P02991279	14	chr14:105425440–105425486	105.42544	–0.5593441	–5.6175857
A_16_P01711866	7	IKZF1	50.385101	0.17555533	–5.6086626
A_16_P17941722	7	IKZF1	50.387437	0.17618023	–5.5908685
A_16_P02991264	14	chr14:105405952–105406011	105.405952	–0.5649764	–5.544837
A_18_P12136727	14	chr14:105451511–105451559	105.451511	–0.5502752	–5.5420628
A_18_P20215609	14	chr14:105518122–105518166	105.518122	–0.539346	–5.5144167
A_16_P01695465	7	chr7:38305379–38305438	38.305379	0.1264447	–5.477569
A_18_P16088644	7	IKZF1	50.42925	0.18619013	–5.435862
A_16_P37994528	7	chr7:38292369–38292428	38.292369	0.12397207	–5.42713

Notes: The differential (Expected—Observed) is a measure of significance.

Comparison of CML lymphoid blast crisis and Ph(+)ALL

84% of the 155 probes differentiating lymphoid blast crisis CML and Ph(+)ALL map to one of two regions of the short arm of chromosome 9, namely 9p21.3–p21.2 and 9p24.1–p23, the latter housing genes PTPRD and MLLT3 among others. A hierarchical cluster analysis shows that five of the 10 Ph(+)ALL cases form a cluster of gains (Fig. 2, in red) although cytogenetic revealed no structural or numerical changes of 9p (Table 1). In contrast, half of the 10 CML BCL cases formed a cluster with extensive genome loss (in green) that had been previously shown to be complex by G-banding and 105 K oligonucleotide array.[6] See Figure 2 and Table 3.

Since many of the significant probes fell on chromosome 9p we repeated the analysis excluding all chromosome 9 loci. The top 10 of 80 probes meeting our significance threshold are revealed by cluster analysis (see Fig. 3 and Table 4). Associated with these loci are known genes such as PDEA4 (cAMP-diestarase) in band 19p13.2 and GSTT1 in band 22q11.23, one of the most commonly reported polymorphic marker (CNV) in man. Genome loss (in green, Fig. 3) dominates the profile of 6 out 10 Ph(+)ALL samples. Surprisingly 5 of the latter cases (297, 299, 300, 301, 303) exhibit gains in the chromosome 9p21–p24 region (Fig. 3, heat map A).

It is suggested from the heat maps in Figures 2 and 3 that the Ph(+)ALL samples split into two

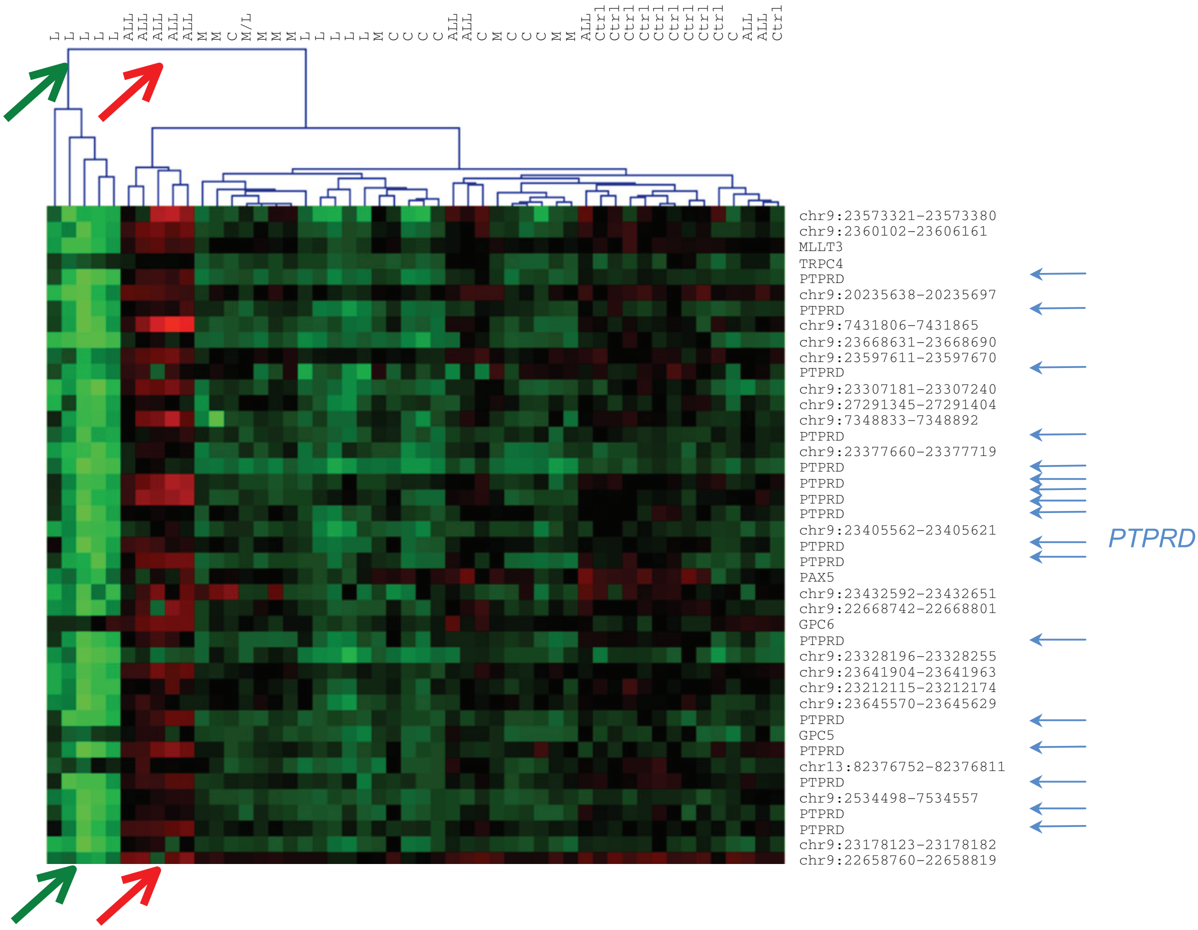

Figure 2. Identification of probes discriminating between ph positive acute lymphoblastic leukemia and CML lymphoid blast transfomation.
Notes: Heat map of the SAM data showing gains (red) and losses (green) for the 40 probes most influential in discriminating between Ph(+)ALL and BC/L CML samples. Altogether 16 of these probes (arrowed) cover the region of the PTPRD gene (protein tyrosine phosphatase, receptor type, D) on 9p24.1–p23. A sub group of 5 Ph(+)ALL samples show gains of chromosome 9p loci (red arrow), whilst the same region in 5 BC/L CML samples is deleted (green arrow) in agreement with their chromosome status.

Table 3. The 40 most significant probes differentiating between Ph(+)ALL and lymphoid blast crisis.

Probe	Chr	Target	Address mbp	Expected score	Observed score
A_16_P18581907	9	chr9:23573321–23573380	23.573321	1.217212	–5.9455347
A_18_P26464352	9	chr9:23606102–23606161	23.606102	1.2239254	–5.305494
A_16_P02074143	9	MLLT3	20.358557	0.870436	–4.711568
A_14_P136633	13	TRPC4	37.12336	–0.8066902	–4.5697494
A_16_P02056848	9	PTPRD	8.670463	0.69564855	–4.4780617
A_18_P16772095	9	chr9:20235638–20235697	20.235638	0.8584173	–4.453213
A_16_P18547080	9	PTPRD	9.114454	0.7288707	–4.4257317
A_18_P26401468	9	chr9:7431806–7431865	7.431806	0.6581821	–4.364759
A_16_P18582203	9	chr9:23668631–23668690	23.668631	1.2362497	–4.320358
A_16_P18581984	9	chr9:23597611–23597670	23.597611	1.2221186	–4.288423
A_16_P18546259	9	PTPRD	8.8232	0.70854175	–4.271888
A_16_P18581300	9	chr9:23307181–23307240	23.307181	1.178406	–4.268987
A_16_P18590554	9	chr9:27291345–27291404	27.291345	1.2408848	–4.2311144
A_16_P18542386	9	chr9:7348833–7348892	7.348833	0.6530274	–4.1917768
A_18_P16743757	9	PTPRD	9.54851	0.7599531	–4.189755
A_16_P38678778	9	chr9:23377660–23377719	23.37766	1.1883749	–4.1894665
A_16_P38644284	9	PTPRD	9.261361	0.73997337	–4.1806855
A_16_P18545920	9	PTPRD	8.706239	0.69865716	–4.108792
A_16_P38647612	9	PTPRD	10.509098	0.83521545	–4.095142
A_16_P38643011	9	PTPRD	8.791208	0.70617497	–4.092501
A_16_P02078229	9	chr9:23405562–23405621	23.405562	1.1912949	–4.0642586
A_16_P18548249	9	PTPRD	9.557477	0.7608688	–4.042481
A_18_P16749071	9	PTPRD	10.587629	0.8423635	–4.0077705
A_16_P02096462	9	PAX5	36.886181	1.439173	–3.9982088
A_16_P02078274	9	chr9:23432592–23432651	23.432592	1.1954948	–3.9890935
A_16_P18579921	9	chr9:22668742–22668801	22.668742	1.1000694	–3.9809976
A_16_P19916466	13	GPC6	93.756923	–0.729778	–3.9487815
A_16_P18551018	9	PTPRD	10.585886	0.8421173	–3.9478562
A_16_P18581339	9	chr9:23328196–23328255	23.328196	1.1817318	–3.9457936
A_16_P38679470	9	chr9:23641904–23641963	23.641904	1.2316526	–3.9433708
A_16_P02077983	9	chr9:23212115–23212174	23.212115	1.1630332	–3.937294
A_16_P18582128	9	chr9:23645570–23645629	23.64557	1.232107	–3.927982
A_16_P38642152	9	PTPRD	8.516555	0.68393993	–3.9103673
A_16_P19909107	13	GPC5	91.029609	–0.735624	–3.8958812
A_16_P38645300	9	PTPRD	9.629695	0.7672258	–3.861492
A_16_P02822684	13	chr13:82376752–82376811	82.376752	–0.74305904	–3.8531454
A_16_P38641641	9	PTPRD	8.373899	0.6731849	–3.8528354
A_16_P18542873	9	chr9:7534498–7534557	7.534498	0.66606975	–3.847243
A_16_P18550697	9	PTPRD	10.463268	0.832524	–3.8354936
A_18_P16742333	9	PTPRD	10.444266	0.83104974	–3.278083

Notes: The differential (Expected—Observed) is a measure of significance.

groups, 5/10 cases showing dominant amplification of loci in the chromosome 9p region and losses elsewhere in the genome, while the remainder (5/10) lack recurrent genome imbalances. However, we were unable to detect any consistent differences in the two groups of Ph(+)ALL samples from an inspection of their chromosome status (see Table 1).

In summary, SAM analysis revealed that while the lymphoid blast stage CML and Ph(+)ALL samples share common losses within the IGH, TCR, and *Ikaros* gene regions together with loci within the 9p21–p24 region, they form separate clusters at other sites on the genome thus suggesting that these acute malignant conditions may represent separate biological entities.

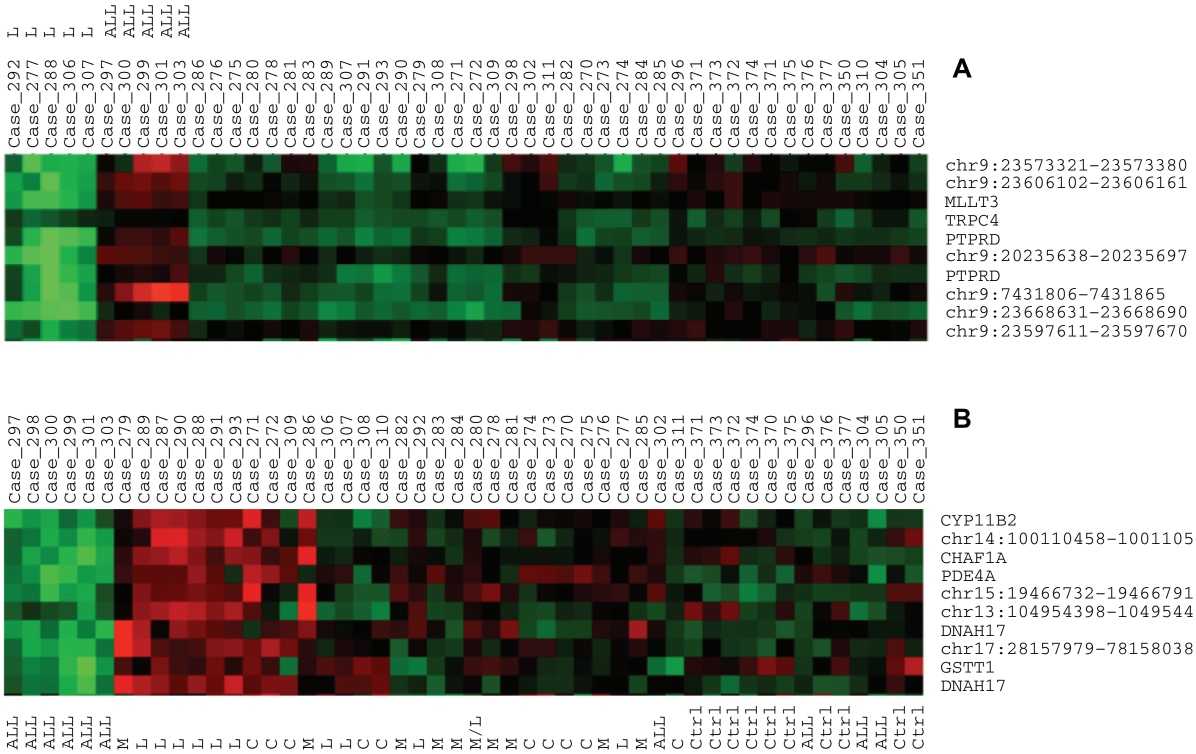

Figure 3. Ph positive all with gains at 9p21–p24 share common losses elsewhere in the genome.
Notes: Cluster analysis of SAM data identified cryptic gains within the 9p21–p24 region in a sub-group of Ph(+)ALL samples (framed in heat map A, losses in green and gains in red). When cluster analysis was performed on SAM results of the genome excluding chromosome 9 data, some 115 probes were shown to discriminate between Ph(+)ALL and BCL CML, the top 10 of which are shown on the heat map B (losses in green and gains in red). All but one (arrowed) of these Ph(+)ALL samples with gains in the 9p21–p24 region share cryptic loss elsewhere in the genome (framed, heat map B), which involves relevant genes such as CYP11B2 (cytochrome P450), GCSTT1 (member of the glutathione transferase gene family with known role in carcinogenesis) and CHAF1A (chromatin assembly factor) among others.

Discussion

Whilst huge progress has been made in the analysis of the genome and the identification of genes associated with malignant disease, there is still much work to be done evaluating the function of coding and non-coding regions[13]. We have identified numerous short 60 mer sequences that appear to play a significant role in the evolution of Philadelphia positive hematological malignancy. We offer no explanation of their function, but provide convincing evidence that their involvement is not a random event.

SAM is used for the analysis of expression arrays to classify samples into groups according to phenotype using false discovery rate (FDR) as a test for significance.[1,14] Here we use SAM to study DNA from a cohort of CML and Ph(+)ALL patients to identify sequences that may help to distinguish between these Philadelphia positive diseases and enlighten their pathogenesis.

Numerous software packages are available for the detection of genomic gains and losses across a range of array technologies, reviewed by Shah,[15] but high-resolution array data presents special problems as typified by a wide variance making detection of small features complicated. Individual signals are rarely if ever considered to be significant on their own but only in the context of a contiguous collections of gains and losses. However if an individual locus is compared across a number of similarly processed arrays, the probability of a random single signal exceeding a 3SD threshold for ***n*** arrays is reduced to approximately (0.003)^**n**. FR signals meeting these criteria could be attributable to artifacts of the array and laboratory procedures or could be *bona fide* data with clinical significance. Since all the samples in our study were prepared under the same conditions and hybridized to the same batch of arrays, we believe that the results reflect recurrent genome copy number aberrations. Our data shows consistent, recurrent gains and losses although in many cases the FR data falls well short of the theoretical values suggesting the presence of clonal cell populations—a common phenomenon

Table 4. The 40 most significant probes differentiating between Ph(+)ALL and lymphoid blast crisis excluding chromosome 9 probes.

Probe	Chr	Target	Address mbp	Expected score	Observed score
A_16_P02042251	8	CYP11B2	143.994134	–0.49485415	–3.6447732
A_16_P02984473	14	chr14:100110458–100110502	100.110458	–0.3787858	–3.4324715
A_18_P12974340	19	CHAF1A	4.361883	0.73060995	–3.3821597
A_16_P41139830	19	PDE4A	10.4373	0.588393	–3.3786511
A_16_P02992588	15	chr15:19466732–19466791	19.466732	–0.3425412	–3.357904
A_14_P127778	13	chr13:104954398–104954457	104.954398	–1.219279	–3.2696884
A_18_P12746992	17	DNAH17	73.949223	0.72273785	–3.236283
A_16_P20748371	17	chr17:78157979–78158038	78.157979	0.23752803	–3.197567
A_16_P41491725	22	GSTT1	22.713973	0.59368485	–3.112421
A_16_P03295625	17	DNAH17	73.998258	–0.27195758	–3.1080885
A_18_P20749354	16	chr16:88235603–88235662	88.235603	1.1027136	–3.094033
A_16_P40556488	16	chr16:10502024–10502083	10.502024	0.5570203	–3.0405986
A_16_P20957257	19	PDE4A	10.434299	0.2495674	–3.021277
A_16_P02380585	11	chr11:1822652–1822711	1.822652	–0.4423064	–2.9679623
A_16_P03204073	17	P2RX5	3.525384	–0.3101799	–2.9653873
A_16_P03605977	22	GSTT1	22.710302	–0.22612213	–2.9506812
A_16_P20690191	17	GDPD1	54.658506	0.22053438	–2.9496646
A_16_P20568636	17	chr17:3283949–3284008	3.283949	0.19664884	–2.8752325
A_14_P132357	17	38595	72.827808	–1.1807562	–2.8733869
A_16_P02738683	13	SPATA13	23.761279	–0.41567326	–2.8507361
A_18_P10967751	11	chr11:1252096–1252155	1.252096	0.63722944	–2.8453183
A_16_P03601865	22	RANBP1	18.494441	–0.23352942	–2.8380888
A_16_P35654245	2	MTA3	42.787245	0.34639606	–2.8232927
A_16_P17507488	6	chr6:35880489–35880548	35.880489	–0.12230495	–2.8113792
A_16_P39158050	10	CNNM2	104.799592	0.49181423	–2.798708
A_16_P39157984	10	CNNM2	104.769703	0.49056116	–2.7931175
A_18_P20982233	17	chr17:72013234–72013278	72.013234	1.128489	–2.7917495
A_16_P19045973	10	CNNM2	104.701305	0.07244033	–2.7840393
A_18_P12745203	17	DNAH17	73.995598	0.72024244	–2.7796998
A_16_P15021468	1	RERE	8.650802	–0.21963291	–2.7703693
A_16_P02991501	14	chr14:105737025–105737080	105.737025	–0.35576868	–2.7698457
A_16_P01729453	7	CALN1	71.497024	–0.6459507	–2.767063
A_16_P35654127	2	MTA3	42.736918	0.34621632	–2.7501495
A_14_P120530	21	AIRE	44.541215	–1.3012085	–2.7431655
A_18_P12659257	17	KIAA1267	41.493236	0.70809025	–2.7361684
A_16_P02873949	14	chr14:22054604–22054663	22.054604	–0.38419548	–2.7211492
A_16_P20660234	17	NSF	42.162	0.21277878	–2.7152972
A_16_P03422686	19	chr19:10389923–10389975	10.389923	–0.24484406	–2.7065868
A_16_P19930395	13	chr13:99371784–99371843	99.371784	0.1376494	–2.7023811
A_16_P40862888	17	GDPD1	54.689835	0.5801528	–2.6963766

Notes: The differential (Expected—Observed) is a measure of significance.

in haematological disease. SAM does not guarantee that the list of ‘significant’ loci are involved in the various classifications discussed here, but it does offer a list of candidates for cluster analysis or other investigative methods. Many gene copy number changes irrespective of the length of the affecting sequences such as those recorded here can contribute directly to monogenic diseases[16]. In recessive diseases, hemizygosity due to deletion of a gene, or part of a gene, could unmask a mutation on the other gene copy, while duplication of a healthy gene copy on one chromosome could theoretically mask the effects of a disease- causing mutation in the gene on the other chromosome, thus rescuing the phenotype.

We designed a high resolution array (~1 kb intervals) designed to explore regions of the genome shown previously at low (33 kb) resolution to display gains or losses in a cohort of 35 samples from CML patients in blast phase[6]. This necessitated sacrificing large areas of the genome to concentrate on these areas for detailed inspection. Using this set of ~15,000 genetic loci enabled us to confirm that lymphoid phenotypes formed a single group characterized primarily by unique deletions within the IgH regions consistent with an early VDJ rearrangement as part of the B cell receptor formation occurring in per-B cells together with loss of the TCR gamma sequences also indicating gene rearrangement. Loss of whole or part of the IKZF1 (*Ikaros*) gene is the third most common feature in the genome profile of these cases. In contrast with a typical CNV that could affect any part of the IgH gene on 14q32.33 the deletions identified by us always involve the sequences 105.41–105.48 mbp and are almost universally accompanied by deletions in the TCR region of chromosome 7. Both IgH and TCR sequences are usually excluded from aCGH analysis as they are reported to be CNVs. We have demonstrated that these deletions are consistent throughout the sample set, suggesting that they are disease specific. These findings could be explained by a chain of events initiated by *BCR*/*ABL1* that leads to compromised V(D)J recombinase machinery thus creating clonal populations of early B-cell progenitors with cross lineage rearrangements.[6] Mullighan et al in their poster presentation "Genome wide analysis of Genetic Aberrations in Chronic Myeloid Leukemia" (Mullighan et al, http://ash.confex.com/ash/2008/webprogram/Paper5715.html) reported results from SNP analysis of 90 CML samples of which 9 were diagnosed as lymphoid blast crisis. This study could not find any genomic features that could differentiate between BC/L and Ph(+)ALL. In contrast we were able to reveal genomic differences in these clinically similar conditions.

Many of the 'significant' probes that distinguish between Ph(+)ALL and BC/L cluster within chromosome 9p21 region, which harbours the CDKN2A/B gene, the loss of which has long been associated with both haematological and solid tumours and shown to result from RAG impairment.[8] Since 5 out of the 10 BC/L CML cases were found to carry imbalances of the short arm of chromosome 9, it is possible that some probes in this location were lost 'by association' and not involved in discriminating between these two diseases. However, we found other loci that did discriminate between Ph(+)ALL and BC/L CML. For example, while half of the BC/L cases had deletions in chromosome 9p, half of the of Ph(+)ALL showed gains at these loci as shown in Figure 2. Omitting chromosome 9 records and reanalyzing the data, the same five Ph(+)ALL samples showed significant losses in 80 loci from other chromosomal locations (Fig. 3). Taken together the tandem CNA—gains at 9p with recurrent losses elsewhere in the genome offer a way to differentiate a Ph(+)ALL from CML lymphoid BC. Whilst we recognize that single aberrant 60 mer sequences could easily be dismissed as random events, the fact that there are more than 80 widely distributed probes not associated with morphological or cytogenetic anomalies but associated with a significant minority of the Ph(+)ALL samples is worth consideration. Further work is required to explore the possible role of these genome aberrations. In conclusion, SAM results offer clues regarding the pathogenesis of *BCR*/*ABL1* positive disorders and furthermore identifies a sets of probes with diagnostic potential.

Authors' Contribution

CG carried out the SAM analysis and co-wrote the manuscript; EPN designed the study and co-wrote the manuscript.

Acknowledgements

We are grateful to Dr. Adele Fielding, Prof. Letizia Foroni and Prof. Anthony Moorman for valuable comments on the manuscript and providing samples; to Dr. Diana Brazma who carried out the aCGH and performed FISH tests and Mrs. J. Howard-Reeves for karyotyping and FISH analysis.

Funding

This work has been supported by LRF grant No. 05098, Jean Coubrough Charitable Trust Project to Dr. EP. Nacheva.

Competing Interests

EPN's institution has received grants from the Leukemia Research Fund and the Jean Coubrough Charitable Trust, and travel or study support from

Cytocell. Research funding was also received from operation of a clinic associated with UCL. CG discloses no conflict of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Tusher V, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98(9): 5116–21.
2. Larsson O, Wahlestedt C, Timmons JA. Considerations when using the significance analysis of microarrays (SAM) algorithm. *BMC Bioinformatics*. 2005;6:129.
3. Sridhar K, Ross D, Tibshirani R, Butte A, Greenberg P. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. *Blood*. 2009.
4. Bacher U, Haferlach T, Hiddemann W, Schnittger S, Kern W, Schoch C. Additional clonal abnormalities in Philadelphia-positive ALL and CML demonstrate a different cytogenetic pattern at diagnosis and follow different pathways at progression. *Cancer Genet Cytogenet*. 2005;157(1):53–61.
5. Brazma D, Grace C, Howard J, et al. Genomic profile of chronic myelogenous leukemia: Imbalances associated with disease progression. *Genes Chromosomes Cancer*. 2007;46(11):1039–50.
6. Nacheva EP, Brazma D, Virgili A, et al. Deletions of Immunoglobulin heavy chain and T cell receptor gene regions are uniquely associated with lymphoid blast transformation of chronic myeloid leukemia. *BMC Genomics*. 2010; 11:41.
7. Khorashad JS, De Melo VA, Fiegler H, et al. Multiple sub-microscopic genomic lesions are a universal feature of chronic myeloid leukaemia at diagnosis. *Leukemia*. 2008;22(9):1806–7.
8. Mullighan C, Miller C, Radtke I, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*. 2008;453(7191):110–14.
9. Moorman AV, Harrison CJ, Buck GA, et al. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood*. 2007;109(8):3189–97.
10. Fielding AK, Rowe JM, Richards SM, et al. Prospective outcome data on 267 unselected adult patients with Philadelphia chromosome-positive acute lymphoblastic leukemia confirms superiority of allogeneic transplantation over chemotherapy in the pre-imatinib era: results from the International ALL Trial MRC UKALLXII/ECOG2993. *Blood*. 2009;113(19):4489–96.
11. Gusnanto A, Calza S, Pawitan Y. Identification of differentially expressed genes and false discovery rate in microarray studies. *Current Opinion in Lipidology*. 2007;18(2):187–93.
12. Saeed AI, Bhagabati NK, Braisted JC, et al. TM4 microarray software suite. *Meth Enzymol*. 2006;411:134–93.
13. Check Hayden E. Human genome at ten: Life is complicated. *Nature*. 2010;464(7289):664–7.
14. Kerr KF. Comments on the analysis of unbalanced microarray data. *Bioinformatics*. 2009;25(16):2035–41.
15. Shah SP. Computational methods for identification of recurrent copy number alteration patterns by array CGH. *Cytogenet Genome Res*. 2008;123(1–4): 343–51.
16. de Smith AJ, Walters RG, Froguel P, Blakemore AI. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res*. 2008;123(1–4):17–26.