Chapter 13

Allelic Imbalance Assays to Quantify Allele-Specific Gene Expression and Transcription Factor Binding

Francesca Luca and Anna Di Rienzo

Abstract

A growing number of noncoding variants are found to influence the susceptibility to common diseases and interindividual variation in drug response. However, the mechanisms by which noncoding variation affects cellular and clinical phenotypes remain to be elucidated. Allele-specific assays allow testing directly the differential properties of the alleles at a regulatory variant, which are detected as an allelic imbalance. Two widely used allelic imbalance assays target cDNA and DNA from chromatin immunoprecipitation (ChIP) experiments, and therefore revealing allele-specific gene expression and transcription factor binding, respectively. The throughput of allelic imbalance assays ranges from single variant to the genome scale, which are made possible by the recent advances in genotyping and sequencing technologies (e.g., genome-wide quantitative cDNA genotyping, ChIP-seq).

Key words Polymorphism, Chromatin immunoprecipitation, RNA, cDNA, Quantitative PCR, Gene expression

1 Introduction

DNA polymorphisms in regulatory regions may account for a large proportion of interindividual differences in common phenotypes. Accordingly, a large number of noncoding SNPs have been associated with diseases in genome-wide association studies (e.g., [1]). The functional relevance of regulatory polymorphisms has been further confirmed by the increasing number of studies establishing an association between genetic variation and cellular phenotypes (e.g., mRNA levels, [2–7]).

Allelic imbalance assays allow the investigator to evaluate the *cis* effect of a putative regulatory variant, by directly assessing the effect of each allele at the site of interest or at a proxy SNP. The power of these approaches relies on the fact that the two alleles at

Federico Innocenti and Ron H.N. van Schaik (eds.), *Pharmacogenomics: Methods and Protocols*,
Methods in Molecular Biology, vol. 1015, DOI 10.1007/978-1-62703-435-7_13,

a site are compared within the same sample (a heterozygous individual), therefore removing the confounding effect of environmental or *trans*-acting factors.

These assays have been largely used to investigate allelic expression and allele-specific transcription factor binding. Ultimately, a combination of both approaches is able to provide direct evidence that variation at a given site results in different levels of gene expression by altering, for example, the interaction between the DNA and a transcription factor.

A cDNA/RNA allelic imbalance assay is based on the notion that *cis*-acting regulatory polymorphisms cause differential expression between chromosomes in heterozygotes. This will result in unequal representation of alleles at coding polymorphisms on the same haplotype in the mRNA of individuals heterozygous for the regulatory polymorphism.

One of the molecular mechanisms by which regulatory polymorphisms affect gene expression is through alterations of DNA–protein binding affinity. This can be detected in a Chromatin immunoprecipitation (ChIP) assay followed by allele specific quantification of the ChIPed DNA (commonly known as HaploChIP). ChIP assays allow in vivo analysis of DNA–protein interaction. Proteins are cross-linked to the chromatin in living cells by formaldehyde treatment; the chromatin is then sheared and incubated with an antibody specific for the protein of interest. Following the immunoprecipitation, the DNA is purified and can be analyzed by a variety of techniques including quantitative real-time PCR.

Both RNA/cDNA and ChIP allelic imbalance assays were originally developed for single gene analyses [8, 9]. In the following protocol, we will describe applications that include TaqMan quantitative genotyping assays. However, alternative approaches to quantifying allelic imbalance can be used (e.g., fluorescent dideoxy terminator-based methods [10], MALDI-TOF-MS [11, 12]). More recently, high-throughput genotyping and sequencing technologies have expanded the potential for allelic imbalance applications to the genome-wide scale [13].

2 Materials

2.1 Cell Culture

The following protocol uses lymphoblastoid cell lines (LCLs). However, allelic imbalance assays can be performed also in other cell lines as well as primary cells. Suggestions on how to modify the protocol when using different cell types are provided throughout the text.

1. LCLs from individuals carrying the heterozygous genotype at the SNP(s) of interest (*see* **Note 1**).
2. RPMI 16490 (Gibco), supplemented with 15 % FBS and 0.1 % Gentamycin.

2.2 Preparation of the Sample for Allelic Imbalance (AI)

In the current protocol the Upstate (Millipore) ChIP Assay Kit reagents are used; however, details to prepare the reagents are also provided.

2.2.1 Preparation of the Sample for ChIP AI

1. 2×10^6 LCLs in mid-log exponential phase.
2. Fresh 18.5 % formaldehyde: 0.925 g paraformaldehyde, 35 μl 1 M KOH, add water to a final volume of 5 ml.
3. Fresh 10× Glycine (1.25 M).
4. Protease Inhibitors (For 1 ml of buffer add 10 μl of PMSF and 1 μl Protease Inhibitors Cocktail).
5. Ice-cold PBS.
6. SDS lysis Buffer: 1 % SDS, 10 nM EDTA, 50 mM Tris, pH 8.1.
7. Protease inhibitors.
8. Sonicator.
9. ChIP Dilution Buffer: 0.01 % SDS, 1.1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8.1, 150 mM NaCl.
10. IgG.
11. Salmon Sperm DNA/Protein A Agarose-50 % Slurry.
12. ChIP-grade antibody specific for the protein of interest.
13. 5 M NaCl.
14. 0.5 M EDTA, pH 8.0.
15. 1 M Tris–HCl, pH 6.5.
16. Low-Salt Immune Complex Wash Buffer: 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8.1, 150 mM NaCl.
17. High-Salt Immune Complex Wash Buffer: 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8.1, 500 mM NaCl.
18. LiCL Immune Complex Wash Buffer: 0.25 M LiCl, 1 % IGEPAL-CA630, 1 % deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.1.
19. 1× TE: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
20. Freshly prepared Elution Buffer: 1 % SDS, 0.1 M $NaHCO_3$.
21. RNase.
22. Qiagen PCR purification kit.
23. 10 μM primers specific to a negative control region (i.e., a region known not to bind the transcription factor).
24. 10 μM primers specific to a positive control region (i.e., a region known to bind the transcription factor).
25. SYBR® Green Master Mix (Applied Biosystems, or any other company).

2.2.2 Preparation of the Sample for RNA/cDNA AI

1. LCLs in mid-log exponential phase (Approx total RNA yield: 3 μg/10^6 cells).
2. Qiagen RNeasy Plus mini kit.
3. High-Capacity cDNA Reverse Transcription Kits from Applied Biosystems.

2.3 AI Assay

1. Taqman® Universal PCR Master Mix with no AmpErase® UNG.
2. Taqman® 40× SNP Genotyping Assay.

3 Methods

Allelic imbalance assays are performed on cells from individuals heterozygous at the site of interest (ChIP AI assays) or at a coding site in linkage disequilibrium with the site of interest (RNA/cDNA AI assays). Here we describe the protocols for these two assays, starting from cell cultures. Different methods have been developed to assay allelic imbalance, both at the single gene and at the genome-wide level. The method we describe uses quantitative TaqMan genotyping, but could be substituted by other methods depending on the resources available to each investigator.

3.1 Cell Culture

This method describes the protocol for LCLs grown in suspension; however, other cell types can be used (*see* **Note 2**). LCLs are seeded at 0.5×10^6 and, once in mid-log exponential phase, may be stimulated according to experimental design (for example with dexamethasone) (*see* **Note 3**) prior to harvesting.

3.2 Preparation of the Sample for AI

Alternative ChIP protocols are successfully used in other laboratories and can replace the one described here.

3.2.1 Preparation of the Sample for ChIP AI

Formaldehyde Cross-linking

1. In the tissue culture hood, add formaldehyde to culture flasks (final concentration of 1 %). Swirl and incubate for 20 min at 37 °C in the tissue culture incubator (*see* **Note 4**).
2. Add 1.25 M (10×) Glycine to the flasks, to a final concentration of 0.125 M (1×). Glycine quenches the cross-linking reaction. Swirl and incubate at room temperature for 5 min.
3. Meanwhile, add protease inhibitors to PBS (need about 20 ml PBS per 75 cm^2 flask) and cool the centrifuge to 4 °C.
4. Collect the cells by centrifugation at 290 × *g* for 7 min at 4 °C.
5. Wash cells twice with 10 ml of ice-cold PBS containing protease inhibitors.
6. Cell pellet can be stored at −80 °C.

Cell lysis and Sonication

7. (If continuing with the sonication on the same day) Add protease inhibitors to SDS Lysis Buffer at room temperature.
8. (If cell pellet was frozen) Thaw cell pellet on ice, take half of the volume and collect the pellet at 700 × *g* at 4 °C for 5 min (store at −80 °C the remaining aliquot).
9. Meanwhile, add protease inhibitors to SDS Lysis Buffer
10. Remove the supernatant and add 700 µl of the SDS Lysis Buffer to the pellet, resuspend and split into two 2 ml tubes with flat bottom and incubate for 10 min on ice.
11. Sonicate the samples in a QSonica S4000 (or a Q700) Sonicator with a cup horn immersed in ice-cold water. The sonication program includes 50 cycles (each cycle is 30 s ON and 1 min OFF, amplitude set at 90) (*see* **Note 5**). To avoid overheating of the sample, replace the water with new ice-cold one every 10 min.
12. Pellet in tabletop centrifuge (14,000 × *g*) at 4 °C for 10 min, place the supernatant into a new 1.7 ml tube.
13. The sonication should result in DNA fragments approx 200–400 bp in size. Remove 20 µl (tester) in order to determine if the appropriate fragment size has been obtained.
14. Store the majority of the supernatant at −80 °C until the sonication has been checked.
15. To reverse the cross-links in the tester, add 2 µl 5 M NaCl and incubate at 65 °C for 4 h or overnight.
16. Add 1 µl RNase to each tester and incubate at room temperature for 30 min.
17. Add 1 µl 0.5 M EDTA, 2 µl 1 M Tris–HCl, pH 6.5, and 1 µl of 10 mg/ml Proteinase K to the testers and incubate for 1 h at 45 °C.
18. Run the samples on a 1 % TBE gel at 80 V for 30 min. If the DNA smear is within 200–400 bp size range proceed with the ChIP assay protocol (*see* **Note 6**).

ChIP Assay

19. Add protease inhibitors to the ChIP Dilution Buffer.
20. Split each sample into two 1.7 ml tubes (one of the two samples can be used to obtain an IgG ChIP control). Dilute the sonicated cell supernatant ~10-fold in ChIP Dilution Buffer by adding 1.2 ml of the ChIP Dilution Buffer to the 150 µl sonicated cell supernatant to a final volume of 1.35 ml.
21. Preclear, by adding 60 µl of Salmon Sperm DNA/Protein A Agarose-50 % Slurry and 1 µg of IgG for 1 h at 4 °C with rotation (*see* **Note 7**).

22. Pellet agarose (5,000 × *g* for 1 min or less at 4 °C) and place supernatant in new tubes.
23. Remove 20 µl for the input control sample and store at −80 °C.
24. Add antibody (1–5 µg depending on antibody of choice) to the supernatant and incubate overnight at 4 °C with rotation.
25. Add 45 µl Salmon Sperm DNA/Protein A Agarose Slurry for 1 h at 4 °C with rotation.
26. Pellet agarose (5,000 × *g* for 1 min at 4 °C). Aspirate out the supernatant and wash for 3–5 min on a rotating platform with 1 ml of each of the buffers listed in the order given below:
    (a) Low-Salt Immune Complex Wash Buffer, one wash
    (b) High-Salt Immune Complex Wash Buffer, one wash
    (c) LiCl Immune Complex Wash Buffer, one wash
    (d) 1× TE, two washes
27. Meanwhile, prepare the Elution Buffer.
28. Elute by adding 250 µl elution buffer to the pelleted agarose/antibody/protein complex. Vortex briefly and incubate at room temperature for 15 min with rotation. Spin down the agarose, and carefully transfer the supernatant fraction to another tube and repeat elution. Combine eluates (total volume ~ 500 µl).
29. Thaw the input control samples.
30. Add 20 µl 5 M NaCl (2 µl for input) to the 500 µl eluates and reverse cross-links by heating at 65 °C for 4 h or overnight (*see* **Note 8**). One may store the sample at −20 °C and continue the next day.
31. Add 1 µl RNase to each sample and incubate at room temperature for 30 min.
32. Add 10 µl 0.5 M EDTA (1 µl for input), 20 µl 1 M Tris–HCl (2 µl for input), pH 6.5, and 2 µl of 10 mg/ml Proteinase K (1 µl for input) to the combined eluates and incubate for 1 h at 45 °C.
33. Recover DNA with the Qiagen PCR Purification Kit, following the manufacturer's instructions.
34. Before performing AI Assays on the ChIPed DNA, the quality of the ChIP experiment should be assessed by performing quantitative real time PCRs targeting a positive and a negative control region.

3.2.2 Preparation of the Sample for RNA/cDNA AI

1. Extract total RNA with the Qiagen RNeasy Plus mini kit following the manufacturer's protocol.

2. Synthesize cDNA from total RNA (100 ng) using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. Dilute cDNA samples 1:30 to perform AI assays.

3.3 AI Assay

While ChIP AI assays target directly the candidate binding variant of interest, cDNA AI assays are designed to target a coding SNP in the gene that is differentially regulated by the two alleles at the candidate regulatory variant.

1. The following criteria should be used to select a coding SNP to be assayed in a cDNA AI assay:
   (a) High linkage disequilibrium with the regulatory variant (if phased genotype data including the regulatory SNP are not available (*see* **Note 9**))
   (b) High heterozygosity
   (c) >40 bp away from exon boundary to allow designing assays that will amplify both gDNA and cDNA thus allowing for the use of a gDNA standard curve
2. In most cases, predesigned TaqMan genotyping assays are available from Applied Biosystems. Alternatively, custom made assays can be designed using the Custom TaqMan Assay Design Tool.
3. Quantitative real-time PCR assays can be performed in either 96- or 384-well plates using any of the ABI systems (e.g., the ABI PRISM 7900HT Sequence Detection System or the ABI StepOnePlus™ Real-Time PCR System). Reactions are typically run in triplicates for each sample.

   PCR mix for a sample run on a 96-well plate

   Total volume: 20 µl

   cDNA/ChIPed DNA: 4 µl

   Taqman® Universal PCR Master Mix with no AmpErase® UNG: 10 µl

   Taqman® 40× SNP Genotyping Assay: 0.5 µl

   Use Applied Biosystems standard recommended PCR cycling conditions.
4. To account for differences between the two fluorochromes, a standard curve should be built for each of the two alleles using serial dilutions (*see* **Note 10**) of genomic DNA from an individual heterozygous at the assayed SNP (Fig. 1). PCR products are quantified for each allele separately in each reaction and ratios between the two different alleles can be calculated (Fig. 2). The results can then be averaged across PCR replicates.

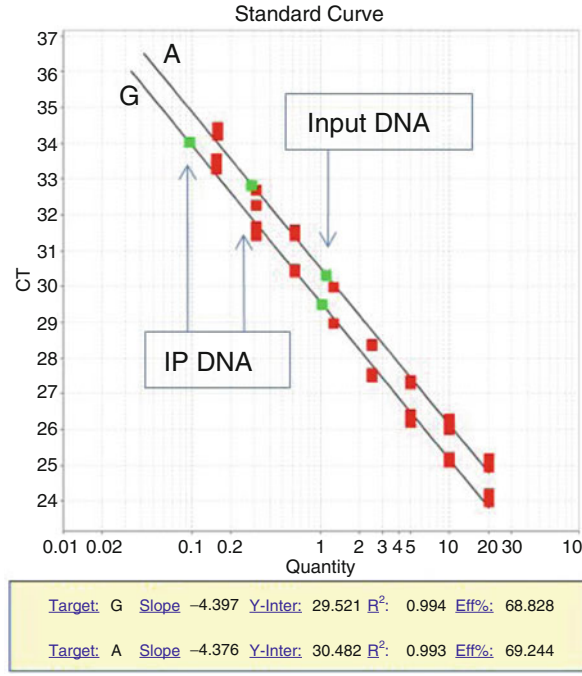

Fig. 1 Example of the results from a ChIP AI assay at a regulatory variant for the *SGK1* gene [14]. The samples are plotted over the standard curves built for the two alleles separately. An imbalance in the ChIPed DNA can be observed

4 Notes

1. For ChIP AI assays, individuals should be heterozygous at the candidate binding variant. For RNA/cDNA AI assays, the assayed individuals should include heterozygotes at a coding SNP in linkage disequilibrium with the candidate regulatory variant (this coding SNP is directly assayed for allelic imbalance in the cDNA) and both homozygotes and heterozygotes at the candidate regulatory variant (a minimum of three individuals at each of these genotype classes is required to perform *t*-test). AI assays can be relatively noisy, especially in genes expressed at low levels; therefore, a large sample size is recommended (e.g., >5). This can be achieved by either performing replicates of the chromatin immunoprecipitation/RNA extraction and PCR in the same individual or by assaying multiple individuals with the same genotype. The latter option has the advantages that it allows testing for robustness of the observed AI in different genetic backgrounds.
2. This same protocol for formaldehyde cross-linking can be applied also to other cell types growing in suspension. For adherent cells, formaldehyde and glycine should be added

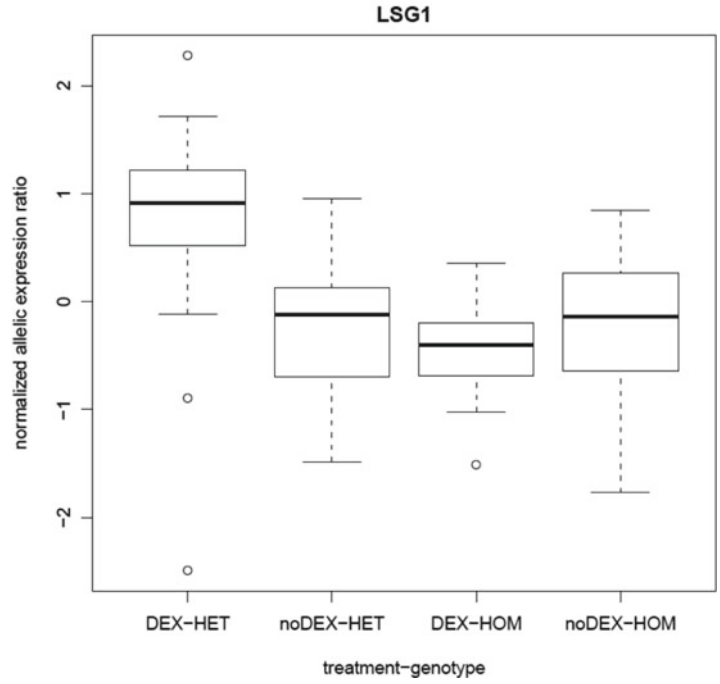

Fig. 2 Example of the results of a cDNA AI assay targeting a coding SNP in linkage disequilibrium with an interaction eQTL at the *LSG1* gene [15]. The assay was performed on samples cultured in two different conditions (with and without dexamethasone). In this assay, the natural log-ratio between the two different alleles was calculated and quantile normalized in each treatment condition separately. Two PCR replicates were performed and the results were averaged. A significant difference between heterozygotes and homozygotes at the candidate regulatory variant in the presence of dexamethasone was observed ($p = 8.38 \times 10^{-5}$). In each box, the horizontal line represents the median and the whiskers represent the first and third quartile

directly to the tissue culture dish. The medium should then be aspirated and cells washed with PBS in the plate. Cells should be harvested using a cell scraper and moved to a conical tube.

3. If the goal of the experiment is to compare allele-specific expression/binding under different conditions, a balanced study design should be employed. Specifically, each experimental unit should be defined as the set of experimental conditions assayed for each sample.
4. The time and concentration of treatment for the cross-linking with formaldehyde should be optimized for the specific cell type analyzed. The conditions described here have been successfully used for MCF-10aMyc and LCLs [14].
5. Depending on the cell type, protein of interest and sonicator available, sonication conditions should be optimized. In general we suggest using a sonicator equipped with a water bath

to prevent the foaming generated by a sonicator equipped with a probe. The use of a water bath also reduces inter-sample variability in sonication size, which is a feature desired when performing comparisons across treatment conditions or between samples.

6. If the desired fragment length distribution has not been achieved, it is possible to repeat the sonication. In our experience if the first sonication has generated a tight fragment distribution of a size larger than the desired one, 10–20 additional cycles of sonication are enough to shift the fragment size distribution to the desired one. Repeated freezing–thawing cycles of the pre-sonicated pellet are not advisable as they could result in disruption of protein–DNA bonds.
7. The preclear step is optional. Its purpose is to remove proteins that interact nonspecifically with the IP components.
8. In our experience, 4 h of incubation are enough for the tester, while overnight incubation is required for the actual IP samples.
9. If the regulatory variant and the coding SNP are in perfect linkage disequilibrium, allelic imbalance can be detected as an overrepresentation of the allele at the coding SNP that occurs on the same chromosome as the allele at the regulatory SNP that results in higher transcript levels. However, when the two variants are not in perfect linkage disequilibrium, different alleles at the coding SNP will be over-represented depending on the haplotype phase in the samples examined. In this case, an allelic imbalance can be detected by comparing the variance of the allelic ratio at the coding SNP across individuals that are heterozygous and homozygous for the regulatory SNP. This test is less powerful than the one performed in the case of perfect linkage disequilibrium.
10. Standard curves should be built using serial dilutions (eight serial dilutions, 1:2 [ChIP AI] or 1:5 [cDNA AI], starting at 20–25 ng/μl)

Acknowledgments

We thank Sonal Kashyap, Allison Richards, and Shaneen Baxter for contributing to the optimization of these protocols and Joseph Maranville for helpful advice. F.L. was supported by an AHA postdoctoral fellowship (11POST5390005).

References

1. WTCCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661–678
2. Morley M et al (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430(7001):743–747
3. Pickrell JK et al (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464(7289):768–772
4. Smirnov DA et al (2009) Genetic analysis of radiation-induced changes in human gene expression. Nature 459:587–591
5. Stranger BE et al (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1(6):e78
6. Stranger BE et al (2007) Population genomics of human gene expression. Nat Genet 39(10):1217–1224
7. Montgomery SB et al (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464(7289): 773–777
8. Yan H et al (2002) Allelic variation in human gene expression. Science 297(5584):1143
9. Knight JC et al (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet 33(4):469–475
10. Matyas G et al (2002) Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. Hum Mutat 19(1):58–68
11. Jurinke C et al (2002) Automated genotyping using the DNA MassArray technology. Methods Mol Biol 187:179–192
12. Braun A et al (1997) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 43(7):1151–1158
13. Pastinen T (2010) Genome-wide allele-specific analysis: insights into regulatory variation. Nat Rev Genet 11(8):533–538
14. Luca F et al (2009) Adaptive variation regulates the expression of the human SGK1 gene in response to stress. PLoS Genet 5(5):e1000489
15. Maranville JC et al (2011) Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. PLoS Genet 7(7):e1002162