Genome Biology

Genome Biology

Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy

Genome Biology 2011, **12**:R61 doi:10.1186/gb-2011-12-6-r61

Michal Ozery-Flato (ozery@il.ibm.com)
Chaim Linhart (chaiml@post.tau.ac.il)
Luba Trakhtenbrot (Luba.Trakhtenbrot@sheba.health.gov.il)
Shai Izraeli (Shai.Izraeli@sheba.health.gov.il)
Ron Shamir (rshamir@post.tau.ac.il)

ISSN	1465-6906
Article type	Research
Submission date	10 February 2011
Acceptance date	29 June 2011
Publication date	29 June 2011
Article URL	http://genomebiology.com/2011/12/6/R61

Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy

Michal Ozery-Flato[1,2], Chaim Linhart[1], Luba Trakhtenbrot[3,4], Shai Izraeli[3,5,6], and Ron Shamir[1,§]

[1]The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, 69978, Israel

[2]Machine learning and data mining group, IBM Haifa Research Lab, ISRAEL BMC Software Building, 6 Habarzel St., Mount Carmel, Haifa, 31905, Israel

[3]Chaim Sheba Cancer Research Center, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Ramat Gan, 52620, Israel

[4]Institute of Hematology, and Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Ramat Gan, 52620, Israel

[5]Department of Pediatric Hemato-Oncology, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Ramat Gan, 52620, Israel

[6]Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

[§]Corresponding author

Email addresses:

MOF:	ozery@il.ibm.com
CL:	chaiml@post.tau.ac.il
LT:	Luba.Trakhtenbrot@sheba.health.gov.il
SI:	Shai.Izraeli@sheba.health.gov.il
RS:	rshamir@post.tau.ac.il

Abstract

Background

Chromosomal aneuploidy, that is to say the gain or loss of chromosomes, is the most common abnormality in cancer. While certain aberrations, most commonly translocations, are known to be strongly associated with specific cancers and contribute to their formation, most aberrations appear to be non-specific and arbitrary, and do not have a clear effect. The understanding of chromosomal aneuploidy and its role in tumorigenesis is a fundamental open problem in cancer biology.

Results

We report on a systematic study of the characteristics of chromosomal aberrations in cancers, using over 15,000 karyotypes and 62 cancer classes in the Mitelman Database. Remarkably, we discovered very high co-occurrence rate of chromosome gains with other chromosome gains, and of losses with losses. Gains and losses rarely show significant co-occurrence. This finding was consistent across cancer classes and was confirmed on an independent comparative genomic hybridization dataset of cancer samples. The results of our analysis are available for further investigation via an accompanying website.

Conclusions

The broad generality and the intricate characteristics of the dichotomy of aneuploidy, ranging across numerous tumor classes, are revealed here rigorously for the first time using statistical analyses of large-scale datasets. Our finding suggests that aneuploid cancer cells may use extra chromosome gain or loss events to restore a balance in their altered proteins ratios, needed for maintaining their cellular fitness.

Background

Most cancer genomes undergo large scale alterations that dramatically alter their content and structure [1]. This phenomenon of genomic instability is responsible for the wide repertoire of chromosomal aberrations observed in cancer genomes. While the role of most aberrations in the carcinogenesis process remains to be determined, the common perception [2] is that some of these aberrations are functionally important to the initiation and growth of cancer (drivers), while others merely represent random somatic changes that carry no selective advantage to the cancer cell (passengers). The identification of strong associations among aberrations, i.e. associations that are observed significantly more than expected by chance, may help in the detection of driver aberrations or point to mechanisms that promote the selection of certain aberrations. As data on chromosomal aberrations in cancer accumulate, the detection of such strong associations can become more accurate and powerful.

Following the four-step model for colorectal cancer evolution suggested by Vogelstein et al.[3, 4], several computational methods were developed for reconstructing common evolutionary paths of chromosomal aberrations in specific cancers. Some of these methods used tree models [5-7], later extended to acyclic networks [8-10]. These evolutionary models enable recognition of aberrations that occur at early stages of cancer; often referred to as "primary", they are suspected of being cancer drivers. As these methods were designed to analyze samples from the same cancer type, they were applied to relatively small datasets, each containing a few hundred samples. More recently, a statistical method named GISTIC [11] was developed for identifying copy-number aberrations whose frequency and amplitude are higher than expected. This method was used in [12] for studying copy number alterations (CNVs) appearing at significant frequency across several cancer types. In another recent study [13], profiles of frequent deletion events were analyzed in order to distinguish between driver and passenger deletion events. The latter two studies focus on CNVs of focal regions, derived by high-resolution techniques, from a heterogeneous pool of cancers with several hundreds to few thousands samples.

The Mitelman database [14] is the largest depository of chromosomal aberrations in cancer. Although the aberrations are described using karyotypes of low resolution, these methods are widely used, notably in hospital labs where the database is the leading source of information for clinicians who diagnose and treat cancer. The large number of samples in the database makes it ideal for statistical analyses, which are capable of overcoming random errors. In this study we present the results of large-scale analysis of chromosomal aberrations from over 15,000 karyotypes of the Mitelman database. By exploiting the huge number of karyotypes, reconstructing the aberrations in them, and developing appropriate statistical tests, we were able to recognize significant cross-cancer associations among aberrations and to identify correlations among tumor types.

Most observed alterations include chromosome gains / losses and translocations. As translocations directly affect a small number of genes, the role of many translocations in cancer causation has become much clearer over the years [15]. Chromosome gains and losses, on the other hand, are broad alterations affecting numerous genes whose significance to the carcinogenesis process is much less understood. In this study we demonstrate strong associations involving chromosome gain and loss aberrations, suggesting selection preferences for aneuploid cells.

The results of our analysis, including the computed associations and links to their underlying karyotypes, are publicly available for further investigation via our website [16]. Each karyotype is linked to its original record in the Mitelman database, thus allowing browsing of its full details. To the best of our knowledge, this is the first resource providing statistical results on such associations among cancer karyotypes.

Results

Figure 1 summarizes our karyotype analysis. Starting from 59,579 karyotypes in the Mitelman database (November 2009 version), we used only 34,107 karyotypes that were annotated as unselected in order to avoid over- or under-estimation of aberration frequencies due to biases in sample selection [17]. We then filtered out any partially characterized or possibly redundant karyotypes, as well as karyotypes that were not near diploid. Tumor classes were defined according to tissue morphology and organ. Karyotypes belonging to classes with small representation (<50 karyotypes) in the remaining dataset were omitted from analysis, resulting in a total of 62 classes and 15,445 karyotypes (Table 1).

Each class was assigned to one of four sets: lymphoid disorders, non-lymphoid hematological disorders, benign solid tumors, and malignant solid tumors (Table 1). Due to its higher rate of successful karyotypic analyses, the group of hematological disorders dominated our dataset, with 11,324 (73%) karyotypes, of which 6,913 (45%) belong to non-lymphoid hematological disorders. We computed for each karyotype a set of most likely aberrations involved in its formation using 11 types of chromosomal rearrangement, deletion, and duplication events

(Materials and methods, Table S1 in Additional file 1). Of those events, chromosome gain / loss and translocation were most frequent (Figure S1 in Additional file 2). An aberration was identified by its causing event and the chromosomal locations it involved. For example, the translocation involving bands 9q34 and 22q11 was identified by t(9;22)(q34;q11), following the ISCN terminology [18].

Aberrations characteristic of specific tumor classes

The karyotypes in our dataset contained 5,179 distinct aberrations, including all possible chromosome gains and losses. We computed the significance of the correlation of each aberration-class pair using the hypergeometric test. Out of 9,208 distinct observed aberration-class pairs, 1705 were found to be significantly correlated at false discovery rate (FDR) of 5% (website). These correlations encompassed all 62 tumor classes in our dataset, involving 1,360 distinct aberrations, where more than half of these correlations (907, 53%) involved translocations. Many of these strong correlations, notably the ones involving translocations, have been well documented in the literature: for example, t(9;22) in chronic myelogenous leukemia [19] and t(11;22) in Ewing sarcoma [20]. This supports the use of our dataset as a valid sample of karyotypes from the considered classes, as well as the soundness of our results.

Two distinct paths to aneuploidy?

We now address a question that can be answered only by complex analysis of a large database: which aberrations tend to co-occur? We seek pairs of aberrations that appear together in karyotypes significantly more than expected by chance. Such associations may reveal either cooperation between different oncogenic events or common mechanisms creating chromosomal aberrations. To answer this question we tested the significance of co-occurrence for 7,202 aberration pairs in our dataset that satisfied the following two conditions: each aberration appeared in at least 10 karyotypes, and the pair appeared together in at least one karyotype. We first filtered pairs with hypergeometric P-value >0.001, leaving 623 pairs whose significance was further evaluated by a permutation test. Our analysis yielded 218 significantly co-occurring aberration pairs (P<0.05, after Bonferroni correction), of which 154 (71%) were chromosome gain pairs, and 47 (22%) were chromosome loss pairs. The induced network split clearly into two disjoint parts: one dominated by chromosome gains and one by chromosome losses (Figure 2a). We carried out the same analysis separately for lymphoid disorders, non-lymphoid hematological disorders, solid tumors, and carcinomas (Figure S2a-d in Additional file 3). Each of these groups showed the same clear strong co-occurrence of specific gain-gain and loss-loss pairs, with almost no cases of significant co-occurrence for any mixed gain-loss pairs. We also detected the trisomy of 1q [21], which appeared in all tumor categories in the associations involving gain of chromosome 1 (Figures 2a, S2a-d in Additional file 3).

We repeated this test on an extended dataset of 42,763 karyotypes, which included selected and partially-characterized karyotypes (omitting non-characterized fragments). The two disjoint clusters of chromosome gains and losses are still clearly evident in the obtained results (Figure S2e in Additional file 3). The major observed change in the results is the addition of many new significant associations that involve aberrations other than chromosome gains and losses. This addition is explained by the growth in the amount of data, which increased the power of the statistical test, allowing it to uncover weaker associations. To confirm this, we examined an extended set of significant co-occurring aberrations (FDR 5%) in the original (filtered) dataset and obtained essentially the same results (not shown).

To test our result on independent data obtained using a different technology, we used data from comparative genomic hybridization (CGH), a laboratory method to measure gains and losses in the copy number of chromosomal regions in tumor cells. We analyzed an independent dataset of 1084 samples obtained by CGH, downloaded from the NCI and NCBI's SKY/M-FISH and CGH database (March 16, 2009 version). This database contains CGH records contributed by molecular cytogeneticists for open investigation. Each sample was assigned a corresponding set of whole chromosome gain/loss aberrations, yielding 648 (60%) samples with non-empty aberration sets. Using a permutation test similar to the one used for karyotypes data (Materials and Methods), we computed a P-value for the co-occurrences of specific aberration pairs in the CGH dataset. Out of 856 distinct co-occurring aberrations pairs, 47 were significantly co-occurring at FDR of 5%. The picture obtained by these pairs (Figure 2b) is strikingly similar to the one produced by the karyotype data. This reaffirms our observation that the progression of aneuploidy in cancer is driven by either multiple chromosomal gains or multiple chromosomal losses.

A similarity map of tumor classes

Which tumor classes have highly similar aberrations? Using the set of significant (FDR 5%) aberration-class correlations, we assessed the statistical significance of the overlap in aberrations for every pair of tumor classes. Of all 1891 possible class pairs, 56 pairs were found to significantly share common aberrations at an FDR of 5% (Figure S3a1 in Additional file 4).

Considering benign and malignant solid tumors as one category, all but three (53, 95%) of these pairs belong to the same category, with two of the three exceptions linking between lymphoid disorders and (malignant) solid tumors. We repeated the analysis, expanding the set of correlative aberrations by considering also weaker correlations with (uncorrected) P-value <0.05. The results show a remarkably similar partition, with 86 significant class pairs (FDR 5%), forming three distinct clusters, with only six links between the sets of lymphoid disorders and solid tumors (Figure S3a2 in Additional file 4). The fact that the categories were very well separated serves as confirmation of the data and of our methodology.

For more in-depth study of similarity among classes, we defined a similarity measure between classes based on the significance of their common aberrations (Materials and Methods) and used it to hierarchically cluster the classes (Figure 3). As before, classes of the three sets – non-lymphoid-hematological disorders, lymphoid disorders and solid tumors – clustered separately. A deeper look into each cluster (Figure 3) revealed that many closely clustered classes were histologically related. For example: diffuse large B-cell lymphoma, follicular lymphoma, and mature B-cell neoplasm (B-cell lymphomas); adenoma and adenocarcinoma in the large intestine; and AML M5 and AML M5a. The correlated aberrations shared by two similar classes can be viewed through our website. One of the interesting results was the close proximity of three embryonic cancers: Wilms' tumor (kidney), Ewing sarcoma (skeleton) and Hepatoblastoma (liver).

The website

All the associations described above can be viewed via the website [16], which contains summary tables for the different types of associations: aberration-class, class-class, and aberration-aberrations. Table rows can be filtered textually and numerically, allowing investigations of associations for a specific group of cancer types, a set of aberrations of interest, or both. For example, the user can view all aberrations whose correlation with a certain tumor class is below some specified P-value. Alternatively, all aberrations significantly co-occurring with a specified aberration can be examined, with their P-values. For aberration-class and aberration-aberration associations, researchers can examine the karyotypes that led to these associations, where each karyotype is linked to its corresponding record in the Mitelman database website.

To demonstrate the utility of the website, we focused on hyperdiploid multiple myeloma (H-MM), a subtype of multiple myeloma (MM) with better prognosis, characterized by having 48-74 chromosomes [22-24]. There were 385 MM karyotypes in the database, and 110 (29%) of which were hyperdiploid. H-MM is associated with recurrent gains of chromosomes 3, 5, 7, 9, 11, 15 and 19 [22]. Indeed, the website's class-aberration table, filtered for MM associations, confirmed this observation: +3, +5, +9, +11, +15, and +19 were the aberrations most associated with MM, and the 142 karyotypes involved in these associations spanned all H-MM karyotypes (hyper-geometric $P < 1E-76$). Chng et al. [25] suggested a FISH-based trisomy index test for identifying H-MM, employing probes for chromosomes 9, 11 and 15, and designating a tested MM cell as H-MM if it contains two or more trisomies in these chromosomes (see Materials and Methods). They reported specificity of 0.98 and sensitivity of 0.69 for that index. The corresponding F-Score (a measure combining sensitivity and specificity, see Materials and

Methods) was 0.8. We analyzed the 385 MM karyotypes in the same fashion as [25]; the criterion of any two trisomies in 9, 15, 19 was best with specificity 0.996 and sensitivity 0.88 [F-Score 0.93]. In fact, the same combination has the highest F-Score on the data of [25] as well (0.83). Thus, the criterion of two or more trisomies of chromosomes 9, 15, 19 should be considered for identifying H-MM.

Discussion

In this study we computationally analyzed a large number of cancer karyotypes from the Mitelman database, the largest available compendium of cancer karyotypes. Based on statistical analysis of more than 15,000 karyotypes, our results provide strong additional evidence for the non-randomness of many chromosomal aberrations in cancer. Our approach is validated by the demonstration of known relationships, including associations between specific aberrations and specific tumor types, and similarities among certain tumors (e.g. adenoma and adenocarcinoma of the large intestines). More importantly, the analysis led to new discoveries, most notably that chromosomal aneuploidy tends to consist of either a pattern of chromosomal gains or a pattern of chromosomal losses. This discovery was verified on an independent CGH database. A similar tendency was observed by Höglund et al. [9] for a small number of specific solid cancers. The karyotypic evolution models of [9] contained two converging paths, one dominated by gains of chromosomal fragments and the other by losses. To the best of our knowledge, our results provide the first rigorous demonstration of this widespread association within chromosomal aneuploidy in cancer cells.

To avoid ambiguities and reduce potential biases in the results, we excluded from our dataset karyotypes that were not random samples (i.e., reported because of a specific/unusual karyotypic feature), and those with missing information. Inclusion of partially-characterized karyotypes (omitting non-characterized fragments) and karyotypes marked as selected (i.e. non-random samples) increased the number of karyotypes to 42,763 (83% of the Mitelman database). The results on that set closely matched those reported here (Figures S2e in Additional file 3, S3b in Additional file 4), indicating the robustness of both the results and our statistical methods.

Chromosome gains/losses and translocations were the most abundant aberrations in our dataset. While many translocations were shown to contribute to carcinogenesis, the role of chromosomal aneuploidy in cancer has been debated for almost a century. Aneuploidy generally interferes with cellular growth and proliferation, but is frequently associated with the disease of uncontrolled proliferation, cancer. In yeast, aneuploid cells show a transcriptional response similar to that described in yeast cells grown under many different stress conditions [26]. As protein expression levels largely scale with chromosome copy numbers [27], this may reflect the aneuploid cell's effort to reestablish protein stoichiometry [26]. The detrimental role of accumulated proteins in aneuploid cells is supported by a recent report demonstrating that mutations accelerating protein degradation increased the tolerance for anueploidy [28]

These observations may explain the striking chromosome gain/loss dichotomy that we observed and suggest a partial explanation for the following conundrum: A germline or experimentally acquired single chromosome gain/loss is usually detrimental, both at the cellular and the organism levels, while the abundance of chromosome gains/losses in cancer cells implies that aneuploidy is beneficial, or at least not harmful, to their vitality [29-33]. As most chromosomes contain dosage-sensitive genes, the strong gain-gain and loss-loss correlations may imply a mechanism for balancing the ratios of proteins that function in complexes. Such balancing may be required to protect the cancer cell from the detrimental effects of partially assembled protein complexes or free subunits by molecular chaperones caused by prior chromosome gain / loss events.

An alternative explanation for these observations is that chromosomal gains and losses are caused by different mechanisms of genomic instability. This is less likely however, as it implies that defects in mitotic checkpoint result in non random distribution of the aneuploidy chromosomes between two daughter cells. There is no experimental data to support that hypothesis. A third possible explanation is that the correlation of gains with other gains and losses with other losses is driven by catastrophically failed mitoses, where many chromosomes fail to separate during anaphase. In this scenario one daughter cell wound up with many more, and the other with many fewer chromosomes. However, this scenario does not explain why many *specific* chromosome pairs are significantly co-gained / co-lost, even when the statistical test is corrected for chromosome gains-losses dichotomy (results not shown). Additional experimental data are needed to substantiate or refute these hypotheses.

Interestingly, gain-gain correlations are more prevalent and more significant than loss-loss co-occurrence (compare Figure 2 and the website). There may be two explanations why gains of chromosome pairs are more common than losses. The first is simply mathematical: Trisomy means 30% more dosage for a set of genes, while a loss implies a more dramatic 50% drop. The second is experimental – Rancati et al. have shown that the higher the ploidy the better the adaptation to aneuploidy is [34]. Hence gains of multiple chromosomes may be advantageous in the evolution of human cancer karyotypes.

One limitation of the use of the Mitelman database is the very low resolution of the karyotypes, disallowing identification of low-level and focal events. On the other hand, the huge number of karyotypes allowed us to carry out rigorous statistical analysis on a very large scale. Another limitation is its inherent bias towards hematological cancers. However, the number of solid karyotypes in the database is still substantial, and allowed us to obtain results on class similarity among solid cancers (Figure 3). Moreover, the results on aberration co-occurrence tendency were similar using the full data (Figure 2) and the solid karyotypes only (Figure S2c in Additional file 3). Cytogenetic techniques are still widely used in cancer studies, and have some advantages over current high-resolution techniques. Cytogentic methods allow distinguishing between different clones that co-exist in a cancer sample, and are often used in verifying the existence of specific aberrations, We emphasize that in our analysis we analyzed all types of aberrations identifiable by cytogentic techniques, including translocations, iso-chromosomes, partial deletions, and more. Nevertheless, the strongest associations we revealed among aberrations involved mainly whole chromosome gains and losses, most likely since other aberrations (e.g. specific translocations, or deletions) are less common and more difficult to detect in cytogentic techniques.

The methodologies developed in this study can be used on other large datasets describing genetic events. As high resolution genetic information on tumors (e.g. from array-CGH and deep sequencing) accumulates, similar analysis can be applied to it. For example, Beroukhim et al. [12] demonstrated that a large majority of somatic focal copy-number alterations (SCNAs) identified in individual cancer types are present in several cancer types. Our method can be used to assess whether common focal SCNAs tend to be shared by related cancers, as has been the case for cytogentic aberrations. The main challenges in adapting our methods for array-based data are assigning each sample with a set of aberrations (aka "aberration calling"), and handling intersecting aberrations (e.g. two deletions with overlapping segments). Another major difficulty in uncovering strong associations in cancer data is the requirement for a large number of cancer samples. To obtain a large dataset, we performed pooled analysis of heterogeneous cancer samples, similarly to [12, 13]. Pooled analysis has the potential of revealing associations possibly pertinent to common cellular mechanisms shared by different cancer types. Recent examples include: cancer-related genes hosted in highly frequent copy-number alterations in cross-cancer data [12], structural signatures of driver / passenger homozygous deletions [13]; and the whole chromosome gain/loss dichotomy phenomenon reported here.

Finally, our website can be useful both for additional global investigations like those reported here and for in-depth analysis of individual associations.

Conclusions

Cancer is a common name for many different diseases: there is large variability among different cancers, and even among cancers of the same morphological and topographical origin.

Nevertheless, different cancers may share similar mechanisms. Analyzing a heterogeneous set of cancers has the potential of uncovering patterns that are related to such common mechanisms. In this study we performed a large-scale analysis of karyotypes from heterogeneous cancer samples. We show that many aberrations, including some whole chromosome gains and losses, are highly specific to certain cancers. Other aberrations exhibiting weaker specificity were shown to be shared among cancers of related morphology. The investigation of aberrations pairs revealed a striking non-random, cross-cancer pattern of aneuploidy, where whole chromosome gains are associated with other gains and whole chromosome losses are associated with other losses. Despite being very common the role of aneuploidy in cancer initiation or progression is unclear, but one explanation of the non-random pattern of aneuploidy that we have found and quantified is that it is necessary for a clonal growth advantage. We hope that this finding will lead to a better understanding of the mechanisms that allow cancer cells to balance the harms with the potential growth advantage caused by aneuploidy.

Materials and methods

Karyotype selection and analysis. Starting from all 59,759 karyotypes present in the Mitelman database on November 17, 2009, we carried out several aggressive filtering steps aiming to reduce ambiguity and avoid any possible bias. (See Text S1 in Additional file 5 for the full details) Briefly, we evaluated all 34,107 karyotypes marked as unselected (i.e. chosen in a non-biased manner). Karyotypes were parsed using the CyDAS ISCN parser [35], and any karyotype detected as invalid during the parsing was excluded, leaving 29,911 (88%) valid karyotypes. We filtered all karyotypes that are not well-defined. For a multiclonal karyotype, we avoided dependency between its karyotypes by choosing only the first well-defined karyotype it contained. In case of multiple karyotypes from the same patient (“case” in the Mitelman database), only one karyotype was taken into account. To avoid potential biases in chromosome gain/loss aberrations, we excluded any karyotype that was not near-diploid (i.e., we omitted karyotypes whose total chromosome number was less than 35 or more than 57). Altogether, 18,813 karyotypes were selected for analysis.

Aberrations reconstruction. We previously identified 11 frequent chromosomal events in tumor karyotypes (chromosome gain/loss, translocation, deletion, duplication and more, see Table S1 in Additional file 1), and developed an algorithm for reconstructing a most plausible set of these events leading to a given karyotype [36]. Briefly, our algorithm mimics the intuitive way a researcher would perform this task manually: Starting with the cancer karyotypes, the algorithm selects the simplest and most evident step of “undoing” one event at a time, bringing the karyotype closer to the normal one. We applied the algorithm to all relevant karyotypes from the Mitelman database, obtaining unambiguous reconstruction in 99% (18,600) of the karyotypes. We recorded each karyotype’s set of aberrations, where an *aberration* is defined by an event and the chromosomal locations involved. See Text S1 in Additional file 5 for further details.

Karyotypes classification. We classified karyotypes by their tissue morphology and topography as specified in the Mitelman database. To permit robust statistical analysis, we omitted all karyotypes whose class had less than 50 karyotypes. Our final dataset contained 15,445 karyotypes.

CGH data. To validate our results for co-occurrence of chromosome gains and losses, which were obtained using karyotypes data, we searched for an alternative independent dataset, We used the NCBI's SKY/M-FISH and CGH database ([37], version March 16, 2009), consisting of 1084 records. Every record has a list of chromosomal segments with abnormal copy number, each classified as a gain or a loss; and the header of the record contains information on the cancer tissue. As most tumor classes in this dataset were relatively small, we ignored the histological classification. For each record we derived chromosome gain / loss aberrations in the following manner: every gained (lost) chromosomal fragment that spanned the centromere was considered a whole chromosome gain (loss). Gain / loss aberrations that were internal to a chromosome arm (i.e. not spanning the centromere) were ignored.

Computing P-values for aberration-class correlations. For an aberration *Ab* and a tumor class *C*, we calculated the significance of the enrichment of karyotypes with *Ab* in *C* using the hypergeometric test.

Computing P-values for classes sharing common aberrations. We say that an aberration *Ab* is *t*-correlative to a tumor class C if the enrichment of karyptypes with *Ab* in *C* had a hypergeometric P-value ≤ *t*. For a fixed *t*, we developed the following method for evaluating the significance of shared aberrations between tumor classes. We constructed a binary matrix M_t whose rows and columns correspond to aberrations and classes, respectively. We set $M_t[Ab,C]=1$ if *Ab* is *t*-correlative to *C*, and otherwise $M_t[Ab,C]=0$. For $t = 0.05$, the maximal *t* used in our analysis, the matrix M_t was already quite sparse, with less than 2% 1's.

For two classes, *C1* and *C2*, we computed a P-value for their number of shared events as follows. Let $n_{t.C1,C2}$ be the number of *t*-correlative aberrations that *C1* and *C2* shared. More formally, $n_{t.C1,C2} = \Sigma_{Ab}\ M_t[Ab,C1]\cdot M_t[Ab,C2]$. For every pair of classes, *C1* and *C2*, we estimated the probability of having at least $n_{t,C1,C2}$ *t*-correlative aberrations by chance by sampling N=10^7 randomized permutations of M_t that preserve row and column sums. Every such permutation corresponds to an assignment of aberrations to tumor classes that keeps the general properties of the original data: Aberrations that occur in few (or many) cancer classes remain so, and tumor classes that had many (or few) correlative aberrations preserve this property. The randomization is done by a long sequence of edge swaps [38]. The P-value for *C1* and *C2* is defined as *r(C1,C2,N,t) / N* where *r(C1,C2,N,t)* is the total number of M_t permutations in which the number of aberrations that *C1* and *C2* shared was ≥ $n_{t,C1,C2}$. In case *r(C1,C2,N,t)* = 0, we defined the P-value to be 1/*N*. Therefore, the minimal P-value we could achieve was 10^{-7}.

Hierarchical clustering of classes. We performed average-linkage hierarchical clustering of the classes using the Expander software package [39]. The similarity measure between classes was defined as follows. We first built a symmetric matrix, *S*, satisfying $S[C_1,C_2] = -log(p)$, where p is the P-value described above for the significance of the number of t-correlative aberrations that C_1 and C_2 share, $n_{t.C1,C2}$. For each class *C*, we set $S[C,C]=log(N)$, where N=10^7 as above. The similarity between classes was now defined as the Pearson correlation between their rows of S. Classes showing no significant similarity at false discovery rate (FDR) 5% to any other class were removed from this analysis.

Computing P-values for co-occurring aberration pairs. For two aberrations, *Ab1* and *Ab2*, let *n*(*Ab1*, *Ab2*) be the total number of karyotypes that contain both aberrations. We estimated the significance of *n*(*Ab1*, *Ab2*) for all pairs of distinct aberrations using a permutation test (similar to the one described above) as follows. We constructed a binary matrix, *M*, whose rows correspond to aberrations that occur in at least 10 karyotypes, and columns to karyotypes. For an aberration *Ab* and karyotype *K*, we set *M*[*Ab,K*]=1 if *K* contained *Ab*, and *M*[*Ab,K*]=0 otherwise. We randomly sampled permutations of M that preserved row and column sums. Thus each permutation corresponds to a random set of karyotypes with the same distributions of (i) number of aberrations per karyotype, and (ii) number of karyotypes per aberration. Moreover, to account for the different distributions of aberrations within each tumor class, the sampled permutations were also required to preserve (sub-)row sum corresponding for each class. We enhanced the performance of this test by filtering aberration pairs whose hypergeometric test P-value was above 0.001, and removing from M any aberration that did not appear in the remaining pairs.

We performed a similar test for the CGH dataset, but since it was smaller in size we used all aberrations (i.e. irrespective of the number of samples in which they were found), and without the step of filtering pairs by the hypergeometric test.

Trisomy index test. To demonstrate the utility of our website, we used it to define *a trisomy index test* (TI-T), a test that uses specific trisomies (i.e. chromsome gains) in order to distinguish between prognostically different subgroups of a certain disease. Similar to Chng et al. [25], we focused on hyperdiploid multiple myeloma (H-MM), a subtype of MM. For a given TI-T, the sensitivity (respectively, specificity) was calculated as the percentage of H-MM (respectively, non-H-MM) karyotypes that are correctly identified as such by the TI-T. The positive predictive value (PPV) was calculated as the percentage of H-MM karyotypes among all karyotypes identified as H-MM by TI-T. The F-score was calculated as the harmonic mean of sensitivity and PPV: F = 2×PPV×sensitivity/(PPV+sensitivity).

Abbreviations

CGH, comparative genomic hybridization; CNV, copy number variation; FDR, false discovery rate; MM , multiple myeloma.

Authors' contributions

RS and MOF designed research. MOF performed research and built the website. CL and MOF developed the statistical scores. MOF, RS, SI and LT analyzed and interpreted the data. MOF, RS and SI drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Gideon Rechavi, Avi Orr-Urtreger, and Uta Francke for helpful discussions, and Annelyse Thevenin and Tamar Lavee for helpful comments on the manuscript. We are grateful to Lior Mechlovich for programming an early version of the analysis code and to Igor Ulitsky for help with the hierarchical clustering code. RS was supported in part by the Raymond and Beverly Sackler Chair in Bioinformatics, the Israel Science Foundation (Grant 802/08) and IBM Research Open Collaborative Research grant. SI was supported by the Israel Science Foundation (Legacy program).

References

1. Bayani J, Selvarajah S, Maire G, Vukovic B, Al-Romaih K, Zielenska M, Squire JA: **Genomic mechanisms and measurement of structural and numerical instability in cancer cells.** *Semin Cancer Biol* 2007, **17:**5-18.
2. Haber DA, Settleman J: **Cancer: drivers and passengers.** *Nature* 2007, **446:**145-146.
3. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL: **Genetic alterations during colorectal-tumor development.** *N Engl J Med* 1988, **319:**525-532.
4. Fearon ER, Vogelstein B: **A genetic model for colorectal tumorigenesis.** *Cell* 1990, **61:**759-767.
5. Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA: **Inferring tree models for oncogenesis from comparative genome hybridization data.** *J Comput Biol* 1999, **6:**37-51.
6. Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA: **Distance-based reconstruction of tree models for oncogenesis.** *J Comput Biol* 2000, **7:**789-803.
7. von Heydebreck A, Gunawan B, Fuzesi L: **Maximum likelihood estimation of oncogenetic tree models.** *Biostatistics* 2004, **5:**545-556.
8. Radmacher MD, Simon R, Desper R, Taetle R, Schaffer AA, Nelson MA: **Graph models of oncogenesis with an application to melanoma.** *J Theor Biol* 2001, **212:**535-548.
9. Hoglund M, Frigyesi A, Sall T, Gisselsson D, Mitelman F: **Statistical behavior of complex cancer karyotypes.** *Genes Chromosomes Cancer* 2005, **42:**327-341.
10. Hjelm M, Hoglund M, Lagergren J: **New probabilistic network models and algorithms for oncogenesis.** *J Comput Biol* 2006, **13:**853-865.
11. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, Du J, Kau T, Thomas RK, Shah K, Soto H, Perner S, Prensner J, Debiasi RM, Demichelis F, Hatton C, Rubin MA, Garraway LA, Nelson SF, Liau L, Mischel PS, Cloughesy TF, Meyerson M, Golub TA, Lander ES, Mellinghoff IK, et al: **Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.** *Proc Natl Acad Sci U S A* 2007, **104:**20007-20012.
12. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, et al: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463:**899-905.
13. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, Widaa S, Hinton J, Fahey C, Fu B, Swamy S, Dalgliesh GL, Teh BT, Deloukas P, Yang F, Campbell PJ, Futreal PA, Stratton MR: **Signatures of mutation and selection in the cancer genome.** *Nature* 2010, **463:**893-898.
14. **Mitelman Database of Chromosome Aberrations in Cancer** [http://cgap.nci.nih.gov/Chromosomes/Mitelman]

15. Mitelman F, Johansson B, Mertens F: **The impact of translocations and gene fusions on cancer causation.** *Nat Rev Cancer* 2007, **7:**233-245.

16. **Statistical Associations in Cancer Karyotypes (STACK)** [http://acgt.cs.tau.ac.il/stack/]

17. Mitelman F, Mertens F, Johansson B: **Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders.** *Genes Chromosomes Cancer* 2005, **43:**350-366.

18. Shaffer L, Tommerup N: *ISCN 2005: an international system for human cytogenetic nomenclature (2005): recommendations of the International Standing Committee on Human Cytogenetic Nomenclature.* S Karger Pub; 2005.

19. Nowell P, Hungerford D: **A minute chromosome in chronic granulocytic leukemia.** *Science* 1960, **132:**1497.

20. Turc-Carel C, Aurias A, Mugneret F, Lizard S, Sidaner I, Volk C, Thiery JP, Olschwang S, Philip I, Berger MP, Philip T, Lenoir GM, Mazabraud A: **Chromosomes in Ewing's sarcoma. I. An evaluation of 85 cases and remarkable consistency of t(11;22)(q24;q12).** *Cancer genetics and cytogenetics* 1988, **32:**229-238.

21. Ghose T, Lee CLY, Fernandez LA, Lee SHS, Raman R, Colp P: **Role of 1q Trisomy in Tumorigenicity, Growth, and Metastasis of Human Leukemic B-Cell Clones in Nude Mice.** *Cancer Res* 1990, **50:**3737-3742.

22. Smadja NV, Fruchart C, Isnard F, Louvet C, Dutel JL, Cheron N, Grange MJ, Monconduit M, Bastard C: **Chromosomal analysis in multiple myeloma: cytogenetic evidence of two different diseases.** *Leukemia* 1998, **12:**960-969.

23. Smadja NV, Bastard C, Brigaudeau C, Leroux D, Fruchart C, Hematol GFC: **Hypodiploidy is a major prognostic factor in multiple myeloma.** *Blood* 2001, **98:**2229-2238.

24. Fonseca R, Barlogie B, Bataille R, Bastard C, Bergsagel PL, Chesi M, Davies FE, Drach J, Greipp PR, Kirsch IR, Kuehl WM, Hernandez JM, Minvielle S, Pilarski LM, Shaughnessy JD, Jr., Stewart AK, Avet-Loiseau H: **Genetics and cytogenetics of multiple myeloma: a workshop report.** *Cancer Res* 2004, **64:**1546-1558.

25. Chng WJ, Van Wier SA, Ahmann GJ, Winkler JM, Jalal SM, Bergsagel PL, Chesi M, Trendle MC, Oken MM, Blood E, Henderson K, Santana-Davila R, Kyle RA, Gertz MA, Lacy MQ, Dispenzieri A, Greipp PR, Fonseca R: **A validated FISH trisomy index demonstrates the hyperdiploid and nonhyperdiploid dichotomy in MGUS.** *Blood* 2005, **106:**2156-2161.

26. Torres EM, Williams BR, Amon A: **Aneuploidy: cells losing their balance.** *Genetics* 2008, **179:**737-746.

27. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R: **Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast.** *Nature* 2010, **468:**321-325.

28. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, Dunham MJ, Amon A: **Identification of aneuploidy-tolerating mutations.** *Cell* 2010, **143:**71-83.

29. Ganmore I, Smooha G, Izraeli S: **Constitutional aneuploidy and cancer predisposition.** *Hum Mol Genet* 2009, **18:**R84-93.

30. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, Amon A: **Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells.** *Science* 2008, **322:**703-709.

31. Weaver BA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW: **Aneuploidy acts both oncogenically and as a tumor suppressor.** *Cancer Cell* 2007, **11:**25-36.

32. Roper RJ, Reeves RH: **Understanding the basis for Down syndrome phenotypes.** *PLoS Genet* 2006, **2:**e50.

33. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A: **Effects of aneuploidy on cellular physiology and cell division in haploid yeast.** *Science* 2007, **317:**916-924.

34. Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, Perera A, Staehling-Hampton K, Seidel CW, Li R: **Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor.** *Cell* 2008, **135:**879-893.

35. Hiller B, Bradtke J, Balz H, Rieder H: **CyDAS: a cytogenetic data analysis system.** *Bioinformatics* 2005, **21:**1282-1283.

36. Ozery-Flato M, Shamir R: **On the frequency of genome rearrangement events in cancer karyotypes.** *Technical Report.* Tel Aviv University; 2007. [http://acgt.cs.tau.ac.il/papers/cancerGR_11b_report-1.pdf]

37. **NCI and NCBI's SKY/M-FISH and CGH Database** [http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi]

38. Sharan R, Ideker T, Kelley B, Shamir R, Karp RM: **Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data.** *J Comput Biol* 2005, **12:**835-846.

39. Shamir R, Maron-Katz A, Tanay A, Linhart C, Steinfeld I, Sharan R, Shiloh Y, Elkon R: **EXPANDER--an integrative program suite for microarray data analysis.** *BMC Bioinformatics* 2005, **6:**232.

40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13:**2498-2504.

TABLES

Table 1: Tumor classes and categories in the dataset

Class	Number of Karyotypes
Benign solid tumors	**1567**
Ad-Large intestine	100
Ad-Salivary gland	191
Ad-Thyroid	66
Benign-Breast	69
Ch hamartoma-Lung	99
Leiomyoma-Uterus	214
Lipoma-ST	269
Mnng-Brain	508
Oncocytoma-Kidney	51
Non-lymphoid hematological disorders	**6913**
AML	1026
AML M0	144
AML M1	315
AML M2	776
AML M3	525
AML M4	621
AML M5	266
AML M5a	52
AML M6	133
AML M7	168
BBL	137
CMD	69
CML at	409
CML t(9;22)	808
CMML	147
Id myelofibrosis	115
JML	50
MDS	187
Polycythemia Vera	166
Rf anemia	374
Rf anemia EB	344
Rf anemia RS	81
Lymphoid disorders	**4411**
ALL	1817
Adult T-Cell lymphoma	64
Ang T-Cell lymphoma	71
Burkitt lymphoma	248
CLL	884
DL B-Cell lymphoma	197
Follicular lymphoma	274
HCL	57
M B-Cell neoplasm	166
MCL	78
Multiple myeloma	385
Per T-Cell lymphoma	62
SMZ B-Cell lymphoma	108
Malignant solid tumors	**2554**
AdC-Breast	323
AdC-Kidney	610
AdC-Large intestine	125
AdC-Ovary	56
AdC-Prostate	124
AdC-Thyroid	84
AdC-Uterus	62
Astrocytoma-Brain [a]	234
BCC-Skin	87
Ewing-Skeleton	181
Giant cell-Skeleton	60
Hpblastoma-Liver	65
Liposarcoma M-ST	59
Melanoma-Eye	72
SqCC-Larynx	58
SqCC-Lung	64
Synovial sarcoma-ST	58
Wilms-Kidney	232

Ad: Adenoma; Adc: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; AML: Acute myeloid leukemia; Ang: Angioimmunoblastic; BBL: Bilineage or biphenotypic leukemia; BCC: Basal cell carcinoma; Ch hamartoma: Chondroid hamartoma; CLL: Chronic lymphocytic leukemia; CMD: Chronic myeloproliferative disorder; CML: Chronic myeloid leukemia; CML at: CML aberrant translocation; CMML: Chronic myelomonocytic leukemia; DL: Diffuse large; HCL: Hairy cell leukemia; JML: Juvenile myelomonocytic leukemia; Liposarcoma M: liposarcoma myxoid/round cell; M B-Cell: mature B-Cell; MCL: mantle cell lymphoma; MDS: myelodysplastic syndrome; Mnng: meningioma; Rf anemia: refractory anemia; Rf anemia EB: refractory anemia with excess of blasts; Rf anemia RS: refractory anemia with ringed sideroblasts; SMZ: splenic marginal zone; SqCC: squamous cell carcinoma; ST: soft tissue.

[a] Astrocytoma grade III-IV

Figures

Figure 1: Overview of karyotypes analysis and the STACK website. A large fraction of the karyotypes in the Mitelman database was removed to avoid potential bias in the analysis. These included partially characterized karyotypes, multiple karyotypes from the same individual, and karyotypes that were not randomly selected in the original report. Tumor type and location were used to classify karyotypes into tumor classes, and classes with small representation (< 50 karyotypes) were removed from the dataset. An algorithm was used to reconstruct the set of aberrations leading to each remaining karyotype. Three types of statistical correlations were computed: aberration co-occurrence, association between class and aberration, and class similarity (based on their common aberrations). All computed correlations, with their P-values, are available for further investigation via our website and are directly linked to the full description of the relevant karyotypes in the Mitelman database. Repeating the analysis without filtering ambiguities and selected karyotypes (yielding 42,763 karyotypes, 83% of the Mitelman database) led to essentially the same conclusions.

Figure 2: Highly co-occurring aberration pairs. Highly co-occurring aberrations in the entire karyotype dataset are connected by lines. Aberrations that are involved only in expected links (e.g. a link between a translocation and a gain /loss of one of its derivative chromosomes; a link between two (two-break) translocations originating from one three-break [18] rearrangement) are not shown. For explanation on aberration names, see Table S1 in Additional file 1. (a) Highly co-occurring pairs in the Mitelman Database karyotypes (links are significant at $P<0.05$, after Bonferroni correction). (b) Highly co-occurring pairs in the CGH dataset (links are significant at FDR 5%). The only gain-loss link is (+1, -16), which has the second worst (i.e. highest) P-value among the 47 pairs that passed the FDR 5% criterion. The figure was drawn using Cytoscape [40].

Figure 3: Hierarchical clustering of classes based on class similarity in sharing common aberrations. The square at the intersection of each two diagonals shows the similarity of their classes, as measured by the aberrations associated with them (Materials and Methods). (An aberration was associated with a tumor class if their correlation had (uncorrected) P-value < 0.05.) Names of cancer classes are colored as follows: orange: lymphoid disorders; red: non-lymphoid hematological disorders; light green: benign solid tumors; dark green: malignant solid tumors. Classes that showed no significant similarity to any other class at FDR 5% were not included in the clustering.

Additional files

Additional file 1:

Table S1

Chromosomal events allowed in the reconstruction algorithm

Additional file 2:

Figure S1

Event frequencies.

Additional file 3:

Figure S2

Highly co-occurring aberration pairs. Highly co-occurring aberrations ($P<0.05$ after Bonferroni correction) are connected by lines. Aberrations that are involved only in expected links are not shown. See Supplemental Table 1 for aberration name abbreviations. (a) Lymphoid disorders, (b) non-lymphoid hematological disorders, (c) solid tumors, (d) carcinomas, (e) all karyotypes, results were obtained on a dataset that includes partially defined and selected karyotypes (83% of the Mitelman database). Legend is as in Fig. 2 for (a)-(d), and for (e) with the addition of light red and light green colors corresponding to partial deletions and partial duplications respectively.

Additional file 4:

Figure S3

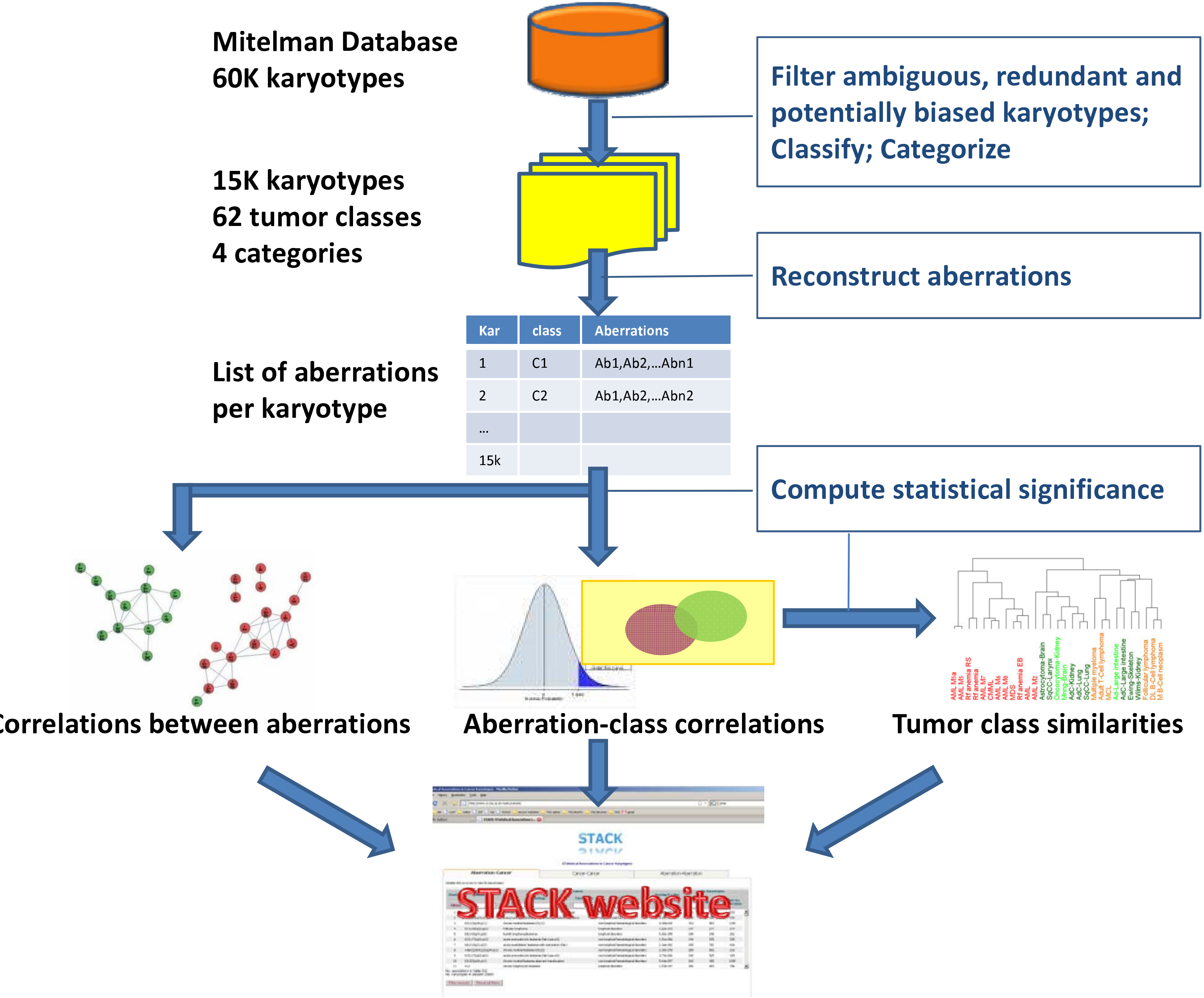
Tumor classes with similar common aberrations. (a) Tumor class pairs with significantly high number of common aberrations are connected by lines (FDR 5%). Aberrations assigned to tumor classes are: (a1) significantly correlated at FDR 5%, (a2) correlated with P-value < 0.05 (uncorrected). (b) Hierarchical clustering of classes based on class similarity in sharing common aberrations. Results were obtained on a dataset that includes partially defined and selected karyotypes (83% of the Mitelman database). Legend is as in Fig. 3.

Additional file 5:

Text S1

Description of the algorithm for reconstructing aberrations from karyotypes.

Figure 1



Mitelman Database
60K karyotypes

Filter ambiguous, redundant and potentially biased karyotypes; Classify; Categorize

15K karyotypes
62 tumor classes
4 categories

Reconstruct aberrations

List of aberrations per karyotype

Kar	class	Aberrations
1	C1	Ab1,Ab2,…Abn1
2	C2	Ab1,Ab2,…Abn2
…		
15k		

Compute statistical significance

Correlations between aberrations

Aberration-class correlations

Tumor class similarities

STACK website

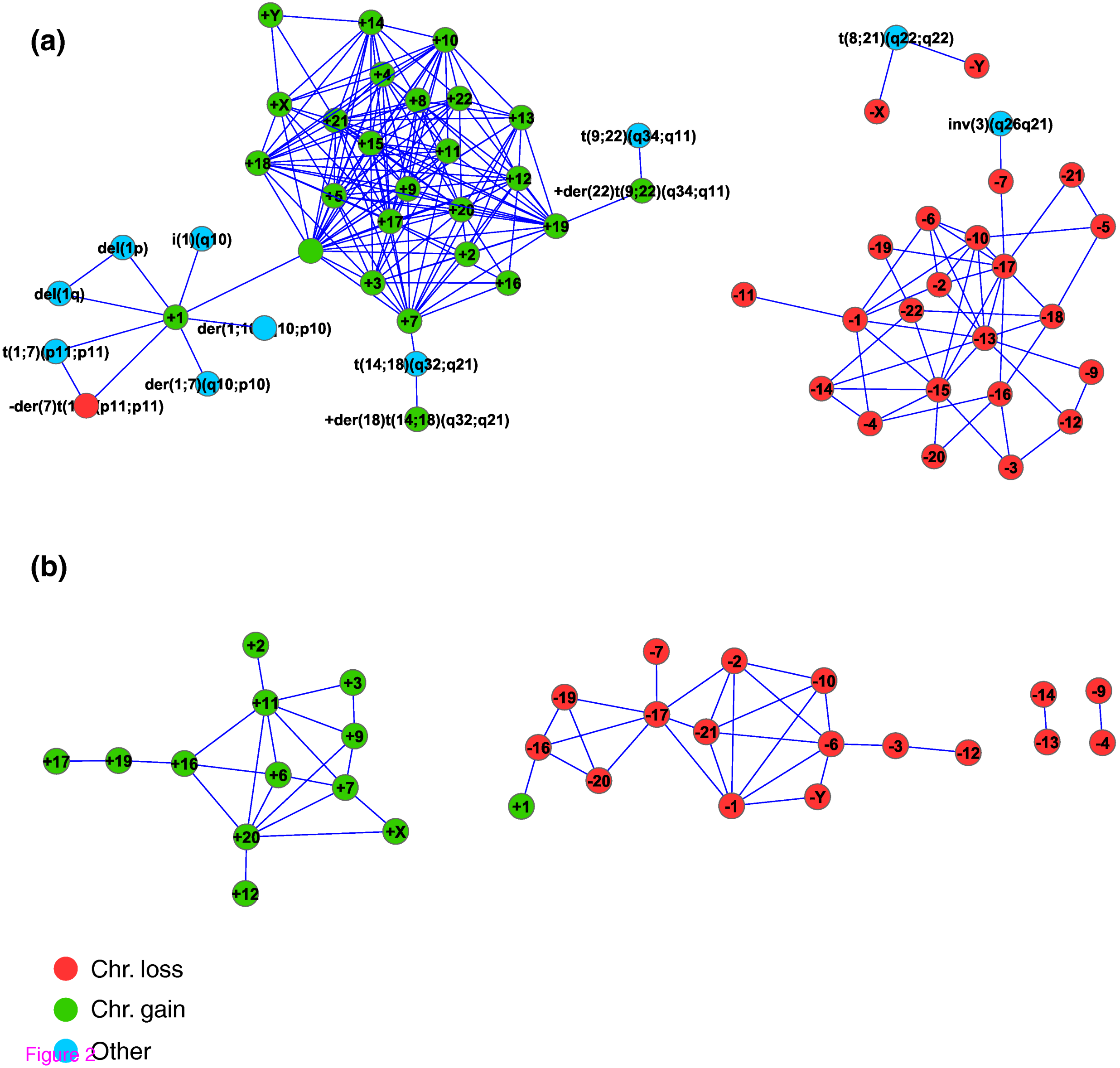
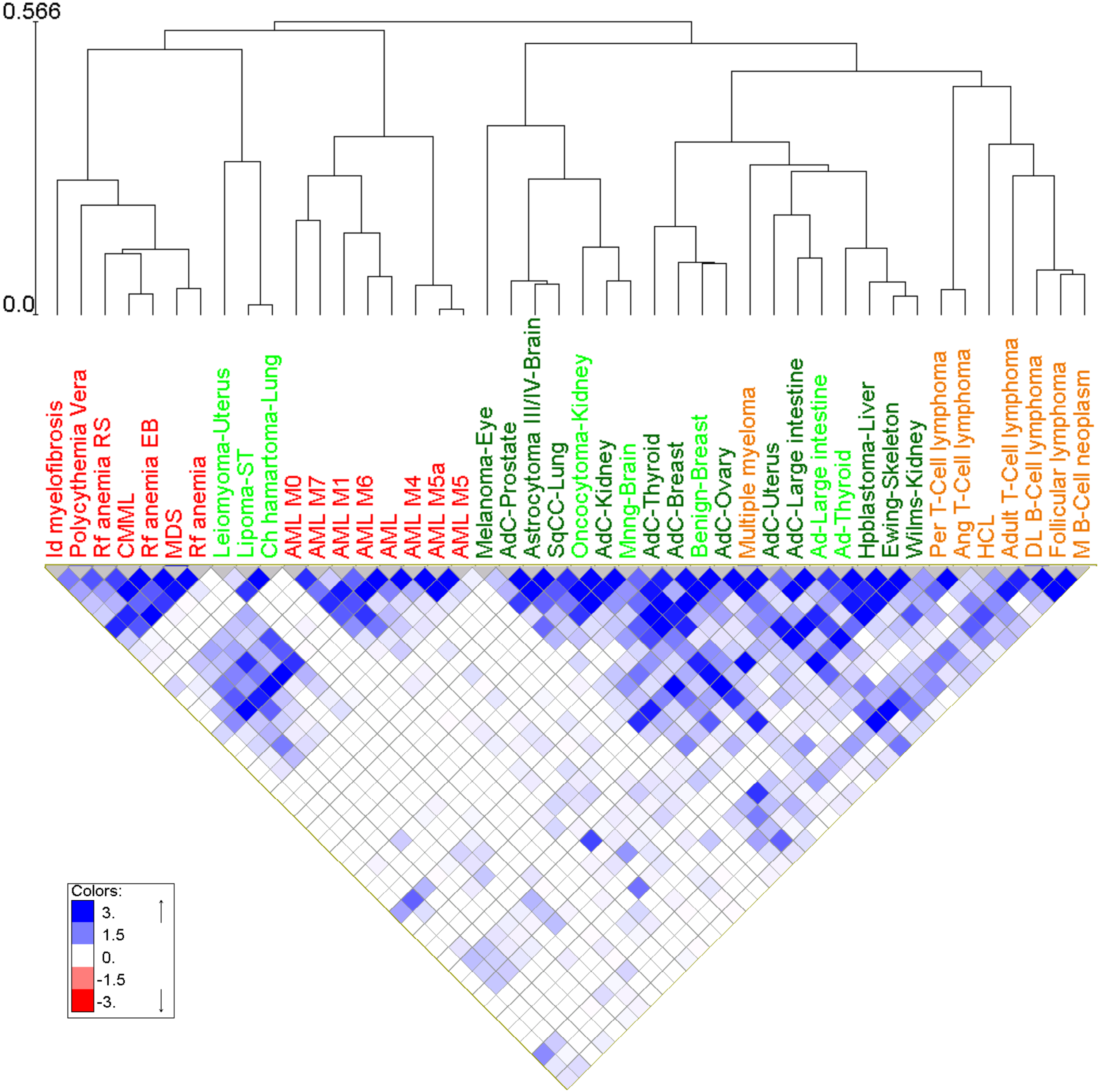

(a)
+Y
+14
+10
+4
+X
+8
+22
+21
+13
+15
+11
+18
+12
+5
+9
+20
+19
+17
+2
+3
+16
+7
+1
t(9;22)(q34;q11)
+der(22)t(9;22)(q34;q11)
del(1p)
i(1)(q10)
del(1q)
der(1;1 10;p10)
t(1;7)(p11;p11)
der(1;7)(q10;p10)
-der(7)t(1 (p11;p11)
t(14;18)(q32;q21)
+der(18)t(14;18)(q32;q21)
t(8;21)(q22;q22)
-Y
-X
inv(3)(q26q21)
-7
-21
-6
-10
-5
-19
-17
-2
-11
-22
-18
-1
-13
-9
-14
-15
-16
-4
-12
-20
-3
(b)
+2
+11
+3
+9
+17
+19
+16
+6
+7
+X
+20
+12
-7
-2
-10
-19
-17
-21
-16
-6
-3
-12
-20
+1
-1
-Y
-14
-9
-13
-4
● Chr. loss
● Chr. gain
● Other

Figure 2

Figure 3


0.566
0.0
Id myelofibrosis
Polycythemia Vera
Rf anemia RS
CMML
Rf anemia EB
MDS
Rf anemia
Leiomyoma-Uterus
Lipoma-ST
Ch hamartoma-Lung
AML M0
AML M7
AML M1
AML M6
AML
AML M4
AML M5a
AML M5
Melanoma-Eye
AdC-Prostate
Astrocytoma III/IV-Brain
SqCC-Lung
Oncocytoma-Kidney
AdC-Kidney
Mnng-Brain
AdC-Thyroid
AdC-Breast
Benign-Breast
AdC-Ovary
Multiple myeloma
AdC-Uterus
AdC-Large intestine
Ad-Large intestine
Ad-Thyroid
Hpblastoma-Liver
Ewing-Skeleton
Wilms-Kidney
Per T-Cell lymphoma
Ang T-Cell lymphoma
HCL
Adult T-Cell lymphoma
DL B-Cell lymphoma
Follicular lymphoma
M B-Cell neoplasm
Colors:
3.
1.5
0.
-1.5
-3.

Additional files provided with this submission:

Additional file 1: sup1.doc, 37K
http://genomebiology.com/imedia/9011729975653037/supp1.doc
Additional file 2: sup2.tiff, 35K
http://genomebiology.com/imedia/1499063581565303/supp2.tiff
Additional file 3: sup3.jpeg, 487K
http://genomebiology.com/imedia/4709967455653031/supp3.jpeg
Additional file 4: sup4.jpeg, 427K
http://genomebiology.com/imedia/1386284447565303/supp4.jpeg
Additional file 5: sup5.doc, 67K
http://genomebiology.com/imedia/2113527081565304/supp5.doc