Modifier genes and non-genetic factors reshape anatomical deficits in Zfp423-deficient mice

Wendy A. Alcaraz1,3,4,5,6, Edward Chen2,†, Phoebe Valdes7, Eunnie Kim2, Yuan Hung Lo2, Jennifer Vo2 and Bruce A. Hamilton1,3,4,5,6,∗

1Biomedical Sciences Graduate Program, 2Division of Biological Sciences, 3Department of Medicine, 4Department of Cellular and Molecular Medicine, 5Moores UCSD Cancer Center, 6Institute for Genomic Medicine and 7Department of Bioengineering, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92093-0644, USA

Received April 29, 2011; Revised June 22, 2011; Accepted June 30, 2011

Development of neural circuitry depends on the integration of signaling pathways to coordinate specification, proliferation and differentiation of cell types in the right number, in the right place, at the right time. Zinc finger protein 423 (Zfp423), a 30-zinc finger transcription factor, forms alternate complexes with components of several developmental signaling pathways, suggesting it as a point of signal integration during brain development. We previously showed that mice lacking Zfp423 have reduced proliferation of cerebellar precursor cells, resulting in complete loss of vermis and variable hypoplasia of cerebellar hemispheres. Here, we show that Zfp423−/− hemisphere malformations are shaped by both genetic and non-genetic factors, producing distinct phenotype distributions in different inbred genetic backgrounds. In genetic mapping studies, we identify four additive modifier loci (Amzn1–4) and seven synthetically interacting loci (Smzn1.1–3.1) that together explain approximately one-third of the phenotypic variance. Strain-specific sequence polymorphism and expression data provide a reduced list of functional variant candidate genes at each modifier locus. Environmental covariates add only modest explanatory power, suggesting an additional stochastic component. These results provide a comprehensive analysis of sources of phenotype variation in a model of hindbrain malformation.

INTRODUCTION

Human genetic disorders often present with variable phenotypes, even among individuals with the same allelic composition. Environmental influences and modifier genes are commonly invoked to account for these differences. Systematic assessment of their relative contributions, however, is often complicated by correlation of genetic and environmental backgrounds among limited numbers of available subjects.

The cerebellum is well suited to study variability in brain developmental disorders because of its evolutionary plasticity, dispensability for survival in laboratory models and access to genetic mutations that mark important molecular components (1–4). Abnormalities of cerebellum development are also a source of many human disorders that differ in degree and regionalization of defect (5–7). Malformation disorders like these are often subject to modifier genes. Few such modifiers have yet been explored (8), although modifier genes identified in experimental organisms often highlight unexpected interactions, novel molecular properties or pathways with therapeutic potential (9–13).

Zinc finger protein 423 (Zfp423)-deficient mice provide a compelling model of cerebellar vermis malformations. Mice homozygous for chemically induced or targeted null alleles completely lack cerebellar vermis, while hypomorphic alleles show dosage-sensitive reduction (14–16). Null embryos have a pronounced loss of granule cell precursor (GCP) proliferation at the midline, and abnormal differentiation and migration of ventricular zone cells (14) in addition to defects in forebrain, olfactory and adipose development (14,17,18). Zfp423-deficient mice also provide an interesting model of developmental variability. In contrast to the vermis, cerebellar hemisphere phenotypes in these mice range from modest size reductions to complete loss (14,15). All Zfp423-deficient mice reported to date have been on mixed genetic backgrounds, but the basis for phenotypic variation has not been investigated.

†Present address: Taipei Medical University, Taipei City, Taiwan
∗To whom correspondence should be addressed. Tel: +1 8588221055; Email: bah@ucsd.edu

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Zfp423 encodes a 30-zinc finger transcription factor whose physical interactions suggest a role in the integration of developmental signaling pathways. Zfp423 binds early B-cell factor 1 (Ebf1) to repress Ebf1 target genes (19) that otherwise promote cell cycle exit and neuronal differentiation (20,21). Zfp423 binds bone morphogenetic protein-activated SMADs as a coactivator. EBF and SMAD usage of Zfp423 are mutually inhibitory, suggesting that one pathway could modulate the other through competition (22,23). Zfp423 binds cleaved Notch intracellular domain to regulate Hes5 transcription and this interaction is also antagonized by EBFs (24). Zfp423 also binds retinoic acid receptors in neuroblastoma cells, where this interaction is required for retinoic acid-induced cell cycle exit and differentiation (25). These results place Zfp423 at a nexus of intercellular signaling pathways and internal gene expression programs.

To understand the sources of variation in clinically relevant phenotypes of Zfp423-deficient mice—and their potential implications for integrative functions of Zfp423 in development—we developed fully inbred stocks and mapped modifier loci in an intercross. We report different distributions of non-genetic variation in distinct genetic backgrounds and identify modifier loci that account for approximately one-third of the variation between these strains.

RESULTS

Variation in Zfp423<sup>nur12</sup> cerebellar hypoplasia is strain dependent

Zfp423<sup>nur12</sup> (null) mice on the original mixed genetic background showed a range of cerebellar phenotypes (Fig. 1). The vermis was absent in all mutant mice, but hemispheres varied dramatically in both degree of hypoplasia and mediolateral position, always in a bilaterally symmetrical fashion. Hemispheres when present were either very small or nearly full and either laterally diverged with substantial space between them or, among animals with larger hemispheres, adjacent at the midline. These characteristics defined four discrete phenotypic categories: (i) no evident cerebellum, (ii) rudimentary cerebellar tissue; 3, foliated hemispheres separated from the midline; and 4, foliated hemispheres joined at the midline. Phenotype 1, no cerebellum; 2, rudimentary cerebellar tissue; 3, foliated hemispheres separated from the midline; and 4, foliated hemispheres joined at the midline.

To test whether variation in Zfp423<sup>nur12</sup> cerebellar phenotype is dependent on strain background, we constructed a series of fully congenic lines. The Zfp423<sup>nur12</sup> mutation was serially backcrossed to BALB/cAnNHsd (BALB) and 129S1/SvImJ (129S1), two strains that contributed to the original mixed background, for at least 10 generations. We assayed markers flanking Zfp423 to minimize linked variation in the resulting congenic lines. Zfp423<sup>nur12</sup>/nur12 mice from these congenic lines were sacrificed at 2 weeks, dissected and assessed for cerebellar gross anatomy. In each line, mutant phenotypes included all four discrete categories we defined in the mixed stock, but each strain had a quantitatively different distribution of phenotypes (Fig. 1C). Among BALB-nur12 mutant animals (N10F1-N19F1 generations, n = 47), the majority (44/47) had cerebellar hemispheres and few (3/47) had no cerebellum. Among 129S1-nur12 mutant mice (N10F1-N16F1, n = 36), in contrast, most had either no cerebellum (13/36) or medially located hemispheres (16/36), with relatively few (7/36) having laterally separated hemispheres. The difference between these two distributions was significant (P = 2.72 × 10<sup>−5</sup>, Fisher’s exact test), indicating that at least one component of the variation in cerebellar hypoplasia was strain dependent.

Four BALB × 129S1 loci modify Zfp423<sup>nur12</sup> cerebellar phenotype

To determine the number and strength of genetic variants that modify Zfp423<sup>nur12</sup> cerebellar phenotype, we performed a BALB × 129S1 F2 intercross (Fig. 2). Two BALB-nur12/+ founder males were each bred to multiple 129S1 stock females. Full or half-sibling F1 heterozygotes were bred inter se. Among 1022 F2 progeny, 246 were homozygous mutant (χ<sup>2</sup> = 0.26, P = 0.61). Two hundred twenty-two homozygotes survived to 2 weeks and were dissected, photographed and assigned categorical phenotypes as defined above prior to genotyping; of these, 45 had phenotype 1, 51 had phenotype 2, 56 had phenotype 3 and 70 had phenotype 4. Large litters had a somewhat higher proportion of homozygote deaths prior to collection, though the difference across the mean litter size was not significant.

We performed a whole-genome linkage scan on DNA from these mice using simple sequence length polymorphism (SSLP) markers with an average spacing of 14.8 cM (largest gap, 22.3 cM). Linkage was evaluated for the categorical trait score using a non-parametric model implemented in R/qtl (26) and 10 000 permutations of the data to define empirical genome-wide significance thresholds (27). This analysis identified three significant linkage peaks (Fig. 2A),...
Figure 2. Significant linkage in 129S1 × BALB-nur12 F2 intercross. (A) Non-parametric linkage for categorical trait model identifies loci on chromosomes 2, 3 and 17 above genome-wide threshold. (B) Dichotomous trait models recapitulate categorical linkage peaks, identify additional linkage peaks on chromosomes 12 and 15 and substantially increase the LOD score on chromosome 3. Dichotomous trait 1 in blue, trait 2 in red and trait 3 in green. Significance of the linkage on chromosome 15 survives correction for testing four phenotype models, while the peak on chromosome 12 does not. Top line in each panel represents significant LOD score (genome-wide $P < 0.05$) and bottom line represents suggestive LOD score ($P < 0.63$), determined empirically by 10,000 permutations of the data.

Bayes credible intervals (BCIs) (27) at each linkage peak, and considering phenotypic direction and dominance relationships at each locus (Fig. 3). Three of the four loci exceed the conservative threshold for significant linkage under the assumption of an infinitely dense map with two degrees of freedom (LOD $\geq 4.3$) (28). We number these Amzn loci 1–4 in order of maximum significance level of refined intervals (Table 1).

For Amzn1 on chromosome 3, dense typing increased the peak LOD score to $4.62 (P = 0.0019)$ for categorical trait and $6.5 (P < 10^{-5})$ for dichotomous trait 3 (Fig. 3A and B). The best marker was D3Mit25, within a 95% BCI of 19.5 cM for the categorical and 22.0 cM for the dichotomous trait. The allele effect showed reduced hemispheres in 129S1 genotypes. At this locus, heterozygote phenotypes are significantly different from both 129S1 ($P = 0.031$, Fisher’s exact test) and BALB ($P = 1.2 \times 10^{-3}$) homozygotes, suggesting substantial but incomplete dominance of the 129S1 allele (Fig. 3C).

Refined analysis of Amzn2 on chromosome 17 for the categorical trait slightly reduced the peak LOD score (Fig. 3A). However, the locus remained significant for both categorical (LOD $4.82, P = 0.0013$) and dichotomous trait 2 (LOD 4.99, $P = 0.0012$), with credible intervals of 34.0 and 28.0 cM, respectively (Fig. 3A and B). Strongest linkage was at D17Mit139 for both traits. The allele effect was towards reduced hemispheres in BALB homozygotes. The 129S1 allele appeared dominant (Fig. 3C), with heterozygote phenotypes significantly different from BALB ($P = 3.9 \times 10^{-3}$), but not 129S1 ($P = 0.61$).

Amzn3 on chromosome 15 was significant only for dichotomous trait 1 (refined LOD 4.63, $P = 0.0054$). The marker with the best linkage was D15Mit270. This locus was centered at 31.6 cM with a 12 cM credible interval (Fig. 3A and B). The effect was toward reduced hemispheres in 129S1. Heterozygotes were significantly different from 129S1 ($P = 4.9 \times 10^{-4}$) but only nominally different from BALB ($P = 0.046$) suggesting a BALB-dominant model (Fig. 3C).

Amzn4 on chromosome 2 has a maximum LOD 4.01 ($P = 0.013$). For both categorical trait (38.0 cM credible interval) and dichotomous trait 2 (40.0 cM credible interval), D2Mit151 was the marker with highest linkage (Fig. 3A and B). The effect was toward reduced hemispheres in 129S1 genotypes. Although the slope of the effect plot appeared semi-dominant, heterozygote phenotypes were significantly different from 129S1 ($P = 0.017$), but not BALB ($P = 0.21$) and thus consistent with a BALB-dominant model at this locus (Fig. 3C).

Covariate loci interact with Amzn modifier genes

We next sought to detect synthetic interactions between loci that might explain a larger proportion of the phenotypic variance than each locus acting independently (Fig. 4 and Table 2). A two-dimensional linkage scan of the F2 intercross data detected only additive effects among the four identified loci with no significant epistasis among these or other loci (Fig. 4A and B). As a potentially more sensitive approach, we repeated the one-dimensional linkage analysis using each of the previously identified loci as a covariate and modelspecific permutations to set significance thresholds (Table 2).

Refinement of Amzn locus intervals

We refined the four significant modifier loci by typing finely spaced genetic markers through each interval, calculating
For the categorical phenotype model, we identified two such synthetic loci, one each for *Amzn1* and *Amzn2* as covariates (Fig. 4C and D). A similar approach applied to significant loci for each dichotomous trait found five additional interacting loci (Fig. 4E–G). These results identify loci that showed significant interactions with identified *Amzn* loci. We refer to these as synthetic modifiers of *Zfp423* nur12 (*Smzn*).

QTL models explain between 21 and 34% of the variance

To determine the proportion of variance explained by these loci, significant linkages and interactors were fit to a quantitative trait locus (QTL) model for each phenotype comparison (Table 2). For the categorical trait, three primary loci (*Amzn1*, *Amzn2* and *Amzn4*) and two interacting QTLs (*Smzn1.1* with *Amzn1* and *Smzn2.1* with *Amzn2*) were modeled. Within this model, *Amzn1* contributed 10.8%, *Amzn2* contributed 12.3% and *Amzn4* contributed 4.59% of the phenotype variance. The full model including synthetic interacting loci gave an overall LOD score of 20.0 and explained 33.9% of the phenotypic variance.

For dichotomous trait 1, we modeled two trait-specific loci (*Amzn3* and a chromosome 12 locus that did not survive experiment-wide correction for multiple trait models) and one interacting QTL (*Smzn3.1* with *Amzn3*). This model had a LOD score of 11.7, with 21.6% of the variance explained. *Amzn3* contributed 14.2% of the variance in this model. For dichotomous traits 2 and 3, similar models incorporating significant independent and interacting QTLs explained 32.0% (21.0% attributed to *Amzn2*) and 21.3% (18.7% attributed to *Amzn1*) of the variance, respectively. Thus, full models accounting for both additive and interactive effects increased the explanatory power of identified loci by 10–50% compared with purely additive models.

Environmental covariates have modest impact on cerebellum phenotypes

Although environmental variation was minimized as much as possible, we tested whether identifiable factors relevant to fetal or peri-natal environment might contribute to either main effects on phenotype or covariate (gene $\times$ environment) effects in our QTL models. Sex, litter size, maternal age and season at birth had no significant primary effects on phenotype in either the F2 cross progeny or the congenic lines. Sex and maternal age also had no significant covariate interactions with QTLs. Litter size, however, showed interaction with *Amzn2* for the categorical phenotype ($P = 0.011$) and especially trait 3 ($P < 10^{-4}$). Adding litter size to the

---

**Table 1.** Significant linkage locations, significance and variance explained

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Locus</th>
<th>Chr</th>
<th>Peak position (cM)</th>
<th>95% BCI (cM)</th>
<th>LOD</th>
<th>$P$-value</th>
<th>% variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td><em>Amzn4</em></td>
<td>2</td>
<td>18.2</td>
<td>11.7–49.7</td>
<td>4.01</td>
<td>0.0131</td>
<td>6.63</td>
</tr>
<tr>
<td></td>
<td><em>Amzn1</em></td>
<td>3</td>
<td>26.2</td>
<td>18.7–38.2</td>
<td>4.62</td>
<td>0.00192</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td><em>Amzn2</em></td>
<td>17</td>
<td>30.2</td>
<td>8.2–42.2</td>
<td>4.82</td>
<td>0.00128</td>
<td>7.61</td>
</tr>
<tr>
<td>Trait 1</td>
<td><em>Amzn3</em></td>
<td>12</td>
<td>24.1</td>
<td>17.5–33.5</td>
<td>3.58</td>
<td>0.0494</td>
<td>6.19</td>
</tr>
<tr>
<td></td>
<td><em>Amzn4</em></td>
<td>15</td>
<td>31.6</td>
<td>23.8–35.8</td>
<td>4.63</td>
<td>0.00543</td>
<td>9.33</td>
</tr>
<tr>
<td>Trait 2</td>
<td><em>Amzn4</em></td>
<td>2</td>
<td>18.2</td>
<td>11.7–51.7</td>
<td>4.09</td>
<td>0.0143</td>
<td>6.19</td>
</tr>
<tr>
<td></td>
<td><em>Amzn2</em></td>
<td>17</td>
<td>32.2</td>
<td>12.2–40.2</td>
<td>4.99</td>
<td>0.00117</td>
<td>7.75</td>
</tr>
<tr>
<td>Trait 3</td>
<td><em>Amzn1</em></td>
<td>3</td>
<td>26.2</td>
<td>14.2–36.2</td>
<td>6.50</td>
<td>$&lt;10^{-5}$</td>
<td>13.8</td>
</tr>
</tbody>
</table>

**Figure 3.** Four *Amzn* loci with divergent effects. (A) Refined linkage statistics with added markers for *Amzn1–4* for categorical phenotype. (B) Linkage scores for dichotomous models. Dichotomous trait 1 in blue, trait 2 in red and trait 3 in green. Ninety-five percent BCIs are represented by color-coded bars below each peak. In both (A) and (B), top line represents genome-wide significant and bottom line represents suggestive LOD scores, determined empirically by 10,000 permutations of the data. (C) Effect plots show bias toward hemispheres with BALB alleles at three of the four major effect loci. The 129S1 allele is incompletely dominant at *Amzn1*, 129S1 is dominant at *Amzn2* and BALB is dominant at *Amzn3* and *Amzn4*. The 129S1 allele is incompletely dominant at *Amzn1*, 129S1 is dominant at *Amzn2* and BALB is dominant at *Amzn3* and *Amzn4*.
models as a non-genetic covariate increased the variance explained to 37.8% for the categorical phenotype (with litter size contributing 1.88%) and 25.7% for trait 3 (litter size contributing 0.84% and founder contributing 9.01%). Season at birth showed an interaction with Amzn1 for trait 2 ($P = 0.011$). For the trait 2 QTL model, seasonality contributed only 2.94% of the variance, increasing the variance explained to 37.0%. Thus, while environmental covariates may be significant even in the vivarium setting, they contributed a relatively small proportion of explained phenotypic variance in this cross.

Figure 4. Covariate interactions with Amzn loci. (A and B) Two-dimensional linkage scan identifies additive effects (A), but no interactive effects (B) on nur12 phenotypes. Heat maps indicate two-dimensional LOD scores indicated by the color bar to the right of each plot. (A) Upper triangle, LOD$_a$, score for the additive QTL model; lower triangle, LOD$_{a+}$, score for the additive model compared with the single QTL model, assuming no epistasis. (B) Upper triangle, LOD$_e$, interactive model for epistasis; lower triangle, LOD$_{a+e}$, full QTL model, allowing for epistasis, compared with the single QTL model. (C–G) One-dimensional scans using each Amzn locus as a covariate identifies synthetically interacting loci. (C and D) For the categorical phenotype, significant interactions are seen with (C) Amzn1 (chromosome 2; LOD = 3.46, $P = 0.016$) and with (D) Amzn2 (chromosome 13; LOD = 3.42, $P = 0.033$) as the covariate. (E) For dichotomous trait 1, Amzn3 shows a significant interaction with chromosome 2 (LOD = 3.40, $P = 0.044$). (F) For trait 2, Amzn2 shows interactions with chromosomes 2 (LOD = 3.50, $P = 0.029$), 13 (LOD = 3.25, $P = 0.048$) and 14 (LOD = 3.30, $P = 0.044$). (G) For trait 3, Amzn1 interacts with chromosome 7 (LOD = 3.40, $P = 0.033$). Top line represents significant LOD score ($P < 0.005$), bottom line suggestive LOD score ($P < 0.63$), determined empirically by 10 000 permutations of the data.

Sequence and expression variants prioritize candidate modifier genes

We systematically identified potentially functional variants in genes within the 95% BCI for each Amzn categorical phenotype (Fig. 5 and Table 3). We reasoned that if modifier loci reflect functional variations in one or more protein coding genes, then modifier genes within these loci should include either non-synonymous coding sequence polymorphisms or RNA expression differences between the divergent strains. In addition, the loci identified here show no indication of linkage in preliminary data from a BALB/c–nur12 × C57BL/6j (B6) cross (Alcaraz et al., unpublished data), suggesting divergence between 129S1 and B6 as an additional filter for candidate polymorphisms. We used the Wellcome Trust Sanger Institute Mouse Genomes project data (http://www.sanger.ac.uk/resources/mouse/genomes) to identify sequence variants and profiled RNA abundance from E12.5 cerebella of nur12 homozygotes and littermate controls from three Zfp423nur12 congenic lines in triplicate to identify expression variants that distinguish 129S1 from BALB and B6 (Fig. 5A). Here, we focus on significant changes between strains only within the four primary linkage intervals; full analysis of this data set will be presented elsewhere.

The Amzn1 interval includes 199 annotated genes, from which we identified 43 functional variant candidate genes, an ~5-fold reduction (Table 3). The Mouse Genomes Project identified 96 894 single nucleotide polymorphisms (SNPs) that differ between BALB and 129S1, of which 56 872 are also divergent between B6 and 129S1. These included 69 non-synonymous substitutions in annotated coding sequences of 26 genes. In addition, RNA expression differences between BALB and 129S1 (corrected genome-wide $P$-value $\leq 0.05$ and $\geq 2$-fold change) identified 19 genes. Of 43 genes tagged with a potentially functional variation, a subset associated with developmental signaling pathways commanded particular attention. Siah2, which regulates GCP exit from the external granule layer (EGL) (29), had a non-synonymous SNP that replaced the conserved serine immediately after the initiating methionine with an asparagine (29), had a non-synonymous SNP that replaced the conserved serine immediately after the initiating methionine with an asparagine (29), which could in principle increase protein turnover through the N-end rule pathway (30). Sequence data from the Mouse Genomes project showed the asparagine allele only in 129S1 and 129S5 strains. Comparison across 29 vertebrate orthologs found only serine residues at this position. Strain-dependent expression differences
identified within Amzn1 included the Ssr3 somatostatin receptor, the Ppm1l regulator of stress-activated protein kinase signaling pathways (31) and transcriptional regulators Mynn, Phf17 and Zfp639 (Fig. 5C).

Similar analysis at the other Amzn loci (Table 3) reduced 907 annotated genes in the Amzn2 interval to 137 with potentially functional variants (7-fold); 451 genes in the Amzn3 interval to 110 functional candidates (4-fold); and 1057 genes in the Amzn4 interval to 125 functional candidates (8-fold reduction). These positional candidates include expression variation in Phf10, which modulates fibroblast proliferation (32), at Amzn2; Csnk1e, a component of the β-catenin destruction complex in the Wnt signaling pathway (33); and Ctnnd2, a target of Hes1 in the Notch pathway (34), at Amzn3. Multi-dimensional filtering of variants thus provided a comprehensive list of positional candidate genes and prioritized targets for future functional studies.

**DISCUSSION**

Complexity in genetic disorders occurs on several levels. Even in nominally single gene disorders, substantial variation is

---

**Table 2.** QTL models including interaction loci and variances

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>QTL</th>
<th>Interaction</th>
<th>Chr</th>
<th>Peak position (cM)</th>
<th>95% BCI (cM)</th>
<th>LOD</th>
<th>P-value</th>
<th>% variance explained*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>Amzn1</td>
<td>Smzn1.1</td>
<td>2</td>
<td>69.7</td>
<td>57.7–105.7</td>
<td>3.46</td>
<td>0.0163</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>Amzn2</td>
<td>Smzn2.1</td>
<td>13</td>
<td>23.1</td>
<td>17.3–49.3</td>
<td>3.42</td>
<td>0.0332</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>Amzn4</td>
<td>Smzn4.1</td>
<td>13</td>
<td>54.3</td>
<td>46.7–62.9</td>
<td>3.50</td>
<td>0.0292</td>
<td>4.59</td>
</tr>
<tr>
<td>Trait 1</td>
<td>Amzn3</td>
<td>Smzn3.1</td>
<td>2</td>
<td>105.7</td>
<td>81.7–105.7</td>
<td>3.40</td>
<td>0.0439</td>
<td>33.9</td>
</tr>
<tr>
<td></td>
<td>Amzn4</td>
<td>Smzn4.2</td>
<td>14</td>
<td>67.3</td>
<td>55.3–70.6</td>
<td>3.30</td>
<td>0.0442</td>
<td>21.6</td>
</tr>
<tr>
<td>Trait 2</td>
<td>Amzn2</td>
<td>Smzn2.2</td>
<td>2</td>
<td>51.7</td>
<td>35.7–63.7</td>
<td>3.50</td>
<td>0.0292</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>Amzn3</td>
<td>Smzn3.2</td>
<td>13</td>
<td>25.3</td>
<td>17.3–45.3</td>
<td>3.25</td>
<td>0.0475</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>Amzn4</td>
<td>Smzn4.3</td>
<td>14</td>
<td>67.3</td>
<td>55.3–70.6</td>
<td>3.30</td>
<td>0.0442</td>
<td>21.0</td>
</tr>
<tr>
<td>Trait 3</td>
<td>Amzn1</td>
<td>Smzn1.2</td>
<td>7</td>
<td>76.9</td>
<td>69.5–76.9</td>
<td>3.40</td>
<td>0.0327</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>Amzn2</td>
<td>Smzn2.3</td>
<td>13</td>
<td>25.3</td>
<td>17.3–45.3</td>
<td>3.25</td>
<td>0.0475</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>Amzn3</td>
<td>Smzn3.3</td>
<td>14</td>
<td>67.3</td>
<td>55.3–70.6</td>
<td>3.30</td>
<td>0.0442</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>Amzn4</td>
<td>Smzn4.4</td>
<td>14</td>
<td>67.3</td>
<td>55.3–70.6</td>
<td>3.30</td>
<td>0.0442</td>
<td>21.0</td>
</tr>
</tbody>
</table>

*% variance explained by the Amzn and its interacting Smzn loci or by the full model including non-interacting Amzn loci.

---

**Figure 5.** Candidate functional variants within Amzn loci. (A) Schematic for prioritization of Amzn candidate genes. Application of this scheme to the four Amzn loci identified 415 out of 2614 genes with high probability functional variants. (B) Siah2 variant in 129S1 is a non-conservative substitution at a highly conserved residue. (C) Heat map of strain-dependent expression levels for the 19 genes at Amzn1 identified as having 129S1-specific expression level significantly different from both BALB/c and B6. Each column represents average of three samples, with each sample comprising cerebella from three embryos. Color bar indicates normalized expression values; with red indicating higher and blue lower expression.
often seen in severity or penetrance, but sources of variation are typically not well understood (12). Here we have systematically analyzed sources of phenotypic variation in the Zfp423<sup>−/−</sup> mouse model of cerebellum malformation. We identified substantial variation in genetically homogeneous cohorts, different probability distributions between genetic backgrounds and significant linkage to multiple loci with distinct effects on cerebellum development in the absence of Zfp423 activity.

### Genetic variation may target distinct components of cerebellum development

The proportionate distribution of phenotypes was markedly different between two congenic backgrounds. Genetic mapping in intercross progeny provided strong statistical support for four distinct Anatomical modifier of Zfp423<sup>−/−</sup> (Amzn) loci. Covariate analysis suggested an additional seven synthetically interacting modifier (Smzn) loci. Full QTL models incorporating these loci explain approximately one-third of the variance in several trait models.

Interestingly, each of the dichotomous trait models supported different subsets of Amzn loci. Three Amzn loci were significant under more than one trait model, but each trait had distinct genetic contributions. Midline proximity of hemispheres (trait 3) was primarily attributable to Amzn1. This might suggest a role in lateral to medial migration of granule neurons, which is necessary to fill out the cerebellar midline (35) and could fit with the reported role of candidate gene Siah2 in mediating precursor cell exit from the EGL (29) or with either of two Bardet–Beidl syndrome genes in the interval, whose function in cilia may be important for signaling by Shh or other spatially organized signals in developing hindbrain (36,37). Sizes of hemispheres (traits 1 and 2) were primarily attributable to Amzn2–4. This may indicate activities in different phases of progenitor or precursor cell proliferation for these two traits. Both magnitude and timing of these events are important for shaping the cerebellum, including degree of foliated hemispheres (38–40).

### Functional variation in positional candidate genes

The availability of both expression analysis and deep coverage sequencing of the strains allowed us to prioritize genes within the Amzn credible intervals, without bias for presumed biological roles. This reduced the number of likely candidate genes by ~6-fold. Availability of mapping data from an additional strain (B6) was important for including or excluding local haplotypes shared by either primary strain (BALB or 129S1). Prioritizing the gene list by known developmental roles reduces the high-priority number an additional 16-fold (100-fold overall). This level of structured assessment should permit direct hypothesis tests for genetic factors that, like the Amzn loci we describe, act as probabilistic rather than deterministic modifiers of phenotype.

### Genetic disorders are often subject to several levels of noise in penetrance and severity of disease phenotypes

Our results provide a systematic dissection of the sources of variation in a model of hindbrain malformation and a structured approach for identifying causal and contributing variants that act, in the context of environmental or stochastic noise, to influence the distribution of traits in a probabilistic fashion. By minimizing environmental covariates, we identify modifier loci that account for a large fraction of the phenotypic variation. By examining residual environmental components, we demonstrate that Zfp423-dependent phenotypes are most likely subject to environmental effects. By accounting for both environmental and genetic variation in congenic strains, we further illustrate likely stochastic effects. Identification of distinct molecules and networks underlying these modifying events may provide further insights into this clinically important category of developmental disorder.
MATERIALS AND METHODS

Mice

All mice in this study were housed in matching cages on hepa-filtered air in an environmentally controlled room to minimize environmental variation. All procedures were approved by the UCSD Institutional Animal Care and Use Committee.

Derivation and mapping of the nur12 mutation on a mixed genetic background (with C57BL/6, C3H, 129S1 and BALB/c contributions) have been described (14,41). Zfp423\textsuperscript{nur12} was genotyped using two three-primer assays specific for the single nucleotide substitution. Separate reactions are performed for each allele, with common primers nur12-5\textsuperscript{′}: CTGCAGATGGTGATGACGAC, nur12-3\textsuperscript{′}: GAGCTGGT GGAGGAGAAGC, and one of two allele-specific primers: nur12-5\textsuperscript{SNP-wt(3)}: GAGCTACTTGAAGGCATGAAC and nur12-5\textsuperscript{SNP-mt(3)}: GAGCTACTTGAAGGCAAT. Allelic discrimination was optimal after 32 cycles of 94°C for 15 s, 66°C for 15 s and 72°C for 30 s. Expected fragment lengths are 258 bp for each allele-specific product and 436 bp for the allele-independent internal control.

Zfp423\textsuperscript{nur12} mice were backcrossed to BALB/cAnNHsd (BALB), 129S1/SvImJ (129S1) or C57BL/6J (B6) for at least 10 generations. SSLP markers from MIT Mouse Genetic and Physical Mapping project (42) and selective breeding were used to minimize linkage variation to the nur12 mutation on chromosome 8 during strain construction.

Linkage mapping was initiated by mating a BALB–Zfp423\textsuperscript{8}\textsuperscript{nur12} male to multiple 129S1 females. Full or half-sib F1 heterozygotes were bred inter se to produce 133 homozygous animals among 67 litters. A second cross of identical design initiated ~1 year later produced 89 homozygotes among 42 litters. The average litter size was 9.4 with a standard deviation of 1.8; litter size and proportion of recovered mutant animals was not different between founders.

Statistical analysis

Statistical analyses of phenotypes and genetic linkage were conducted in open source R packages (R version 2.10.1 http://www.r-project.org/). Fisher’s exact test was conducted in the R base environment. Linkage analysis was performed using the R/QTL package (26), version 1.16-6, using a non-parametric model based on the Kruskal–Wallis test for the categorical phenotype and a binary model for the dichotomous traits. Significant (genome-wide \( P < 0.05 \)) and suggestive (genome-wide \( P < 0.63 \), equivalent to one false positive per genome scan) LOD thresholds (28,43) were determined empirically by 10,000 permutations of the data for the autosomes and 151,860 X-chromosome-specific permutations. Interactions were identified with covariates under a normal model with multiple imputation. QTL interval locations were estimated by calculating 95% BCI at each locus. Variances were determined by fitting a QTL model to the data using the \textit{fitqtl} function (44).

Expression array samples and analysis

Zfp423\textsuperscript{nur12} mutant and littermate control embryos were collected from congenic BALB, 129S1 and B6 stocks at E12.5. Cerebella were dissected from littermates and homogenized immediately. After typing Zfp423 alleles, RNA was prepared in pools of three cerebella for matched samples for each genotype. Samples were purified using QIAGEN RNeasy mini kits with DNase. Three replicate experiments were performed for three matched pools per genotype, for each of the three congenic strain backgrounds (total of 18 array samples, comprising 54 type-selected embryos).

Labeled cRNA was hybridized to Illumina MouseWG-6 v2 BeadChip at the UCSD Biomedical Genomics laboratory (BIOGEM). Expression data were analyzed in GeneSpringGX 11.0 (Agilent). Differentially expressed genes were identified by two-factor analysis of variance, with \( P \)-values assigned by asymptotic method and multiple testing correction by the Benjamini–Hochberg false discovery rate method (45).

ACKNOWLEDGEMENTS

We thank Dr Nicholas Schork for helpful discussions and Dr Xiang-dong Fu for comments on a draft manuscript.

FUNDING

This work was supported by a research grant from the National Institutes of Health (R01 NS05487). W.A.A. was supported in part by an institutional training grant to the UCSD Genetics Training Program (T32 GM008666) from the National Institute of General Medical Sciences and by a National Research Service award (F31 NS061513) from the National Institute of Neurological Disorders and Stroke.

REFERENCES


