PART 3

The Zebrafish Genome and Mapping Technologies

CHAPTER 13

Single Nucleotide Polymorphism (SNP) Panels for Rapid Positional Cloning in Zebrafish

Matthew D. Clark[*], Victor Guryev[†], Ewart de Bruijn[†], Isaac J. Nijman[†], Masazumi Tada[‡], Catherine Wilson[§], Panos Deloukas[*], John H. Postlethwait[§], Edwin Cuppen[†] and Derek L. Stemple[*]

[*]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[†]Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

[‡]Department of Cell and Developmental Biology, University College London, Gower Street, London, UK

[§]Institute of Neuroscience, 1254 University of Oregon, Eugene, OR, USA

Abstract
I. Introduction
II. Extraction of Zebrafish Genomic DNA
  A. Adult Fin Clips
  B. Zebrafish Larvae
III. Affymetrics Molecular Inversion Probe Panel
  A. SNP selection
  B. MIPs Protocol
  C. Strain Polymorphism
  D. Genetic Mapping
  E. Pooling Tests
  F. Bulk Segregant Mapping
IV. Suggested Mapping Strategy
V. General Considerations
  A. Genetic Diversity
  B. Usefulness of the SNP Panel
References

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00013-6

Abstract

Despite considerable genetic and genomic resources the positional cloning of forward mutations remains a slow and manually intensive task, typically using gel based genotyping and sequential rounds of mapping. We have used the latest genetic resources and genotyping technologies to develop two commercially available SNP panels of thousands of markers that can be used to speed up positional cloning.

I. Introduction

Even with a dense map of 3842 simple sequence length polymorphisms (SSLPs) on the MGH meiotic panel (Shimoda *et al*., 1999), 4073 predominately coding markers on the doubled haploid mapping panel (Woods *et al*., 2005) and a near complete genome sequence (http://www.sanger.ac.uk/Projects/D_rerio/) the positional cloning of mutations can still be time consuming. SSLPs while highly polymorphic between strains do not lend themselves to a highly parallel approach, and only a small subset are useable as agarose scorable markers rather than on radioactive polyacrylamide gels (Geisler *et al*., 2007). The T51 Radiation Hybrid (RH) panel provides the densest and highest resolution map (Geisler *et al*., 1999). However, doubled haploid and RH panel markers are, however, based largely on coding sequences and as such are not as polymorphic as SSLP markers. Furthermore, in assembling the genome we have noted that the T51 panel, while accurate over short distances, disagrees with meiotic panels and the physical maps over longer distances. It is possible that the AB9 cell line used to make both RH panels (LN54 and T51) (Hukriede *et al*., 1999; Kwok *et al*., 1998) had undergone rearrangements. Indeed a large part of the improvements in genome assembly from Zv8 onwards were due to the implementation of a marker weighting scheme that took this into account, giving more weight to meiotic map placements than to RH map placements (see http://tinyurl.com/5trm9uz). Thus while there are significant marker resources, there is a continuing need for more markers, especially those that are polymorphic, meiotically mapped, and are amenable to high throughput genotyping, both for better genome assembly and to accelerate the cloning of mutants.

Well over 650,000 candidate SNPs have been mined from sequence traces from the Washington University Zebrafish EST project (Clark *et al*., 2001) and the Wellcome Trust Sanger Institute's zebrafish genome sequence, with confirmed polymorphism rates of 65–86% (Bradley *et al*., 2007; Guryev *et al*., 2006). In addition to providing many more markers, high throughput SNP genotyping technologies with over one million markers typed in parallel (Frazer *et al*., 2007) (see also http://tinyurl.com/66knryp) are now widely available. Most researchers have access to these genotyping platforms through core facilities, or commercial services. A previous Zebrafish SNP microarray contained only 599 SNPs (optimized for the C32 and SJD strains) at 234 unique map positions with a mean sex averaged distance between markers of 9.8 cM (Stickney *et al*., 2002) but required many PCR amplifications before pooling and SNP detection, and was not widely adopted.

We have constructed a SNP panel where all 3212 SNPs are assayed simultaneously, using molecular inversion probes (MIPs) technology (Hardenbol *et al*., 2005). A single experiment can be used to screen all 3212 SNPs at once for coarse mapping. In addition we have designed and validated a subset of 1072 SNPs, which can be used individually as PCR based assays using an allele specific amplification method known as KASPar (Cuppen, 2007). Both panels are commercially available from Affymetrix (MIPs) and KBiosciences (KASPar), respectively. Data files for the SNP assays described here are available from the Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv).

Using these panels we have characterized the genotypes and genetic relationships of nine common zebrafish strains, have more than doubled the number of confirmed zebrafish SNPs, and added 971 markers to the Heat Shock genetic map (Kelly *et al*., 2000; Woods *et al*., 2005). Finally we have also used these methods to map a mutation with the commonly used bulk segregant approach. We believe that with this commercially available SNP panel, a single researcher could roughly (5-10cM resolution) map up to 96 mutations a week and identify SNPs polymorphic in their cross for finer mapping on individuals, for example by KASPar.

II. Extraction of Zebrafish Genomic DNA

MIPs assays use more concentrated DNA than standard genotyping PCRs, thus any contaminants would also be more concentrated. For MIPs, we use a phenol chloroform isoamylalcohol (PCI) extraction step to make cleaner DNA. For KASPar, since less DNA is needed per reaction, we have used standard proteinase K digestions, heat killing the proteinase and diluting before use, though we obtain better results after precipitation and re-suspension. Both protocols can easily be carried out in 96-well microtiter plates.

A. Adult Fin Clips

This protocol scales well in our hands and has better yields than kits we have tried. Using a 96-tip pipetting robot (e.g., Agilent Bravo or Robbins Hydra) allows easy removal of the supernatant from above the phenol phase. This can also be achieved by multi-channel pipette. Starting at an outer row, and using bevelled tips, judge by eye (through the plate wall) the level of the phenol, and move the tips ~2 mm above this to pipette the supernatant – now note the position of the plate edge relative to the tips, and use the same tip height for rows where you cannot view the tips and meniscus through the plate wall.

Briefly, caudal fin clips are collected into 2 ml Corning 96 Well Clear V-Bottom 2 ml Polypropylene Blocks (Corning Costar Catalogue # 3961) containing 400 μl of ProtK Buffer: (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.2% SDS, 5 mM EDTA pH 8.0 and 100 μg/ml proteinase K). After heat-sealing the plates (Agilent PlateLoc Thermal Plate Sealer), fins are digested overnight at 55 °C floating in a large water bath. The proteinase K is heat inactivated by 30 min at 80 °C in the water bath. The

plate is cooled to room temperature, and 400 μl of PCI is added to each well, the plate is re-sealed and mixed by inversion. After 10 min at 4000 rpm, 300 μl of the supernatant is added to a new Corning plate already containing 300 μl of isopropanol. Plates are re-sealed, mixed by inversion and precipitated for 30 min at 4000 rpm. The pellets are washed and spun twice for 10 min at 4000 rpm with 500 μl of 70% ethanol, the wash solution is removed by inversion, and draining on tissue paper (in these Corning plates the pellets stick well) finally the pellet is air dried including a final 15 min at 70 °C in a hot air oven to remove the last traces of ethanol, and dissolve overnight at 4 °C in 20 μl of T0.1E (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0), heat-sealed, and heated for 10 min in a 70 °C oven. DNA concentrations must be calculated using Picogreen fluorescence (Invitrogen) against standards using a Qubit (Invitrogen) or in a fluorescent plate reader (e.g., Pherastar from BMG labtech). A spectrophotometer can give inaccurate readings due to organic solvents, RNA, and protein contaminants. We routinely obtain 1–5 μg depending on fin size. The concentration is adjusted accordingly to the Picogreen readings to a final of 75 ng/μl, either by hand, or results can be fed into a pipetting robot.

B. Zebrafish Larvae

Assemble single embryos in a 96-well-rigid skirted PCR plate, for example an ABgene Thermo-fast 96 skirted (Catalogue # AB-0800), in which each well is prefilled with 100 μl of 100% methanol and stored at -20 °C. Remove the methanol when ready to prepare DNAs. Dry embryos completely on heating block (until they jump in their wells). Add 25 μl 2.0 mg/ml proteinase K in TE (10 mM Tris-HCl pH 8.0 and 1 mM EDTA pH 8.0). Digest at least 4 h at 55 °C. Heat inactivate for 30 min at 75 °C. DNA yields vary especially depending on developmental stage; for example, one 3-day fry gives about 100–150 ng DNA. For KASPar assays, 5–50 ng will be sufficient (approximately 4 μl of a 4x to 20x dilution in ddH2O). If assays are new or problematic, then DNAs can be cleaned by precipitation in plates by the addition of 2.5 μl 3 M sodium acetate pH 5.2, and 20 μl isopropanol, mixed by pipetting, precipitated for 30 min at 4 k rpm, washed with 100 μl of 70% ethanol, wash solution is drained off by inversion and traces removed by heating in a PCR block for 2 min at 70 °C, the pellet is dissolved in 10 μl of T0.1E (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA pH 8.0 to a final concentration of ~10 ng/μl) and heated to 70 °C for 2 min.

For MIPs, 1 μg is needed for each bulk segregant pool. Take half the DNA from each fry after digestion, assemble mutant and sibling pools of 12–48 fry, extract the pools with an equal volume of PCI (Phase Lock Gel Tubes from Eppendorf or MaXtract tubes from Qiagen make this easier) and then precipitate with 0.6 volumes of isopropanol and 0.1 volumes of sodium acetate pH 5.2, wash pellet with 1 ml 70% ethanol, air dried and re-dissolve in 20 μl T0.1E (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and heat to 70 °C for 2 min. Take 1 μl to quantify using Picogreen assay, and adjust the concentration to 75 ng/μl.

III. Affymetrics Molecular Inversion Probe Panel

The MIPs technology is a development of padlock probes (Nilsson *et al*., 1994) that allows up to 10,000 SNPs to be assayed simultaneously (Hardenbol *et al*., 2005) and is an attractive platform for the rapid design and testing of new SNP panels. In MIPs, a long probe with both ends targeting a SNP's flanking sequence and containing a unique 25 bp "barcode" is manufactured, these can then be pooled together (up to 10,000 probes) and hybridized to the same genomic sample overnight in a PCR machine. After hybridization, the 5' and 3' ends of each long oligonucleotide are designed to match the genomic sequence flanking a SNP such that the SNP base is a 1 bp gap between the ends of the oligonucleotide. At this point, the hybridized samples are split into four (all subsequent steps from hereon are carried out on these split reactions until ready to be hybridized to the Affymetrix Array), and four single base polymerase extension reactions each with just one of the four bases A, C, G, and T is carried out. In a correctly extended probe, the two ends of the probe (still hybridized on the genomic DNA) now abut, and a ligation reaction generates a circular single stranded DNA molecule. At this point all non-circular probes are destroyed by exonuclease digestion. Within the circularized probes there are two universal primer sites which point towards each other (this is the molecular inversion as before circularization they pointed away from each other and towards the ends of the DNA molecules), small amplicons (also containing the unique 25 bp barcodes) can now be PCR amplified only from circularized probes. Four amplifications are carried out on the split reactions, which were extended with the four separate nucleotides, and each amplicon is labeled with a primer specific to that base. The reactions can then be fragmented, pooled, and hybridized to the Affymetrix Universal Tag Arrays, to which the barcodes hybridize. The barcodes identify the SNP assays, and single base extensions and subsequent four-color labeling identifies the base(s) added at that position.

The MIPs technology is highly specific because of the controlled hybridization, extension by a high fidelity polymerase, and by the specificity of the ligase. More details on the MIPs technology, and the use of MIPs in genotyping project are available here: (Absalan and Ronaghi, 2007). In addition, the design requirements for the MIPs probes is very similar to that for the allele specific amplification Kaspar assay, which also uses a single base extension and PCR and maximizes the likelihood that the multiplexed MIPs assays will work as single assays across many individuals for fine mapping.

A. SNP selection

At the time of selection, there were two previous publications on Zebrafish SNPs, both identifying SNPs in gene transcripts (Guryev *et al*., 2006; Stickney *et al*., 2002). Stickney *et al*. characterized SNPs in genes already mapped by single stranded conformation polymorphism (SSCP) on the basis of polymorphisms between C32 and SJD strains. The Guryev *et al*. data set contains over 50,000 predicted SNPs by comparing EST traces from the Washington University Zebrafish EST project

(Clark *et al*., 2001), generally taken from known strain cDNA libraries (e.g., SJD, C32, and AB), to genomic sequence traces from the Sanger Zebrafish genome project (Tübingen strain). More recently a set of SNPs has been defined by comparing whole genome shotgun reads to finished clones from the zebrafish genome sequencing project (Bradley *et al*., 2007). Since the genome project is based on Tübingen strain this should largely identify intra-Tübingen SNPs and the majority will be intergenic SNPs.

In our SNP selection, we prioritized experimentally confirmed SNPs over predicted SNPs (confirmed SNPs are those that are experimentally polymorphic in at least one comparison of the AB, C32, TL, Tu, and WIK strains). We also required that SNPs could not overlap or be within 50 bp of an already placed SNP, finally we attempted to generate an even spread across the genome.

The selection consists of four SNP subsets:

1. Confirmed SNPs from Guryev *et al*. (190 SNPs).
2. Confirmed SNPs from Stickney *et al*. (66 SNPs).
3. High-quality predicted SNPs from Guryev *et al*, with each allele confirmed by two reads with a Phred quality score (Ewing and Green, 1998; Ewing *et al*., 1998) of at least 20 (1245 SNPs).
4. High-quality predicted SNP from Guryev *et al*., where one of the alleles was confirmed with only one read, that has a Phred quality over 30 (6837 SNPs).

We used all of subsets 1, 2, and 3 (1501 SNPs), and then we sampled through subset 4 to fill up to a total of 3501 and to get an even distribution of SNPs throughout the zebrafish linkage groups, filling as many gaps above 158 kbp as possible (Fig. 1). We added another set of SNPs, 20 ENU-induced STOP codons from the Sanger Zebrafish Mutant Resource project, which we reasoned would not occur in wild type populations. We submitted 3521 SNPs with flanking sequences to Affymetrix for assay design and 3212 were designed and manufactured by Affymetrix.

Based on the genotyping data from the MIPs panel a 1078 SNPs subset was selected for KASPar assays. These SNPs were selected based on performance in the MIPs assay across a selection of diverse strains (call rates differ across strains, especially when more divergent from the Tübingen reference genome). The criteria were high call rate in MIPs (>90%), a high reproducibility rate (>95%), a minor allele frequency (MAF) of greater than 0.05, and were called and polymorphic between at least two of the commonly used strains (AB, Tübingen LongFin, Hubrecht LongFin and Wik). Assays were designed and tested by KBiosciences, genotyping four fish for each of the four strains with duplicate reactions.

B. MIPs Protocol

We carry out MIPs assays according to the manufacturer's instructions except we use 1 μg of DNA instead of the suggested 2 μg. Since the zebrafish genome is about

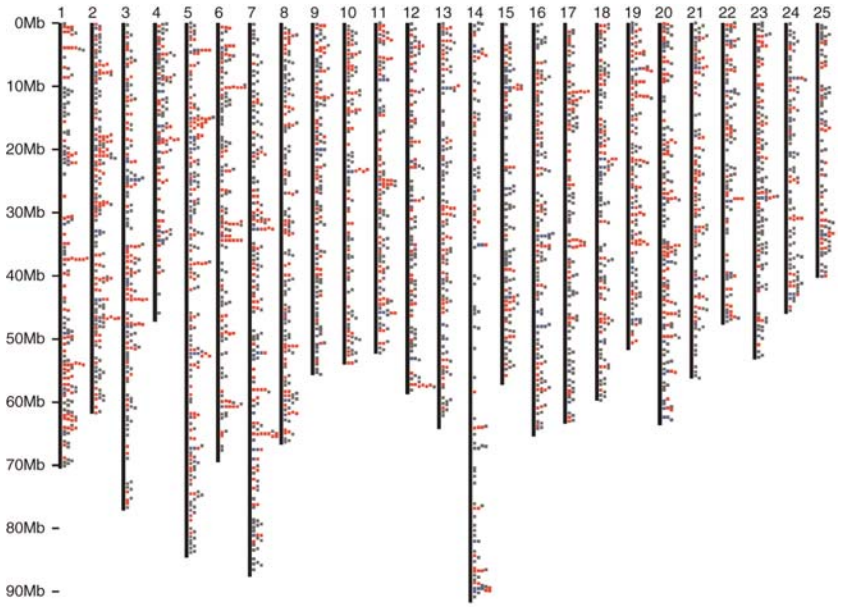

Fig. 1 SNPs in the MIPs panel displayed in relation to zebrafish chromosomes. Experimentally confirmed SNPs from Set 1 (Stickney *et al.*, 2002) and Set 2 (Guryev *et al.*, 2006) are blue, high-quality predicted SNPs from Set 3 are red, and lower-quality predicted SNPs from Set 4 are gray. Original colour pictures are available from Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv). (See color plate.)

half the size of human the concentration of any single copy locus will be the same. A control DNA from a single Hubrecht Long Fin male was included in each plate.

For analysis, we use the standard Affymetrix Targeted Genotyping software with default settings. For quality control:

(1) The coefficient of variation of control elements must be less than 30%.
(2) The median ratio of allelic signals to non-allelic signal must be greater than 20. For each assay, there are only two pre-defined bases. This can cause problems in tri- or quad- allelic systems but these are rare.
(3) Greater than 80% of the 3212 SNPs must be called on a sample for it to be included.
(4) Less than 10% of SNPs should be marginally called (the genotype appears between homozygous and heterozygous, see below for more details on calling genotypes). Pooled samples often fail this, but can be manually passed if they pass the other criteria.

After QC and normalization calls are assigned to homozygous and heterozygous genotypes, for example AA, AB, and BB calls, based on the ratio of signals for each

barcode (SNP assay) in each of the four colors (bases). Only the two known alleles (bases) designated during design are considered. A confident homozygous call has >80% of the signal in one channel, and a confident heterozygous call has >90% of all the signal in the two channels and at a ratio of close to 1:1. Many assays cannot be confidently assigned and are marked as marginal. Such marginal calls are then clustered together with the confident calls for each SNP assay, and marginal calls are assigned to the close genotypes using the allele ratio to measure distance. Because of this clustering poor experiments (or pooled experiments) if manually passed increase the misassignment of calls. Thus, it is worth deleting poor experiments, and keeping a copy of a core set of high quality experiments across wild type strains with which to cluster your experiments.

Note: This system requires the use of an Affymetrix four-color GeneChip Scanner 3000 7G, which is less common than the standard one-color models. Check with your genotyping core or service provider.

C. Strain Polymorphism

Our first tests were on the common strains. We fin clipped, prepared DNA, and genotyped 12 fish of each of the common strains using our MIPs assay: Tübingen, Tübingen LongFin, AB, and Wik as well as some less common inbred strains SJD and C32, and the G0 founders of the MGH cross (Knapik *et al*., 1996). The result of the genotyping of an inbred (SJD) and outbred line (Tübingen) as seen by Affymetrix's supplied software is shown in Fig. 2.

Using the 2221 SNPs that passed the quality control criteria described above, we discretized the calls into A (Homozygous for A allele), B (Homozygous for B allele), and C (Heterozygous). This makes the data anonymous with respect to the allele type, and the nucleotide types. The tree in Fig. 3 was generated using this data and fpars, which is part of the Phylip package http://evolution.genetics.washington.edu/phylip.html (Felsenstein, 1989). Due to the results of these experiments, we split two different strains that we originally thought were both the Tübingen Long Fin strain into two groups, Tübingen Long Fin and Hubrecht Long Fin. The Hubrecht Long Fin strain was used to generate ENU libraries for Zebrafish Mutation Resource and was originally obtained from a Hubrecht pet shop.

D. Genetic Mapping

We had already tested the G0s of the two main meiotic panels MGH and Heat Shock. We decided to genotype the other 42 members of the heat shock panel. On this cross 1072 SNPs appeared to be polymorphic, of which 997 could be assigned a map position with high confidence when merged with the extant markers (Kelly *et al*., 2000; Woods *et al*., 2005). These markers were added to the data used in the genome assembly from Zv8 onwards, and are present in ENSEMBL (Flicek *et al*., 2011) in the marker DAS track.

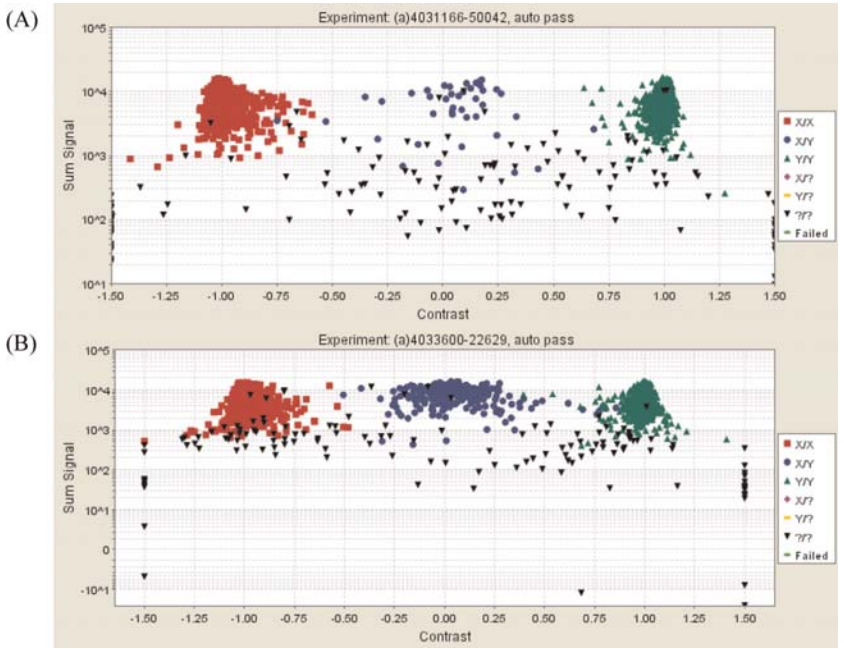

Fig. 2 Results of genotyping 3212 SNPs on individual fish. Each position is the result of one of the 3212 SNP assays. Red is homozygous XX allele, green is homozygous YY, blue is heterozygous (XY), and black are assays that could not be called. (A) The genotype of a single male of the SJD strain, a highly inbred line (1.3% heterozygous); (B) the genotype of a single male from the Tübingen strain, which is not an inbred line (8.3% heterozygous). Original colour pictures are available from Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv). (See color plate.)

E. Pooling Tests

To be able to use bulk segregant mapping it is important to be able to identify alleles within a pooled population. To test this requirement, we picked rare induced ENU alleles from the Zebrafish Mutation Resource (TILLING) project, as these ENU induced alleles were seen only once in over 6000 sequenced fish. In serial twofold dilutions, we diluted these ENU allele carrying fish with the control DNA (a Hubrecht Long Fin male used as a control on each plate), and also with a more stringent test, a pool of 96 mixed fish (24 fish of four strains: Tübingen, Tübingen Long Fin, AB, and Wik). The pool is a particularly demanding test since a second SNP within the primer sequences for the assay can lead to poor annealing and even subsequent allele bias when genotyping the target SNP.

Figure 4 shows the SNP genotyping of the induced Gli3^{hu2965} allele in wild type, heterozygous, and homozygous fish. Figure 5 shows the same SNP, increasingly diluted by the wild type allele: from homozygous, to heterozygous, and heterozygote diluted with an equal amount of control DNA (C) or pooled wild type strains (P). The

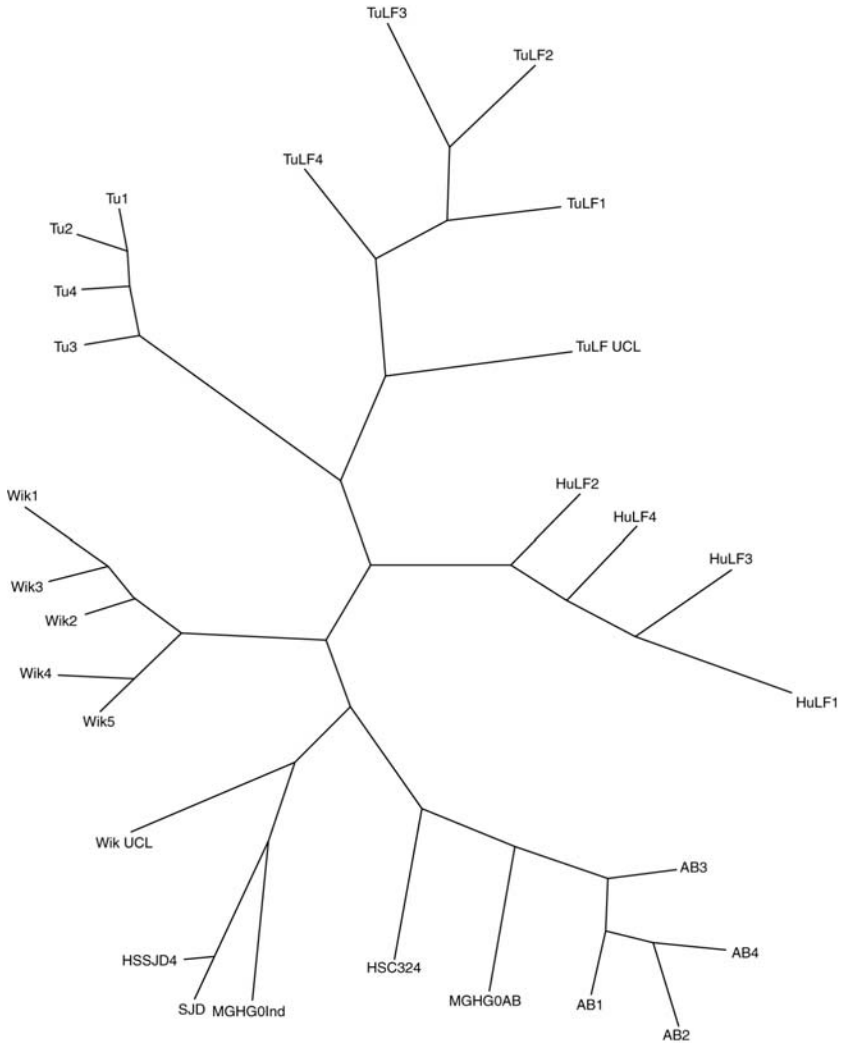

Fig. 3 Phylogenetic tree of strains. Genotypes were discretized to calls of A, B, and C, and used fpars, part of the Phylip package. Tu is Tübingen, TuLF is Tübingen LongFin, TuLF UCL is a Tübingen LongFin individual from the University College of London, HuLF is Hubrecht LongFin, AB is AB strain, MGHG0AB is the AB strain G0 individual used to make the MGH mapping panel, HSC324 is the C32 strain (inbred isolate of AB) G0 individual used to make the Heat Shock mapping panel, MGHG0Ind is the India strain G0 individual used to make the MGH mapping panel, SJD is Steve Johnson Darjeeling (inbred Indian wild type), HSSJD4 is the SJD strain G0 individual used to make the Heat Shock mapping panel, Wik1 to Wik4 are individuals of the Wik strain, and Wik UCL is a Wik individual from the University College of London. All fish are from the Sanger facility and originally sourced from ZIRC unless otherwise stated. The individuals from the mapping panels correspond to the strain as it was at that time, for example 10–15 years ago.

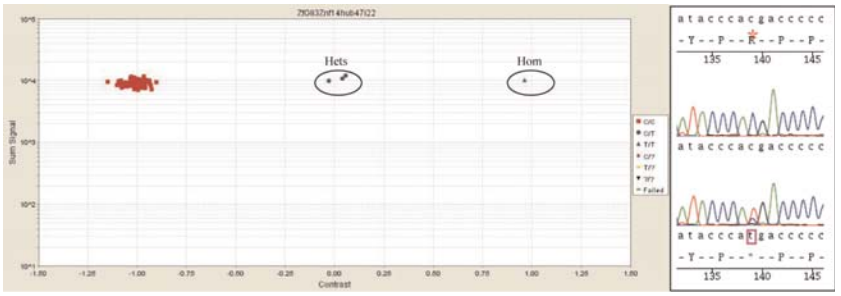

Fig. 4 MIPs assay of homozygous wild-type individuals (C/C: red squares), three heterozygous carriers of a nonsense mutation (C/T: blue circles), and one homozygous mutant individual (T/T: green triangle). The X-axis indicates the spectral contrast between the two dyes of the assay, and the Y-axis is the log of the fluorescence signal. To the right is shown capillary sequencing traces indicating the induced mutation in allele *gli3*^{hu2965} (lower) in comparison with wild type (upper). The MIPs homologous flanking sequence is**: gli3_hu2965 Zv9_24_11311698** atggtcctgaagcccatgtcaccaaaaaacaacgtggggagacatacccca [**C/T**] gacccccacctcagccacgagaacctggagcaaccactgaccaaaaggag. (See color plate.)

full range allele dilutions are: 1/1 (undiluted mutant), 1/2 (heterozygote), 1/4 (heterozygote diluted with equal amount of wild type DNA), 1/8, 1/16, 1/32, 1/64, and 1/128. The 1/8 dilution is scored as containing the ENU allele with both control and pooled dilutions, and the 1/16 dilutions still appear separate from the Wild type samples. Since bulk segregant mapping normally involves comparing the genotypes of mutants (homozygotes) to siblings (a mix of heterozygotes and wild types) a 1/16 sensitivity is more than is needed for to map a monogenic Mendelian recessive

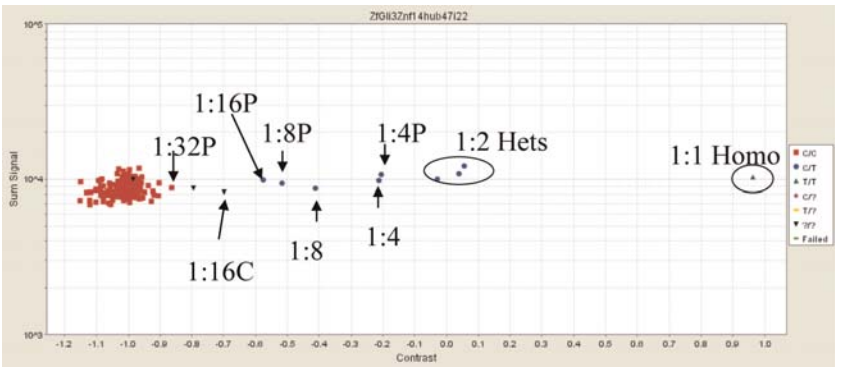

Fig. 5 The *gli3* MIPs was tested by combining DNA from a single wild-type fish (e.g., 1:16C) with heterozygous *gli3*^{hu2965} DNA at different ratios or by combining the pooled DNA of 100 adults (e.g. 1:16P) with *gli3*^{hu2965} DNA different ratios. Also shown is the signal from many wild-type (C/C: red squares) fish, as well as three heterozygous *gli3*^{hu2965} carriers (C/T: blue circles) and one homozygous mutant individual (T/T: green triangle). The X-axis indicates the spectral contrast between the two dyes of the assay, and the Y-axis is the log of the fluorescence signal. (See color plate.)

mutant. It should be possible to screen pools for the presence of a given allele, for example, a specific induced allele.

F. Bulk Segregant Mapping

Next we mapped a mutation found to enhance the *silberblick* phenotype (Heisenberg *et al.*, 2000). The allele was isolated on a TuLF background carrying *silberblick* and crossed onto Wik for mapping. We tested the grandparents, parents of the cross, and ools of mutants and siblings: 6, 12, and 24 embryos deep. In addition, we tested a further improvement using low DNA input. In this case, we amplified 10 ng of DNA from single embryos using the GenomiPhi V2 kits from GE Life Sciences according to the manufacturer's protocol: the amplifications were cleaned with Qiagen Qiaquick columns and Picogreen quantified using 1 μg inputs for MIP assays.

The results were promising: we saw a good signal that was around the *siberblick* locus, and a very strong signal of four SNPs showing linkage on LG21 that on Zv9 (http://www.ensembl.org/Danio_rerio) spans less than a 2.6 Mb region. The signal was strong in pools (Table I) and in individuals (Table II). In a separate experiment we also use GenomiPhi amplification of embryos that had been fixed and hybridized with *in situ* probes, to generate sufficient DNA for a MIPs assay 1 μg (data not shown).

Table I
Bulk segregant mapping

SNP	Zv9 pos	G1	G2	P1	P2	24M	12M	6M	24S	12S	6S
ss49818686	21:2913135	BB	AA	AB	AB	BB	BB	BB	AB	AB	AB
ss49817351	21:4563832	AA	BB	AB	AA	AA	AA	AA	AB	AB	AB
ss49835121	21:3548874	AB	BB	AB	AB	AB	AA	AA	AB	AB	AB
ss49818143	21:2002074	AB	BB	AB	AB	AA	AA	AA	AB	AB	AB

There are four SNPs that show strong linkage with the *silberblick* enhancer; the SNPs are within a 2.6-Mb region of LG21 in Zv9 (between bases 2,002,074 and 4,563,832). G1 and G2 are grandparents – TLF carrier and WIK map outcross, respectively, P1 and P2 are parents, and 24M, 12M, and 6M are mutant pools of 24, 12, and 6 fry, respectively, whereas 24S, 12S, and 6S are the equivalent sibling pools. The pools of 24, 12, and 6 mutants are homozygous (with the exception of ss49835121 in 24M a likely genotyping error), whereas the pools of siblings are all heterozygous.

Table II
Single embryo mapping

SNP	Mut1	Mut2	Mut3	Mut4	Mut5	Mut6	Sib1	Sib2	Sib4	Sib5	Sib6
ss49818686	BB	BB	BB	BB	BB	BB	AB	AB	AA	AB	AB
ss49817351	AA	AA	AA	AA	AA	AA	AB	BB	BB	AB	AB
ss49835121	AA	AA	AA	AA	AA	AA	AB	AB	BB	AB	AB
ss49818143	AA	AA	AA	AA	AA	AA	BB	AB	BB	AB	BB

The same four SNPs as in Table I, on DNA from single larvae amplified using GenomiPhi V2. There are six mutants and five Siblings (one failed to pass QC filters). All mutants are homozygous; no sibling embryos are homozygous for the same allele as a mutant.

Interestingly, on further investigation and cDNA sequencing it was found this *silberbick* enhancer is a mutation in *hmgcrb,* which was isolated previously (D'Amico *et al.*, 2007).

IV. Suggested Mapping Strategy

For a complete mapping strategy, we suggest a two-step approach, rough mapping followed by fine mapping. For rough mapping, we suggest the use of the 3212 SNP MIPs array and bulk segregant pools ideally no more than 24 fry deep, in this way just two chips should establish linkage for about $2–300. Since four sets of 48 samples can be processed in the MIPs protocol per person per week 96 mutants per week could be roughly mapped in this way. This compares well with the time and total cost for the SSLP approach (Geisler *et al*., 2007). For even greater confidence one could genotype two mutant and two sibling pools, or add the parents. However, linkage should be obtained using just two pools (one mutant and one phenotypically wild type) and can be quickly tested by fine mapping. This may not be true in some cases, especially if closely related strains or carrier fish of an admixture of strains are used. If linkage fails to be established, additional pools or parental carriers could be subsequently genotyped to help.

For fine mapping, candidate SNPs can be assayed using the KASPar panel across large numbers of mutants, for example 372 and a few siblings, for example 12. The candidate SNPs should clearly be polymorphic on the MIPs, and preferably in the extant KASPar panel, though it is relatively inexpensive to order new assays.

V. General Considerations

A. Genetic Diversity

In the design of any mapping strategy the correct choice of strains is critical in maximizing the polymorphic marker rate and the number of individuals that need to be genotyped. Essentially a mapping strain distant from the mutant line should be chosen. We have genotyped the commonest stains, several individuals from each strain, often from independent labs, and established the alleles within strains and polymorphisms between them. This is largely summarized by the tree in Fig. 3, which allows the quick identification of distant strains. If an admixture line is used then genotyping the grandparents and parents helps identify the SNP alleles origins and makes subsequent analysis easier.

B. Usefulness of the SNP Panel

After initial rough mapping using the MIPs panel, the emphasis switches from many markers on few pools, to a few closely linked markers on many individuals.

With a large number of extant cloned and well characterized mutants, many laboratories are performing suppressor or enhancer screens. A combined screening and cloning strategy has already been described (Rawls *et al*., 2003) combined with a quick and cost effective method of high-resolution genotyping, such screens become even more attractive.

More recently RAD-tag marker microarrays containing 13,824 elements have been used to map mutations (Miller *et al.*, 2007); however, the sequence, genetic, and genomic positions of these elements is unknown: in addition such RAD markers assayed by microarrays are dominant and non-quantitative and therefore heterozygosity can only be seen after sequencing the products and uncovering additional SNPs. Combining RAD-tags and Illumina sequencing, that is RAD-seq makes this more attractive (Baird *et al.*, 2008), especially for organisms without many existing markers or finished genomes, but still requires considerable bioinformatics analysis.

With the increasing numbers of predicted SNPs, the ability to identify new SNPs using RAD-seq, and other sequencing-based experiments, SNP panels will become increasingly used for mapping zebrafish mutations and QTLs.

References

Absalan, F., and Ronaghi, M. (2007). Molecular inversion probe assay. *Methods Mol. Biol.* **396**, 315–330.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Bradley, K. M., Elmore, J. B., Breyer, J. P., Yaspan, B. L., Jessen, J. R., Knapik, E. W., Smith, J. R. (2007). A major zebrafish polymorphism resource for genetic mapping. *Genome Biol.* **8**, R55.

Clark, M. D., Hennig, S., Herwig, R., Clifton, S. W., Marra, M. A., Lehrach, H., Johnson, S. L., Group t, W. (2001). An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Res.* **11**, 1594–1602.

Cuppen, E. (2007). Genotyping by allele-specific amplification (KASPar). *CSH Protoc. 2007* pdb prot4841.

D'Amico, L., Scott, I. C., Jungblut, B., and Stainier, D. Y. (2007). A mutation in zebrafish hmgcr1b reveals a role for isoprenoids in vertebrate heart-tube formation. *Curr. Biol.* **17**, 252–259.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., *et al*. (2011). Ensembl 2011. *Nucleic Acids Res.* **39**, D800–D806.

Frazer, K. A., Eskin, E., Kang, H. M., Bogue, M. A., Hinds, D. A., Beilharz, E. J., Gupta, R. V., Montgomery, J., Morenzoni, M. M., Nilsen, G. B., *et al*. (2007). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053.

Geisler, R., Rauch, G. J., Baier, H., van Bebber, F., Bross, L., Dekens, M. P., Finger, K., Fricke, C., Gates, M. A., Geiger, H., *et al*. (1999). A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23**, 86–89.

Geisler, R., Rauch, G. J., Geiger-Rudolph, S., Albrecht, A., van Bebber, F., Berger, A., Busch-Nentwich, E., Dahm, R., Dekens, M. P., Dooley, C., *et al*. (2007). Large-scale mapping of mutations affecting zebrafish development. *BMC Genomics* **8**, 11.

Guryev, V., Koudijs, M. J., Berezikov, E., Johnson, S. L., Plasterk, R. H., van Eeden, F. J., Cuppen, E. (2006). Genetic variation in the zebrafish. *Genome Res.* **16**, 491–497.

Hardenbol, P., Yu, F., Belmont, J., Mackenzie, J., Bruckner, C., Brundage, T., Boudreau, A., Chow, S., Eberle, J., Erbilgin, A., *et al*. (2005). Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* **15**, 269–275.

Heisenberg, C. P., Tada, M., Rauch, G. J., Saude, L., Concha, M. L., Geisler, R., Stemple, D. L., Smith, J. C., Wilson, S. W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81.

Hukriede, N. A., Joly, L., Tsang, M., Miles, J., Tellis, P., Epstein, J. A., Barbazuk, W. B., Li, F. N., Paw, B., Postlethwait, J. H., *et al*. (1999). Radiation hybrid mapping of the zebrafish genome. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9745–9750.

Kelly, P. D., Chu, F., Woods, I. G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y. L., Zhou, Y., *et al*. (2000). Genetic linkage mapping of zebrafish genes and ESTs. *Genome Res.* **10**, 558–567.

Knapik, E. W., Goodman, A., Atkinson, O. S., Roberts, C. T., Shiozawa, M., Sim, C. U., Weksler-Zangen, S., Trolliet, M. R., Futrell, C., Innes, B. A., *et al*. (1996). A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. *Development* **123**, 451–460.

Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, R., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N., Schmitt, K. (1998). Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. *Nucleic Acids Res.* **26**, 3562–3566.

Miller, M. R., Atwood, T. S., Eames, B. F., Eberhart, J. K., Yan, Y. L., Postlethwait, J. H., Johnson, E. A. (2007). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol.* **8**, R105.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B. P., Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088.

Rawls, J. F., Frieda, M. R., McAdow, A. R., Gross, J. P., Clayton, C. M., Heyen, C. K., Johnson, S. L. (2003). Coupled mutagenesis screens and genetic mapping in zebrafish. *Genetics* **163**, 997–1009.

Shimoda, N., Knapik, E. W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., Fishman, M. C. (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**, 219–232.

Stickney, H. L., Schmutz, J., Woods, I. G., Holtzer, C. C., Dickson, M. C., Kelly, P. D., Myers, R. M., Talbot, W. S. (2002). Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Res.* **12**, 1929–1934.

Woods, I. G., Wilson, C., Friedlander, B., Chang, P., Reyes, D. K., Nix, R., Kelly, P. D., Chu, F., Postlethwait, J. H., Talbot, W. S. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* **15**, 1307–1314.