Supplemental Methods Table of Contents

Supplemental Protocol 1: Lab 1, Brown Lab

FileName	3_Spike-in_Rep1_MCF7_LMPCR_4X.CEL.txt	2_Spike-in_Rep2_MCF7_LMPCR_4x-1.CEL.txt	3_Spike-in_Rep3_MCF7_LMPCR_4X.CEL.txt	4_Control_Rep1_MCF7_LMPCR_4X.CEL.txt	5_Control_Rep2_MCF7_LMPCR_4X-1.CEL.txt	6_Control_Rep3_MCF7_LMPCR_4X-1.CEL.txt
Lab Name	Brown	Brown	Brown	Brown	Brown	Brown
Array Platform	720K v1.0	720K v1.0	720K v1.0	720K v1.0	720K v1.0	720K v1.0
Array Resolution	35 bp	35 bp	35 bp	35 bp	35 bp	35 bp
Probe Length	25 mer	25 mer	25 mer	25 mer	25 mer	25 mer
Amount of DNA added to the arrays	2 ug	2 ug	2 ug	2 ug	2 ug	2 ug
Amount of DNA started in the PCR amplification	10 ng	10 ng	10 ng	10ng	10ng	10ng
Amplification Method	LM-PCR	LM-PCR	LM-PCR	LMPCR	LMPCR	LMPCR
PCR cycles	25 cycles	25 cycles	25 cycles	25 cycles	25 cycles	25 cycles
Hybridization site	DFCI array core	DFCI array core	DFCI array core	DFCI array core	DFCI array core	DFCI array core
Sample Name (one color array)	Spike-in	Spike-in	Spike-in	Input	Input	Input
Sample Name (two color array)	NA	NA	NA	NA	NA	NA
Replicate	1	2	3	1	2	3

Ligation-mediated PCR protocol:

End filling and blunt ended ligation

- On ice, add:

11ul 10x T4 DNA pol buffer
0.5ul BSA (NEB 10mg/ml)
1ul 10mM dNTP mix
0.2ul T4 DNA pol (NEB 3U/ul)
up to 110ul with dH_2O

- Straight from ice, add to pre-cooled PCR machine and run the program: 12oC for 20 min

- Add 11.5ul 3M NaAc and 10ug glycogen

- Extract once with phenol:chloroform:IAA

- Add 230ul 100% EtoH. 30 min -80^oC

- Spin max 15 min at 4^oC

- Wash 150ul 70% EtoH. Spin max 10 min

- Resuspend dried pellet in 25ul dH_2O

- All on ice (include thawing of ingredients), add:

7.8ul H_2O
10ul 5x ligase buffer (Gibco)
6.7ul of annealed linkers (15 uM) –see below
0.5ul T4 DNA ligase (NEB)
Total: 25 ul to each sample

- Incubate 16^{o}C overnight

- Next day, add 6ul NaAc and 130ul EtoH. 30min at -80^{o}C

- Spin max 15 min, wash in 150ul 70% EtoH and spin max 10 min.

  - Resuspend dried pellet in 25 ul of dH_2O

Annealing of linkers:

Oligo 102- GCGGTGACCCGGGAGATCTGAATTC
Oligo 103- GAATTCAGATC

Make up: 250ul Tris-Hcl (1M) pH 7.9
375ul oligo 102 (40uM stock)
375ul oligo 103 (40uM stock)

Heat at 95^{o}C for 5 min, then immediately into a 70^{o}C heat block (with water in the holes) and remove heat block and leave on the bench to slowly cool. Freeze aliquots

Ligation Mediated-PCR

- To the 25ul samples, add 15ul of:
  4.75ul dH_20
  4ul 10x Thermopol buffer (NEB)
  5ul 2.5mM dNTP mix
  1.25ul of 40uM oligo (the longer oligo in the linker)

- Place the 40ul samples in PCR machine and run program:
  55^{o}C 4 min
  72^{o}C 3 min
  95^{o}C 2 min
  95^{o}C 30 sec
  60^{o}C 30 sec
  72^{o}C 1 min
  Go to step 4, 24 more times
  72^{o}C 4 min
  4^{o}C forever

- Midway through the first 55^{o}C incubation, add:
  8ul dH_2O
  1ul 10x Thermopol buffer (NEB)
  1ul Amplitaq (Perkin Elmer….not Amplitaq Gold!)
  0.01ul PFU turbo (what sticks to the end of a pipette tip)

- Mix whilst at 55^{o}C.

- Clean up with Qiagen PCR kit.

- Elute in 60ul of elution buffer.

- Do real-time PCR of validation target to check that LM-PCR has not biased the target ratio to the input samples.

Labeling

On ice:

- Dilute DnaseI (Ambion) to appropriate amount based on optimized conditions. We do a 1:100 dilution in water and use 1ul of this for each sample. Include the buffer. Efficient digestion can be checked by running the DNA on a gel. The optimal smear should range from 50bp to 100bp. A little higher is OK.

- So, add buffer, Dnase I and buffer to final volume of 55ul. Run the program:

37oC 30 min
95oC 15 min
4oC hold

- Run 5ul on a gel to check for good fragmentation. Smear should be between 50bp and 100bp, although a little larger is OK

On ice:

- Add: 13ul TdT buffer (Promega)
1ul Biotin (1mM stock)
1ul TdT (30U/ul)
50ul of fragmented DNA

-Run the program:

37oC 16 hours
95oC 10 min
4oC hold

- Spin briefly and prepare for hybridization

Microarray protocol :

Hybridization is standard Affymetrix protocol.

Supplmental Protocol 2: Lab 3, Farnham and Green Labs

Amplifications were performed using the following protocol.

From O'Geen et al., BioTechniques 41(5), (November 2006)

Farnham Lab Whole Genome Amplification Protocol for ChIP-chip
(adapted from protocol provided with Sigma GenomePlex Kit)

**[Since the input material is sonicated chromatin, the initial series of fragmentation steps is skipped and one can go right to library preparation]

A. Library Preparation

1. Add 2 ul 1X Library Preparation Buffer to 10 ul of input material

[For the "input" sample, measure the concentration of reverse crosslinked, QIAquick purified DNA and add 10 ng to a total volume of 10 ul with H2O. For the ChIP sample, the concentration of nucleic acid is usually too low to get an accurate quantitation. Typically the entire 50ul of reverse crosslinked, QIAquick purified DNA is lyophilized and resuspended in 10 ul of H20]

Transfer samples to strip tubes or individual thin walled 0.2 ml PCR tubes

2. Add 1 ul Library Stabilization Solution, vortex or mix by pipetting. Quick spin and place at 95° for 2 minutes in thermal cycler

3. Immediately cool on ice, quick spin again

4. Add 1 ul Library Preparation Enzyme, vortex or mix by pipetting and quick spin if necessary

5. Incubate in thermal cycler as follows:
   16° for 20' (cycler should be precooled to this temperature)
   24° for 20'
   37° for 20'
   75° for 5'
   4° hold

9. Quick spin if necessary and either proceed to first amplification or freeze at –20° for up to three days

B. Amplification (round 1)

10. Prepare master mix for each sample containing:

7.5 ul of 10X Amplification Master Mix
47.5 ul Nuclease-free H20
5 ul WGA DNA polymerase

[For multiple samples, multiply above volumes by the number of samples then add 1/10 volume extra of each component]

11. Add 60 ul master mix to each sample, vortex or mix by pipetting and quick spin if necessary

12. Incubate in thermal cycler as follows:
    95° for 3' , then 14 cycles of
    94° for 15"
    65° for 5', then
    4° hold

At this point, amplified material is stable and can be stored at –20 ° indefinitely

13. Purify samples using QIAquick PCR cleanup columns or analogous product. It is important to elute the samples in water so that the subsequent labeling reactions are efficient.

[Since the amplified material contains both single- and double-stranded DNA that can be effectively labeled, the column purification method used should recover both.]

[At this stage, the purification column eluates for total and immunoprecipitated samples should be readily quantifiable by nanodrop, spectrometer, or dye intercalation, eg, picogreen (dye intercalation may underestimate amount due to single strand product). Optimally, total recovery for immunoprecipitated samples will be in the 1-4 ug range. This gives enough material for several labelings for downstream microarray analysis. If yields are less, or more product is desired, re-amplify material using Sigma GenomePlex WGA Reamplification Kit]

C. Reamplification (round 2)

1. Add 15 ng purified amplification product in 10 ul volume to strip tubes or individual thin walled 0.2 ml PCR tubes

[For input material start with the high concentration primary amplified stock]

2. Prepare master mix for each sample containing:

7.5 ul of 10X Amplification Master Mix
47.5 ul Nuclease-free H20
5 ul WGA DNA polymerase

For multiple samples, multiply above volumes by the number of samples then add 1/10 volume extra of each component

3. Add 60 ul master mix, vortex or mix by pipetting and quick spin if necessary

From O'Geen et al., BioTechniques 41(5), (November 2006)

4. Incubate in thermal cycler as follows:
   95° for 3' , then 14 cycles of
   94° for 15"
   65° for 5', then
   4° hold

At this point, amplified material is stable and can be stored at –20 ° indefinitely

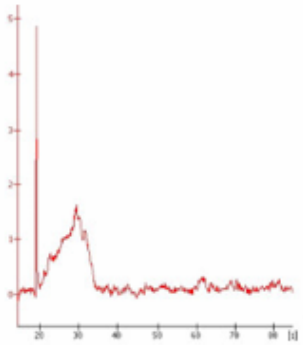
5. Purify samples using QIAquick PCR cleanup columns or analogous product.

[Since the amplified material contains both single- and double-stranded DNA that can be effectively labeled, the column purification method used should recover both.]

Hybridizations were performed at the Madison, WI NimbleGen facility with the following protocol:

NimbleGen SYSTEMS, INC.

Microarray Analysis of ChIP Samples



Example of a degraded sample trace

Step 3. Sample Labeling

Most researchers performing ChIP-chip include a total input DNA reference as a control. Some include a non-specific binding control, along with their IP sample.

- Total input DNA – a small aliquot of the starting DNA pool is carried through the labeling protocol and used as an array reference sample
- Non-Specific Binding (NSB) Control – NSB is a control IP reaction carried out in parallel to the specific IP reaction where the primary antibody has been omitted or substituted with pre-immune or non-specific immunoglobulin G. These controls are used to test the level of enrichment that is specific to the IP antibody.

Pairs of samples intended for hybridization to the same array should be labeled in parallel. Test samples should be labeled with Cy5 and reference samples (either total input DNA or NSB control) should be labeled with Cy3.

1. Prepare the following three solutions:

50X dNTP Mix		
VWR deionized water	250µl	Aliquot 50X dNTP mix into single-use amounts. Avoid freeze-thaw cycles, which can result in diminished yields.
10X TE	50µl	
100 mM dATP	50µl	
100 mM dGTP	50µl	
100 mM dTTP	50µl	
100 mM dCTP	50µl	
Total	**500µl**	
Random 9mer Buffer		
VWR deionized water	8.6ml	Make fresh each time primers are diluted.
1M Tris-HCl	1.25ml	
1M $MgCl_2$	125µl	
β-Mercaptoethanol	17.5µl	
Total	**10ml**	
10X TE		
1M Tris-HCl	1.5ml	Mix and store at room temperature.
0.5M EDTA	300µl	
VWR deionized water	13.2ml	
Total	**15ml**	

 |

NimbleGen SYSTEMS, INC.

Protocols for

ChIP-chip Hybridization

Outline

This protocol describes the process for hybridization of samples prepared by chromatin immunoprecipitation and amplified by ligation-mediated PCR (LM-PCR) on NimbleGen DNA microarrays. The protocols for chromatin crosslinking, immunoprecipitation and LM-PCR are described in the Chromatin Immunoprecipitation and Amplification protocol, available from NimbleGen.

The first section of this protocol describes the QC of samples prior to microarray hybridization to verify sample quality. The protocol then describes the labeling of the sample, followed by hybridization, and stringent washing.

Protocol Information & Safety

- Wear gloves and take precautions to avoid sample contamination.
- Cy dyes are light sensitive. Minimize light exposure of the dyes during use and store in the dark immediately after use.
- Cy dyes are ozone sensitive. See http://genomics.princeton.edu/dunham/ozone.html for information on monitoring and remediation for atmospheric ozone levels exceeding 5ppb.
- We have found using VWR water and DTT for all post-hybridization washes useful in obtaining higher signal from Cy dyes.

Required Apparatus & Labware*

* See last page for a reagent supplier list

- Thermocycler
- Microcentrifuge with 12,000 x g capability
- Heat block(s) at 37°C, 45°C, and 95°C
- Agilent 2100 Bioanalyzer
- Vortexer
- Spectrophotometer (A NanoDrop Spectrophotometer is recommended)
- Speed Vac
- MAUI Hybridization System (BioMicro Systems, Model # 02-A002-03)
- MAUI Mixer SL Low Temp. Hybridization Chambers (BioMicro Systems, Catalog # 02-A008-03)
- Method of drying arrays after washing (e.g., custom NimbleGen Array-Go-Round unit, ArrayIt Microarray Highspeed Centrifuge, Cat# MHC110V)
- Dark Desiccator

2. Dilute Cy3 and Cy5 dye-labeled 9mers to 1 O.D./42µl Random 9mer Buffer. Aliquot to 40 µl individual reaction volumes in 0.2 ml thin-walled PCR tubes and store at –20 °C.

3. Assemble the following components in a 0.2ml thin-walled PCR tube:

	1. Test Sample	2. Reference Sample
LM-PCR Sample (1µg)	1µg	1µg
Cy3-9mer Primers		40 µl
Cy5-9mer Primers	40 µl	
VWR Water (to 80 µl)	To volume	To volume
Total	**80µl**	**80µl**

Important! Snap-chilling step after denaturation is critical for high efficiency labeling.

4. Heat denature samples in thermocycler at 98 °C for 10 minutes. Quick chill in ice water bath. Add the following (for larger quantities of samples, you may use a master mix for this step):

	1. Test Sample	2. Reference Sample
Reaction volume from Step 3.3	80µl	80µl
50X dNTP mix	10µl	10µl
VWR deionized water	8µl	8µl
Klenow (50 U/µl)	2µl	2µl
Cumulative Total	**100µl**	**100µl**

Important! Do not vortex after addition of Klenow.

5. Mix well by pipetting 10 times.

6. Spin down at low RPM to force contents to bottom of tube.

7. Incubate at 37 °C for 2 hours in a thermocycler protected from light.

8. Stop the reaction by addition of EDTA:

	1. Test Sample	2. Reference Sample
Reaction volume from step 3.4	100µl	100µl
Stop Solution (0.5 M EDTA)	10µl	10µl
Cumulative Total	**110µl**	**110µl**

Important! Vortex briefly after addition of 5M NaCl prior to addition of isopropanol.

9. Transfer the reaction to a 1.5 ml tube. Precipitate the labeled samples by adding NaCl and isopropanol to each tube as indicated.

	1. Test Sample	2. Reference Sample
Reaction volume from step 3.7	110µl	110µl
5M NaCl	11.5µl	11.5µl
Isopropanol	110µl	110µl
Cumulative Total	**231.5µl**	**231.5µl**

10. Vortex. Incubate 10 minutes at room temperature in the dark.

11. Centrifuge at maximum speed for 10 minutes. Remove supernatant with pipette.

12. Rinse pellet with 500µl 80% ice-cold ethanol. Dislodge pellet from tube wall.

13. Centrifuge at maximum speed for 2 minutes. Remove supernatant with pipette.

14. Speed-Vac on low heat for 5 minutes until dry, protected from light.
15. Store labeled samples at -20° C, protected from light.

Important!

Spin tubes down briefly prior to opening.

16. Rehydrate dried pellets in 25µl VWR deionized water.
17. Vortex for 30 seconds and quick spin to collect contents at bottom of tube. Continue to vortex or let sit in the dark at room temp until the pellet is completely rehydrated, then vortex again and quickly spin.
18. Measure the A_{260} in each sample.

    It is important to consume as little sample as possible when performing this measurement. NimbleGen recommends the use of a NanoDrop spectrophotometer (www.nanodrop.com) that allows the measurement of a 1µl sample without dilution. Typical yields range from 10µg to 30µg per reaction.

Step 4. Hybridization of Cy-labeled ChIP Samples

1. Set MAUI hybridization unit to 42°C and allow time for the temperature to stabilize.

Important!

Protect tube from light during handling to prevent photobleaching of the light-sensitive Cy dyes.

2. Based on the A_{260} measurement performed in step 3.18, combine 13µg each of the Test and Reference Samples into a single 1.5ml microcentrifuge tube.
3. Dry the combined contents in a Speed-Vac on low heat.
4. Resuspend the sample in 10.9µl VWR water and vortex to completely dissolve the sample. Spin the tube down briefly to collect the contents in the bottom

Important!

CPK6 48mer Oligos are included n the hybridization as controls that hybridize to alignment features on NimbleGen arrays. They are required for proper extraction of array data from the scanned image.

Note: All volumes in this step have changed due to a change in the volume of the MAUI Hybridization Chamber.

5. Using the NimbleGen Array Reuse Kit, add the following to the resuspended sample:

Cy labeled Test & Reference samples	10.9µl
2X Hybridization Buffer	19.5µl
Hybridization Component A	7.8µl
Cy3 CPK6 50mer Oligo, 50nM	0.4µl
Cy5 CPK6 50mer Oligo, 100nM	0.4µl
Total	**39µl**

6. Mix the tube briefly then spin down to collect the contents in the bottom and place at 95°C for 5 minutes.
7. Immediately transfer the tube to the MAUI 42°C sample block and hold at this temperature until you are ready for sample loading.
8. Place the MAUI Mixer SL Hybridization Chamber on the array using the provided assembly/disassembly jig and carefully follow MAUI setup instructions. Use the braying tool to remove all air bubbles from the adhesive gasket around the outside of the hybridization chamber.
9. Put the array and Hybridization Chamber on the MAUI and allow 30 seconds for the chips to come up to temperature.

Important!

Hybridization chambers have a finite shelf life. Check the expiration date. Use the chambers within 30 minutes of opening the package.

10. Load the sample using the pipet supplied with the MAUI Station and following manufacturer instructions. During loading, a small amount (3-7µl) of the sample may flow out of the outlet port.

11. Confirm that there are no bubbles in the chamber.

    If there are, very gently massage any bubbles to either of the ends, away from the center of the array. Avoid applying too much pressure since this will force liquid out of the ports.

12. Place the loaded array into one of the four MAUI bays and let equilibrate for 30 seconds. Wipe off any sample leakage at the ports with a Kim-Wipe, and adhere MAUI stickers to both ports.

13. Close the bay clamp and select mix mode B. Hold down the mix button to start mixing. Confirm that the mixing is in progress before closing the cover.

14. Hybridize the sample overnight (16-20 hours).

Step 5. Washing of Arrays

1. Prior to removing the array from the MAUI Hybridization Station, prepare the following solutions. You will need **two** 250 ml dishes of Wash I, and one each for Wash II and Wash III. One dish for Wash I should be shallow and be wide enough to accommodate the array and mixer loaded in the MAUI assembly/disassembly jig. The lid from a 1000 µl pipet tip box works well. Also, the buffer in the first Wash I dish should be heated to 42°C to help soften the adhesive on the hybridization chamber; this will help prevent braking the array. Place the remaining three wash solutions in 300ml Tissue-Tek slide staining dishes.

	Wash I x 2	Wash II	Wash III
VWR Water	225ml	225ml	225ml
10X Wash Buffer I, II, or III	25ml	25ml	25µl
1M DTT	25µl	25µl	25µl
Total	**250ml**	**250ml**	**250ml**

Important!

Peel the hybridization chamber off very slowly to prevent the slide from cracking. Do not let the surface of the slide dry out at any point during washing.

The first volume of Wash I may be heated to 42°C to enable the removal of the mixer.

2. Remove chip from MAUI Hybridization Station, load it back into the MAUI assembly/disassembly jig, and immerse in the shallow 250ml Wash I at 42°C. While the chip is submerged, carefully peel off the mixer. Gently agitate the chip in Wash 1 for 10-15 seconds.

3. Transfer the slide into a slide rack in the second dish of Wash I and incubate 2 minute with agitation.

4. Transfer to Wash II and incubate 1 minute with agitation. Rock the dish to move the wash over the tops of the arrays.

5. Transfer to Wash III and incubate for 15 sec with agitation.

6. Remove array and spin dry in an array-drying unit (e.g. the NimbleGen Array-Go-Round) for 1 minute.

7. Store the dried array in a dark desiccator and proceed immediately to the protocol for NimbleGen Two-Color Scanning of NimbleGen Arrays.

Component and Supplier List

Component	Vendor	Package Size	Item Number
0.5M EDTA	Sigma Aldrich	100ml	E-7889
Absolute Ethanol	Sigma Aldrich	500mL	E702-3
Water, Reagent Grade, ACS, Nonsterile, Type 1	VWR	2.5 gallon	RC915025
100mM dNTP's	Invitrogen	4 X 25umole	10297-018
1M Tris HCL pH 7.4	Sigma Aldrich	1L	T-2663
1M $MgCl_2$	Sigma Aldrich	100ml	M-1028
β-Mercaptoethanol	Sigma Aldrich	100 mL	6010
Klenow Fragment 3' ->5' exo-	NEB	50,000u/ml	M0212M
5M NaCl	Sigma	500ml	9759
Isopropanol	Sigma Aldrich	500ml	I-9516
Cy3 9mer Wobble	TriLink Biotechnologies	50 O.D.	N46-0001-50
		200 O.D.	N46-0001-200
Cy5 9mer Wobble	TriLink Biotechnologies	50 O.D.	N46-0002-50
		200 O.D.	N46-0002-200
Cy3 CPK6 50mer (see below)	IDT	250nmole	Custom oligo synthesis
Cy5 CPK6 50mer (see below)	IDT	250nmole	Custom oligo synthesis
MAUI Mixer SL	BioMicro	12 Lids	02-A008-03
Labchip RNA or DNA Reagent and Supply kit	Agilent	25 Chips	5065-4476
NimbleGen Array Reuse Kit 40	NimbleGen	40 Rxn	KIT001-2
NimbleGen Hybridization Kit 40 – CGH & ChIP	NimbleGen	40 Rxn	KIT002-2
NimbleGen Hybridization Kit 40 Refill – CGH & ChIP	NimbleGen	40 Rxn	KIT005-2

Oligo Ordering Information

Source: Integrated DNA Technologies (800-328-2661)

Scale: 250 nmole, HPLC purified

3' Cy3 CPK6-48mer:

5' TTC CTC TCG CTG TAA TGA CCT CTA TGA ATA ATC CTA TCA AAC AAC TCA 3'

3' Cy5 CPK6-48mer:

5' TTC CTC TCG CTG TAA TGA CCT CTA TGA ATA ATC CTA TCA AAC AAC TCA 3'

Supplemental Protocol 3: Lab 4, Myers Lab

Richard M. Myers Lab Unamplified Spike-in Protocol

For each replicate array, we labeled 1µg of spike-in sample and 1µg of control sample. The Bioprime Plus Array CGH labeling Module (Invitrogen Catolog number 18095-014) was used. The spike-in was labeled with the red channel dye (Alexa Fluor-647) and the control was labeled with the green channel dye (Alexa Fluor-555) in two of the replicates. A dye swap was performed for the third replicate. These samples were then competitively hybridized to three replicate 244k arrays from Agilent, (AMADID 25150451). We hybridized the samples to the arrays using the Agilent Array CGH protocol, with some modifications. Briefly, labeled DNA was combined with Cot-1 DNA, blocking reagent, and hybridization buffer. The hybridizations were carried out at 60°C. The arrays were scanned using an Agilent dual laser scanner, and the images were processed using Agilent Feature Extraction Software.

Supplemental Protocol 4: Lab 5, Ren Lab

Protocol for Spike-in DNA Amplification

Finish Ligation, LM-PCR

Methods and Materials:

PCR Labeling Mix:

Components	1X
10X ThermoPol Reaction Buffer NEB	4uL
2.5mM dNTP	5uL
40uM oJW102	1.25uL
dH2O	4.75uL
	15uL

Procedure:

Blunt and Ligate (cont'd):

15. The next day, add 6uL of 3M NaOAc to each tube, and 130uL of 100% EtOH. Votex.
16. Freeze at -80°C for 30 min (O/N), spin 14K for 15min in cold room. Remove supernatant.
17. Wash with 200uL 70% EtOH, spin in cold room for 5min @14k rpm, remove all traces of EtOH, air dry 5 minutes (pellet turns clear).
18. Resuspend well in 25uL of cold dH2O (on thermomixer) or by hand (Make sure to completely resuspend).***.

LM-PCR:

1. Place ligated DNA on ice.
2. Add 15uL of PCR labeling mix to tube.
3. Transfer to PCR tubes on ice, warm up PCR machine, pause machine, place samples in machine for heated start.
4. Add 10uL of polymerase mix and mix by pipetting

Polymerase Mix:

Components	1X
10X ThermoPol Reaction Buffer NEB	1uL
dH2O	8uL
Taq Polymerase 5U/uL Qiagen	1uL
PFU Turbo 2.5U/uL Stratagene #600250-51	0.01uL
	10uL

5. Start PCR machine and start program
Step 1: 55°C for 2';
Step 2: 72°C for 5';
Step 3: 95°C for 2';
Step 4: 95°C for 1';
Step 5: 60°C for 1';
Step 6: 72°C for 2'; (3' if the DNA fragments are 2-3kb on average)
Step 7: go to step 4 for 22 times; (* the number of cycles is only empirically determined.)
Step 8: 72°C for 5';
Step 9: 4°C for ever;
6. Purify with Qiaquick PCR purification kit. Elute in 60uL elution buffer. OD samples.

Sample	[ug/mL]
#1a Diluted spike(1)	125
#2a Diluted control(2)	111
#1b Diluted spike(1)	95.9
#2b Diluted control(2)	111
#1c Diluted spike(1)	90.7
#2c Diluted control(2)	96.4

Protocol for Spike-in DNA Blunting and Ligation

Methods and Materials:

Blunting Mix:

Components:	1X
10X T4 DNA polymerase buffer NEB #007-203 (buffer 2)	11uL
10mg/mL BSA NEB #007-BSA	0.5uL
20mM dNTP	0.5uL
T4 DNA polymerase 3U/uL NEB #203L	0.2uL
	12.2uL

NaOAc/glycogen mix:

Components:	1X
3M NaOAc	11.5uL
20mg/mL Glycogen Roche #85343271-31	1uL

Procedure:

1. Add 20ng DNA and bring the total volume to 100uL with dH2O. (see the table for correct amounts)

Sample	[ug/mL]	vol = 20ng	Vol dH20 to 100ul	sample volume	sample mass
#1a Dilute spike(1)	3ng/ul	6.67ul	93.33ul	100ul	300ng
#1b Dilute spike(1)	3ng/ul	6.67ul	93.33ul	100ul	300ng
#1c Dilute spike(1)	3ng/ul	6.67ul	93.33ul	100ul	300ng
#2a Dilute control(2)	3ng/ul	6.67ul	93.33ul	100ul	300ng
#2b Dilute control(2)	3ng/ul	6.67ul	93.33ul	100ul	300ng
#2c Dilute control(2)	3ng/ul	6.67ul	93.33ul	100ul	300ng

2. Add 12.2 l of Blunting Mix
3. Mix by pipetting and incubate at 12°C for 20 minutes in a PCR machine.
4. Transfer blunted DNA to new 1.5mL tubes and place on ice.
5. Add 12uL of NaOAc/glycogen mix. Vortex
6. Add 120uL of phenol/chloroform/isoamyl alcohol (25:24:1, Sigma P-3803).
7. Vortex, spin 5 min @ 14k, pipet off top (carefully) and put in new tube.
8. Transfer 110uL to a new 1.5 ml Eppendorf tube and add 230 uL cold EtOH (100%). Vortex. Store at -80°C for 30 minutes. Spin for 14K rpm 15 minutes at 4°C.

(Take out linkers & thaw on ice for steps below)

9. Remove supernatant and wash the pellet with 500uL cold 70% EtOH.
10. Spin for 5 minutes at 4°C.
11. Pipet off supernatant, spin briefly and remove any remaining liquid with pipette. Allow the pellet to dry briefly (turns clear).
12. Resuspend pellet in 25uL dH20 and place on ice.**(Make sure resuspend well)**

Ligase Mix: **Mix by pipetting, the ligase is kind of fragile.**

Components:	1X
5X Ligase Buffer (Invitrogen)	10uL
15mM annealed linkers	6.7uL
T4 DNA ligase NEB #202L	0.5uL
dH2O	8uL
	25.2uL

13. Add 25uL of cold ligase mix to each tube
14.Mix by pipetting and incubate overnight at 16°C.

Protocol for ChIP-chip Hybridization

Methods and Materials:

0.5M EDTA (Sigma Aldrich 100ml E-7889)
Absolute (Ethanol Sigma Aldrich 500mL E702-3)
100mM dNTP's (Invitrogen 4 X 25umole 10297-018)
Isopropanol(Sigma Aldrich 500ml I-9516)
Klenow Fragment 3' ->5' exo- (NEB 50,000u/ml M0212M)
5M NaCl (Sigma 500ml 9759)
1M MgCl2(Sigma Aldrich 100ml M-1028)
®-Mercaptoethanol (Sigma Aldrich 100 mL 6010)
Cy3 9mer Wobble (TriLink Biotechnologies 200 O.D. N46-0001-200)
Cy5 9mer Wobble (TriLink Biotechnologies 200 O.D. N46-0002-200)
Cy3 CPK6 50mer (see below) IDT 250nmole Custom oligo synthesis
Cy5 CPK6 50mer (see below) IDT 250nmole Custom oligo synthesis
Oligo Ordering Information
Source: Integrated DNA Technologies (800-328-2661)
Scale: 250 nmole, HPLC purified
3' Cy3 CPK6-48mer:
5' TTC CTC TCG CTG TAA TGA CCT CTA TGA ATA ATC CTA TCA AAC AAC TCA 3'
3' Cy5 CPK6-48mer:
5' TTC CTC TCG CTG TAA TGA CCT CTA TGA ATA ATC CTA TCA AAC AAC TCA 3'

Prepare the following three solutions:

50X dNTP Mix
(VWR deionized water 250 uL, 100 mM dATP 50 uL ,100 mM dGTP 50 uL ,
100 mM dTTP 50 uL, 100 mM dCTP 50 uL ,Total 500 uL)
Aliquot 50X dNTP mix into single-use mounts. Avoid freeze-thaw cycles, which can result in diminished yields.

Random 9mer Buffer
(VWR deionized water 8.6ml,1M Tris-HCl 1.25ml,1M $MgCl_2$ 125 uL,®-Mercaptoethanol 17.5 uL, Total 10ml)
Dilute Cy3 and Cy5 dye-labeled 9mers to 1 O.D./42uL Random 9mer Buffer.
Aliquot to 40 uL individual reaction volumes in 0.5 ml thin-walled PCR tubes and store at –20°C.

Procedure:

Labeling

1. Assemble the following components in a 0.5ml thin-walled PCR tube (Total 80 uL) :

Sample	[ug/mL]	vol = 1ug	dH20=40ul	Cy5-9mer/cy3-9mer Primers
#1a Diluted spike(1)	125	8uL	32ul	Cy5-9mer 40uL
#2a Diluted control(2)	111	9uL	31ul	Cy3-9mer 40uL
#1b Diluted spike(1)	95.9	11.6uL	28.4ul	Cy5-9mer 40uL
#2b Diluted control(2)	111	9uL	31ul	Cy3-9mer 40uL
#1c Diluted spike(1)	90.7	11uL	29ul	Cy5-9mer 40uL
#2c Diluted control(2)	96.4	10.4uL	29.6ul	Cy3-9mer 40uL
#0a Undiluted spike(1)	68.8	14.5uL	25.5uL	Cy5-9mer 40uL
#3 Undiluted control(2)	224.9	4.5uL	35.5uL	Cy3-9mer 40uL
#0b Undiluted spike(1)	68.8	14.5uL	25.5uL	Cy5-9mer 40uL
#3 Undiluted control(2)	224.9	4.5uL	35.5uL	Cy3-9mer 40uL
#0c Undiluted spike(1)	68.8	14.5uL	25.5uL	Cy5-9mer 40uL
#3 Undiluted control(2)	224.9	4.5uL	35.5uL	Cy3-9mer 40uL

2. Heat denatures samples in thermocycler at 98°C for 10 minutes. Quick chill in ice water bath for 5 minutes.
3. Add 10uL of 50X dNTP mixes
4. Add 8uL of VWR deionized water
5. Add 2uL of high concentration Klenow (50 U/ uL). Mix well by pipetting 10 times.
6. Spin down at low RPM to force contents to bottom of tube.
7. Incubate at 37°C for 3 hours in a thermocycler protected from light.
8. Stop the reaction by adding 10 uL Stop Solution (0.5 M EDTA).
9. Transfer the reaction to a 1.5 ml tube. Precipitate the labeled samples by adding 11.5 uL 5M NaCl and 110uL isopropanol to each tube.
10. Vortex. Incubate 10 minutes at room temperature in the dark.
11. Centrifuge at maximum speed for 10 minutes. Remove supernatant with pipette.
12. Rinse pellet with 500 uL 80% ice-cold ethanol. Dislodge pellet from tube wall.
13. Centrifuge at maximum speed for 2 minutes. Remove supernatant with pipette.
14. Air dries for 5 min.
15. Rehydrate dried pellets in 13uL VWR deionized water.
16. Vortex for 30 seconds and quick spin to collect contents at bottom of tube.

Continue to vortex or let sit in the dark at room temp until the pellet is completely rehydrated, then vortex again and quickly spin.

17. Measure the concentration of each sample.

Hybridization

1. Set MAUI hybridization unit to 42°C and allow time for the temperature to stabilize.

2. Based on the measurement, combine 13ug each of the Test and Reference Samples into a single 1.5ml micro centrifuge tube.

Sample	[ug/uL]	vol. = 13ug	Slide No.
#1a Diluted spike(1)	2.092	5.7uL	90899-1
#2a Diluted control(2)	2.958	4.0uL	
#1b Diluted spike(1)	2.244	5.3uL	90931-1
#2b Diluted control(2)	1.295	9.2uL	
#1c Diluted spike(1)	2.072	5.8uL	90951-1
#2c Diluted control(2)	2.637	4.5uL	
#0a Diluted spike(1)	1.945	6.2uL	90899-2
#3 Diluted control(2)	2.891	4.2uL	
#0b Diluted spike(1)	2.268	5.3uL	90951-2
#3 Diluted control(2)	2.891	4.2uL	
#0c Diluted spike(1)	2.561	4.7uL	91991-1
#3 Diluted control(2)	2.891	4.2uL	

3. Dry the combined contents in a Speed-Vac on low heat.
4. Resuspend the sample in 10.9 uL VWR water and vortex to completely dissolve the sample. Spin the tube down briefly to collect the contents in the bottom
5. Use the NimbleGen Array Reuse Kit; add the following to the resuspended sample:

	1X
2X Hybridization Buffer	19.5 uL
Hybridization Component A	7.8 uL
Cy3 CPK6 50mer Oligo, 50nM	0.4 uL
	28.1 uL

6. Mix the tube briefly then spin down to collect the contents in the bottom and place at 95°C for 5 minutes.
7. Immediately transfer the tube to the MAUI 42°C sample block and hold at this temperature until you are ready for sample loading.
8. Place the MAUI Mixer SL Hybridization Chamber on the array using the provided assembly/disassembly jig and carefully follow MAUI setup instructions. Use the braying tool to remove all air bubbles from the adhesive gasket around the outside of the hybridization chamber.
9. Put the array and Hybridization Chamber on the MAUI and allow 30 seconds for the chips to come up to temperature.
10. Load the sample using the pipette supplied with the MAUI Station and following manufacturer instructions. During loading, a small amount (3-7 uL) of the sample may flow out of the outlet port.
11. Confirm that there are no bubbles in the chamber.
If there are, very gently massage any bubbles to either of the ends, away from the center of the array. Avoid applying too much pressure since this will force liquid out of the ports.
12. Place the loaded array into one of the four MAUI bays and let equilibrate for 30seconds. Wipe off any sample leakage at the ports with a Kim-Wipe, and adhere
13. Close the bay clamp and select mix mode B. Hold down the mix button to start mixing. Confirm that the mixing is in progress before closing the cover.10Hybridize the sample overnight (16-20 hours).

Washing of Arrays

1. Prior to removing the array from the MAUI Hybridization Station, prepare the following solutions. You will need two dishes of Wash I, and one each for Wash II and Wash III. One dish for Wash I should be shallow and be wide enough to accommodate the array and mixer loaded in the MAUI assembly/disassembly jig. The lid from a 1000 uL pipette tip box works well. Also, the buffer in the first Wash I dish should be heated to 42°C to help soften the adhesive on the hybridization chamber; this will help prevent braking the array.
Place the remaining three wash solutions in 300ml Tissue-Tek slide staining dishes.

	Wash I	Wash I	Wash II	Wash III
VWR Water	225ml	99 ml	225ml	225 ml
10X Wash Buffer I, II, or III	25ml	11 ml	25ml	25 ml
1M DTT	25 uL	11 uL	25 uL	25 uL
Total volume	250ml	110 ml	250ml	250 ml

2. Remove chip from MAUI Hybridization Station, load it back into the MAUI assembly/disassembly jig, and immerse in the shallow 250ml Wash I at 42°C. While the chip is submerged, carefully peel off the mixer. Gently agitate the chip in Wash 1 for 10-15 seconds.
3. Transfer the slide into a slide rack in the second dish of Wash I and incubate 2 minute with agitation.
4. Transfer to Wash II and incubate 1 minute with agitation. Rock the dish to move the wash over the tops of the arrays.
5. Transfer to Wash III and incubate for 15 sec with agitation.
6. Remove array and spin dry in an array-drying unit (e.g. the NimbleGen Array-Go-Round) for 1 minute.
7. Store the dried array in dark desiccators and proceed immediately to the protocol for NimbleGen Two-Color Scanning of NimbleGen Arrays.

Supplemental Protocol 5: Lab 6, Snyder Lab

FileName	85403_635_over_85403_532_norm_ratios.gff	85923_635_over_85923_532_norm_ratios.gff	86888_635_over_86888_532_norm_ratios.gff
Lab Name	Snyder	Snyder	Snyder
Array Platform	2005-03-08_ENCODE_0307_50mer	2005-03-08_ENCODE_0307_50mer	2005-03-08_ENCODE_0307_50mer
Array Resolution	38 bp	38 bp	38 bp
Probe Length	50 mer	50 mer	50 mer
Amount of DNA added to the arrays	10 _g per channel	10 _g per channel	10 _g per channel
Amount of DNA started in Klenow reaction with Cy5 or Cy3 random primers	5 ug added to start of Klenow reaction; had ~60- 70 ug post Klenow, of which 10 ug was hybridized	5 ug added to start of Klenow reaction; had ~60- 70 ug post Klenow, of which 10 ug was hybridized	5 ug added to start of Klenow reaction; had ~60- 70 ug post Klenow, of which 10 ug was hybridized
Amount of DNA started in the PCR amplification	NA	NA	NA
Amplification Method	No Amplification	No Amplification	No Amplification
PCR cycles	NA	NA	NA
Amount of Cot-1DNA added per Hyb	10 _g	10 _g	10 _g
Hybridization site	NimbleGen Iceland site	NimbleGen Iceland site	NimbleGen Iceland site
Sample Name (one color array)	NA	NA	NA
Sample Name (two color array)	Cy3: Input; Cy5: Spike-in	Cy3: Input; Cy5: Spike-in	Cy3: Input; Cy5: Spike-in
Replicate	1	2	3

Lab 5 does not amplify ChIP DNA samples prior to labeling. Our ChIP DNA is random primed with Cy-coupled random primers and Klenow. So the most analogous tubes received appeared to be the samples: ENCODE spike 77 ng/ul and ENCODE control (Input) 200 ng/ul.

1) For each array to be hybridized, NimbleGen-Iceland labeled 5 ug (65 ul) of spike with Cy5 and 5 ug (25 ul) of control/Input with Cy3. The labeling reactions were done in parallel (3 x spike and 3 x control/input) to try and mimic biological reps.
2) Post-labeling the yields were ~60 ug for each of the spike-ins and ~70 ug for each of the control/inputs, (ie so 3 x 60 ug and 3 x 70 ug). For each pair (3 mock biological replicates) of labeling reactions, 10 ug of spike and 10 ug of control/input were mixed and hybridized to each of 3 arrays. 3) 10 ug of unlabeled COT-1 DNA was added per hybridization along with herring sperm DNA, as we use these in our standard ChIP-chip protocol.

Supplemental Protocol 6: Lab 7, Struhl Lab

FileName	EC2_AS_Conc_Spike_B1_T1.CEL.bz2	EC2_AS_Conc_Spike_B2_T1.CEL.bz2	EC2_AS_Conc_Spike_B3_T1.CEL.bz2	EC2_AS_Conc_Ctrl_B1_T1.CEL.bz2	EC2_AS_Conc_Ctrl_B2_T1.CEL.bz2	EC2_AS_Conc_Ctrl_B3_T1.CEL.bz2
Lab Name	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl
Array Platform	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R
Array Resolution	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM
Probe Length	25 mer	25 mer	25 mer	25 mer	25 mer	25 mer
Amount of DNA added to the arrays	3.6 ug	3.6 ug	3.6 ug	3.6 ug	3.6 ug	3.6 ug
Amount of DNA started in the PCR amplification	NA	NA	NA	NA	NA	NA
Amplification Method	No Amp	No Amp	No Amp	No Amp	No Amp	No Amp
PCR cycles	NA	NA	NA	NA	NA	NA
Hybridization site	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab
Sample Name (one color array)	Spike-in	Spike-in	Spike-in	Input Control	Input Control	Input Control
Sample Name (two color array)	NA	NA	NA	NA	NA	NA
Replicate	1	1	1	1	1	1
Notes	These represent six independent labelings and hybridizations.					

Amplified Cel Files

FileName	EC2_AS_AY_1A_B1_T1.CEL.bz2	EC2_AS_AY_1B_B1_T1.CEL.bz2	EC2_AS_AY_1C_B1_T1.CEL.bz2	EC2_AS_AY_2A_B1_T1.CEL.bz2	EC2_AS_AY_2B_B1_T1.CEL.bz2	EC2_AS_AY_2C_B1_T1.CEL.bz2
Lab Name	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl	Gingeras/ Struhl
Array Platform	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R	Affymetrix GeneChip ENCODE 2.0R
Array Resolution	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM	7 bp, 5uM
Probe Length	25 mer	25 mer	25 mer	25 mer	25 mer	25 mer
Amount of DNA added to the arrays	3.6 ug	3.6 ug	3.6 ug	3.6 ug	3.6 ug	3.6 ug
Amount of DNA started in the PCR amplification	20 ng	20 ng	20 ng	20 ng	20 ng	20 ng
Amplification Method	Random Primed PCR	Random Primed PCR	Random Primed PCR	Random Primed PCR	Random Primed PCR	Random Primed PCR
PCR cycles	4 + 27	4 + 27	4 + 27	4 + 27	4 + 27	4 + 27
Hybridization site	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab	Gingeras Lab
Sample Name (one color array)	Spike-in	Spike-in	Spike-in	Input Control	Input Control	Input Control
Sample Name (two color array)	NA	NA	NA	NA	NA	NA
Replicate	1	1	1	1	1	1
Notes	These represent six independent amplifications, labelings, and hybridizations.					

Data files also available at: http://transcriptome.affymetrix.com/download/EC_Spike-in_Sample/2ndRound/
Map files are at: http://transcriptome.affymetrix.com/download/EC_Spike-in_Sample/ENCODE_v2_map_files/

Amplified DNA was ethanol-precipitated prior to labeling. The non-amplified spike-ins were labeled directly from the provided sample. Five microgram of DNA (amplified and un-amplified) was fragmented to an average size of 50-100 nt under the following conditions:

DNA	-	5µg
1-for-All buffer (Amersham)	-	1X
DNAse I (1Unit/1µl , Epicentre), 1:16 dilution	-	1µl
Total volume	-	20µl

Incubation was at 37^{o}C for 8 minutes, followed by boiling for 10 minutes. Digestion was checked by gel using 1µl. The samples were cooled on ice and used directly for end labeling using terminal transferase (TdT).

Fragmentation reaction	-	19µl
5X TdT reaction buffer (Roche)	-	7µl
25mM CoCl2 (Roche)	-	3.5µl
7.5 mM DNA labeling reagent (Affymetrix)	-	0.75µl
Terminal transferase (400 Unit/1µl , Roche)	-	0.5µl

Water	-	4.25µl

Labeling was conducted at 37^{o}C for 3 hours, after which samples were used directly for hybridization.

Labeling reaction	-	35µl
12X MES buffer (See below)	-	20.8µl
5M tetramethyl ammonium chloride TMAC (Sigma)	-	150µl
Herring sperm DNA 10mg/ml (Invitrogen)	-	2.5µl
5nM Oligo B2 (Affymetrix)	-	2.5µl
1% Triton X100	-	5µl
water	-	34.1µl

The hybridization cocktails were boiled for 10 minutes, allowed to cool to 45^{o}C for 10 minutes and 200 µl of it was directly applied to the ENCODE arrays. In total, 3.6 µg of labeled DNA was hybridized per array. The hybridizations were performed at 45^{o}C for 18 hours. Washes were performed using standard Affymetrix expression protocol. Scans were performed on the GCS3000 scanner (Affymetrix) with 0.7µm pixel resolution.

12X MES (1000 ml):
-70.4 g MES free acid monohydrate
-193.3 g MES-Na
-800 ml water, mix and adjust to 1000 ml

Supplemental Protocol 7: Lab 8, McCuine Lab

FileName	061017_251504510001_S01_CGH-v4_91.txt	061017_251504510002_S01_CGH-v4_91.txt	061017_251504510003_S01_CGH-v4_91.txt	061017_251504510004_S01_CGH-v4_91.txt
Lab Name	Young	Young	Young	Young
Array Platform	Agilent ENCODE 244k	Agilent ENCODE 244k	Agilent ENCODE 244k	Agilent ENCODE 244k
Array Resolution	70 bp	70 bp	70 bp	70 bp
Probe Length	60 mer	60 mer	60 mer	60 mer
Amount of DNA added to the arrays	3.5 _g	3.5 _g	3.5 _g	3.5 _g
Amount of DNA started in the PCR amplification	NA	NA	~150ng	~150ng
Amplification Method	No Amplification	No Amplification	LM-PCR	LM-PCR
PCR cycles	NA	NA	15 cycles, diluted, then 25 cycles	15 cycles, diluted, then 25 cycles
Hybridization site	Young lab - Whitehead Institute	Young lab - Whitehead Institute	Young lab - Whitehead Institute	Young lab - Whitehead Institute
Sample Name (one color array)	NA	NA	NA	NA
Sample Name (two color array)	Cy3: Concentrated Input (Sample 3); Cy5: Concentrated Spike-in (Sample 0)	Cy3: Concentrated Input (Sample 3); Cy5: Concentrated Spike-in (Sample 0)	Cy3: Diluted Input (Sample 2A); Cy5: Diluted Spike-in (Sample 1A)	Cy3: Diluted Input (Sample 2B); Cy5: Diluted Spike-in (Sample 1B)
Replicate	1	2	1	2

ENCODE Spike-Ins Project Summary – Whitehead Institute

Experimental Methods – Concentrated Samples

Both concentrated samples – the input DNA (tube 3), and the input plus spiked-in DNA (tube 0) – were processed identically.

Upon receipt, concentrated samples were measured using a NanoDrop ND-1000 spectrophotometer to ensure that DNA quality/quantity was not adversely affected by shipping conditions.

DNA from each of these samples was directly labeled in duplicate reactions (2ug input per reaction) as outlined in the Agilent Mammalian ChIP-on-chip manual (fifth edition, August 2006). Labeling was performed using a BioPrime random primer labeling kit (Invitrogen); input/control DNA was labeled with the Cy3 fluorophore, while the input plus spiked-in DNA was labeled with the Cy5 fluorophore. Reactions were cleaned up on the purification spin columns supplies with the BioPrime kit. Labeled and purified DNA from one of each of the replicate labeling reactions (e.g. one Cy3 and Cy5 sample) was combined (3.5ug each channel) and hybridized to Agilent arrays (AMADID 015045) using an Agilent aCGH hybridization kit and Human Cot-1 DNA (Invitrogen). Arrays were hybridized in Agilent chambers in an Agilent oven for 40 hours at settings of 65°C and 10rpm. Arrays were washed for five minutes at RT in aCGH wash buffer 1, one minute in aCGH wash buffer 2 pre-warmed to 31C, one minute in acetonitrile, and finally for 30 seconds in Agilent Stabilization and Drying solution. Arrays were then scanned on an Agilent scanner using the extended dynamic range (XDR) feature, with scans at PMT values of both 100 and 10 per array. Array images were extracted and processed with Agilent's Feature Extraction software (v9.1).

Experimental Methods – Diluted Samples

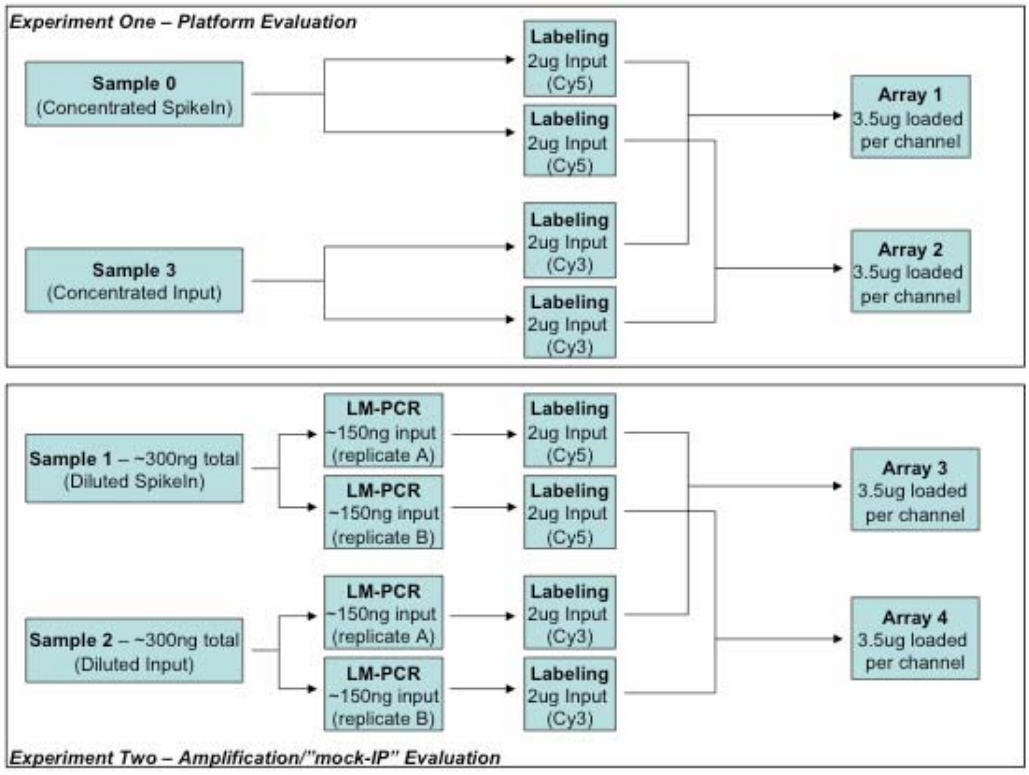
As with the concentrated samples, both diluted samples – the input DNA (tube 2), and the input plus spiked-in DNA (tube 1) – were processed identically.

Samples were too dilute to reliably measure upon receipt using a NanoDrop ND-1000 spectrophotometer; expected values of 300ng per sample relayed from David Johnson were used as the base estimate for DNA quantity per sample, assuming no degradation during shipping.

DNA from each of these samples (150ng per reaction) was blunted, ligated to a universal linker and amplified using a two-stage LM-PCR protocol outlined in the Agilent Mammalian ChIP-on-chip manual. Amplifications were done in duplicate, parallel experiments (resulting in samples 1A and 1B from tube 1, and in samples 2A and 2B from tube 2). Amplified DNA from each of the duplicate amplification reactions (2ug per reaction) was directly labeled and purified in single labeling reactions. Array hybridizations were performed as above and set up so that

3.5ug of purified, labeled Sample 1A was hybridized against 3.5ug of purified, labeled Sample 2A on one array, and 3.5ug of purified, labeled Sample 1B was hybridized against 3.5ug of purified, labeled Sample 2B on a second array. Arrays were processed through the array scanning and data extraction stages as detailed above.

ENCODE Spike-Ins Project Summary – Whitehead Institute Experimental Set-up



ENCODE Spike-Ins Labeling

October 14, 2006

BioPrime Labeling Module Lot:	1350048
Cy3 Lot:	313207N
Cy5 Lot:	318829
BioPrime Purification Module Lot:	1333164

1. Prepare DNA Master Mixes

Cy3 Sample:	**Concentrated Input (3)**		
DNA Concentration	244.2 ng/ul		
	1x	2x	
DNA/LM-PCR Product	8	16	ul
2.5x Random Primers	35	70	ul
H2O	32	64	ul
Total	75	150	ul

Cy5 Sample:	**Concentrated SpikeIn (0)**		
DNA Concentration	90.8 ng/ul		
	1x	2x	
DNA/LM-PCR Product	22	44	ul
2.5x Random Primers	35	70	ul
H2O	18	36	ul
Total	75	150	ul

Cy3 Sample:	**Diluted Input (2A)**	
DNA Concentration	102 ng/ul	
	1x	
DNA/LM-PCR Product	20	ul
2.5x Random Primers	35	ul
H2O	20	ul
Total	75	ul

Cy5 Sample:	**Diluted SpikeIn (1A)**	
DNA Concentration	97 ng/ul	
	1x	
DNA/LM-PCR Product	21	ul
2.5x Random Primers	35	ul
H2O	19	ul
Total	75	ul

Cy3 Sample:	**Diluted Input (2B)**	
DNA Concentration	87 ng/ul	
	1x	
DNA/LM-PCR Product	23	ul
2.5x Random Primers	35	ul
H2O	17	ul
Total	75	ul

Cy5 Sample:	**Diluted SpikeIn (1B)**	
DNA Concentration	104 ng/ul	
	1x	
DNA/LM-PCR Product	19	ul
2.5x Random Primers	35	ul
H2O	21	ul
Total	75	ul

2. Aliquot 75ul into separate tubes
3. Incubate at 95C for 5 minutes
4. Immediate Transfer to ice-water bath for 5 minutes
5. Make Cy3 and Cy5 Mixes

Cy3 Mix	1x	6x	
10x dUTP Nucleotide Mix	8.2	49.2	ul
Cy3-dUTP	1.5	9	ul
Klenow (40U/ul)	1.5	9	ul
H2O	1.8	10.8	ul
Total	13	78	ul

Cy5 Mix	1x	6x	
10x dUTP Nucleotide Mix	8.2	49.2	ul
Cy5-dUTP	1.5	9	ul
Klenow (40U/ul)	1.5	9	ul
H2O	1.8	10.8	ul
Total	13	78	ul

6. Add 13ul aliquots of the appropriate labeling mixes to the DNA samples
7. Incubate at 37C for 3 hours
8. Add 9ul stop buffer to each tube and mix
9. Clean up samples on Invitrogen's columns (supplied in kit)
10. Elute samples in 60ul H2O

ENCODE Spike-Ins Array Hybridizations

October 16, 2006

Agilent CGH Hybridization Kit Lot: GK0196
Human Cot-1 DNA Lot: 1325007

1. Prepare DNA Master Mixes

Array 1:	**Concentrated Input (3) vs. Concentrated Spikeln (0)**		**Array 2:**	**Concentrated Input (3) vs. Concentrated Spikeln (0)**	
Cy3 Sample	89.6	ng/ul	Cy3 Sample	94.1	ng/ul
Cy5 Sample	66.6	ng/ul	Cy5 Sample	63.7	ng/ul
	1x			1x	
Cy3 Sample (3.5ug)	39	ul	Cy3 Sample (3.5ug)	37	ul
Cy5 sample (3.5ug)	51	ul	Cy5 sample (3.5ug)	55	ul
H2O	60	ul	H2O	58	ul
Total	150	ul	Total	150	ul

Array 3:	**Diluted Input (2A) vs. Diluted Spikeln (1A)**		**Array 4:**	**Diluted Input (2B) vs. Diluted Spikeln (1B)**	
Cy3 Sample	103.2	ng/ul	Cy3 Sample	107.8	ng/ul
Cy5 Sample	100.4	ng/ul	Cy5 Sample	102.2	ng/ul
	1x			1x	
Cy3 Sample (3.5ug)	34	ul	Cy3 Sample (3.5ug)	32	ul
Cy5 sample (3.5ug)	35	ul	Cy5 sample (3.5ug)	34	ul
H2O	81	ul	H2O	83	ul
Total	150	ul	Total	150	ul

2. Aliquot 50ul Agilent 10x Controls into each tube and mix
3. Aliquot 50ul Human Cot-1 DNA Controls into each tube and mix
3. Incubate at 95C for 3 minutes
4. Immediate transfer to 37C for 30 minutes
5. Spin down reactions at 17,000g for 1 minute
6. Pipette samples onto slides, and assemble hybridization chambers
7. Hybridize at 65C for 40 hours

Preparation and validation of the simulated ChIP sample

The simulated ChIP sample was prepared by mixing human genomic DNA (Roche) with plasmid clones (Promega pGL3 vector) containing human DNA sequences (Cooper et al., 2007). At least 10µl of liquid was transferred in each step of the dilution series, with the intention of reducing shot noise during transfer. The mixtures were all sonicated on ice according to standard chromatin sonication procedures (Johnson et al., 2007). The simulated ChIP sample was validated in three ways: (1) sequencing of the original clone preps before dilution, (2) sequencing of the diluted clones with PCR preamplification using universal primers, and (3) insert specific PCR of the diluted clones, followed by agarose gel electrophoresis. Our experimental validation revealed no anomalies in the spike-in mixtures, and our analysis of the array predictions adds extra evidence that the libraries were mixed at the proper stoichiometries and that the clone identities were correct.

All undiluted control samples were supplied at 200ng/ul. Participating labs were instructed to add the DNA quantity specified in their respective protocols.

Bootstrapping estimate of AUC significance

We used bootstrap re-sampling method to assess the statistical variance associated with our evaluation metric AUC of each prediction.

The bootstrap involves repeated estimation of AUC using random samples with replacement from the predicted sites in the original ROC-like curve. Please note that we only re-sampled from the top N sites with the ratio of called false positives over total number of true positives less than 10% (as shown in the original ROC-like curve), not from all the sites in the original prediction. Because the re-sampling is with replacement, some sites were selected more than once while others were not selected at all. We repeated this procedure many times to get a thousand AUCs for each prediction. The AUC standard deviation of each prediction ranges from 0.02 to 0.12 with the average of 0.07. Therefore, readers of this paper are advised to be cautious when comparing with the AUCs of different predictions, as small AUC differences may not reflect statistically significant differences.

Description of analysis algorithms

MAT

Model-based Analysis of Tiling arrays first standardizes each individual Affymetrix tiling arrays by modeling the effect of probe's 25-mer sequence and genome copy number on its signal as follows:

$$log(PM_i) = \alpha n_{iT} + \sum_{j=1}^{25} \sum_{k \in \{A,C,G\}} \beta_{jk} I_{ijk} + \sum_{k \in \{A,C,G,T\}} \gamma_k n_{ik}^2 + \delta log(c_i) + \varepsilon_i$$

where

- PM_i is the perfect match probe value of probe i
- n_{ik} is the nucleotide k count in probe i
- α is the baseline value (intercept or constant) based on the number of T's on the probe. E.g. 25α is the baseline when the probe sequence is a run of 25 T's
- I_{ijk} is an indicator function such that $I_{ijk} = 1$ if the nucleotide at position j is k in probe i, and $I_{ijk} = 0$ otherwise
- β_{jk} is the effect of each nucleotide k (except T which is already modeled in α) at each position j
- γ_l is the effect of nucleotide count squared
- c_i is the number of times probe i's sequence appears in the genome. Affymetrix tiling array libraries provide the 25-mer sequence of every probe, which we mapped to the non-repeat-masked newest (May 2004) version of the human genome assembly.
- δ is the effect of the log of the probe copy number
- ε_i is the probe-specific error term, assumed to follow a normal distribution

With the model, probes with similar 25-mer sequences and copy numbers (thus similar hybridization affinities) are grouped in bins and normalized together as:

$$t_i = \frac{log(PM_i) - \hat{m}_i}{s_{i\ affinity\ bin}}$$

where $\hat{m}_i$ is the baseline intensity predicted by the model based on the sequence and copy number of probe i, and $s_{i\ affinity\ bin}$ is the standard deviation of the affinity bin to which probe i belongs.

After probe standardization, a window sliding approach is applied to the spike-in replicates and control replicates separately. Within each window, a MATscore is calculated as:

$$MATscore(region) = \sqrt{n_p} \times TM(t's\ in\ region)$$

where TM is the trimmed-mean (remove probes with the top and bottom 10% signals and average the rest) of all the probe t's in the region, and n_p is the number of observation points in the region used to calculate the TM. Spike-in regions are predicted if the MATscore in the spike-in samples are much higher than that in the control samples.

TiMAT

Processing of the Affymetrix microarray data was performed in three basic steps using TiMAT (http://sourceforge.net/projects/timat2): data normalization, sliding window summaries, and enriched region identification.

All cel file perfect match intensities were group together, quantile normalized and individual replicas median scaled to 100. A sliding window of 275bp was advanced across each chromosome. Windows containing > 17 oligos were scored by first calculating all the ratio pairs between the spike-in sample replicas and the control input sample replicas and second, taking a pseudo median of these window ratios. To estimate confidence in the window scores, Richard Bourgon's symmetric null R application was used to assign a p-value to each window score. John Storey's q-value R application was then applied to correct for multiple testing and assign FDRs to each p-value. Overlapping high scoring windows (< 5% FDR) were merged into enriched regions and then ranked by their best window score.

TAS

Affymetrix Tiling Array Software first uses quantile normalization to normalize probes on all the arrays. Then a Mann-Whitney U test (also known as Wilcoxon rank-sum test) is used across 500bp sliding windows to identify windows where the spike-in probes has higher signals than the control probes. All the probes (all replicates in controls and spike-in samples) in each window are ranked based on probe signal (higher signals receive lower rank), and the sum of ranks (U) in spike-in and sum of ranks in control are calculated. Spike-ins windows are identified if the U in the spike-in is much lower than the U in the control.

Weighted Average (WA)

To detect enriched regions, we used an approach that judged the significance of ratios of a contiguous set of probes defining a region by comparing a score based on their weighted average to the distribution of scores of all sets of probes taken in windows of the same predefined size (500bp in this case.)

Probe ratios were computed from lowess-normalized median probe intensities if the "core" pixes of each feature ("cookie cutter.") The score for each windowed region was computed by first taking an average of probe ratios in which high ratios were down-weighted, and then by multiplying this average by a factor that penalizes regions with high variance, as well as regions containing fewer probes.

The weighting factor w_i for each probe ratio was computed according to Eq 1, below.

$$w_i = \frac{1}{1 + abs(\log_2 r_i)} \quad (1)$$

Where r_i was the lowess-normalized ratio of each probe. In the case that replicate arrays were analyzed together, probes from different replicates were considered as distinct. The weighted average was therefore:

$$\bar{r} = \frac{\sum_{i=0}^{k} r_i w_i}{\sum_{i=0}^{k} w_i} \quad (2)$$

Where *i* was the index of the probe within the region (and across replicates), and *k* was the total number of probe within the region. The largest ratio within each group of probes was discarded from the weighted average.

The standard deviation was used for computing the score, and it was calculated without weights:

$$\sigma = \sqrt{\sum_{i=0}^{k} \frac{(\bar{r} - r_i)^2}{k}} \quad (3)$$

Then the score (*s)* was computed according to Eq 4:

$$s = \bar{r} \cdot 2^{\frac{\sqrt{\frac{k}{n}-1}}{\sigma+1}} \quad (4)$$

Here, *n* was the number of replicate arrays used in the analysis. The quantity *(k/n-1)* is small when there are few probes, so it penalized sparse regions. Similarly, the quantity $(\sigma+1)$ was large in noisy regions, and it served to penalize them. The penalties were raised to the power of 2 because the statistics on the scores were analyzed in $\log_2$-scale.

To determine the significance of a region, the distribution of $\log_2(s)$ aws computed over all regions within an ENCODE segment. P-values were computed for each score assuming that the distribution is nearly Gaussian. For regions found to be significant ($P < 0.001$), the "true," unweighted average ratio was computed and reported as the predicted ratio.

Splitter

A web implementation of the Splitter algorithm was developed to process the outputs of ChIP-chip experiments. It predicts the genomic regions that are bound by a protein on which a ChIP-chip experiment was performed. ChIP-chip stands for chromatin immunoprecipitation (ChIP) followed by detection with oligonucleotide tiling arrays (chip). These regions are commonly referred to as ChIP-chip hits. The main advantage of Splitter is that it is able to determine a cutoff for the ChIP-chip intensity automatically, to separate signal (hits) from background noise. In addition, the web server allows the user to perform standard preprocessing steps, such as normalization and replica combination, to facilitate the determination of cutoff and hits.

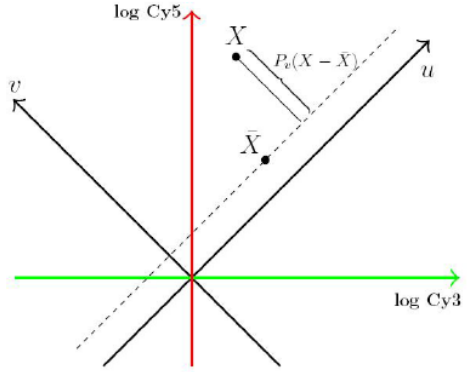
The input file required by the Splitter server consists of the coordinates of the probes on the tiling array and their intensities in each replica. The intensity cutoff can be defined statistically by percentile or standard deviation, or dynamically by the Splitter algorithm.

Parameters for Splitter include an intensity range that will be divided by the total number of splits, which results in the step size to search for the best cutoff, i.e., the cutoff value will be increased from the lower bound of the range by the step. Splitter compares the total numbers of hits before and after each increment. If the ratio between the number of hits at the current level and the number of hits at the next level is smaller than a "break-ratio", all hits at the current level are reported as the final output. Clusters of probes that are located closer than the gap parameter ("maxgap") are merged into one cluster. Clusters of probes that have number of probes less than the clustering parameter ("minrun") are marked as noise and discarded without further consideration.

For the Spike-in experiment the next parameters were selected: “replica combination” – mean; “from” – 0; “to” – initial guess about histogram separation based on visual observation; “splits” – 5, except Agilent, Myers and Ren_dil where 10 was used ; “break ratio” – 2; “minrun” – 2; “maxgap” – 200.

MA2C

Model-based Analysis for 2-Color arrays normalizes a probe’s log (Cy5 / Cy3) ratio by checking against all other probes with the same GC count. Within each GC group, MA2C first subtracts the baseline from log intensity vectors and then projects the adjusted vectors onto v-axis, yielding log mean-scaled ratios of the Cy5 and Cy3 signals within each GC bin. Finally, the projected values are adjusted for variance, so that probes in bins with higher variance (i.e. low GC bins showing low correlation between Cy5 and Cy3 channels) are given lower weight.



After normalization, each 500 bp sliding window is examined, and a median normalized probe is identified for each window. Windows with high median probe values are predicted as spike-ins.

Aberration Detection Method (ADM)

For the ENCODE Spike-in data we applied the algorithm that we created to detect changes in copy numbers along the genome. This algorithm is sensitive to small variations in copy number along the genome. The method is described briefly in the following paragraphs.

Prior to detecting any variation in copy number, a proper normalization method must correctly identify the regions of log ratio zero, and the error model must correctly evaluate the significance of deviations from log ratio zero. Under the reasonable assumption that most pairs of consecutive probes are probing regions having the same copy number (i.e. that the number of probes exceeds the number of breakpoints), the true copy numbers of the targets of most pairs of consecutive probes are expected to be strictly equal. Therefore, variations in the measured log ratios of consecutive probes result from noise, and a good estimate of the probe-to-probe noise can be computed from the robust standard deviation of differences between log ratios of consecutive probes. We report this error estimate as the dLRsd (Derivative of Log ratio Spread) for an array. The dLRsd is used in estimating the significance of observed deviations of log ratios from zero.

Assigning a statistical "deviation score" to an interval requires the assumption of a zero value for the data - the mean around which we expect the data to be distributed when no change in copy number is present. This specific normalization adds a constant to the $\log_2$ ratios such that the median reported log ratio of probes targeting diploid (or otherwise "normal") regions of the sample, after correction, is exactly zero[1]. This normalization correction step is referred to as centralization. To estimate the value of the constant to be added, a *centralization curve* is generated, which shows the distribution of $\log_2$ ratios of regions of the genome.

The next step is to identify all regions of statistically significant copy number changes, using ADM (aberration detection module) algorithms[2]. The ADM algorithms identify genomic regions with copy number differences between the sample and the reference based on $\log_2$ ratios of fluorescent signals from probes in the interval. In brief, ADM algorithms use an iterative procedure to rank all genomic intervals according to the mean $\log_2$ratios of probes in the intervals. Intervals with mean $\log_2$ratios greater than a specified threshold are reported as copy number changes. At each iteration, the region with the most significant score is reported.

The statistical score in the ADM model, for each interval I is calculated as:

$$S(I) = \frac{h}{\sigma_I}$$

where h is the average log ratio of all probes in the interval, and σ_I is the estimated error of the interval. See Ref 2 for further details.

[1]Doron Lipson PhD Thesis,http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-get.cgi/2007/PHD/PHD-2007-05.pdf

[2]Lipson, D, Aumann, Y, Ben-Dor, A, Linial, N, and Yakhini, Z. "Efficient Calculation of Interval Scores for DNA Copy Number Data Analysis", Proceedings of RECOMB '05, LNCS 3500, p.83, Springer-Verlag, 2005. Also in Journal of Computational Biology, Vol.13, No. 2: 215-228, 2006.

TAMAL

TAMAL is based on the previously published TAMALPAIS algorithm (Bieda et al., 2006, Genome Res.16:595-605). The TAMAL algorithm proceeds in two basic steps. First, peaks are found using the TAMALPAIS. Then, the enrichment is estimated within the peak by using the maxfour approach described in Krig et al. (2007, J Biol Chem 282:9703). Bieda et al. (2006) describe four levels of stringency, called L1, L2, L3, L4, with L1 being the most stringent set of detection parameters and L4 the least stringent. For TAMALg, the L2 and L3 peak sets are merged and the enrichment is calculated. For TAMALs, the L1 peak set is used for the peak detection.

NimbleGen's Permutation

Peaks are determined using a sliding window with probe-height (represented by score of the probe, typically log base 2 ratio) cutoff and in-window-number cutoff. If the height of a probe is above probe-height cutoff, this probe is a potential peak probe. If the number of potential peak probes in the window is equal to or larger than in-window-number cutoff, a peak is recognized. As the window slides along the data track, the recognized peak can extend. Peak detection is repeated with decreasing probe-height cutoff values. The probe-height cutoff value with that a peak is first recognized is used as score of the peak. A parameter **p** is used to represent probe-height cutoff, which is directly proportional to probe score (log base 2 ratio) and is scaled to be a value between 0 and 100. Larger number means more stringent peak recognition.

Width of sliding window: Size of the sliding window.
p_start: Value of parameter **p** for the peak detection to start with
p_step: The decrement of parameter **p** for the peak detection to proceed (go from most to least stringent)
n_step: The number of steps for peak detection to go through

The default values of the above parameters are 90, 1 and 76. Under this setting peak detection will be done 76 times. Each time the sliding window will scan the whole data track, using the parameter **p** of 90, 89, 88, … 15. Peaks recognized during the first scan will be in the category of **p**=90, peaks recognized during the 2nd scan but not recognized during the 1st scan will be in the category of **p**=89, … peaks that are recognized only during the last scan will be in the category of **p**=15. This categorization is used in FDR calculation.

Min probe > cutoff in peak: The in-window-number cutoff, when there are probes in the window that are below the probe-height cutoff.
When all probes in peak > cutoff: The in-window-number cutoff used when all probes in the window are above probe-height cutoff.

FDR are calculated by applying the same peak detection procedure to simulated noise data that are made by scrambling the original data. Ratio of the average number of peaks recognized across 20 scrambled data tracks to the number of peaks recognized in the

original data track in a detection scan with probe-height cutoff parameter **p**=p is used as FDR for peaks fall in the category of **p**=p.

Permute data with: track or chip. This indicates whether or not data points of a track are still within that track after scrambling. A track typically consists data points regarding a particular sample pair for a particular chromosome or genome region. Data of a chip typically form multiple tracks. If data characteristics are different from track to track then scrambling should be done by track.
Calculate FDR within: track or chip. The number of peaks before and after scrambling could be counted by track or by chip. FDR is then determined accordingly. Again, if data characteristics are different from track to track then the counting should be done by track.
FDR estimator: 0, 1 or 2. FDR is determined by number of peaks of the same category before and after permutation. For 0, category is determined solely by value of p under that the peak is just detected, as is described above. For 1 and 2, category is determined by peak score as well as number of probes in the peak (1), or number of probes in the peak that are above probe-height cutoff associated to the peak (2).

TileScope

Tilescope is a fully integrated data processing pipeline for high-density tiling array data analysis. Tilescope first performs intra- and inter-slide scaling using quantile normalization, the results of which are then integrated using a sliding window approach, with a window of size 400 bp in genomic space used to integrate neighboring probes from replicate arrays. For each window centered at the genomic coordinate of each oligonucleotide probe, both the pseudo-median signal as well as a p value measuring the likelihood that the region was bound by the transcription factor were calculated. The pseudomedian was represented by the median of pair-wise averages of the log2 ratio of test to reference signals for all oligonucleotide probes within the window, while the p value was calculated using a Wilcoxon paired signed rank test comparing test signal against reference signal for all oligonucleotide probes in the window. Regions corresponding to putative spike-in regions were then determined using a threshold for both the signal and p value, as well as requiring the region have MaxGap (the maximum gap allowed between probes above threshold) and MinRun (the minimum length region with probes above threshold).

Wavelet

The Wavelet algorithm uses wavelet transform of the signals from the red and green channels of the tiling array. From the approximation coefficients of the wavelet transform we obtain clear intensity and length-scale separation between the background signal and the signal coming from the regions of the biochemical activity. Non-parametric thresholding is applied to the log2 ratio of the approximation coefficients of the wavelet transform at different resolution levels in order to delineate spike-in regions at the same confidence level for all the relevant length-scales.

MPeak

The model-based Mpeak method is used to identify peaks in ChIP-on-chip data. However, we used the SIMPLE_CLUSTER option of Mpeak to identify regions of enrichment. Log2 Ratios of the signal intensities were first normalized using the loess method from the limma package within the R Project for Statistical Computing. Probe values and positions were converted to GFF format and Mpeak was run using the SIMPLE_CLUSTER option instead of the model based method. Essentially, clusters were identified which met the following conditions:

1. each probe intensity was greater than or equal to the a THRESHOLD (defined as the mean of the data + some multiple of the standard deviation of the data) (2 sd)
2. the gap between probes was less than or equal to MAXGAP (500 bp)
3. the number of probes in a cluster was greater than or equal to MINPROBES (4 probes)

Because the number of identified clusters was only 18, the Mpeak parameters were relaxed to achieve > 100 hits.

Consequently, the following criteria were required to exclude noise:

1. a cluster size < 600bp,
2. M >= 1.04,
3. R >= 9

where

M = max log ratio / median log ratio of hits
R = max log ratio / median of background log ratio

(From http://www.stat.ucla.edu/~zmdl/mpeak/)

Zheng M, Barrera LO, Ren B, Wu YN. ChIP-chip: data, model, and analysis. Biometrics. 2007 Sep;63(3):787-96.