

Benchmarks

A simple approach for classifying new mutations as somatic or germinal in DNA samples lacking paired tissue

Ignacio Erquiaga, Cristina Hurtado, Paula Aranaz, Francisco J. Novo, and José L. Vizmanos

Department of Biochemistry and Genetics, University of Navarra, Pamplona, Spain

BioTechniques 56:327–329 (June 2014) doi 10.2144/000114180

Keywords: mutation detection; somatic mutation; cancer; loss of heterozygosity

Supplementary material for this article is available at www.BioTechniques.com/article/114180.

When studying mutations in DNA samples, determining whether novel sequence changes are somatic mutations or germline polymorphisms can be difficult. Here we describe a novel and very simple approach for identification of somatic mutations and loss of heterozygosity (LoH) events in DNA samples where no matched tissue sample is available. Our method makes use of heterozygous polymorphisms that are located near the putative mutation to trace both germinal alleles.

Somatic mutations, along with other genetic and epigenetic aberrations, are important in the development of cancer as well as non-neoplastic diseases (1). These mutations can arise in normal cells through various mechanisms, conferring proliferative advantages that lead to the development of a neoplastic clone (2).

The identification of new mutations is an important part of cancer research (3–4) since it allows for the determination of clonality and can improve diagnosis and prognosis (5–7), thus allowing the development of rational treatments, such as TK inhibitors (8–9). In addition, determining the somatic status of a sequence change is necessary for the proper selection of candidates for subsequent functional assays.

Today, the gold standard technique for somatic mutation identification is the use of another tissue from the same individual for comparative analysis. Unfortunately, such paired tissue samples are not always available to researchers.

In these cases, efforts are usually made to seek the newly identified sequence change in healthy controls and/or databases, or to estimate the ratio of the new sequence change versus the normal sequence. However, these approaches can lead to the misidentification of very rare polymorphisms or sequence artifacts as somatic mutations (10), and do not demonstrate the somatic nature of the change.

An additional challenge is that the complexity of the human genome makes

the number of healthy samples required to screen for rare germline variants increasingly large in the first approach, and even these large numbers can be insufficient to rule out very rare variants (11–12).

The ratio of the sequence change versus the normal sequence can be quantified by sequence chromatogram dropping factor analysis (13), pyrosequencing, or RQ-PCR. Any significant deviation from the expected 1:1 ratio of mutated to non-mutated DNA molecules (pure heterozygosity) would suggest a somatic change. However, depending on the abundance of cells carrying the mutation and on the allelic burden of mutant alleles in these cells, this ratio can vary significantly in neoplastic tissue, even in germinal variants.

In our experience, analyses of more than 350 healthy samples have proven to be insufficient to detect rare polymorphisms or singletons; dropping factor analysis also showed inconclusive results.

For these reasons, we developed a simple method for assessing the somatic or germline nature of these new sequence variants. Our approach requires the presence of a known polymorphism in heterozygosity near the sequence variant to allow identification of both alleles. Using this polymorphism as a reference, we can check to see if the new variant is present in all of the copies of one allele or only in some of them. In addition, this is a useful approach for detecting loss of heterozygosity (LoH) events.

The suitability of our method depends on the distance between the heterozygous reference polymorphism and the sequence change; the longer the distance, the higher the probability of recombination occurring between both variants. Recombination would result in the somatic change being present on both alleles, not only on the allele where it arose.

Experimentally, fragments of DNA containing the new variant and a reference heterozygous polymor-

METHOD SUMMARY

Using a known polymorphism in heterozygosity adjacent to a sequence change of interest, we were able to classify sequence variants as somatic mutations or germline polymorphisms by cloning and sequencing individual molecules to determine if the variants are present in different proportions on one or both germline alleles.

Table 1. Allele count for clones carrying a fragment of *JAK2* containing the p.V617F mutation and SNP rs10283730

	Allele	Patient 1 (ET)			Patient 2 (ET)			Patient 3 (PV)		
		p.V617F			p.V617F			p.V617F		
		G (WT)	T (mut)	Total	G (WT)	T (mut)	Total	G (WT)	T (mut)	Total
rs10283730	G	8	0	8	9	0	9	4	10	14
	A	8	1	9	5	5	10	5	0	5
	Total	16	1	17	14	5	19	9	10	19

Patient 3 has an imbalance between alleles G and A of rs10283730, suggesting a loss of heterozygosity (LoH) event where allele A, carrying the wild-type V617 allele, has been lost in some or all the cells carrying the p.V617F mutation ($P = 0.035$, goodness-of-fit test). ET, essential thrombocythemia; PV, polycythemia vera.

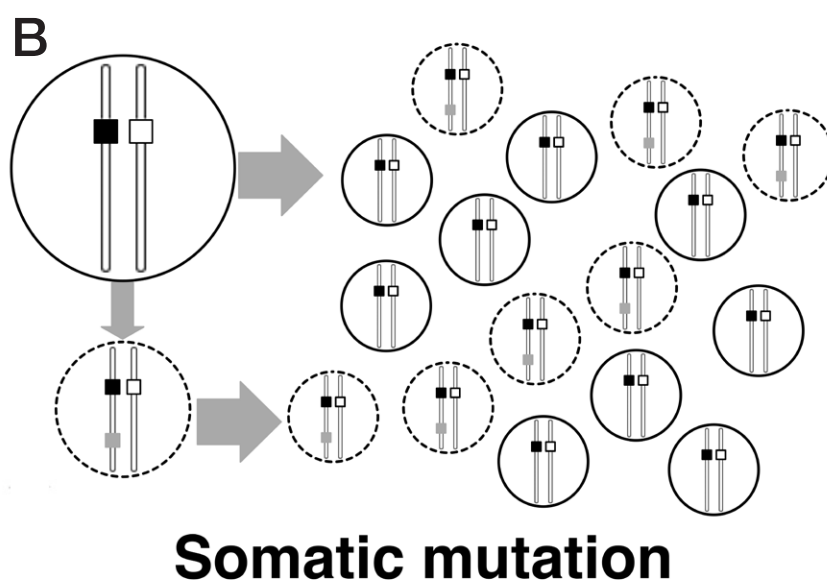
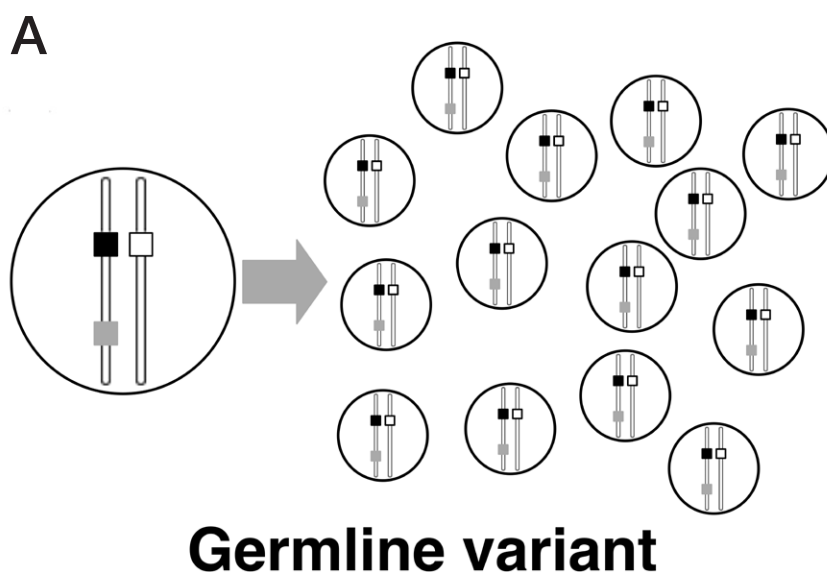


Figure 1. Cell populations carrying a novel sequence variant next to a known polymorphism in heterozygosis. (A) If the new sequence change (gray square) is a germline variant, all of the DNA molecules carrying the black allele (black square) of the control polymorphism will carry the new sequence change. (B) If the new sequence change is a somatic mutation, some of the molecules will carry the new sequence variant next to the black allele, while other molecules will carry the wild type allele next to the black allele.

phism must first be amplified. Close candidate polymorphisms can be identified using database searches. Given the number of polymorphisms in the human genome (estimated at 9.5 million on dbSNP 138; www.ncbi.nlm.nih.gov/SNP/), these are expected to occur once every 300–350 nucleotides. As recombination rates for the human genome are on the order of 1% for 1.2 Mb (14), the recombination rate for a region 2 kb in length would be expected to be <0.002%.

If good candidate polymorphisms are lacking, 2 DNA fragments containing the new variant and 1.5–2 kb upstream and downstream sequence, respectively, can be amplified and sequenced in order to look for unknown nearby heterozygous polymorphisms (see Supplementary Materials).

After amplification, the resulting amplicon is cloned into a suitable plasmid. Isolated clones will carry copies of a single molecule with one allele of the reference polymorphism and either the wild type or the novel sequence of the variant to study. Plasmid sequencing of individual bacterial clones will show if the new sequence variant is present in every clone of one of the alleles, as would occur in the case of a germline variant (Figure 1A), or only in some of them, as in the case of a somatic mutation (Figure 1B).

Using this approach, we were able to classify two new sequence variants as germline (see Supplementary Materials): p.P166S in *CSF2RA* and p.W226X (not described as a polymorphism at the time of the experiment, currently rs143118009) in *IL3RA*.

In addition, our method also allowed us to properly classify two cases of essential thrombocythemia carrying a heterozygous *JAK2* p.V617F mutation

and to identify an LoH event in a *JAK2* p.V617F positive polycythemia vera patient (Table 1), a frequent occurrence in these neoplasms (15). The three patients were heterozygous for the nearby SNP rs10283730, which allowed us to identify both alleles. Patients 1 and 2 were both found to carry allele A of rs10283730 and either the wild type (G) or mutant (T) allele of p.V617F, demonstrating that this sequence change is not in the germline.

Interestingly, patient 3 showed a 9:10 ratio of wild-type (G) to p.V617F mutant (T) molecules, similar to a pure 1:1 heterozygosity condition. Analysis of the ratio of alleles using RQ-PCR, pyrosequencing, or sequence chromatogram would have suggested that this was a germline polymorphism, but further analysis of individual molecules reveals that there are molecules with the allele G of the reference SNP carrying either the wild type (G) or mutant (T) p.V617F allele. This nearly 1:1 ratio for the p.V617F alleles is in fact due to an LoH event that makes allele G of rs10283730 (carrying mutant allele T) more frequent than allele A. A goodness-of-fit test for the allelic distribution of SNP rs10283730 (14 G, 5 A) shows a significant difference to the expected 1:1 ratio ($P = 0.035$), supporting the existence of LoH.

Assuming a diploid state, ratios that significantly differ from 1:1 would indicate the existence of an imbalance in gene copy number due to either gains or losses. Here, we successfully detect two such events: LoH in patient 3 and monosomy in the case with *CSF2RA* p.P166S (see Supplementary Material).

Author contributions

I.E. and J.L.V. designed the experiments. I.E., C.H., and P.A. conducted the experiments. All of the authors contributed to the writing of the manuscript.

Acknowledgments

This work was funded with the help of the PIUNA program of the University of Navarra (PIUNA 2011-12). Ignacio Erquiaga is a recipient of a predoctoral grant from the Government of Navarra. Paula Aranaz is a recipient of a postdoctoral grant from the Government of Navarra.

Competing interests

The authors declare no competing interests.

References

- Poduri, A., G.D. Evrony, X. Cai, and C.A. Walsh. 2013. Somatic mutation, genomic variation, and neurological disease. *Science* 341:1237758.
- Hanahan, D. and R.A. Weinberg. 2000. The hallmarks of cancer. *Cell* 100:57-70.
- Levine, R.L., M. Wadleigh, J. Cools, B.L. Ebert, G. Wernig, B.J. Huntly, T.J. Boggon, I. Wlodarska, et al. 2005. Activating mutation in the tyrosine kinase *JAK2* in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7:387-397.
- Stephens, P.J., P.S. Tarpey, H. Davies, P. Van Loo, C. Greenman, D.C. Wedge, S. Nik-Zainal, S. Martin, et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486:400-404.
- Pietrantonio, F., F. De Braud, V. Da Prat, F. Perrone, M.A. Pierotti, M. Gariboldi, G. Fanetti, P. Biondani, et al. 2013. A review on biomarkers for prediction of treatment outcome in gastric cancer. *Anticancer Res.* 33:1257-1266.
- Taylor, J.W., A.S. Chi, and D.P. Cahill. 2013. Tailored therapy in diffuse gliomas: using molecular classifiers to optimize clinical management. *Oncology (Williston Park)*. 27:504-514.
- Xing, M., B.R. Haugen, and M. Schlumberger. 2013. Progress in molecular-based management of differentiated thyroid cancer. *Lancet* 381:1058-1069.
- Deininger, M., E. Buchdunger, and B.J. Druker. 2005. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* 105:2640-2653.
- Pytel, D., T. Sliwinski, T. Poplawski, D. Ferriola, and I. Majsterek. 2009. Tyrosine kinase blockers: new hope for successful cancer therapy. *Anticancer. Agents Med. Chem.* 9:66-76.
- Abdel-Wahab, O., O. Kilpivaara, J. Patel, L. Busque, and R.L. Levine. 2010. The most commonly reported variant in *ASXL1* (c.1934dupG;p.Gly646TrpfsX12) is not a somatic alteration. *Leukemia* 24:1656-1657.
- Tennessen, J.A., A.W. Bigham, T.D. O'Connor, W. Fu, E.E. Kenny, S. Gravel, S. McGee, R. Do, et al. 2012. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science* 337:64-69.
- Nelson, M.R., D. Wegmann, M.G. Ehm, D. Kessner, P. St. Jean, C. Verzilli, J. Shen, Z. Tang, et al. 2012. An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science* 337:100-104.
- Ernst, T., A. Chase, K. Zoi, K. Waghorn, C. Hidalgo-Curtis, J. Score, A. Jones, F. Grand, et al. 2010. Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms. *Haematologica* 95:1473-1480.
- Jensen-Seaman, M.I., T.S. Furey, B.A. Payseur, Y. Lu, K.M. Roskin, C.F. Chen, M.A. Thomas, D. Haussler, and J. Jacob. 2004. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. *Genome Res.* 14:528-538.
- Kralovics, R., F. Passamonti, A.S. Buser, S.S. Teo, R. Tiedt, J.R. Passweg, A. Tichelli, M. Cazzola, and R.C. Skoda. 2005. A gain-of-function mutation of *JAK2* in myeloproliferative disorders. *N. Engl. J. Med.* 352:1779-1790.

Received 10 December 2013; accepted 11 April 2014.

Address correspondence to Ignacio Erquiaga, Department of Biochemistry and Genetics, University of Navarra, Pamplona, Spain. E-mail: ierquiaga@alumni.unav.es

To purchase reprints of this article, contact: biotechniques@fosterprinting.com





STEM-CELLBANKER

Chemically defined and animal product free

Consistent high cell viability & proliferation

Long term cell storage at -196°C & -80°C

Ready-to-use formulation with a simple protocol

No programmed freezer or liquid nitrogen required

COME VISIT US AT ISSCR, VANCOUVER. 18 - 21 JUNE 2014, BOOTH 600.