FuseFISH: Robust Detection of Transcribed Gene Fusions in Single Cells

Citation: Semrau, Stefan, Nicola Crosetto, Magda Bienko, Marina Boni, Paolo Bernasconi, Roberto Chiarle, and Alexander van Oudenaarden. "FuseFISH: Robust Detection of Transcribed Gene Fusions in Single Cells." Cell Reports 6, no. 1 (January 2014): 18–23.

As Published: http://dx.doi.org/10.1016/j.celrep.2013.12.002

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/101634

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Cell Reports
Report

FuseFISH: Robust Detection of Transcribed Gene Fusions in Single Cells

Stefan Semrau,[1,2,3,9] Nicola Crosetto,[1,2,3,9,11] Magda Bienko,[1,2,3,10,11] Marina Boni,[4,10] Paolo Bernasconi,[4] Roberto Chiarle,[5,6,7,*] and Alexander van Oudenaarden[1,2,3,8,*]
[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Division of Hematology, Fondazione IRCCS Policlinico San Matteo, University of Pavia, 27100 Pavia, Italy
[5]Department of Pathology, Children's Hospital, Boston, MA 02138, USA
[6]Harvard Medical School, Boston, MA 02138, USA
[7]Department of Molecular Biotechnology and Health Sciences, University of Torino, 10126 Torino, Italy
[8]Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands
[9]These authors contributed equally to this work
[10]These authors contributed equally to this work
[11]Present address: Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands
*Correspondence: roberto.chiarle@childrens.harvard.edu (R.C.), a.vanoudenaarden@hubrecht.eu (A.v.O.)
http://dx.doi.org/10.1016/j.celrep.2013.12.002

SUMMARY

Transcribed gene fusions are key biomarkers in many hematologic and solid tumors, often representing the primary oncogenic driver mutation. Here, we report an experimental and computational pipeline for detecting fusion transcripts using single-molecule RNA FISH and unbiased correlation analysis (FuseFISH). We constructed a genome-wide database of optimal oligonucleotide sequences, enabling quick design of FuseFISH probes against known and novel fusions. We implemented FuseFISH in cell lines, tissue sections, and purified RNA, reliably detecting one *BCR-ABL1* positive in 10,000 negative cells. In 34 hematologic samples, we detected *BCR-ABL1* transcripts with high specificity and sensitivity. Finally, we measured *BCR-ABL1* expression heterogeneity and dynamics in single CML cells exposed to the kinase inhibitor Nilotinib. Our resource and methods are ideal for streamlined validation of fusions newly identified by next-generation sequencing, and they pave the way to studying the impact of fusion expression variability on clinical outcome.

INTRODUCTION

Cytogenetic abnormalities such as translocations, inversions, and insertions are characteristic attributes of cancer cells, and they often result in the formation of chimeric genes consisting of segments of two different genes fused together (Fröhling and Döhner, 2008). In most cases, the chimeric gene is transcribed into a fusion transcript encoding parts of a tyrosine kinase or a transcription factor, which become deregulated as a consequence of the fusion (Fröhling and Döhner, 2008). Often a gene fusion represents the primary oncogenic driver mutation in a tumor and hence an ideal pharmacologic target, as demonstrated by the archetypical case of BCR-ABL1 and its selective inhibitor, Imatinib, in chronic myeloid leukemia (CML) (Melo and Barnes, 2007; Ren, 2005; Schiffer, 2007). Therefore, detecting and monitoring in time the expression levels of specific gene fusions in cancer has become common practice in molecular pathology. While recurrent fusions have long been known in hematologic tumors and sarcomas (Mitelman et al., 2007; Rowley, 2009), recent progress in next-generation sequencing technologies has fueled the discovery of new fusions in solid tumors (Maher et al., 2009; Rabbitts, 2009), as exemplified by EML4-ALK in a subset of non-small cell lung cancers (NSCLCs; Soda et al., 2007). Thus, demand for simple and quantitative assays to detect a broad spectrum of fusions will likely emerge in the future.

Detection of fusions is typically achieved at the DNA level by karyotype analysis and DNA fluorescence in situ hybridization (FISH), or at the RNA level by RT-PCR. In spite of the effectiveness and broad use of these techniques, several limitations call for new complementary methods. For instance, even though considerable progress in image processing automation has been done (Alpár et al., 2008; Lerner et al., 2001; Shirley et al., 2011), data analysis of DNA FISH remains difficult to standardize and automate because colocalization of dual-fusion probes or splitting of break-apart probes is usually assessed in a subjective manner. Importantly, DNA FISH is unable to provide information about expression levels of the fusions, which is clinically relevant information (Baccarani et al., 2009). On the other hand, RT-PCR

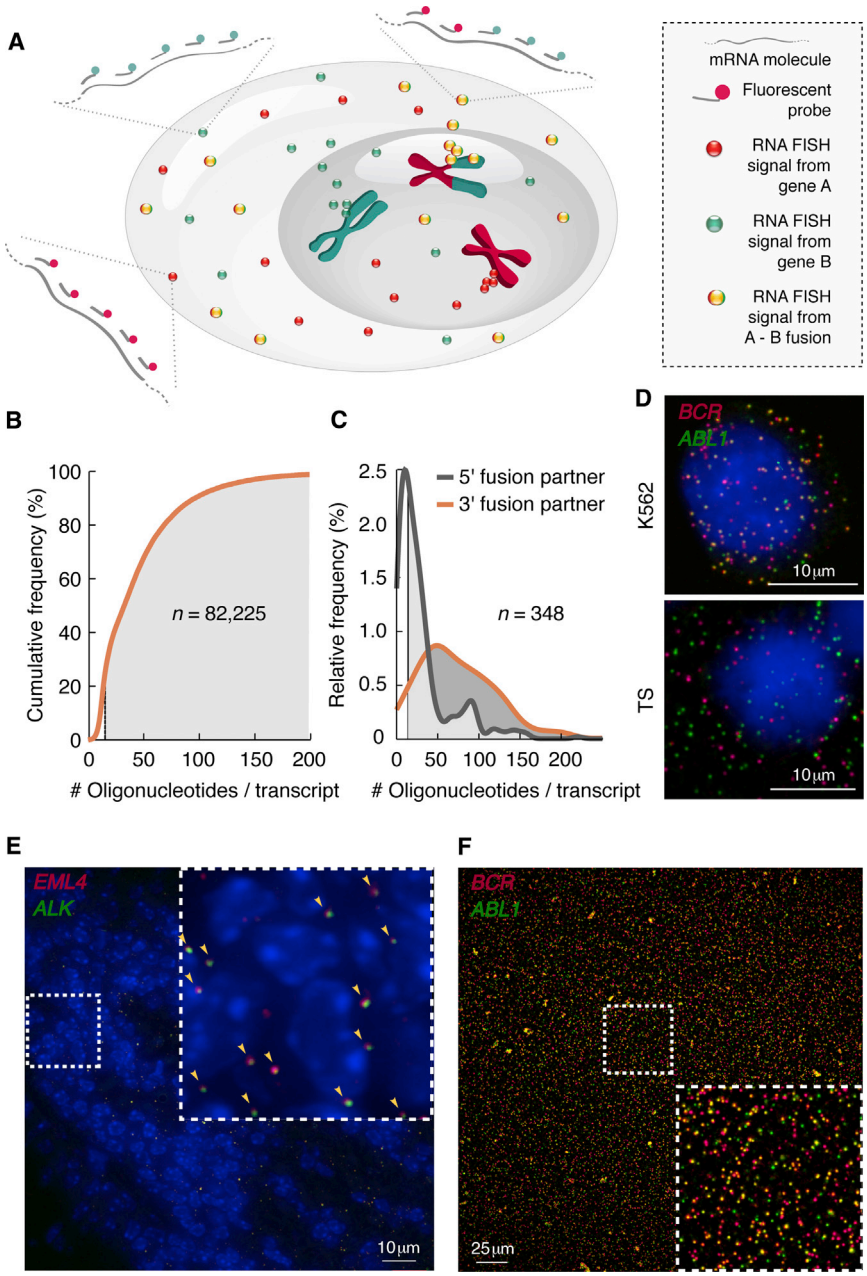

Figure 1. FuseFISH Method

(A) A gene fusion originating from a translocation (magenta-green chromosome) produces fusion transcripts (yellow dots) that are detected as colocalized fluorescent spots by probe sets consisting of differently labeled oligos targeting each fusion partner (zoom-in).

(B) Cumulative frequency of smFISH oligos per coding transcript in the human transcriptome. A total of 79% of transcripts contain ≥15 oligos (gray area). *n*, transcript number.

(C) Relative frequency of smFISH oligos for either the 5′ (gray curve) or 3′ (orange curve) moiety of the shortest variant of 348 unique fusions cataloged in Table S1. Gray area: fraction of transcripts with ≥15 oligos coverage (57% for 5′ and 92% for 3′ partners). *n*, number of fusions analyzed.

(D) *BCR-ABL1* transcripts (yellow spots) in positive (K562) versus negative (TS) cells.

(E) *EML4-ALK* transcripts (yellow spots) in mouse xenografts of lung adenocarcinoma H3122 cells. Yellow arrowheads in the inset indicate examples of fusions.

(F) *BCR-ABL1* fusion transcripts (yellow spots) in spotted purified RNA extracted from K562 cells. See also Figure S1.

is a powerful method to quantify expression, but the development of standardized and reproducible assays for absolute quantification of fusion transcripts can be challenging, especially in formalin-fixed, paraffin-embedded (FFPE) tissue sections. In addition, fusion transcripts often involve different exons in independent clinical samples, thus requiring multiple PCR reactions and controls for their detection. Another limitation relates to the use of RT-PCR in single cells to monitor intratumor expression heterogeneity, which seems clinically informative (La Thangue and Kerr, 2011; Marusyk et al., 2012). Though technically feasible, routine clinical application of single-cell RT-PCR in the clinical context remains challenging, especially in solid tumors, and is associated with high costs for a relatively moderate throughput.

Here, we sought to develop a robust and unbiased experimental and computational framework for detecting specific fusion transcripts in situ or using purified RNA. We demonstrate the feasibility and simplicity of our approach for a variety of fusion transcripts in cell lines, tumor sections, and hematologic specimens. Our resource and methods can be readily applied to biological studies of gene fusions and integrated into clinical cytogenetics.

RESULTS

Method and Probe Resource

In order to detect fusion transcripts at single-molecule resolution, we capitalized on a method for single-molecule RNA FISH (smFISH) previously developed by our group (Raj et al., 2008) based on earlier work (Femino et al., 1998) and on a recent method to detect different mRNA isoforms (Waks et al., 2011). We devised an approach by which each fusion partner is labeled with a set of oligonucleotides coupled to a specific fluorophore, so that the resulting fusion can be detected as two spectrally distinguishable, colocalized diffraction-limited spots (Figure 1A; see Experimental Procedures). We named this approach FuseFISH.

To facilitate smFISH probe design, we compiled a genome-wide list of oligonucleotides with optimal thermodynamic properties targeting 82,225 annotated human protein-coding

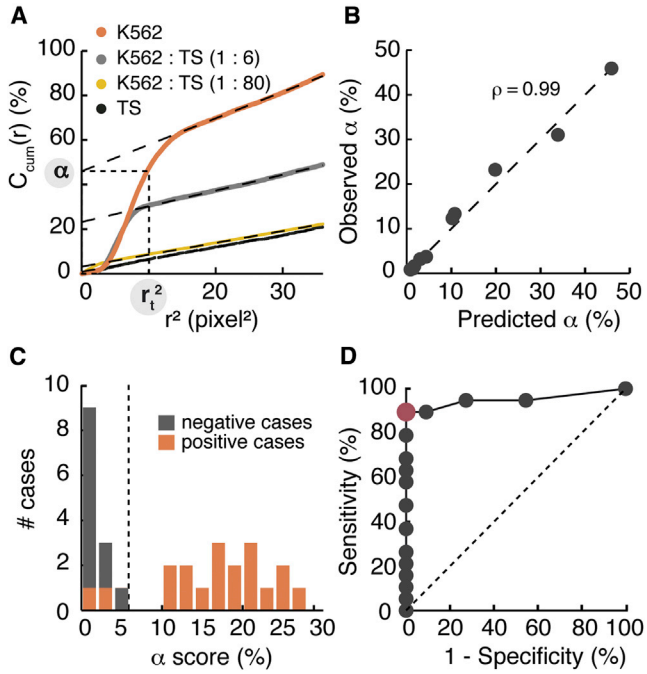

Figure 2. FuseFISH Validation

(A) PICCS correlation analysis of *BCR-ABL1* in K562 cells serially diluted with TS cells. Fitting of the linear portion of each curve yields the correlation fraction α as the offset of the linear fit. The optimal distance r_t threshold balances false positives and negatives such that correlating signals closer than r_t results in the correlation fraction α.

(B) Measured versus predicted scaling of α in the same cell dilutions as in (A).

(C) Distribution of α in hematologic samples. Optimal cut-off is shown as dashed line.

(D) ROC curve of FuseFISH versus DNA FISH and/or RT-PCR for detection of *BCR-ABL1* in hematologic samples. Optimality is achieved for $\alpha \geq 5\%$ (red). Dashed line, line of no discrimination.

See also Figure S2.

transcripts (see Experimental Procedures). For 79% of transcripts, at least 15 oligos can be synthesized, which is sufficient to produce a reliable smFISH signal (Figure 1B). This oligo library is a powerful resource for versatile and rapid design of smFISH probes and is readily accessible at our website (http://www.fusefish.eu).

To estimate coverage for transcript fusions, we separately calculated the number of oligos available for the 5′ and 3′ fusion partner of 348 fused transcripts for which we managed to retrieve the exact breakpoint coordinates by comprehensive literature and database screening (Table S1; see Supplemental Experimental Procedures). To calculate the minimum number of oligos available, we selected the shortest variant when multiple fusion variants were annotated. On average, 30 and 76 oligos are available for 5′ and 3′ fusion partners, respectively, the difference reflecting the fact that 3′ UTRs tend to be longer than 5′ UTRs in most transcripts (Figure 1C). We note that this coverage represents a very stringent estimate, in the sense that it is based on the shortest fusion variants described in the literature, even though their frequency is low in most of cases.

FuseFISH Implementation

As a proof of principle, we constructed probes targeting the most frequent variants of *BCR-ABL1*, *NPM1-ALK*, and *EML4-ALK* fusions in CML, anaplastic large cell lymphoma, and NSCLC, respectively (Figure S1A; Table S2). In positive cell lines, we observed fusion-specific colocalized transcripts mixed with wild-type transcripts (Figure 1D). Next, we applied FuseFISH to FFPE tissue sections of mouse xenografts of human *EML4-ALK* positive NSCLC cells. Similar to cultured cells, FuseFISH was able to detect fusion transcripts amidst wild-type ones (Figure 1E).

While detection of fusion transcripts in cells and tissues can be powerful in cytogenetic diagnostics of well-characterized fusions, detection in purified RNA would be desirable for rapid validation of newly discovered fusions. To maximize the versatility of FuseFISH, we devised a protocol for smFISH using purified RNA spotted on microscopy slides (Figures S1B and S1C; see Supplemental Experimental Procedures). We spotted RNA extracted from the CML cell line K562 and used probes targeting *BCR* and *ABL1* transcripts. As expected, *BCR-ABL1* fusion transcripts appeared as bright, colocalized fluorescent spots (Figure 1F).

Automated Signal Detection

Though spatial colocalization of fusion transcripts is immediately apparent in the above experiments, unbiased quantification is unfeasible by eye. To achieve fully automated and unbiased fusion detection, we implemented a computational pipeline that determines the average fraction α of signals from a given transcript species A (e.g., *BCR*) with a correlated (i.e., nonrandomly colocalized) signal from a different transcript species B (e.g., *ABL1*), based on particle image cross-correlation spectroscopy (PICCS; Semrau et al., 2011; Figures 2A and S2A; see Experimental Procedures). Briefly, we constructed a cumulative correlation function $C_{cum}(r)$ by counting the average number of B signals neighboring an A signal depending on the distance r of separation. Linear fitting of $C_{cum}(r)$ versus r^2 at distances exceeding the typical correlation length gave the correlation fraction α as the offset of the fitted line. We initially processed 3D image stacks to reduce the incidence of random colocalization. However, at observed transcript densities, we obtained similar correlation fractions and accuracies using maximum projections and 2D analysis (Figure S2B). Since the computational effort is substantially smaller for 2D analysis, we applied this approach in all subsequent analyses. We compared α in cells expressing a particular fusion with negative cells. As expected, α was systematically higher in positive cells (Figure S2C). We also compared formalin fixation to fixation in methanol-acetic acid—the standard fixative in cytogenetics—and obtained very similar α values (Figure S2C). To prove that FuseFISH is quantitative, we calculated $\alpha_{BCR-ABL1}$ for serial dilutions of K562 with *BCR-ABL1*-negative TS cells. We observed strong linear scaling of $\alpha_{BCR-ABL1}$ with our prediction based on pure population measurements (Figure 2B). We then determined the sensitivity of FuseFISH by estimating the accuracy of the measured α by bootstrapping. A mix of K562 and TS cells at 1:80 ratio yielded a significantly higher $\alpha_{BCR-ABL1}$ value compared to TS cells alone (3.2% versus 0.8%, $p < 10^{-3}$). Finally, we estimated the minimal amount of signals that need to be measured to determine α with a given accuracy. We

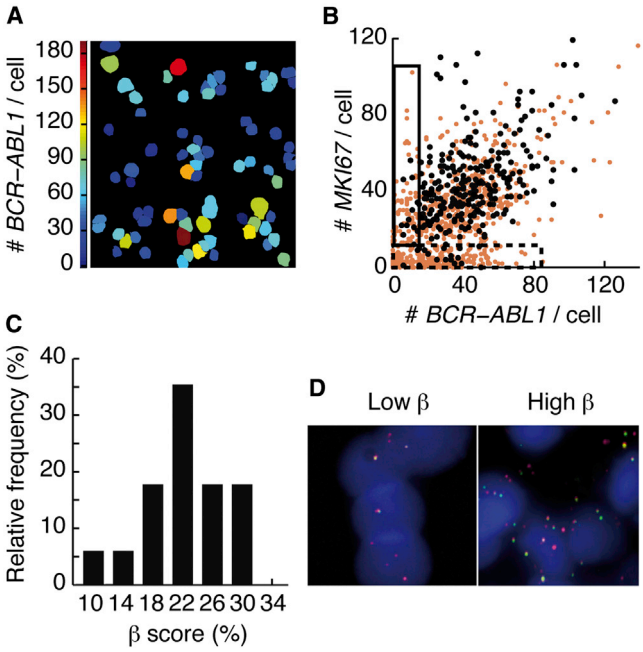

Figure 3. Applications of FuseFISH

(A) Cell-to-cell *BCR-ABL1* expression variability in K562 cells.

(B) *BCR-ABL1* and *MKI67* expression in K562 cells exposed to Nilotinib (orange) or vehicle (black) for 48 hr. Dashed box: heterogeneous *BCR-ABL1* expression in Nilotinib-treated $MKI67^{Low}$ cells. Thick box: $BCR\text{-}ABL1^{Low}$-$MKI67^{High}$ Nilotinib-treated subpopulation.

(C) Intertumor variability of local correlation fraction β.

(D) *BCR-ABL1* expressing cells representative of the left (Low β) and right (High β) portion of the histogram in (C).

See also Figure S3.

modeled PICCS accuracy as a function of several parameters (Figure S2D). For example, if $\alpha = 5\%$ is to be measured with an error margin of 1% at a density of ten transcripts per cell (100 μm^2 cell area), the model predicts that approximately 1,000 signals (100 cells) have to be imaged.

Detection of rare fusion events holds important biological and clinical implications. For example, monitoring of decreasing *BCR-ABL1* levels in CML patients is crucial to assess therapeutic response (Kantarjian et al., 2008; Radich, 2009). However, as the frequency of a fusion decreases in a population of cells, the imaging time required to obtain statistically significant results becomes quickly unpractical. We reasoned that performing FuseFISH in concentrated purified RNA would allow us to detect rare fusions, maximizing the sensitivity of our method. Thus, we devised a protocol to quickly and easily immobilize purified RNA on microscope slides (see Supplemental Experimental Procedures) and imaged thousands of transcripts from a mixture of K562 and TS cells. By this approach, we reliably detected *BCR-ABL1* fusions derived from as little as one K562 cell per 10,000 TS cells (Figure S2E).

Validation of FuseFISH

To validate our approach, we compared FuseFISH with gold standard methods used for *BCR-ABL1* diagnostics (see Experimental Procedures). We retrospectively analyzed 34 hematologic samples fixed in methanol-acetic acid, including 17 cases of CML and 5 cases of acute lymphoblastic leukemia expressing *BCR-ABL1* (age: median, 56 years; range, 52–63 years; see Table S3). We derived a receiver operating characteristic (ROC) curve by comparing FuseFISH α to the diagnosis (*BCR-ABL1* positivity or negativity) previously formulated based on DNA FISH and/or RT-PCR. A score $\alpha \geq 5\%$ correctly identified positive cases with 100% specificity and 89% sensitivity, and α positively correlated ($r^2 = 0.62$, $p = 0.05$) with RT-PCR results (Figures 2C, 2D, and S2F). We also performed systematic cross-validation and determined that our algorithm achieves maximal sensitivity and specificity when trained on $\geq$25 cases (Figure S2G).

BCR-ABL1 Cell-to-Cell Variability

Until now, robust quantification of fusion transcript in single cells has been challenging. Such measurement would advance our understanding of tumor heterogeneity and its implications on disease prognosis and therapeutic response. As a proof of principle, we applied FuseFISH to study fusion transcript expression variability and dynamics in single cancer cells upon fusion-targeted therapy. We first assessed the expression of *BCR-ABL1* in K562 cells and found that it was noisy (coefficient of variation [CV] = 0.48; Figure 3A). Next, we simultaneously measured *BCR-ABL1* and the proliferation marker *MKI67* in hundreds of individual K562 cells undergoing treatment with the specific tyrosine kinase inhibitor, Nilotinib. As expected, upon treatment the fraction of proliferating cells decreased substantially (Figures 3B and S3A). Based on *MKI67* expression upon Nilotinib treatment, we classified cells into $MKI67^{High}$ and $MKI67^{Low}$. After 24 hr exposure to Nilotinib, the $MKI67^{High}$ subpopulation exhibited increased expression variability compared to vehicle control (CV = 0.6, $p = 5 \times 10^{-3}$, F test). Interestingly, we also observed a subpopulation of Nilotinib-treated $MKI67^{High}$ cells with lower *BCR-ABL1* expression compared to untreated cells (Figures 3B and S3A–S3C).

Finally, we measured *BCR-ABL1* variability in our set of hematologic samples. Since cell segmentation can be challenging in these specimens, we computed a local correlation fraction, β, reporting on the *BCR-ABL1*/*BCR* ratio in the neighborhood of *BCR-ABL1* signals (see Experimental Procedures). In serial dilutions of K562 with TS cells, β was constant over a wide range of dilution rates ($\alpha \geq 5\%$) and close to the α value calculated for a pure population of K562 cells (Figure S3D). Thus, β is a faithful proxy of the level of *BCR-ABL1* expression in positive cells. In *BCR-ABL1*-positive specimens the β score varied substantially from cases expressing low *BCR-ABL1* per cell (Low β) to cases expressing high amounts (High β) (Figures 3C and 3D). Interestingly, in seven *BCR-ABL1*-positive cases with known progression-free survival (PFS), we found that PFS tended to be shorter in patients with high α and β; however, without reaching significance ($p > 0.05$) due to low statistical power (Table S3). In the future, ad hoc prospective trials should be designed to clarify the impact of fusion expression and variability on clinical outcome.

DISCUSSION

We have devised a readily applicable and versatile experimental and computational framework that can serve multiple purposes

and complements existing technologies for gene fusion detection. Our genome-scale oligo database is a powerful resource thanks to which FuseFISH probes (and, more generally, smFISH probes) can be easily and rapidly designed, making the method particularly suitable to confirm newly discovered fusions.

Since it can be applied to methanol-acetic acid fixed cells—the standard fixation procedure in cytogenetics—FuseFISH can be integrated in diagnostics practice, accelerating assay time (hybridization, imaging, and automated signal quantification can be performed on the same day) and eliminating subjective biases in scoring colocalized or split-apart DNA FISH signals. We expect that the fully automated and unbiased correlation analysis achievable with our computational pipeline will facilitate fusion transcript detection and help improve interlaboratory reproducibility.

The ability to visualize individual transcripts using purified RNA is particularly beneficial for quick and reliable validation of fusions discovered by next-generation sequencing techniques such as RNA sequencing. Purified RNA FuseFISH might also be applied when the fraction of fusion-expressing cells is low, such as for monitoring *BCR-ABL1* levels in CML patients during targeted therapy. While FuseFISH sensitivity is currently ten times lower than for RT-PCR (Kantarjian et al., 2008; Radich, 2009)—the current standard to monitor minimal residual disease in CML patients—implementation of the method on a high-throughput imaging system could further improve sensitivity.

Investigation of transcriptional heterogeneity in early developing cancers is in its infancy, and experimental and computational frameworks to measure fusion transcript variability are missing. FuseFISH opens the possibility to investigate the dynamics of fusion expression in response to therapeutic agents, relate expression variability to clinical outcomes, and simultaneously assess specific gene expression signatures with defined fusions in individual cells. In the long term, it will be important to design prospective trials to test whether these features predict patients' response to fusion-targeted drugs and whether measuring fusion expression in single cells should be implemented in clinical practice.

EXPERIMENTAL PROCEDURES

smFISH Probe Database

All bioinformatic analyses were performed using custom scripts written in MATLAB. We retrieved the sequence of all human protein-coding transcripts by linear concatenation of the exons of each transcript in the ENSEMBL annotated gene database (release 70, January 2013). We scanned each transcript except the last 19 nt using 20 nt windows sliding in 1 nt steps, computed the guanine-cytosine (GC) content, and then assembled all windows with GC content comprised between 40% and 60%. We designed as many 20 nt oligos as possible separated by at least 2 nt in each region with optimal GC content, starting from the 5′ end of each transcript. We maximized the number of oligos with GC = 45% or GC = 50% by computing the GC content of 20 nt oligos starting from each of five consecutive nucleotide positions and selecting the most 5′ position with GC = 45% or GC = 50%, if available. We screened the uniqueness of designed oligos by performing a local BLAST search against the NCBI RefSeq_RNA database (http://www.ncbi.nlm.nih.gov/refseq/) using *blastn*.

smFISH

Probes targeting *BCR* (48 oligos), *ABL1* (48 oligos), *NPM1* (16 oligos), *ALK* (48 oligos), and *EML4* (48 oligos) transcripts consisted of amine-labeled oligos labeled with Cy5 (GE Healthcare, catalog no. Q15108), Alexa Fluor 594 (Molecular Probes, catalog no. A20004), or 6-TAMRA (Molecular Probes, catalog no. C6123) (Table S2). Hybridizations and washes were done according to modified protocols based on previously described procedures (Bienko et al., 2013; Raj et al., 2008; see Supplemental Experimental Procedures for details).

Microscopy

We performed imaging as described earlier (Bienko et al., 2013) using an inverted epi-fluorescence microscope (Nikon) equipped with a high-resolution charge-coupled device camera (Pixis, Princeton Intruments). Magnification of 100× and 40× oil immersion, high-numerical-aperture Nikon objectives were used for cells and tissues, and purified RNA, respectively. Per region of interest, we typically acquired an image stack consisting of 40 (cells and tissues) or 20 (purified RNA) image planes spaced 0.2 μm apart. Details on image processing are available in the Supplemental Experimental Procedures.

PICCS

Due to chromatic aberration and unavoidable imperfections of microscopy setup, the two signals stemming from the two moieties of a fusion transcript are shifted with respect to each other. A typical solution is defining an arbitrary distance threshold below which two signals are considered colocalized (i.e., correlated). However, especially when expression levels are high (and signals are, therefore, dense), results depend sensitively on the choice of this threshold. To circumvent this limitation, we used PICCS (Semrau et al., 2011) to calculate the global level of correlation, α, between fusion partners in an unbiased way. We constructed the cumulative distribution function, C_{cum}, of distances between signals from differentially labeled species (A and B) by counting the number of all B signals within a distance r of an A signal and dividing it by the number of A signals (Figure S2A). The resulting $C_{cum}(r)$ function reports the average number of B neighbors (of an A signal) closer than r and has the general form $C_{cum}(r) = \alpha_{AB} P_{cum}(r) + \pi c_B r^2$, where α_{AB} is the correlation fraction (i.e., the fraction of A signals which have a correlated B signal), c_B is the average density of B signals, and P_{cum} is the cumulative distribution function of distances between truly correlated signals. When plotted versus r^2, the contribution of uncorrelated B signals $\pi c_B r^2$ can be estimated by linear fitting of $C_{cum}(r)$ at distances exceeding the correlation length. The fit line intercepts the y axis at the correlation fraction α, and its slope is proportional to the density of B signals c_B (Figure 2A).

ROC Analysis and Cross-Validation

We constructed ROC curves by computing the rate of true positives (sensitivity) and false positives (100% − specificity) based on specimen classification by DNA FISH and/or RT-PCR. We obtained separate ROC curves for various signal-to-noise ratio cut-offs. We performed cross-validation by repeated random sampling (without replacement) of a subset of specimens (training set), while the ratio of positive and negative cases was the same as in the full data set. For each sample, we determined an optimal α threshold using three different algorithms and calculated the false and true positive rates for the remaining specimens (test set). Algorithms 1 and 2 force the false positive rate to be below 0.001 while maximizing the true positive rate. While algorithm 1 maximizes the threshold, algorithm 2 tries to minimize the threshold if there is no negative impact on the true positive rate. Algorithm 3 maximizes the absolute difference between the false positive and the true positive rate, which is equivalent to the distance of a point on the ROC curve to the diagonal. Algorithm 1 was chosen because it led to smaller false positive and true positive rates compared to the other algorithms, and it saturated to false positive and true positive rates found for the complete data set (Figure S2G).

Local Correlation Fraction

The optimal threshold r_t is given by $C_{cum}(r) = \alpha_{AB} P_{cum}(r) + \pi c_B r^2$ as indicated in Figure 2A. We considered neighboring *BCR* and *ABL1* signals to be part of the same fusion transcript if they were closer than r_t. If single cells can be identified, *BCR-ABL1* expression can also be determined for individual cells. Unfortunately, abundant cell clumping in clinical samples precluded such analysis. However, having identified *BCR*, *ABL1*, and *BCR-ABL1* separately, we again used PICCS to determine a local correlation fraction, β, between *BCR-ABL1* and all *BCR*-containing transcripts. To do so, we started from *BCR-ABL1*

transcripts and counted the number of *BCR-ABL1* or all *BCR*-containing transcripts in the neighborhood, as shown in Figure S2A. The local correlation fraction, β, is the ratio of the retrieved densities, $\beta = c_{BCR-ABL1}/(c_{BCR-ABL1} + c_{BCR})$, reporting on the ratio of *BCR-ABL1* and total *BCR* expression in *BCR-ABL1*-expressing cells. For a perfectly homogeneous population of cells $\beta = \alpha$, whereas in a mixed population consisting of *BCR-ABL1*-expressing and -nonexpressing cells β is the local correlation fraction of positive cells and, in general, $\beta > \alpha$.

We retrieved hematological specimens in methanol-acetic acid 3/1 v/v from the archive of the Division of Hematology of IRCCS Policlinico San Matteo after approval of the study by the institutional Ethical Committee. Permission to conduct mice experiments was granted by the Italian Ministry of Health (approval number 218-2009-B).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.12.002.

AUTHOR CONTRIBUTIONS

N.C. conceived FuseFISH and the protocol for smFISH on purified RNA, designed the genome-wide probe database, performed experiments, and wrote the manuscript. S.S. developed all computational methods, performed all data analyses, conceived the experiments with Nilotinib, and wrote the manuscript. M. Bienko contributed ideas, performed experiments, and prepared figures. M. Boni and P.B. provided hematologic samples and relevant clinical information. R.C. contributed ideas and provided cell lines and mouse xenotransplants. A.v.O. supported and steered the project, provided ideas, and corrected the manuscript.

ACKNOWLEDGMENTS

This work was supported by MIT NIH/NCI Physical Sciences Oncology Center (U54CA143874 to A.v.O.), NIH Pioneer Award (1DP CA174420 to A.v.O.), NWO Vici Award (to A.v.O.), NWO Rubicon Award (to S.S.), FP7 ERC-2009-StG grant (proposal no. 242965 "Lunely" to R.C.), AIRC IG-12023 and AICR 12-0216 grants (to R.C.), and the Human Frontiers Science Program (to M. Bienko). We thank Sandy Klemm (A.v.O. lab) and Chiara Ambrogio (CNIO, Madrid) for critically reading the manuscript and help with analyses, Maria Stella Scalzo (R.C. lab) for assistance with tissue preparation, and Barbara Rocca and Celeste Calvello (P.B. lab) for RT-PCR data.

Received: August 12, 2013
Revised: November 5, 2013
Accepted: December 3, 2013
Published: December 26, 2013

REFERENCES

Alpár, D., Hermesz, J., Pótó, L., László, R., Kereskai, L., Jáksó, P., Pajor, G., Pajor, L., and Kajtár, B. (2008). Automated FISH analysis using dual-fusion and break-apart probes on paraffin-embedded tissue sections. Cytometry A *73*, 651–657.

Baccarani, M., Cortes, J., Pane, F., Niederwieser, D., Saglio, G., Apperley, J., Cervantes, F., Deininger, M., Gratwohl, A., Guilhot, F., et al.; European LeukemiaNet (2009). Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. J. Clin. Oncol. *27*, 6041–6051.

Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. Nat. Methods *10*, 122–124.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Fröhling, S., and Döhner, H. (2008). Chromosomal abnormalities in cancer. N. Engl. J. Med. *359*, 722–734.

Kantarjian, H., Schiffer, C., Jones, D., and Cortes, J. (2008). Monitoring the response and course of chronic myeloid leukemia in the modern era of BCR-ABL tyrosine kinase inhibitors: practical advice on the use and interpretation of monitoring methods. Blood *111*, 1774–1780.

La Thangue, N.B., and Kerr, D.J. (2011). Predictive biomarkers: a paradigm shift towards personalized cancer medicine. Nat. Rev. Clin. Oncol. *8*, 587–596.

Lerner, B., Clocksin, W.F., Dhanjal, S., Hultén, M.A., and Bishop, C.M. (2001). Automatic signal classification in fluorescence in situ hybridization images. Cytometry *43*, 87–93.

Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A.M. (2009). Transcriptome sequencing to detect gene fusions in cancer. Nature *458*, 97–101.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer *12*, 323–334.

Melo, J.V., and Barnes, D.J. (2007). Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat. Rev. Cancer *7*, 441–453.

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

Rabbitts, T.H. (2009). Commonality but diversity in cancer gene fusions. Cell *137*, 391–395.

Radich, J.P. (2009). How I monitor residual disease in chronic myeloid leukemia. Blood *114*, 3376–3381.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Ren, R. (2005). Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat. Rev. Cancer *5*, 172–183.

Rowley, J.D. (2009). Chromosomes in leukemia and beyond: from irrelevant to central players. Annu. Rev. Genomics Hum. Genet. *10*, 1–18.

Schiffer, C.A. (2007). BCR-ABL tyrosine kinase inhibitors for chronic myelogenous leukemia. N. Engl. J. Med. *357*, 258–265.

Semrau, S., Holtzer, L., González-Gaitán, M., and Schmidt, T. (2011). Quantification of biological interactions with particle image cross-correlation spectroscopy (PICCS). Biophys. J. *100*, 1810–1818.

Shirley, J.W., Ty, S., Takebayashi, S.-I., Liu, X., and Gilbert, D.M. (2011). FISH Finder: a high-throughput tool for analyzing FISH images. Bioinformatics *27*, 933–938.

Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S.-I., Watanabe, H., Kurashina, K., Hatanaka, H., et al. (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature *448*, 561–566.

Waks, Z., Klein, A.M., and Silver, P.A. (2011). Cell-to-cell variability of alternative RNA splicing. Mol. Syst. Biol. *7*, 506.