DISSECTING THE GENETIC ARCHITECTURE OF COMPLEX TRAITS: HOT SPOTS AND VASCULAR COLLATERALS

SHILIANG WANG

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor Of Philosophy in the Department of Cell and Molecular Physiology.

Chapel Hill
2010

Approved by:

Dr. James Faber	Advisor
Dr. Daniel Pomp	Reader
Dr. Timothy Wiltshire	Reader
Dr. Fei Zou	Reader
Dr. Robert Sealock	Reader

ABSTRACT

Shiliang Wang: Dissecting the Genetic Architecture of Complex Traits: Hot Spots and Vascular Collaterals

(Under the direction of Dr. James Faber)

Collateral arteries are endogenous “bypass vessels” that minimize tissue injury during arterial obstruction. Recent studies from our laboratory demonstrate that wide variation exists among individuals in the extent (number and diameter) of native (pre-existing) collaterals in healthy tissue and their outward remodeling (increase in anatomic diameter) in obstructive disease. Evidence suggests this variation contributes importantly to the wide variation in thromboembolic stroke, ischemic heart disease and peripheral arterial disease in humans. We hypothesized that genetic components contribute significantly to this variation. We tested this hypothesis in 243 C57BL/6 X BALB/c (CXB) F2 mice, wherein number and diameter of native cerebral collaterals and collateral remodeling after middle cerebral artery occlusion (MCAO) were measured. Linkage analysis identified a major QTL on chromosome 7 (*Canq1*) responsible for more than 30% of the variation in collateral extent. Three additional QTL were obtained for collateral number. Analysis of chromosome substitution and CXB recombinant inbred strains confirmed the dominant role of the *Canq1* locus. We also identified a QTL on chromosome 11 linked to variation in collateral remodeling. Efficient mix model association mapping (EMMA) of collateral number among 15 inbred strains delineated

172k (p=0.00002) and 290k (p=0.0004) base-pair regions containing 2 and 7 candidate genes, respectively, within *Canq1*. Analysis of six additional inbred strains, chosen according to their haplotype within the 172 kb EMMA region, strengthened and narrowed the locus from 172 kb interval to 2 kb. In candidate gene analyses, we found that collateral extent, infarct volume after MCAO, bleeding and re-bleeding times did not differ between *Itgal*$^{-/-}$, *IL4*$^{-/-}$ or *IL4*-receptor-$\alpha^{-/-}$ and wildtype mice. mRNA expression of 120 genes within the 95% confidence interval of *Canq1*, measured in the pial vasculature of C57BL/6 and BALB/c at embryonic day-14.5, -16.5 and -18.5 when the collateral circulation forms, identified 19 differentially expressed genes.

These results demonstrate that native collateral extent and collateral remodeling are heritable complex traits, with a highly significant QTL on chromosome 7 governing the majority of the variation in these traits. Furthermore, my work prioritizes a set of genes for future analysis as candidates underlying the process of collateral formation and its variation among individuals.

DEDICATIONS

To my wife, Zhuowei, for her endless support, advice and encouragement

To my daughter and son, Megan and Eric, for countless laughs they gave to me and for their understandings

ACKNOWLEDGEMENTS

I would like to acknowledge my doctoral dissertation advisor, Dr. James Faber, for his mentoring and guidance throughout my time in his laboratory. His enthusiasm to science drives me to develop my attitude to research. His knowledge in vascular biology inspired my critical thinking in my research. I would also like to thank my former advisor and current committee member Dr. Daniel Pomp. He gave me opportunities to develop my strengths in Statistics and help me with my writings and presentations. I would like to express my gratitude to other committee members Drs. Robert Sealock, Timothy Wiltshire, Fei Zou for their time, supports, and critical inputs.

I must give my thanks to several members of the Faber laboratory that I have had the good fortune of working with during the past one and a half years. Dr. Sealock, who is my committee chair and heavily involved my dissertation project, is considered as our lab's sibling. He not only gave me many scientific suggestions about my project and revised my writing, but also helped me go through my graduate studies by eliminating variety of obstacles confronting me. I could not have asked more for an outstanding, skillful, and congenial colleague - Hua Zhang. He taught me everything about the surgical skills required by my dissertation project given I knew nothing in the beginning. He also collaborated with me for several projects. Alex Mcfadden provided many supports and help in both academic research and personal life. The friendship with him made my life more joyful in the lab. I also learned from Jennifer Lucitti's insights about embryonic development of vasculature and had fun to share her humors about politics.

My time in the Department of Cell and Molecular Physiology has been enriched

by interacting with the caring, dynamic faculty and staff, including Drs. Ann Stuart and Alan Fanning, Vicki Morgan and Adriana Tavernise. I would especially like to acknowledge Jan McCormick, who gave me numerous help ranging from logistic supports to visa issues.

Thanks also go to several great friends I have met in Chapel Hill, including Yonghong Nie, Melinda DiVito, Deepti Rao, and Mahita Kadmiel. I would especially like to thank Dr. Lindsey Ho and Tiffany Thai, who have offered help in improving my spoken English, suggestions in career development, and friendship. I would like to express my eternal gratitude to my family. Pursuing an advanced degree is always difficult and stressful, specially studies in dual programs. My wife, Zhuowei Li has been my biggest supporter. She continues to encourage me and offer me endless supports and suggestions. My sister Shizheng Wang has always been there motivating me to go forward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACA	anterior cerebral artery
A/J	A/J strain of inbred mice
B6	C57BL/6J strain of inbred mice
Bc	BALB/cByJ strain of inbred mice
CAD	coronary artery disease
Cast/EI	Cast/EI strain of inbred mice
CFI_p	coronary collateral follow index
Chr	chromosome
CI	confidence interval
cis-eQTL	an eQTL that maps to the location of the parent gene that produces the mRNA or protein
CLIC4	chloride intracellular channel 4
cM	centiMorgans
CSS7	Chr substitution strain 7-- C57BL/6J mice with Chr 7 replaced by Chr 7 from the A/J strain
CSS17	Chr substitution strain 17-- C57BL/6J mice with Chr 17 replaced by Chr 17 from the A/J strain
E14.5	embryonic day 14.5
E16.5	embryonic day 16.5
E18.5	embryonic day 18.5
EC	endothelial cell
EMMA	efficient mixed model association algorithm
eQTL	expression quantitative trait loci
FDR	false discovery rate
ICR	its non-obese control line for M16
IL4	interleukin-4

IL4-α	interleukin-4 receptor alpha chain
IPA	Ingenuity pathway analysis
Itgal	integrin alpha L chain
LEWES	LEWES/EiJ strain of inbred mice
LFA-1	lymphocyte function-associated antigen-1
LOD	logarithm of the odds
LRT	likelihood ratio test
M16	a polygenic obesity mouse strain
Mb	megabase pairs
MCAO	middle cerebral artery occlusion
mRNA	messenger ribonucleic acid
PCA	posterior cerebral artery
PWD	PWD/PhJ strain of inbred mice
QTL	quantitative trait locus
RIL	recombinant inbred line
SNP	single nucleotide polymorphism
TF	transcription factors
Th2	type 2 T helper cell
trans-eQTL	an eQTL that does not map to the location of the parent gene that produces the mRNA or protein
VEGF	vascular endothelial growth factor
VEGFR-1	vascular endothelial growth factor receptor-1

CHAPTER I

BACKGROUND AND INTRODUCTION

Occlusive vascular diseases (eg, atherosclerosis of arteries supplying the heart, brain and lower extremities; thrombosis, stroke, and ischemic heart disease) are the leading cause of morbidity and mortality in the developed countries[1]. Upon occlusion of an artery, many physiological pathways are recruited to lessen the severity of ischemic tissue injury. Among these, three vascular anatomic protective mechanisms have been increasingly recognized: (1) the extent (i.e., number and diameter) of native (pre-existing) collaterals, i.e., arteriole-to-arteriole anastomoses present in a healthy tissue before the onset of disease[2, 3]; (2) enlargement of lumen diameter and wall thickness of these collaterals, a process termed collateral remodeling or arteriogenesis that is initiated by increased shear stress[4-8]; and (3) ischemia-induced angiogenesis consisting of sprouting of new capillaries from existing ones[9]. Since it has become recognized that angiogenesis can only improve the distribution for whatever blood flow is supplied by the collaterals after artery occlusion, attention has focused on arteriogenesis in the past twenty years[1, 4]. Arteriogenesis, which requires days-to-weeks to achieve up to an approximately 10-fold increase in collateral diameter depending on tissue and species, is induced by flow-dependent shear stress across the native collaterals when one arterial tree becomes occluded[1, 10, 11].

It has long been known that the severity of ischemic injury varies substantially among individuals and species[12-15]. Given the major role played by collaterals in restoration of blood flow after occlusion, this variation likely depends in large part on variation in native collateral extent and the amount of collateral remodeling[3]. A previous study from our laboratory reported that marked differences exist in native collateral extent, collateral remodeling, and tissue injury between C57BL/6 and BALB/c mouse strains after acute artery occlusion[16]. Based on this study, two subsequent studies identified genetic loci (QTL) associated with variation in ischemic tissue injury in these

same strains[17, 18]. However, the potential contribution of variation in native collateral extent and the amount of collateral remodeling was not examined in these studies. Thus, the physiological basis underlying genetically specified variation in ischemic injury has not been studied. The purpose of this dissertation is to identify genetic sources of variation in extent of the native collateral circulation and its remodeling in ischemic disease.

A. Collateral circulation

Collaterals are arteriole-to-arteriole anastomoses that occasionally are present to cross-connect the outermost branches of neighboring or opposing artery trees. In healthy tissues, they are limited in number, have lumen diameters of 10-40 microns, and have virtually the same arterial blood pressure at both ends, and hence have very low flow with no net direction[1]. Although tiny in size, collaterals offer significant protection against occlusive vascular diseases when abundant[19, 20]. These vessels that make up collateral circulation in tissues are highly unique, as unique as vessels of the general and lymphatic circulations are from each other[1]. Collaterals have at least 6 unique features: 1) During embryonic development, they somehow defy the canonical guidance molecules that ensure, in the general circulation, that arterioles always connect to capillaries that in turn connect to venules[21]. 2) They are rare in tissues – at least 500-fold less numerous than similarly sized arterioles in the general circulation. For example, the thigh of the mouse is estimated to have 10-20 of these vessels, while there are thousands of similarly sized arterioles in the general circulation of the thigh[1]. 3) Owing to the absence of a blood pressure drop between the two arteriolar connections of collaterals in healthy tissues, their flow is normally oscillatory and with little or no net

direction. Yet they persist in this abnormal hemodynamic environment that, if presents in arteries leads to inflammation and thrombosis, and if present in the vessels of the microcirculation leads to their pruning away. 4) Flow is induced in collaterals when the trunk of one of the trees that they cross-connect becomes occluded, causing a drop in blood pressure at that end of the collaterals. 5) This flow causes flow mediated remodeling, in which collateral vessels undergo rapid and remarkable anatomic enlargement of their diameter (up to 10-fold), as well as longitudinal growth over a period of days to weeks leading to tortuosity (the process of diameter enlargement is called arteriogenesis). Thus, native collaterals serve as "endogenous bypass vessels", wherein their number and diameter in healthy tissues determines the level of perfusion and tissue injury immediately after occlusion, and the amount of recovery of perfusion back towards normal as the collaterals remodel[1].

Finally, 6) There is a sixth unique feature of the collateral circulation that our laboratory has discovered. The density and diameter of native collaterals in tissues of healthy mouse strains vary dramatically as a function of genetic background, while the general and lymphatic circulations in the same strains are comparable in both form and function[1, 12]. Despite the uniqueness and importance of the collateral circulation, nothing was known—until the work appearing from our lab in the last several--about when or how collaterals form (ie, signaling molecules and cellular processes) or about the mechanisms responsible for the wide variation in extent of the collateral circulation.

In common with variation in anatomic and physiologic mechanisms among individuals, as well as in the severity of complex diseases such as diabetes, hypertension and cancer, variation in the extent of the native collateral circulation in healthy tissues and its ability to remodel in occlusive disease can be considered a complex trait. The etiologyies of complex traits are, by definition, multifactorial, with inputs from genetic predisposition, environment and interactions between the two. The

goal of this dissertation is to dissect the genetic components underlying variation in extent and remodeling of collateral circulation, using unbiased statistical genetic approaches, and thus pinpoint possible candidate genes and genetic loci governing this variation.

B. Native collateral extent widely varies among individuals

In a classic paper published in 1912, Herrick first noted that intercoronary anastomoses exist in man and other mammalian species, and that their number varies among species[22]. Schaper and Maxwell systematically examined the functional significance of the variation in these anastomoses by comparing infarct sizes and collateral-dependent flow in a wide variety of species using standardized conditions and a single, established techniques[20, 23]. Pig, rabbit, baboon, sheep, pony and wild boar experienced large transmural infarctions after coronary ligation owing to low collateral flow, while cat and dog experienced a slower progression of ischemic injury and had greater collateral flow. Guinea pig, which has abundant coronary collaterals and almost no reduction in blood flow of the tissue area served by the left circumflex artery (LCx) after left anterior descending artery (LAD) ligation, sustained little or no infarction[1, 20]. These studies clearly demonstrated widespread variation in collateral extent among species with important functional consequences.

Variation among individuals within a species is also present. In an important paper, Seiler and his colleagues recruited 106 patients without coronary artery or heart disease[24]. Using a 1-minute angioplasty balloon occlusion in the root of left circumflex artery tree, they measured a pressure-derived coronary collateral follow index (CFI_p), an index reflecting collateral conductance. Strikingly, an over 10-fold variation in CFI_p was found in this cohort. Although CFI_p may be affected by other factors[25], the magnitude of

this variation is remarkable. Because the occlusions in this healthy population were acute, the variation in CFI_p, in large part, reflected variation in extent of native coronary collateral circulation among individuals. This variation presumably reflects differences in genetic and environmental factors. These same investigators followed 739 patients with chronic stable coronary artery disease over a 10-year interval[24]. Those individuals with a good collateral flow index (CFI_p >= 0.25) had 75% lower mortality compared to those with a poor index (CFI_p <0.25)[24, 26]. Since these patients have coronary atherosclerosis, variation in CFI_p involves variation in both native collateral extent and amount of remodeling of the collaterals that occurred as the occlusive disease progressed. These results demonstrate that a wide range of variation in the coronary collateral circulation exists in healthy human subjects and this variation influences differential outcomes when the subjects experience coronary obstructive disease[24, 26].

Additional evidence for collateral variation has been found in patients with acute ischemic stroke cause by thromboembolic occlusion of the middle cerebral artery (MCA). Residual blood flow to the brain area served by the MCA in these patients was evaluated by various techniques: diffusion MRI, dynamic angiography or computed tomographic angiography[27-30]. The data suggested that collateral flow can minimize the degree of hypo-perfusion; those patients with indices of good collateral conductance showed larger areas of only mild hypo-perfusion and smaller infarct size than the patients with evidence of poor collateral conductance. Therefore, they concluded that collateral flow is an important determinant of severity of stroke[27, 28]. The variation in collateral flow cannot be explained by differences in capacity for remodeling in these studies, because of the acute nature of thromboembolic stroke and subsequent neural tissue injury. Hence, this variation is presumably due largely to the anatomic variation in native (pre-existing) collateral extent, although other factors such as individual variation in downstream stagnant thrombosis, thrombolysis or arterial pressure may also have contributed.

Further evidence for variation in the extent of the native collateral circulation in healthy tissues come from the observation that residual flow to the extremities varied widely among young healthy soldiers who sustained acute traumatic interruption of a major artery[31].

C. Genetic predisposition to variation in collateral abundance

Variation in collateral extent among or within a species can be attributed to genetic factors, environmental factors, or combination of the two. Since most species examined are either from outbred populations, such as humans, or domesticated species such as dog, it is difficult to eliminate the confounding variation in environment. Among inbred mouse strains, whose environment can be uniformly controlled, the BALB/c strain was reported to be more susceptible to 24-h focal cerebral ischemia after the middle cerebral artery occlusion (MCAO), compared with C57BL/6[18, 32, 33]. These differences in ischemic injury suggested that variation in collateral extent may exist among inbred strains and play an important role in determining ischemic injury. However, strength of this conclusion was diminished by the possibility that genetic variation in other factors which influence the response to ischemia such as inflammation, thrombosis and thrombolysis among strains may confound the interpretation[16].

More recent evidence has confirmed and amplified the idea that some mouse strains are genetically predisposed to high collateral abundance and others to low abundance. Chalothorn et al. cleared, maximally dilated, and filled the cerebral arterial circulation which resides in the pial membrane overlying cerebral cortex in BALB/c and C57BL/6 mice with a fluorescent vascular casting agent. Viscosity was adjusted to restrict filling to pre-capillary vessels. Collaterals connecting the anterior cerebral (ACA) and middle cerebral (MCA) artery trees in both hemispheres were then counted and

measured for diameter[16] under a microscope. These investigators also examined collateral abundance in the hind limb, level of hindlimb ischemia, and the time course and extent of recovery after femoral artery ligation (FAL). Their results were the first to show a pronounced difference in native collateral extent in the pial circulation, hind limb skeletal muscle, and intestine in these strains[16, 34]. Subsequently, Keum et al. permanently occluded the distal middle cerebral artery (MCAO), thereby inducing focal cerebral ischemia, in the above-studied C57BL/6 and BALB/c strains plus 13 other inbred strains. Infarct volume was measured 24 hours after MCAO. Their results confirmed the much larger infarction in BALB/c (Bc) than C57BL/6 (B6) mice, and further demonstrated that ischemic injury in a surgery-induced, acute focal ischemic mouse model is highly variable among all 15 strains. They created two F2 populations from two crosses with wide differences in infarct volume: B6 X Bc and B6 X SWR, and performed genome-wide linkage analyses in both populations. They found that the extent of ischemic tissue damaged 24 hours after MCAO was mapped to three loci with a predominant locus on Chr 7[18]. Building on previous reports from our lab, the authors postulated that infarct volume assessed 24 hours after MCAO, when only minimal remodeling of pial collaterals could have occurred, depended to a large degree on a genetic difference in native collateral extent governed by these loci[1]. However, collateral extent was not evaluated in their study. Hence, the physiological mechanism for variation in infarct volume remains unclear.

Our lab (Zhang and co-workers) subsequently quantified native collateral extent in the pial circulation in the same 15 inbred mouse strains. Remarkably, a 56-fold variation in native collateral number and 3-fold variation in native collateral diameter were found among the strains. For example, whereas B6 had on average, 9.2 collaterals / hemisphere, Bc had almost none (~0.2). With B6 and A/J as two extremes in native diameter (B6, 23um; A/J, 11um), the variation pattern for diameter across the

15 strains was found to not be congruent with variation in number. Zhang et al. further discovered that collateral number and diameter were highly correlated with the infarct volume measured 24 hours after MCAO by Keum and Marchuk[12, 18]. In fact, after measuring differences in collateral length as well as blood viscosity (hematocrit) and MCA tree size, these investigators were able to predict infarct volume for the 15 strains that closely followed the actual measured infarct volume. These data clearly support the concept that genetic components underlying variation in native collateral extent are dominant factors responsible for variation in ischemic tissue injury. Furthermore, according to their results, B6 and Bc are two extreme strains when both native collateral number and diameter are considered. Thus, these strains are optimal candidate strains to cross for subsequent linkage analysis aimed at dissecting the genetic components underlying this wide variation[35].

D. Collateral remodeling

Arteriogenesis refers to the vascular growth (remodeling) in which collaterals connecting two adjacent arterial trees increase their diameter and length when the supply to one arterial tree is reduced or abolished[8]. As mentioned above, arteriogenesis has long been the focus of much research, given its potential to increase collateral flow with time after occlusion. In occlusive diseases such as atherosclerosis, which progressively narrow an artery's lumen, collaterals can significantly enlarge 2–to–10 fold during disease progression, thus functioning to "bypass" the occluded vessel[4, 6]. The capacity for vascular enlargement is known to be affected by many factors: the mechanical forces of as shear stress and circumferential wall stress; cellular mediators such as endothelial cells, monocytes and smooth muscle cells; and several important

molecules such as monocyte chemotactic protein-1 (MCP-1), vascular endothelial growth factor (VEGF), and endothelial nitric oxide synthase (eNOS)[5, 7, 8, 10]. These studies show that many gene products contribute to arteriogenesis. Studies from our lab have shown that natural genetic variation, presumably affecting these or other factors, contributes to wide differences in collateral remodeling among inbred mouse strains[16, 36]. However, no study has used unbiased genome-wide mapping approaches to detect genetic loci responsible for genetic variation in the process of arteriogenesis. Those studies that did seek genetic explanations were from our lab used gene deletion approaches[34, 37] . Thus, key questions remain regarding regulation of arteriogenesis: Which genetic component(s) is(are) the most important in regulating arteriogenesis and variation in it? Are the genetic components underlying variation in native collateral extent reflecting variation in perinatal formation of these vessels[38] the same as those underlying variation in arteriogenesis?

Dokun et al preformed a genome-wide scan in B6 X BALB/c F2s using a qualitative necrosis score and Doppler-measured recovery of perfusion assessed 21 days after femoral artery ligation (a model of peripheral artery disease, PAD, in humans) as their traits. They identified a major QTL chromosome 7 for both traits (LOD 8, necrosis score; LOD 3.7, recovery of perfusion)[17]. These results demonstrated that a genetic locus links to variation in ischemic injury and recovery of perfusion after femoral artery ligation. As in the case of the pial circulation, multiple mechanisms such as downstream stagnant thrombosis, thrombolysis, arterial pressure and metabolic sensitivity to hypoxia besides variation in the collateral circulation, could affect the traits measured. However, our work has demonstrated that the extent and severity of ischemic injury, plus the amount of recovery of perfusion, in a large part depend on collateral extent and collateral remodeling[1, 36]. The recent findings by Zhang et al. shed further light on this problem: They found that both native (initial) number and diameter of

collaterals, plus the amount of remodeling (fold change in collateral diameter 6 days after MCAO) vary substantially among 15 strains and that the rank order of remodeling for 15 strains is very different with that of collateral number and diameter [12]. These findings suggest that genetic loci that underlie variation in collateral number, diameter and variation in remodeling are different.

E. Genes involved in formation of collaterals

Given the protective mechanism provided by the collateral circulation, important questions are that when and how the collateral circulation forms. Work from our lab has define the when – ie, formation occurs late during embryogenesis, well after the general and lymphatic circulations have formed; and stabilization of the newly formed vessels occurs early postnatally[38]. How, mechanistically, these unique vessels form is only just beginning to be defined. Our group has found that differences in expression of two genes – *Vegfa* and *Clic4* are involved[34, 37]. And ongoing work is pursuing this in depth[39]. A goal of this dissertation is to not only identify genetic loci that determine differences in extent of the native collateral circulation in an individual, but by narrowing down the potential candidate genes in such regions, provide clues into the genes responsible for controlling formation and stabilization of these vessels. Without such fundamental knowledge, the molecular mechanism by which collateral formation is regulated cannot be revealed.

When and how do native collaterals form and develop to their mature density and diameter? Is it in the embryo, the neonate or adult? The best tissue to examine this question is the cerebral circulation, because the arterial circulation is uniquely located on the "outside" of the cortex where it can be clearly imaged in 2D after removal of the calvarium. Earlier studies showed that anastomoses cross-connecting the middle

cerebral artery tree (MCA) or posterior cerebral artery tree (PCA) with anterior cerebral artery (ACA) in the cerebral circulation, are present in new born mice and rat[40, 41]. Such collaterals have also been found in this and most other tissue types of other mammalian species[20, 23]. Clayton and coworkers demonstrated the presence of cerebral collaterals at day one post natal (P1) in mice. They measured pial collateral number in various genetically engineered mice: mice overexpressing (*Vegfa*$^{hi/+}$), underexpressing (*Vegfa*$^{low/+}$), mice haploinsufficient for VEGF-R1 (*Vegfr1*$^{+/-}$; Flt1) or for VEGF-R2 (*Vegf2*$^{+/-}$; Flk1), and the respective wild type littermates at P1, P21, and adult (3 months). They discovered that the expression of VEGF-A has a dose-dependent effect on native collateral density at P1. Measurements of ischemic injury after femoral artery ligation indicated a similar gene dose-dependent effect among strains[34]. Moreover, higher expression of VEGF correlated with diminished neonatal pruning over the first 3 weeks after birth (maturation phase). These findings clearly suggested that native collaterals form during embryogenesis and are then subject to a pruning process in which mice experience a decline in collateral density until 3 weeks after birth.

When do native collaterals start to form during embryogenesis? And what molecules are responsible for this early collaterogenesis? Chalothorn et al. first characterized that wide differences in collateral density and ischemia-induced injury exists between two inbred strains: C57BL/6J and BALB/c[16]. After staining the embryonic cerebral circulation in both strains, they found that the initial collateral anastomoses, or "proto-collaterals", start to form by sprouting from existing arterioles within the crows of opposing cerebral artery trees by E14.5 to E15.5 at points where penetrating arterioles descend from the pia into cortex. Others in the lab have shown that this proto-collateral plexus cross-connects the opposing artery trees by taking a course of formation above the plane in which the cerebral capillary plexus forms[39]. Then, the number of proto-

collaterals keeps increasing through expansion of the proto-collateral plexus until ~E18.5, followed by partial pruning through P21[38].

The BALB/c mouse strain forms much fewer collaterals with smaller diameter, than the C57BL/6J mouse by E18.5; And by P21 when the adult density and diameter of the collaterals has been established, BALB/c have 60% fewer and 50% smaller diameter collaterals than the C57BL/6J strain[16]. What mechanisms account for the wide difference in collateral extent between two strains? Chalothorn et al. quantified the area of the primary embryonic capillary present at ~ E9.0 and E12.5 in the region where the collaterals will form. They eliminated the possibility that fewer endothelial cells in BALB/c than that in B6 are responsible. Nor could they attribute it to differences in brain size or branching morphogenesis of the cerebral artery trees[38]. However, collaterals in BALB/c were found to undergo more pruning compared with those in C57BL/6, starting at E18.5 to three weeks after birth. Collateral diameter in BALB/c mice was also found to be smaller during embryonic formation and to increase less during the maturation process occurring from E18.5 to P21. Furthermore, they reported that pericytes attached to nascent collaterals, which are important in stabilizing newly formed vessels, were comparable in both strains in all developmental stages examined[38].

If the above mechanisms are not responsible for fewer collaterals in BALB/c, can genes be identified that regulate collateral formation? Expression quantitative trait mapping (eQTL) in a cohort of recombination inbred lines (RILs) derived from C57BL/6 and BALB/c parentals suggested that variation at or near the *Vegfa* locus associates with low expression of VEGF-A and PDGF-B but high expression of angiopoietin-2 in BALB/c[42]. This pattern is consistent with less formation and maturation of collaterals in BALB/c. In a report prior to the above study, Clayton et al. used x-ray arteriography in the middle adductor collateral zone to count collaterals in VEGF under- and over-expressing strains and wild type[34]. A positive gene-dose dependent relationship

between VEGF-A and collateral density in skeleton muscle was evident, along with a similar relationship identified in the cerebral cortical circulation as described above[34]. An independent study in which investigators made a transgenic mouse with over-expressed human VEGF165 driven from a transgene in neural cells, supported this concept that VEGF from non-vascular cell types promotes collateral formation. Vogel et al. found that the transgenic mice have 2-3 fold greater vascular density than wild type. However, the cerebral blood flow was only minimally increased. They hypothesized that additional vessels were inserted into the existing vascular network, and postulated that they were arranged as collaterals, since collaterals have no net flow in the normal healthy tissue, thus have no effect on normal flow irrespective of their number or diameter[43]. Although the three-dimension imaging did not support these anatomical changes, this is likely because the investigators failed to examine the surface pial circulation where collaterals vessels reside. A subsequent study finding smaller infarct volume and neurological deficits in these transgenic mice than that in wild type after MCAO is consistent with this expectation: more collaterals lead to smaller ischemic injury[44]. On the other hand, several lines of evidence imply that CLIC4 also plays an important role in angiogenesis in general circulation: Ulmasov et al. have reported that *Clic4*$^{-/-}$ mice form fewer vessels normally and have impaired hypoxia-induced neo-angiogenesis in retina[45], and that endothelial cells from *Clic*4$^{-/-}$ mice cannot form mature vessels due to failure to lumenized the intial endothelial cell cords (ie, failure in tubulogenesis) in vitro. Subsequently, Chalothorn et al. found that *Clic4* disruption leads to greater reduction of hindlimb perfusion immediately after femoral artery ligation, worse ischemia, use impairment, and muscle atrophy over time. Arteriography found 70% and 40% fewer collaterals formed in the pial circulation and hindlimb of *Clic4* null mice by P1 and that a similar reduction was seen in the pia and hindlimb in adults[37].

In summary, these studies suggest that CLIC4, VEGF-A, and possible VEGFR1 are important signaling molecules in the formation and maturation of the proto-collateral plexus in the embryonic and neonate periods[1].

F. Quantitative trait locus analysis and its statistical algorithms

Quantitative trait locus (QTL) analysis, also termed linkage analysis, utilizes the correlation between phenotypes of interests and genotypes of genome-wide genetic markers in a population to infer sites of genetic variation associated with variation in a particular phenotype. QTL analysis provides unbiased inference about the genetic architecture underlying a complex trait[46]. This dissertation research will utilize this unbiased approach to dissect the genetic components underlying variation in extent of native collateral circulation and collateral remodeling in mice. In this section, we will discuss methodologies plus concepts of thresholds and confidence intervals (CI) used in a QTL analysis.

In the past two decades, a variety of statistical approaches have been developed to identify QTL in experimental organisms. These approaches were designed to meet different study purposes. Some were developed to improve statistical power and others to analyze for different types of distributions of traits[47, 48]. These approaches are classified into specific categories: single-marker analysis or interval mapping according the locations of the statistical tests performed; single QTL model vs multiple QTL model mapping according to statistical modeling; likelihood based or Bayesian based approaches[46, 49, 50].

Single-marker analysis does not require estimating the genetic distance between markers and is simple to conduct. And it conveys the key idea of linkage analysis: the phenotypic difference at one QTL is linked to genetic markers[51]. Accordingly, a simple *t*-

test can be performed to detect the marker close to a QTL, or ANOVA can be used in the case of an F2 population[52]. Though simple, single-marker analysis cannot estimate the QTL position. Moreover, estimation of the QTL effect can be confounded by the genetic distance between the QTL and the markers[46].

To overcome these limitations, Lander and Botstein proposed an interval mapping in which a statistical test can be conducted at any location between two adjacent markers[53]. With the assumption that a QTL has a linear effect on the phenotype of interest and that the phenotypes of each genotyped group of individuals follow a normal distribution, the problem of detecting a QTL becomes one of testings whether the coefficient for the QTL equals zero. Treating unobserved genotypes at a QTL as missing data, these investigators used the maximum likelihood-based approach—an expectation-maximization (EM) algorithm[54]—to detect the genetic effect and position of a QTL. To improve computational efficiency, Haley and Knott, in addition to several other researchers[55-57], proposed a linear regression-based interval mapping method, in contrast to maximum likelihood (ML)-based interval mapping. Similarly, they also assumed a linear relationship between the QTL and phenotype and a normal distribution. However, Haley and Knott replaced the unobserved genotypes for the QTL with the expectation of a genotype conditional on the two flanking markers. Using a linear regression framework, this mapping process is much faster, and additional covariates are easily handled. In addition, almost any standard statistical software can be programmed to run the analyses[55].

Single QTL mapping algorithms that detect QTL one-at-a-time are robust but not ideal for most complex traits, since such traits are usually affected by multiple loci each with small effects. This problem becomes evident when two closely linked QTL reside on the same chromosome, often resulting in a "ghost" peak[58]. To remove the confounding effects from the nearby QTL, Zeng et al. developed composite interval

mapping (CIM) in which they introduced additional markers as covariates. Since choosing appropriate markers as covariates is an important issue in this algorithm, Zeng provided detailed guidelines[59]. Although CIM can simultaneously account for multiple QTL effects, it is not capable of detecting interactions between loci, which are very common in most complex traits[46]. Further, the mapping result from CIM is sensitive to the additional markers that are chosen[59].

Simultaneously searching for multiple loci in the whole genome can be viewed as a model-selection problem in which the genotypes at any location in the genome can be potential covariates. The challenge of the mapping process is to identify the best covariate set or interactions between these covariates[60]. Kao et al. extended the likelihood-based interval mapping methods originally developed by Lander and Botstein[61]. They derived variance-covariance matrices for both positions and genetic effects of k putative QTL, in addition to accommodating for interactions. Haley and Knott proposed a model-selection strategy based on linear regression[55], while Churchill et al. developed a selection method based on Monte Carlo Markov Chain (MCMC) imputation for multiple QTL allowing interaction[62]. All these algorithms involved a model-fitting process and a model-selection criterion. While most criteria for selecting a model, such as Akaike information criterion (AIC) or Bayesian information criterion (BIC), are especially liberal in terms of interactions, Speed and Broman improved a penalized likelihood method as a model-selection criterion[63]. They considered strictly additive QTL models and generated the null distribution of genome-wide maximum LOD score by permutation to derive the penalty. By penalizing the LOD score, they apply an empirical control over rate of inclusion of extraneous loci, instead of classifying a prior distribution in Bayesian criterion[63]. Manichaikul and Broman extended this idea and included pair-wise interaction in the model[64]. In this algorithm, the LOD score is first penalized by main effects. The resultant score is further penalized by the interactions among QTL,

which are obtained by permutations to control false-positive rate. The LOD scores penalized by both interactions and main effects are then used as criteria to choose the best model. In summary, the above approach provides a straightforward and versatile QTL model selection. It directly controls the false-positive rates and avoids the uncertainty of choosing a prior[46, 64]. In addition to likelihood-based and regression-based algorithms, Bayesian approaches are also being rapidly developed for multiple QTL mapping[65-69].

Churchill et al. developed a permutation test of estimating empirical threshold for declaring a significant QTL. By shuffling subject numbers and disrupting the association between the phenotype and the genetic markers in an individual, permutations derive a null distribution of maximum LOD scores. Empirical thresholds derived by permutations have proved to be robust and widely accepted by the field[46, 70]. Constructing a confidence interval (CI) for a QTL, on the other hand, is challenging with the need to uncover the underlying genes for complex traits. A QTL region resulting from a classical F2 intercross typically contains hundreds of genes. To further narrow this region down, several algorithms have been proposed to accurately locate a 95% CI. The LOD drop-off method of Lander and Botstein is widely used to construct a CI because of the simplicity. An approximate 95% CI in a backcross or an F2 population equals a 1.5 -1.8 LOD drop on both sides of the peak[53]. Some researchers argue that the drop-off method is very biased when the QTL has a small effect. Mangin et al. proposed a likelihood ratio test framework for locating the 95% CI. Their simulation demonstrated that the 95% CI obtained by their algorithm has a higher probability to contain the true map location[71]. Visscher and Haley developed an empirical CI algorithm using bootstrapping. Simulations showed that calculating CI using bootstrapping performed very well regardless of the population size or effect. However, estimating CIs using this approach is slightly biased towards the conservative when segregation of alleles is

scarce or correlations between adjacent markers are strong[46]. Recently, Manichaikul et al. recommended a Bayesian method to calculate the 95% CI, and their simulation results suggested that the Bayes credible interval performed very well regardless of the effect size of the QTL, marker density, and sample size[72].

In summary, multiple QTL model mapping takes into account the effects from other loci and, and thus increases statistical power[46, 49]. Although more complicated and computationally intensive, it is a superior approach for mapping of QTL for complex traits. The Bayes credible interval derived by Manichaikul et al. for defining the 95% CI is more accurate regardless of conventional confounding factors[72].

G. Association mapping in mice and its statistical algorithms

A key goal of this dissertation is to identify novel gene candidates governing native collateral extent. To this end, association mapping in mice is a powerful tool. In association mapping, one uses a large number of genetic markers and inbred mouse strains to test for association between a trait and a genotype, treating each strain as an individual from a population[73]. Bi-allelic SNP status at a single locus across a panel of inbred strains, in most cases, is insufficient to discern the complex genomic structure. Instead, allele patterns across multiple loci (i.e., haplotypes) are more appropriate to represent the inheritance structure[74]. These observations have led investigators to define multiple adjacent SNPs over a local window as haplotype. Inferred haplotype is then used as a classification variable for haplotype association mapping (HAM) across the genome[74, 75].

Grupe et al. first applied association mapping in mice. They used approximately 3,350 SNPs in 8 inbred strains and found that a proportion of predicted QTL for 10 phenotypes overlapped with previously verified QTL for narrower regions[76]. This

approach initially drew significant criticisms, e.g., for high false-positive rate, limited sample size, and uncertainties about the population structure embedded among the inbred strains[77]. As such, association mapping methodologies are currently under considerable improvement to address these limitations. The simplest association mapping algorithm is to perform a *t*-test at every bi-allelic locus along the genome. This approach has been successfully applied to identify sweet preference loci[74]. However, because of small sample size and frequent switching of group classification along the genome, it is not possible for this approach to meet the *t*-test assumptions of normality and homogeneity of variance. Furthermore, this parametric approach not only cannot meet these assumptions, but it also suffers from the fact that the bi-allelic status of alleles cannot account for the complexity of underlying genetic variations among inbred strains[73, 74]. To address this limitation, a haplotype consisting of a three-SNP sliding window is used as the classification variable, and an F-test is constructed to measure the association[74]. However, this approach still requires normality assumption. A rank-based non-parametric test was therefore introduced. In this approach, a Kruskal-Wallis test statistic is constructed along the genome. The Kruskal-Wallis test statistic, which is based on rank information, discards quantitative information contained in the original dataset and, thus, lowers statistical power to detect association. To alleviate this problem, a bootstrap-based F-test was conducted at every three-SNP window. Instead of parametric Fisher distribution, the distribution of the test statistic is empirically estimated by drawing a bootstrap sample[74]. However, population structure and relatedness among inbred strains due to the genealogical history can cause spurious associations and high false-positive rates. Yu et al. suggested that a mixed linear model can correct this problem by modeling relatedness as random effect[78]. Recently, Kang et al. developed an efficient mixed model for association mapping (EMMA)[79]. They used a simple genetic similarity matrix estimated by a regression approach[80] to serve as the

kinship matrix. By accounting for the relatedness among inbred strains with this kinship matrix, they estimated the random effect using the maximum-likelihood approach instead of using an iterative algorithm. This method is much faster than other commonly used implementations by orders of magnitude[79]. Moreover, applications of EMMA to previous datasets gave almost identical results as those derived from earlier algorithms[73, 78, 79, 81].

Although initially limited by the few strains available, population structure, and inappropriate statistical algorithms[77], hundreds of studies employing association mapping have been published in the past few years with the development of new statistical algorithms[73, 74]. For example, Liu et al. combined 173 phenotypes in the Mouse Phenome Database (MPD) with 148,062 SNPs in order to identify 937QTL. They refined 67% of the experimentally found QTL to a genomic region of less than 0.5Mb with much-increased precision[82]. Bennett et al. recently applied the EMMA algorithm to a hybrid mouse diversity panel consisting of classic inbred and recombinant inbred strains. They verified this algorithm by mapping over 2500 cis-expression QTL[83]. The EMMA algorithm can effectively correct the population structure which previously limited association mapping in model organisms. Ghazalpour et al. applied this algorithm to an outbred mouse stock, the MF1, and successfully replicated expression QTL identified in previous F2 intercrossed mice[84].

Several bioinformatic tools are available to narrow down QTL identified by linkage analyses: comparative genomics, combined cross analysis, and interval-specific haplotype analysis[85, 86]. Compared to other bioinformatics tools, EMMA benefits the goal of narrowing the CIs through its speed, accountability of population structure, precision, and use of maximum information from the inbred strains examined[73]. With the aim toward assessing the utility of association mapping in mice, a recent study conducted association mapping for red blood cell and high lipoprotein cholesterol. The

investigators found that 78% of 54 haplotype association mapping peaks were true-positives and 22% false-positives in significant allelic association with real QTL[87].

Given the discussion of aforementioned algorithms for linkage analysis and association mapping, we will conduct linage analyses utilizing both single- and multiple-QTL model mapping algorithms in a B6 X BALB/c F2 population. The 95% CI will be extracted using Bayes creditable interval. The 95% CI of the most significant QTL will then be narrowed by the EMMA mapping utilizing the collateral extent data collected before[12].

CHAPTER II

EXPRESSION QTL MAPPING IN A DIVERGENT MOUSE CROSS SHOWS THAT MOST EQTL "HOT SPOTS" ARE CAUSED BY TRANSCRIPT CORRELATION STRUCTURE AND SIMPLISTIC STATISTICAL MODELING

This Chapter describes the project that I had done in the Dr. Daniel Pomp's laboratory

A. Abstract

Transcriptional hot spots, defined as chromosomal regions which contain more expression quantitative trait loci (eQTL) than expected by chance, have been detected in many studies. Often, hot spots have been hypothesized to harbor genetic variation (i.e. a master transcriptional regulator) which controls variation in mRNA abundance for hundreds or even thousands of genes. We tested the hypothesis that high pair-wise correlations among transcripts, and oversimplified statistical modeling during eQTL mapping, might create spurious hot spots. We evaluated multi-tissue eQTL data on ~400 mice from a large F2 population created by crossing a polygenic obesity mouse strain (M16) and its non-obese control line (ICR). eQTL mapping with classical modeling was performed in hypothalamus, muscle, liver and adipose. Hot spots were identified based on harboring eQTL for >130 transcripts within a 5 cM window. We first characterized the gene action and allelic effects of the eQTL within hot spots to understand their overall genetic architecture. A permutation test, where genotypes were permuted while conserving gene expression phenotypes and the internal correlation structure of the data, was conducted to test if any of the hot spots were caused by high correlation structure among transcripts. Finally, we explored appropriate statistical modeling for eQTL analysis in the context of hot spot detection, by accounting for effects from transcription of genes other than the one being analyzed, and the network to which a transcript is belonged.

A total of 11 eQTL hot spots were identified, with no overlap across the four tissues. Underlying genetic architecture varies widely across hot spots in regard to additive and dominance effects and direction of allelic effects of the eQTL. Permutation testing revealed that only one of the 11 hot spots, with >2000 eQTL on MMU17 in hypothalamus, showed genome-wide significance ($p < .05$). In addition, by accounting

for highly correlated transcripts and network effects in the eQTL analysis, all eleven hot spots disappeared. We conclude that in this data set most hot spots are artifactual due to the high pair-wise correlation among transcripts and/or oversimplified statistical modeling.

B. Introduction

To understand the transcriptional landscape underlying a complex trait, expression quantitative trait locus (eQTL) mapping systematically identifies genetic variation that controls phenotypic variation in mRNA abundance on a genome wide basis[53, 88, 89]. During eQTL mapping, a commonly observed yet striking phenomenon is the identification of transcriptional hot spots, also known as trans-bands[90, 91]. A transcriptional hot spot is defined as a chromosomal region which contains more eQTL than expected by chance[92], characterized by the apparent clustering of several hundred and even thousands of trans-acting eQTL, within a relatively small chromosomal window. Given the almost ubiquitous nature of hot spots in eQTL studies, it is imperative to gain a better understanding of why they arise and what, if any, biological significance can be attributed to them.

There are several interpretations regarding the existence of eQTL hot spots. First, it has been hypothesized that a regulatory element, disrupted by one or more polymorphisms in a hot spot, leads to widespread changes of downstream transcript abundances[93, 94]. Transcription factors (TFs), tRNA synthetases or other important regulatory elements within a hot spot have been proposed to be responsible for this regulation. The mechanism of this type of regulation was explored by Sun et al (2007) using statistical methods. They identified four eQTL modules which feature a common sequential regulation pattern mediated by TF activity: DNA polymorphisms, cis-linked

genes, TF activity and downstream linked genes. In another study, Mozhui et al (2008) identified several candidate tRNA synthetase pertinent proteins as potential master regulators within a hot spot on the distal part of mouse chromosome 1. Second, one or more cis-acting upstream genes in an enriched pathway has been hypothesized to be an underlying cause of eQTL hot spots. Wu et al. (2008) performed a genome-wide association analyses for expression in adipose and reported over 1600 candidate hot spots in 28 inbred lines. They examined the hot spots for enrichments in functional gene sets representing known biological pathways and discovered several novel upstream regulators in the pathways[95].

In contrast, it has been proposed that hot spots are artifacts or at least that they cannot necessarily be interpreted as master regulators. In a simulation study, Pérez-Enciso (2004) showed how hot spots can be caused by high inter-correlation between transcript levels. Similarly, Breitling et al (2008) hypothesized that most hot spots are artificially generated due to highly correlated gene expression among transcripts or linkage disequilibrium among markers. They proposed permutation strategies, where genotypes are permuted while conserving gene expression phenotypes and the internal correlation structure of the data, to test the significance of hot spots. A second theory is that hot spots are at least partially artificially caused by oversimplified statistical modeling during eQTL mapping[96]. The rationale is that the complex regulatory elements for the expression level of a gene cannot be just confined to underlying genetic factors. Perez-Enciso et al. (2007) found that transcripts other than the one being analyzed explain on average much more variability than the markers themselves using support vector machine.

Given the strong interest in eQTL hot spots, how and why they arise, and what biological meaning they have, the objective of this study was to assess the nature of eQTL hot spots by employing appropriate model-free permutation testing, and more

complex statistical modeling, in their detection. To achieve this goal, we systematically identified and characterized eQTL hot spots in four tissues from a large F2 mice population. Surprisingly, our results demonstrated that, for this specific mouse cross, the vast majority of identified eQTL hot spots appear to be artifactual in nature.

C. Materials and methods

Mice

The F_2 population used has been previously described[96-98]. Briefly, a large F_2 population (*n* = 1,181) was established by intercrossing the M16 and ICR long-term selection lines[97, 99]. Twelve F_1 families resulted from six pair matings of M16 males × ICR females and six pair matings of the reciprocal cross. A total of 55 F_1 dams were mated to 11 F_1 sires in sets of five F_1 full sisters mated to the same F_1 sire. These same specific matings were repeated in three consecutive replicates. All litters were standardized at birth to eight pups, and mice were weaned at 3 weeks of age and provided *ad libitum* access to water and pellet feed (Teklad 8604 rodent chow). Mice were caged individually from 4 to 8 weeks of age. All procedures and protocols were approved by The University of Nebraska Institutional Animal Care and Use Committee prior to experiments.

RNA sample preparation and hybridization

As previously described[98], global expression analysis was determined using the 23,574-feature mouse Rosetta/Merck Mouse TOE 75k Array 1 (Gene Expression Omnibus (GEO) Platform: GPL 3562; Agilent Technology, Palo Alto, CA, USA). Transcriptonal analysis was performed in hypothalamus (n = 451 F_2 mice), liver (n = 441), epididymal (males) or perimetrial (females) adipose (n = 477), and gastrocnemius muscle (*n* = 465). Labeled cRNA from each F_2 animal was hybridized against a pool of

labeled cRNAs constructed from equal aliquots of RNA from 160 F_2 animals for each of the tissues in the cross that was balanced for sex and litter[100]. Microarrays were scanned, and individual feature intensities were pre-processed by background subtraction, normalization to mean intensities of the Cy3 and Cy5 channels, and detrending to fit a linear relationship between channels. Normalized intensities were used to derive expression ratios using the Rosetta error model[98].

Expression QTL mapping

Expression QTL (eQTL) mapping was performed using R code in the following steps. First, we built linear regression models with appropriate fixed effects and covariates, including sex, replicate, and family. This represented the null model. Second, at every cM position throughout the genome, the probability for parental line of origin for each F_2 individual was estimated based on flanking marker genotypes[101]. Third, we modeled the additive and dominance effects of each locus as presented by Haley and Knott (1994). We then added the additive and dominance effects and their interaction with sex into the null model as a full model to analyze the expression phenotypic data. Fourth, the likelihood ratio test (LRT) score for each locus was approximately calculated according to the ratio of two residual of sum squares from the two models[55]. Similarly to Schadt et al (2003), we used an LRT of 20 (LOD = ~4.4) as the significance threshold for an eQTL[102].

Detection of eQTL Hot Spots

To evaluate the distribution of eQTL, we plotted the number of eQTL detected at every cM position along the genome in each tissue (data not shown). At positions where eQTL tended to cluster, we evaluated the distribution of eQTL within extending cM windows. Based on this distribution analysis (data not shown), we defined an eQTL hot

spot as a cluster of >130 eQTL within a 5 cM window. This threshold of eQTL for a hot spot to be present is conservative, being significantly greater than that used in other studies (e.g. Wu et al., 2008), with a goal of evaluating the veracity of only the strongest signals in our study.

General genetic architecture of eQTL hot spots

Within each eQTL hot spot region, additive and dominance effects for each eQTL were estimated by regression. The direction of allelic effects was obtained by the sign of regression coefficients of the additive or dominance effects. If an eQTL has an additive effect on the mRNA abundance of a transcript, the regression coefficient for this additive effect was extracted from the additive model. A positive sign indicated that the M16 allele increases the expression level since we modeled the additive effect as the difference of line origin probabilities between both alleles from M16 and both from ICR. Similarly, directions of dominance effects were determined.

Permutation Testing

We applied permutation testing to artificially disrupt the relationship between transcription level and genotype while maintaining the correlation structure between transcripts. This was achieved by permuting mouse ID for expression phenotypes but maintaining the actual expression data within each mouse. We then performed eQTL mapping and hot spot detection as described above in each of 1000 permuted data sets. The strongest hot spot (with the most eQTL) was recorded in each permutation and ranked within each tissue separately. The genome wide p value for significance of the presence of a hot spot was defined as the probability of the number of eQTL clustered under the null hypothesis being equal to or greater than the observed number clustered.

Gene covariates selection to remove hot spot effects in eQTL mapping

To test whether eQTL hot spots may arise because an effect common to several variables is not accounted for properly in the statistical modelling, we further investigated the covariate modelling proposed in Pérez-Enciso et al (2007). We postulated that modelling strongly influences the eQTL P-value profiles and that expression data for a given gene could be modelled using, potentially, the rest of the gene expression levels. We utilized an efficient and robust modelling strategy based on forward selection and bootstrap model aggregation, bagging[103], in order to model directly these hidden effects prior to eQTL mapping of the expression trait. Given a gene *A*, the effect of all other gene expressions can be estimated from:

$$y_{iA} = \mu_A + \sum_m \beta_{mA} y_{im} + e_{iA},$$

(1)

where y_{iA} is the expression measure for gene *A* and *i*-th individual, μ_A is a global mean, β_{mA} is the effect of a putative covariate *m* on the expression level of the gene *A*, *e* is the residual. Therefore, the set of covariates *m* measures the influence of external gene expression levels (y_{im}) on the expression of the gene studied (*A*). Deciding about the optimum model in this situation poses an enormous challenge computationally because of the almost infinite number of potential models. Here we used a forward selection approach combined with a bootstrap procedure in order to incorporate model uncertainties. At each stage, the most significant gene expression level *j* was included as a new covariate in the modelling of the expression levels of the gene *A* and the effects of the remaining candidates were re-evaluated one at the time successively given the new model

$$y_{iA} = \mu_A + \sum_m \beta_{mA} y_{im} + \beta_{jA} y_{ij} + e_{iA},$$

(2)

where *m* indexes the covariate genes already included in the model and *j*, the remaining covariates to test. The addition of covariates in the model was stopped when no candidates reached the significance threshold of $P = 10^{-5}$. Fitting was done using least squares and the significance of the effects was obtained from an *F*-test.
This variable selection was repeated on 1,000 different bootstrap samples of the data[104] to incorporate uncertainties on the model building and minimize the risks of false positives[103]. The proportion of models where the gene *B* was covariate of the gene *A* is the bootstrap probability (BP) of this relationship between genes *A* and *B*.

The final model can be depicted as a directed network where B being covariate to A can be represented as directed edge between genes *A* and *B* ($B \rightarrow A$) when the BP is not null. It can happen that A is selected to be covariate of B and vice versa, i.e., B is selected as covariate to A. In this case, we retained only the edge with largest BP. Remaining covariates with a BP lower than 0.25 were deemed not to be reliable and also suppressed from the network. Note that some edges may still remain ambiguous if they have equal BP in two directions $B \rightarrow A$ and $A \leftarrow B$. Then the edge was removed if the two genes were located on the same chromosome considering that the ambiguity could result from genetic linkage. Nonetheless ambiguous edges were rare (0.5 to 1%). Next, eQTL mapping for gene *A* was carried out with the NET-model determined by the bootstrap procedure:

$$y_{iA} = \mu_A + \sum_m \beta_{mA} y_{im} + \alpha_{lA} q_{il} + e_{iA} \quad (3)$$

where *m* indexes the covariates of the gene *A* in the network, as described above, α_{lA} is the effect of the locus *l* on the gene *A*, q_l is the additive QTL coefficient at the tested locus *l*. For computational simplicity, no other effects were included in the model, and this analysis was performed for selected hot spots (hypothalamus MMU17, and muscle

MMU15, MMU17). QTL effects were tested against the null model of no QTL with an F-test. Results were reported as the $-\log_{10}$ of the *P*-value. Genome-wide maximum score for each gene was recorded. The location of this maximum is the most likely eQTL for the gene given a one QTL model, and therefore this maximum will be referred as the eQTL for the gene. These eQTL will be classified as *cis* or *trans* whether they were positioned locally (≤ 5 cM) or distant (≥ 5 cM) to the gene position, respectively.

Simulated datasets

To further explore the behavior of the NET modelling, we simulated two scenarios that comprised 1000 genes and 450 F2 individuals. The first scenario (Sim1) consisted of a gene regulatory network, whereas in scenario 2 (Sim2), gene co-expression was caused by a block correlation matrix among gene expressions. Genome simulations were based on the sex-averaged map of Shifman et al.[105] for the 19 autosomal mouse chromosomes. A position in the genome was assigned randomly to each gene. Marker and QTL genotypes in the F2 individuals were simulated using R/qtl[106]. Only one QTL, located in *cis* position, was simulated for every gene. 28 replicates per scenario were simulated.

In Sim1, a topological rank in the network was assigned to each gene; a gene can be regulated only by a gene with a higher topological rank. The expression levels of the k^{th} gene for the i^{th} individual were simulated according to $y_{ik} \sim N(\hat{y}_{ik}, \sigma_e)$ where $\hat{y}_i \sim N(\sum_j c_{jk}\hat{y}_{ij} + \alpha_{lk}q_{il}, \sigma_t)$, where *j* indexes the parents of the gene *k* in the network, c_{jk} is the regulatory effect of the parent *j* on the gene *k*, α_{lk} is the additive effect of the *cis*-QTL *l* of the gene *k*, q_{il} is the indicator variable of the genotypes at *l*, σ_t is an error that will be transmitted to any genes down-regulated by the gene *k*, whereas σ_e is a non transmitted error. We set-up σ_t = 0.02 and σ_e = 0.98. The gene effects *c* were all linear

with absolute effects uniformly drawn from 0.1 – 0.3, with 2.5% of them negative and the rest, positive so the network was strongly directionally biased. Additive effects of QTL were always 5 % of the phenotypic variance conditional on the network component, i.e., $\hat{y}_i - \sum_j c_{jk}\hat{y}_{ij}$. This results in an effect size that is negatively related to the topological rank and ranged between 5% and 1% of the variance in expression. Signs of the QTL effects (σ) were either always positive or 50% positive.

In Sim2, the expression levels were drawn from $y_i \sim MVN(aq, S)$, where $\mathbf{y}_i$ is the vector of gene expression levels of the individual *i*, **a** is the vector of additive effects. Each *cis*-locus had an effect of 5% the phenotypic variance. Two cases were considered, either all effects were positive or 50% of the effects were positive and 50%, negative; **q** is the indicator vector of genotypes of the individual *i* at the *cis*-QTLs. **S** is a ten-block covariance matrix with high correlation among genes within a block (r = 0.8) and null correlation among genes in different blocks, and diagonal values σ_e equals to 1. The number of genes in each block was 100.

We computed power as the proportion of significant linkage ($P \leq 10^{-3}$) at the gene positions. The FDR was estimated as the proportion of significant ($\leq 10^3$) genome-wide maximum scores that were not linked given a genetic distance between the transcript and the eQTL greater than 15 cM.

D. Results

eQTL show strong tissue-specific patterns

We found that 28%, 41%, 28% and 39% of transcripts have eQTL among 23574 transcripts on the arrays in adipose, hypothalamus, liver and muscle, respectively (**Figure 2.1**). Of these eQTL, 1432 (21.47%), 1584 (23.88%), 1983 (20.08%) and 2085 (22.72%) were classified as cis-acting in each respective tissue (Figure 2.1, Diagonal).

Across the tissues, 739 (5.05%) transcripts have eQTL in all four tissues, and 330 of these are in similar positions in every tissue (**Figure 2.2**). The heatmap for eQTL (Figure 2.1) strongly suggests that eQTL distribute across the whole genome in a tissue-specific pattern. This tissue-specific distribution pattern is especially evident for the eQTL hot spots as demonstrated by the vertical lines (Figure 2.1). Strikingly, there are no overlapping hot spot clusters across the tissues.

Characterization of eQTL hot spots within and across tissues

Applying our definition of an eQTL hot spot to the four tissues, we identified a total of 11 independent hot spots (**Table 2.1**). Among them, three were in adipose, three in hypothalamus, one in liver and four in muscle. The number of eQTL within each hot spot vary dramatically, with the strongest encompassing 2112 eQTL on MMU17 in hypothalamus and the weakest having 139 eQTL clustered on MMU19 in adipose (Table 2.1).

To further understand the genetic architecture of eQTL hot spots, we tested whether the transcript eQTL within each hot spot have predominantly additive or dominance gene action. Interestingly, as shown in Table 2.1, some hot spots have eQTL whose effects are nearly all additive in nature (MMU17 in hypothalamus), some are strongly biased towards dominance gene action (MMU8 in muscle), while others have nearly equal representation (MMU7 in hypothalamus). Cumulatively, however, there is significantly more additive eQTL in hot spots than eQTL with dominance effects.

In addition, we characterized the direction of allelic effects for eQTL localized within hot spots, articulating whether an allele inherited from the M16 parental line increases or decreases mRNA abundance relative to an allele from the parental ICR line (Table 2.1). We also evaluated the direction of dominance. In general, direction of allelic effects was relatively evenly distributed across the two parental lines, although this was

hot spot specific. For example, 90% of the additive eQTL in the adipose hot spot on MMU9 had increased expression values due to the M16 allele, while the opposite was seen for the muscle hot spot on MMU15. And for the strongest hot spot in hypothalamus on MMU17, nearly the same number of eQTL had increased expression values due to the ICR and M16 lines. In regard to dominance, the most interesting finding was for the muscle hot spot on MMU8, where nearly all the eQTL showed directional dominance based on the ICR allele.

Most hot spots arise due to high correlations between transcripts

Based on permutation testing (**Figure 2.3**), only one of the hot spots (MMU17 in hypothalamus) was significant. For the hot spot on MMU17 in hypothalamus with 2112 eQTL, the P value was equal to 0.01 that this hot spot did not arise by chance after the association between marker genotype and expression was artificially broken, while maintaining the correlation structure between transcripts. The P values for other hot spots were 0.9, 0.6 and 0.25 for the strongest within-tissue hot spots on MMU7 in liver, MMU15 in muscle and MMU2 in adipose, respectively. These results provide model-free evidence that most hot spots arise because of correlation structures among transcripts. The average pair-wise correlation among the transcripts with eQTL in the hypothalamus hot spot on MMU17 was 0.64. This was significantly higher ($P < 10XE-30$) than the pair-wise correlations calculated for sets of 1000 random transcripts, where the highest average correlation obtained was 0.45 using 10000 permutations.

Evidence from network modelling

In order to reduce computational burden, only the set of genes that make up the same hot spot as that of gene *A* were considered as putative covariates for a gene *A*. As case studies, we analyzed three hot spots: the strongest one on MMU17 in

hypothalamus, and two in muscle (MMU15 and MMU17). For this specific analysis, these two muscles hot spots were evaluated together as a single set of genes (in preliminary analyses modelling was done on the two gene sets independently, but results were similar what is presented here because selected covariates for a gene were mostly from the same set of genes when the two sets are analyzed simultaneously). Modelling was initialized using first-order partial correlations[107] with the null hypothesis of zero correlation rejected at $P \leq 10^{-2}$. Sex-specific expression levels ($P \leq 10^{-2}$) were identified only for a few genes in the hypothalamus hot spot (2.3%), whereas 60.5% of the genes in the muscle hot spots exhibited sex-specific expression (69.7 % for muscle MMU15 and 48.7% for muscle MMU17). A gender effect was therefore incorporated for these genes.

The NET model yielded a dramatically different picture compared to the classical mapping **(Figure 2.4**C, 2.4D versus Figure 2.4A, 2.4B): i) the large hot spots disappeared; ii) the *cis* diagonal was highlighted instead, reflecting the increase in the number and in the significance of *cis* QTL as compared to model 0; iii) few of the remaining *trans*-eQTLs showed highly significant scores. The number of eQTL was reduced by approximately 30% (**Table 2.2**). However, with stricter significance thresholds, the NET model detected up to 4-fold more eQTL than the traditional M0 model in the hypothalamus gene list and in the muscle gene list (**Figure 2.5A**). In contrast, the NET model had no effect whatsoever on the muscle MMU17 gene list, as the ratio of detected eQTL between models M0 and NET was roughly 1. However, of the 111 genes with eQTL at the hot spot, 42 were actually *cis*-eQTL in the classical mapping. About 16% of the genes in the MMU17 gene list are located on chromosome 17, 22 of the genes are physically located within a 2 Mb region of the eQTL hot spot, and 18 additional genes are within 15 Mb of this region.

Overall, the number of *cis*-eQTL dramatically increased (2 to 4-fold) with the NET model in the hypothalamus and muscle MMU15 hot spots, whereas no gain was observed on the muscle MMU17 gene list (Figure 2.5B). In the hypothalamus, this gain was from 128 to 336 in significant *cis*-eQTL, representing an increase from 7.2% to 29.2% out of the total number of significant eQTL (Table 2.2). This increase was consistent whatever the distance used to decide whether the eQTL was proximal or distal to the gene (**Figure 2.6**). Similarly, a two-fold increase was observed in the muscle hot spot on MMU15. The number of *cis*-QTL increased from 21 to 51 (6.5 to 19.7%) in the MMU15 gene list, whereas there was no increase in *cis*-QTL for genes in muscle hot spot on MMU17 in model M0 vs. NET model (72 vs. 68 *cis*-eQTL, or 32.7 vs. 32.8% of the total number of significant eQTL).

Evidence from simulations with high correlations between transcripts

Gene network

Simulations of expression levels of 1000 genes controlled by one *cis*-eQTL per gene and a complex gene network yielded a clearly visible pattern of eQTL clustering when a naïve QTL model is employed (**Figure 2.7A**). Moreover, the NET model consistently removed this hot spot pattern while highlighting the *cis*-signals (Figure 2.7C). The NET approach did not increase the power of the analysis as the power changed from 0.59 ± 0.16 to 0.58 ± 0.16 when the QTL effects were either positives or negatives, and from 0.79 ± 0.16 to 0.65 ± 0.16 when these effects were all positives. The rate of false discovery was, however, consistently reduced from 27.1% ± 1.8 to 24.3% ± 0.99, positive and negative QTL effects, and from 31.3% ± 2.2 to 24.1% ± 1.2 when all QTL effects were positive.

Block correlation matrix among gene expressions

The second type of simulations was based on a block-correlation matrix among gene expressions of 1000 genes (100 genes in each of ten blocks). As in the previous simulation scenario, the traditional mapping resulted in eQTL hot spots whereas the NET approach removed the hot spot structure (Figure 2.7B and 2.7D). Nonetheless, these hot spots were less frequent than when a network was simulated. Modelling did not affect power: it varied from 0.57 ± 0.27 to 0.61 ± 0.24 when the QTL effects were either a mixture of positives and negatives, and from 0.64 ± 0.29 to 0.59 ± 0.23 when they were all positives. The rate of false discovery was only marginally reduced with the NET model when the QTL effects were either positives or negatives, 13.7 % ± 1.4 vs. 11.1 % ± 0.9 with M0 and NET modelling, respectively. No improvement in FDR was observed when all QTL effects were positive (M0: 13.4 % ± 1.6 and NET: 13.7 % ± 1.4).

E. DISCUSSION

Clusters of colocalized trans-eQTL, otherwise known as trans-bands or transcriptional hot spots, have been identified in many studies in variety of organisms, including yeast, arabidopsis, rat and mouse[92]. The presence of such hot spots is exciting for two reasons. First, an eQTL hotspot that overlaps with a QTL for an important clinical trait could provide a biological framework as to how the locus impacts that trait while also identifying potentially causal positional/functional candidates for the underlying genetic variation. Secondly, understanding how eQTL arise, and how genetic variability in a single locus can co-regulate phenotypic variability in mRNA abundance for a large numbers of genes, may shed light on regulation of transcriptional control in general. Several models for how eQTL hot spots function have been proposed (e.g. Sun et al (2007); Mozhui et al (2008); Wu et al (2008)[93-95].

Most eQTL hot spots were discovered by classical QTL mapping techniques in segregating crosses in various different tissues. By only modelling genetic variation and correlated phenotypes as covariates, the classical mapping algorithms ignore regulatory effects among transcripts[96]. However, the regulation mechanisms for gene expression are very complex and dynamic in any biological system. One instance is the regulation of expression levels between one gene and another. While traditional QTL mapping algorithms give us tremendous insight about complex traits and about genetic regulation mechanism[102], the statistical modelling underlying the algorithm is far from perfect because regulation mechanisms such as regulation between one transcript and another, hormonal or neuronal regulations have not been considered at all. As a result, detection of hot spots can be affected by different modelling algorithms. Further, existence of hot spots can also be largely influenced by correlation among genes[92]. The rational is that if one transcript is associated with a SNP, then many transcripts that are highly correlated with this original transcript may be also be represented as eQTL within the same locus.

To further explore this issue, we first performed permutation analysis where genotypes were permuted while conserving gene expression phenotypes and the internal correlation structure of the data, as proposed by Breitling et al (2008). Through this permutation strategy, we tested the hypothesis correlation amongst transcripts can lead to the detection of eQTL hotspots. Our results support this hypothesis, in providing model-free evidence that nearly all hot spots in our segregating F2 cross arise because of correlation structures among transcripts.

The most commonly used mouse resources, including the lines used in the present study, are from Mus musculus *domesticus* and harbor only a fraction of the genetic diversity of Mus musculus, which is not uniformly distributed[108]. It is possible that large regions of identity-by-descent and repeated SNP distribution patterns across the genome due to shared ancestry lead to non-biologically relevant phenotypic correlations

in this and other crosses where eQTL hot spots have been reported. Indeed, an early study with emerging lines of the Collaborative Cross (CC) mouse resource, which captures significantly more genetic diversity with no blind spots and has a much more uniform distribution of genetic variation, shows that many common phenotypic correlations are disrupted relative to their eight progenitor strains (D Pomp, unpublished data). And interestingly, a liver eQTL study in these same emerging CC lines found prevalent strong cis eQTL signals, but a complete lack of any eQTL hot spots (G Churchill and F Pardo Manuel de Villena, personal communication). While this may partially be a result of relatively underpowered studies, it is intriguing to speculate that the lack of eQTL hot spots in the emerging CC lines is caused by the natural lack of correlation structure amongst transcripts in that diverse resource, supporting the present findings where such hot spots are no longer present once this correlation structure is uncoupled from the structure of the underlying genetic variation.

While permutation analysis showed that hot spots are caused by high correlation between transcripts, we also applied NET or structural modelling for an additional robust manner to address this issue. We proposed a systematic, network based approach to deal with modelling in eQTL studies; note that the NET approach can be useful in many other instances where large scale multivariate data are available, but is of particular relevance for eQTL mapping. Interestingly, two simulation strategies mimicked the experimental results, and stress the relevance of more sophisticated models than those currently employed. One of the important advantages of NET over classical modelling is that we were able to obtain an increased frequency of cis-eQTL with more significant *p* values, whereas the main picture of eQTL hot spots obtained using traditional eQTL mapping was removed.

There are drawbacks to the NET modelling, however. Accounting for gene correlations alone could not completely remove ambiguities in edge direction and the

proportion of false direction of edges is likely high. A further step will be to simultaneously integrate gene covariate selection and QTL selection. Integrating QTL in the building of gene networks may dramatically improve eQTL power and decrease FDR. This approach provides anchors with unambiguous edge directionalities and could help to direct the links between phenotypes in the network[109, 110]. However covariate modelling has a profound effect on QTL mapping as we show here.

Bagging, and more generally model averaging techniques, provides a sound basis to incorporate model uncertainties in the building of such networks. Their statistical foundation for the problem of multiple QTL selection has been published recently[111] and their extension to the problem of multiple co-regulated traits as initiated here should be straightforward. Its computational cost is no longer a limitation due to end-user parallelization tools and increasing CPU speed[112].

Structures in data arise from various sources (complex pedigree, historical contingencies, biological correlation or analytical biases), and they can have a deep impact on QTL discoveries. Therefore to be successful, approaches dealing with QTL to decipher gene networks should probably not rigidly fix the QTL found using simple models in the first instance, but should modify them accordingly and simultaneously to the reconstruction of the gene network.

Table 2.1. Descriptive statistics of eQTL hot spots.

Tissues	Locations	Total	Add	Dom	Both	ICRadd	M16add	ICRdom	M16dom
Adipose	MMU2, 63-67cM 144-153Mb	373	127	33	0	47	80	11	22
	MMU9, 17-21cM 45-52Mb	166	100	2	0	90	10	0	2
	MMU19, 0-2cM 3-9Mb	139	36	7	0	19	17	7	0
Hypothalamus	MMU4, 17-21cM 47-57Mb	233	194	3	0	96	98	1	2
	MMU7, 29-33cM 78-90Mb	341	50	49	0	14	36	41	8
	MMU17, 39-43cM 79-83Mb	2112	2054	4	1	1059	995	3	1
Liver	MMU7, 70-72cM 147-151Mb	170	152	0	0	98	54	0	0
Muscle	MMU2, 45-49cM 95 - 107 Mb	288	43	65	1	15	28	40	25
	MMU8, 11-15cM 30-40Mb	167	10	127	0	7	3	117	10
	MMU15, 16-20cM 52-59Mb	393	187	4	0	28	159	2	2
	MMU17, 7-11cM 30-41Mb	298	137	43	1	76	61	22	21

Total: total number of eQTL falling into a 5 cM window during eQTL mapping; Add: number of eQTL with additive effects; Dom: number of eQTL with dominance effects; Both: number of eQTL with both additive and dominance effects; ICRadd: the allele from ICR increases the expression of a transcript; M16add: the allele from M16 increases the expression of a transcript; ICRdom: the allele from ICR is dominant relative to the allele from M16; M16dom: the allele from M16 is dominant relative to the allele from ICR

Table 2.2. Number of eQTLs and *cis*-eQTLs detected at P ≤ 10^{-3}

Tissue	Hot spot	Model	N eQTLs	(%)*	N *cis*-eQTLs	(%)†	NET/M0‡	*cis*-NET/*cis*-M0
Hypo	MMU17	M0	1,785	(84.5)	94	(5.3)		
Hypo	MMU17	NET	1,149	(54.4)	264	(23.0)	0.64	2.81
Muscle	MMU15+ MMU17	M0	539	(78.5)	72	(13.3)		
Muscle	MMU15+ MMU17	NET	462	(67.2)	104	(22.5)	0.86	1.44
Muscle	MMU15	M0	322	(81.9)	14	(4.3)		
Muscle	MMU15	NET	259	(65.9)	43	(16.6)	0.80	3.07
Muscle	MMU17	M0	220	(73.8)	58	(26.4)		
Muscle	MMU17	NET	207	(69.5)	61	(29.5)	0.94	1.05

*Percentage of the number of eQTL detected on the total number of genes (see Table 2.1)

†Percentage of the number of *cis*-eQTL detected among the total detected eQTL. QTL are annotated as *cis*-eQTL if they are at most 5cM away from the gene they act on.

‡Ratio of detected eQTL between the NET model and the M0 model

Figure 2.1. Trans-bands show tissue-specific patterns
Each panel represents eQTL mapping result in a tissue. A red dot indicates a significant QTL for a transcript. The X-axis denotes the QTL location for a QTL while the Y-axis describes the physical location of a transcript associated with this QTL. Diagonal lines indicate the cis-acting QTL and vertical lines suggest the trans-bands.

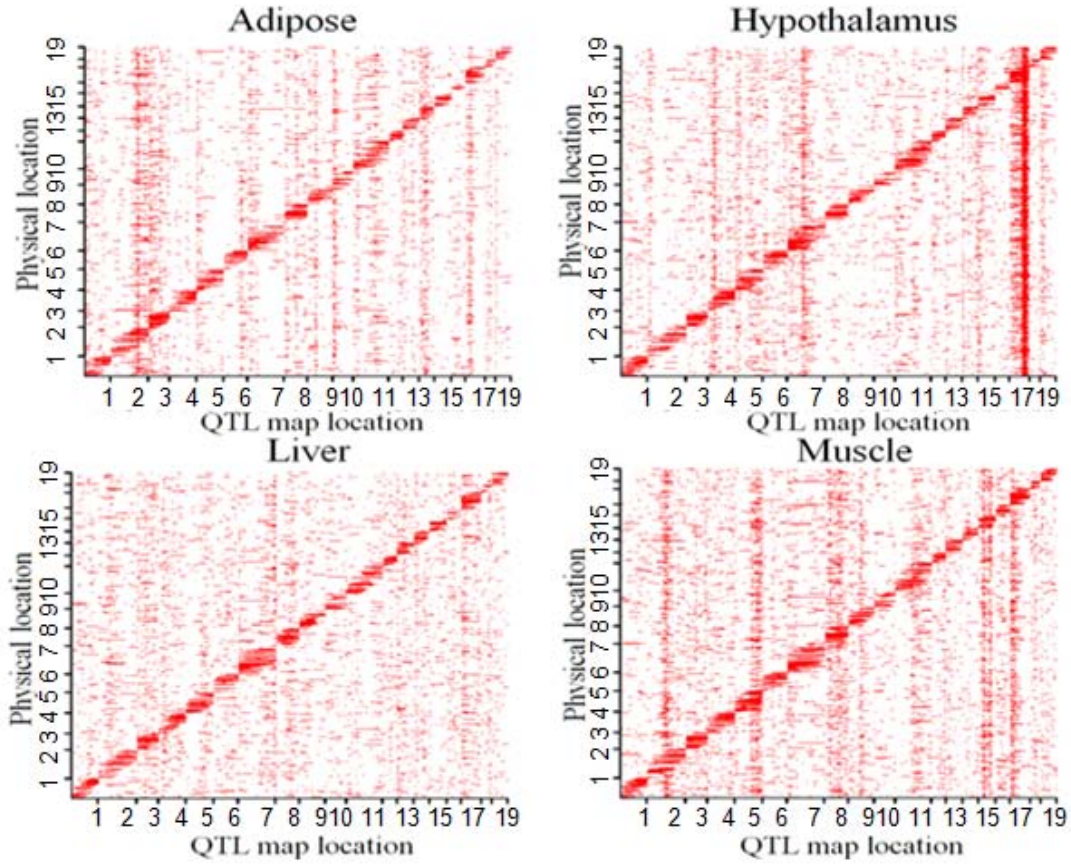


Figure 2.2. QTL sharing among four tissues
There are total 14624 transcripts having QTL in all four tissues. Each bar indicates the number of transcripts of which each transcript has common QTL shared by tissue number denoted by the X-axis.

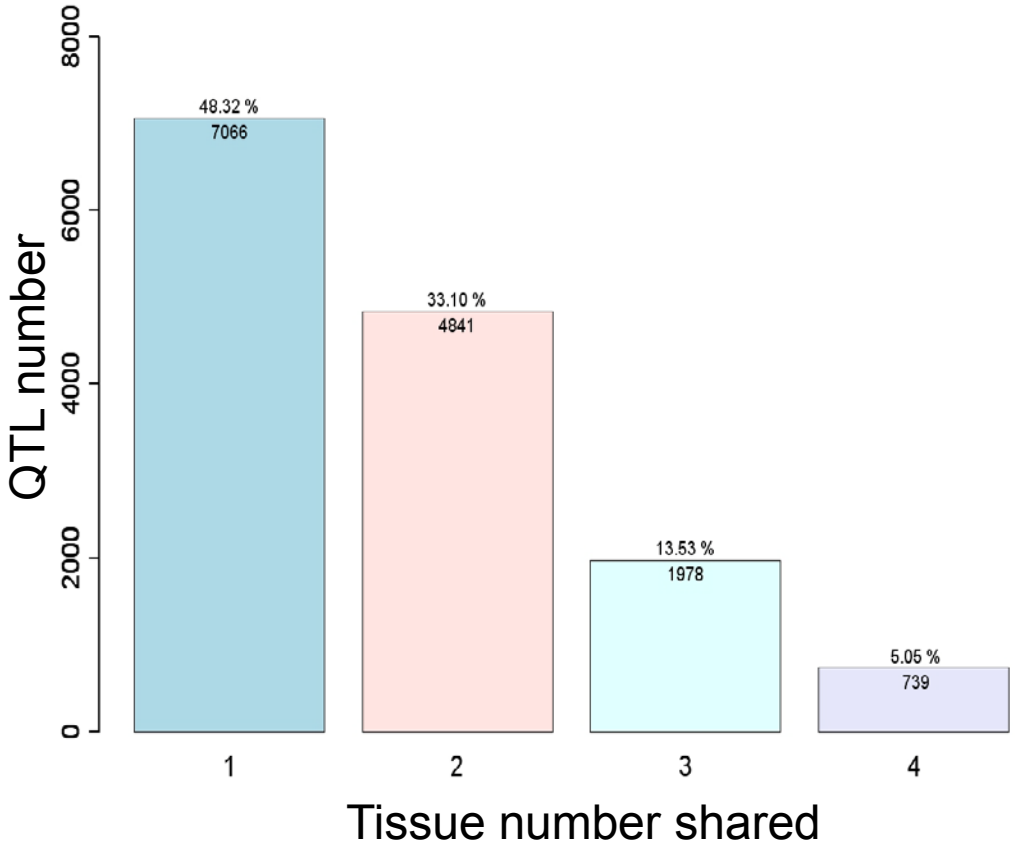


Figure 2.3. Hot spot strength distribution under permutation
We performed 1000 individual permutations. In each permutation, the eQTL mapping was conducted. The X axis is a vector in which each element represents the maximal number of eQTL clustering in a 5 cM window after the eQTL mapping from a permutation. The height of each grey bar indicates the frequency in which the maximal number of eQTL clustering repeats in one thousand permutations. The red bar shows the observed maximal number of eQTL clustering in a 5 cM window without permutation. The inlet table on each panel shows the maximal number of eQTL clustering and the probability to observe this number without permutation.

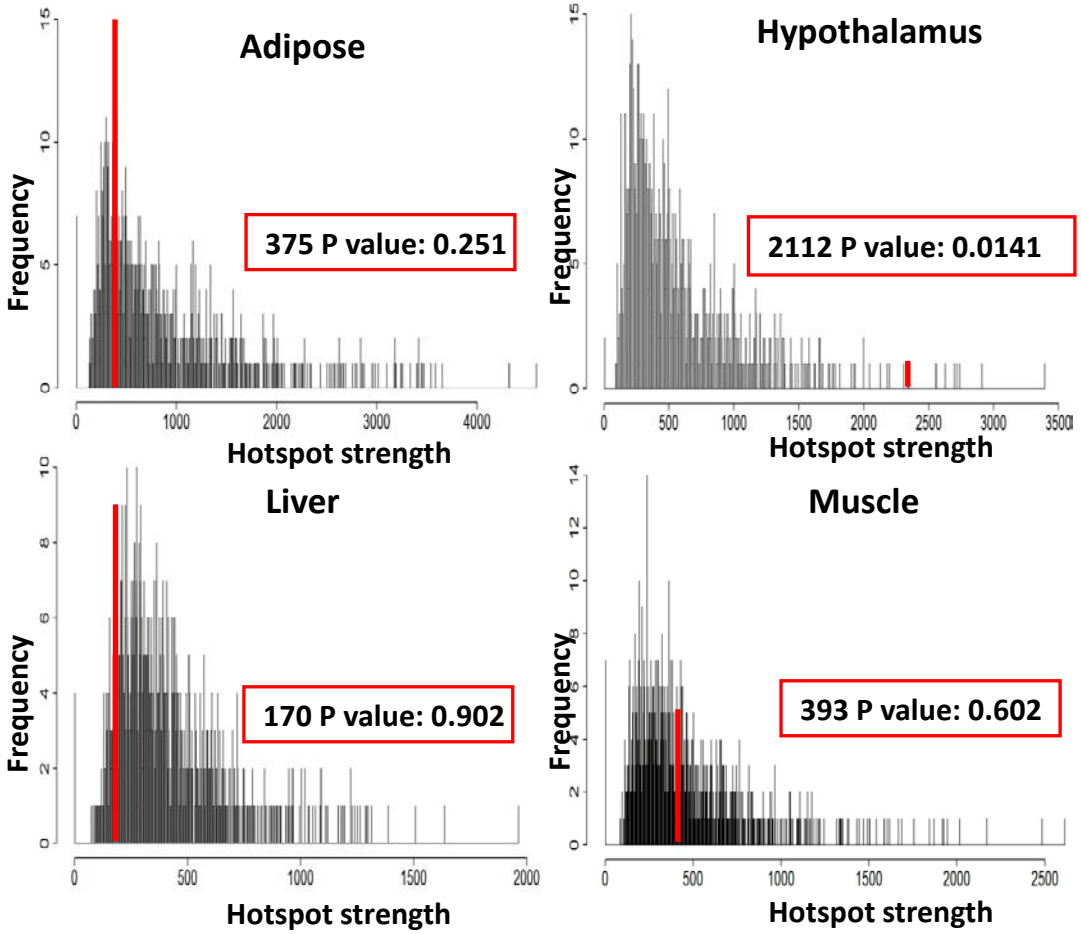


Figure 2.4. Genome-wide profiles of significant $-\log_{10} p$ ($P \leq 10^{-3}$) for the two tissues.

The x-axis is the location in cM of the QTL, and the y-axis, the position of the gene in Mb. Only regions with $P \leq 10^{-3}$ are plotted. The model used here is different to the one used for Fig 2.1 and analyzed only a subset of the data shown in Fig 2.1.

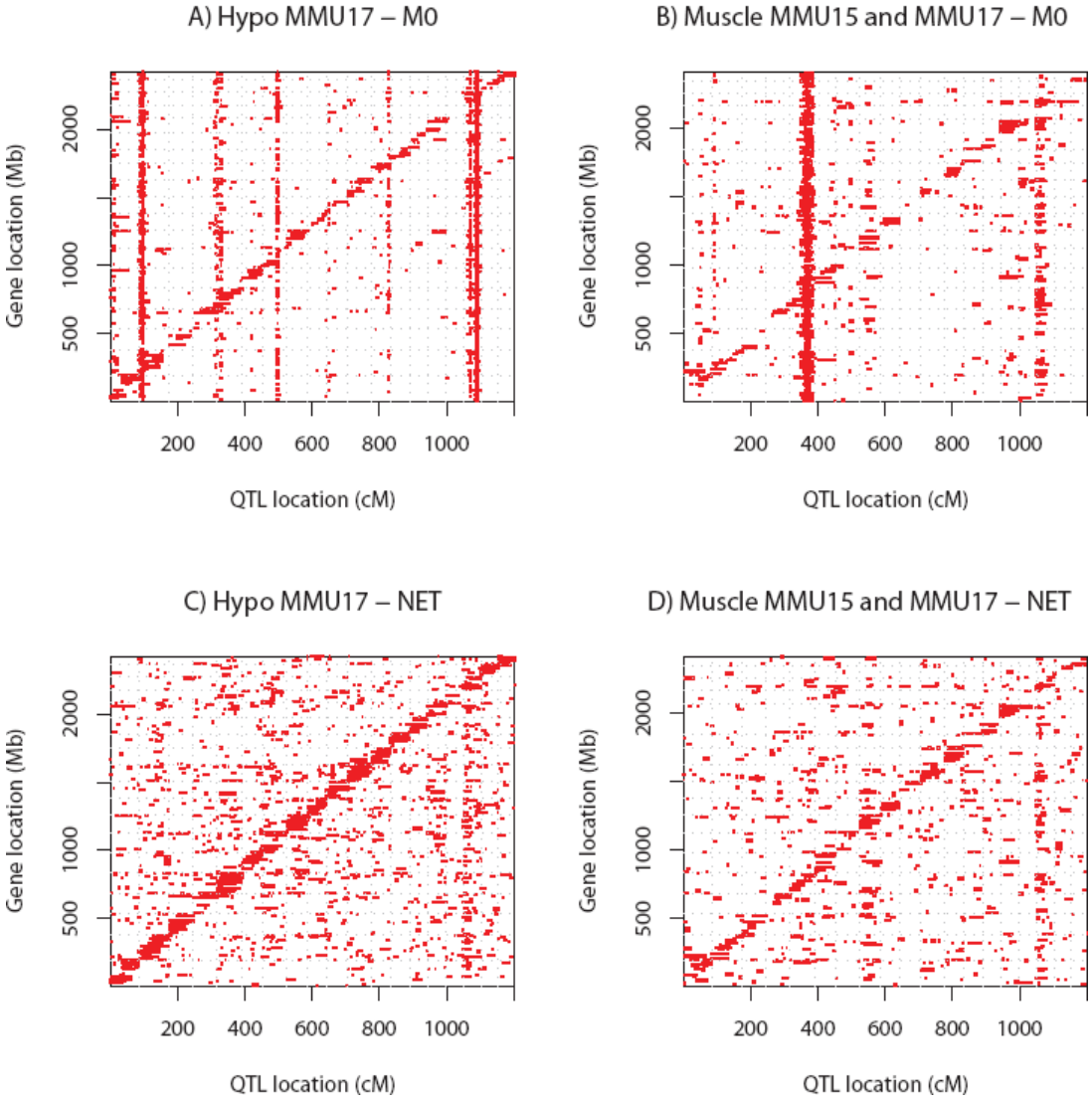


Figure 2.5. Ratio of the number of all eQTLs (A) or *cis*-eQTLs (B) between models M0 and NET given the significance threshold on the eQTL scores.
Hypothalamus results are the plain line, the muscle MMU15 results are the dashed line, and the dotted line refers to the muscle MMU17 results. A ratio of 1 means neither gain nor loss in the discovery of eQTLs. A ratio > 1 means that a larger number of eQTLs was detected with the NET model. The eQTL distant of at most 5 cM of the gene they act on are annotated *cis*-eQTL.

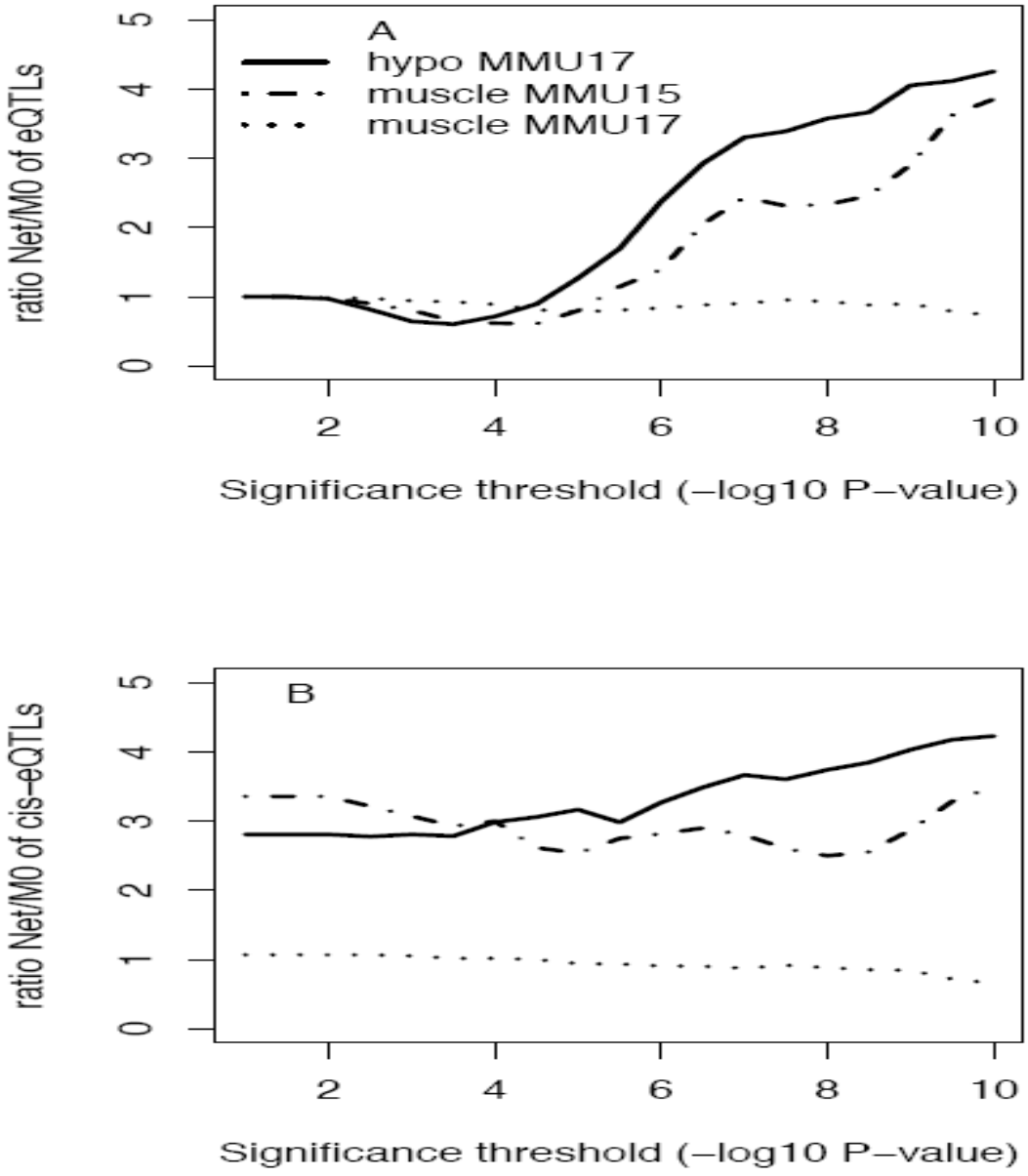


Figure 2.6. Accumulation curves of cis-eQTL given the thresholding distance in M0 and NET models.

The plain line is the M0 model and the dashed line is the NET model. The x-axis is the distance (cM) between the QTL and the gene, and the y-axis, the number of eQTL declared *cis* (The QTL that are located closer than or equal to the distance). The gray vertical bar shows the used threshold of 5 cM. The increase of *cis*-eQTL detected using the NET model compare to the classical M0 model is consistent whatever the linkage distance that would be chosen.

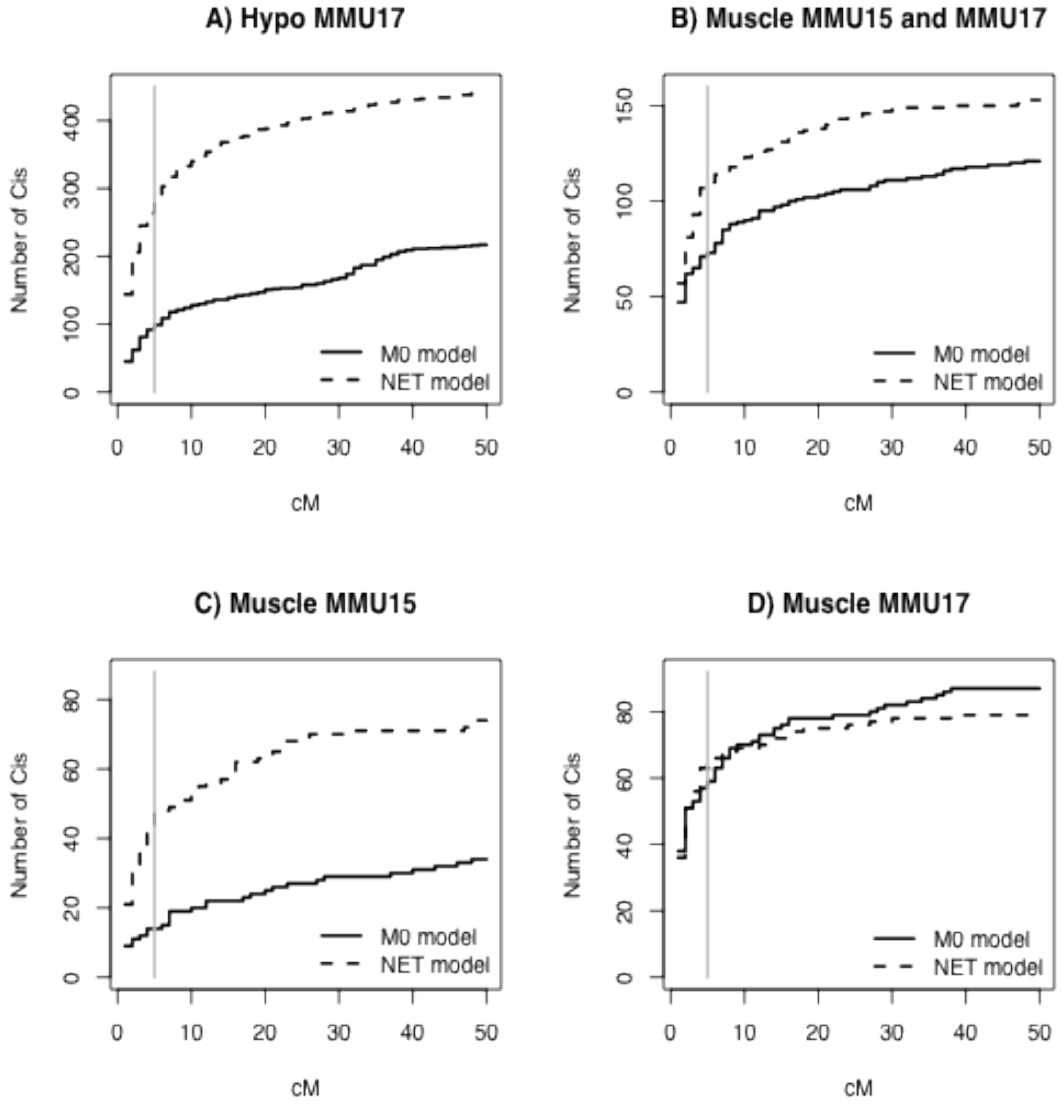
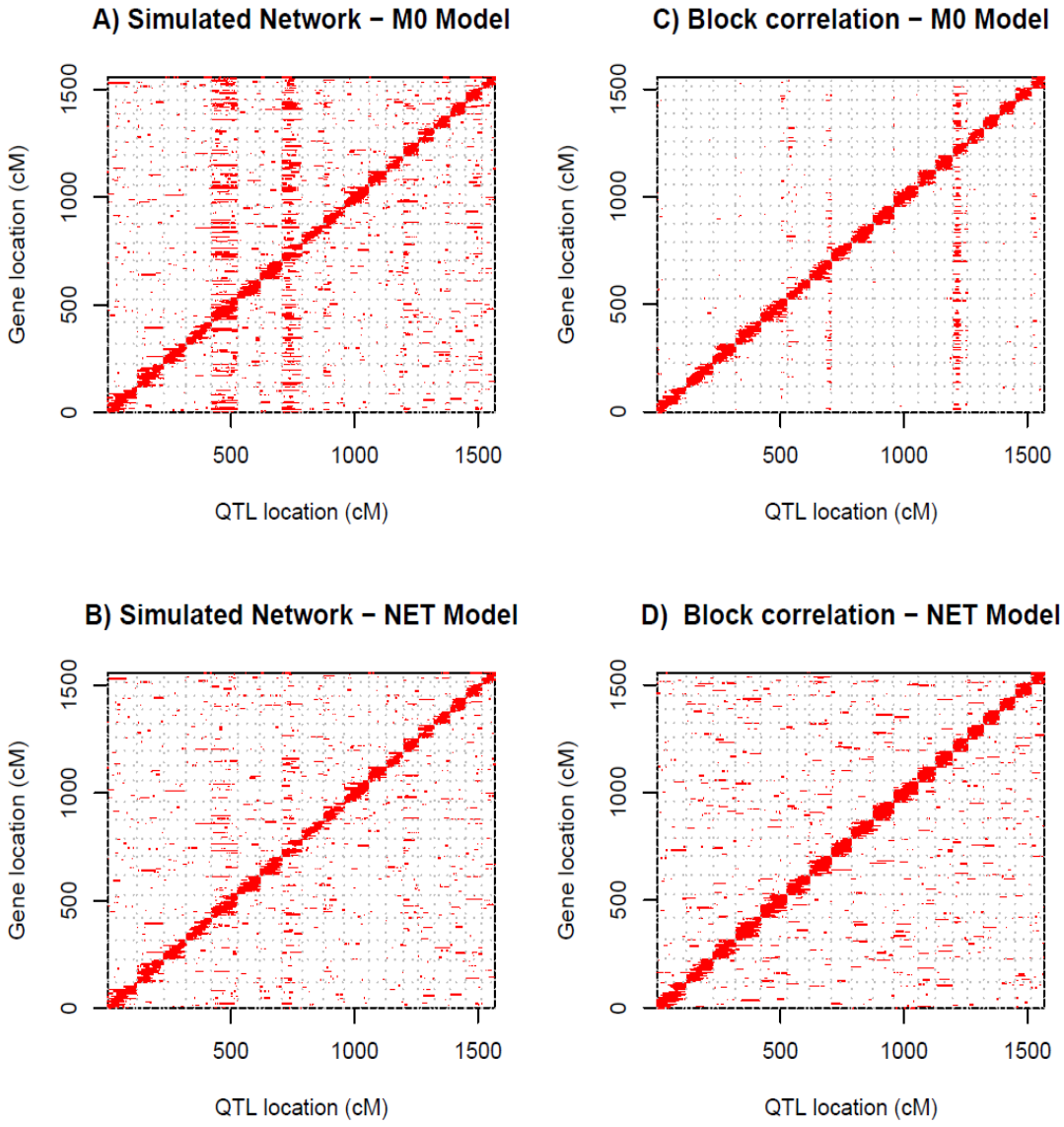


Figure 2.7. Genome-wide profiles of significant $-\log_{10}$ *p* ($P \leq 10^{-3}$) for the 2 simulations.

A and B) Simulation was done with gene regulatory network. C and D) Simulation of 10 blocks of 100 highly correlated gene expression levels. A and C) Classical eQTL mapping. B and D) NET mapping. The x-axis is the location in cM of the QTL, and the y-axis, the position of the gene in cM. Only regions with $P \leq 10^{-3}$ are plotted.



CHAPTER III

GENETIC ARCHITECTURE UNDERLYING VARIATION IN EXTENT AND REMODELING OF THE COLLATERAL CIRCULATION

The contents of the following chapter were originally published by the open access publisher Biomed Central (3)

A. Abstract

Rationale. Collaterals are arteriole-to-arteriole anastomoses that connect adjacent arterial trees. They lessen ischemic tissue injury by serving as endogenous bypass vessels when the trunk of one tree becomes narrowed by vascular disease. The number and diameter ("extent") of native (pre-existing) collaterals, plus their amount of lumen enlargement (growth/remodeling) in occlusive disease, show remarkably wide variation among inbred mouse strains (eg, C57BL/6 and BALB/c), resulting in large differences in tissue injury in models of occlusive disease. Evidence suggests similar large differences exist among healthy humans.

Objective. To identify candidate loci responsible for genetic-dependent collateral variation.

Methods and Results. Cerebral collateral number and diameter were determined in 221 C57BL/6-x-BALB/c F2 progeny, followed by linkage analysis to identify quantitative trait loci (QTL) for collateral number and diameter. Four QTL were obtained for collateral number, including epistasis between two loci. A QTL that was identical to the strongest QTL for collateral number on chromosome 7 (LOD=29, effect size=37%) was also mapped for collateral diameter (LOD=17, effect size=30%). Chromosome substitution strain analysis confirmed this locus. We also obtained a unique QTL on chromosome 11 for collateral remodeling after middle cerebral artery occlusion. Association mapping within the chromosome 7 QTL interval using collateral traits measured for 15 inbred strains, delineated 172k (p=0.00002) and 290k (p=0.0004) base-pair regions on chromosome 7 containing 2 and 7 candidate genes, respectively.

Conclusions. We conclude that collateral extent and remodeling are unique, highly heritable complex traits, with one QTL predominantly affecting native collateral number and diameter.

Key Words: Collateral vessels, genetics, quantitative trait loci, cerebral circulation, arteriogenesis

B. Introduction

Atherosclerotic, atherothrombotic and thromboembolic occlusive vascular diseases constitute the primary cause of morbidity and mortality in developed countries. Many physiological systems are concomitantly recruited, albeit with significant inter-individual variation, which lessen the accompanying ischemic tissue injury. Among these, three vascular protective mechanisms are paramount: (1) the number and diameter of arteriole-to-arteriole anastomoses present in the tissue before the onset of disease that cross-connect occasional distal-most arterioles of adjacent trees (ie, the "native collateral extent"), (2) an anatomic increase in lumen diameter and wall thickness of these vessels caused by obstruction of flow to one of the trees—a process termed collateral remodeling, collateral growth or arteriogenesis, and (3) ischemic angiogenesis, ie, the sprouting of additional capillaries[1, 7, 11, 113-115]. Arteriogenesis, which requires days-to-weeks to achieve up to an approximately 10-fold increase in diameter depending on tissue and species, occurs when perfusion of an adjacent tree is chronically reduced below a critical level. This increases flow-dependent shear stress on collateral endothelial cells, which in turn initiates an inflammatory-like remodeling process. It is becoming recognized that the extent of the native collateral circulation and collateral remodeling have significantly greater impact on restoring blood flow to ischemic tissue than does angiogenesis—which can only increase dispersion of whatever flow is provided by the collateral network[1, 113, 115].

The severity of the clinical manifestations of occlusive vascular diseases, ie, myocardial infarction, stroke and chronic ischemic disease of the heart, brain and lower

extremities, have long been known to vary widely among individuals. This is presumably due to variation in environmental influences and risk factors as well as genetic mutations and polymorphisms affecting the extant pathological and physiological processes. Although much effort has focused on identifying the sources of this variation, the possibility that differences in extent of the native collateral circulation exist among healthy individuals and contribute significantly when disease strikes has, until recently, received little attention. Meier and colleagues[24] found, using an index primarily dependent on collateral density and diameter, that coronary collateral conductance varied by approximately 10-fold in healthy human subjects. Since these subjects lacked coronary artery disease, the variation could not be affected by differences in stenosis-induced collateral remodeling. Studies employing dynamic angiography in patients with acute stroke from middle cerebral artery obstruction, where differences in remodeling can again be ruled out due to the acute nature, suggest that wide variation in native collateral conductance is also present in the cerebral circulation[27-30]. Likewise, evidence exists for significant variation in collateral extent in the peripheral limbs of healthy individuals[116, 117].

Recent studies in mice suggest that a substantial amount of this variability arises from naturally occurring polymorphisms in genetic loci involved in molecular pathways directing formation of the collateral circulation, which occurs late embryonically and early postnatally in mice[38]. Native collateral density and diameter evidenced large differences in multiple tissues in adult C57BL/6 and BALB/c mouse strains[16, 118, 119]. In fact, C57BL/6 and BALB/c mice have the largest extremes in collateral extent in the cerebral circulation among 15 strains, for example exhibiting a greater than 30-fold difference in collateral density[16]. Importantly, these differences in collateral extent in the cerebral circulation are shared qualitatively by other tissues, including skeletal muscle and intestine[16, 34, 37]. Moreover, the rank-order of cerebral collateral extent among 15

mouse strains closely predicts the rank-order of severity of stroke (infarct volume) after permanent middle cerebral artery occlusion[12]. This finding establishes a major role for genetic variation in the extent of the collateral circulation in determining variation in severity of ischemic tissue injury. Interestingly, collateral remodeling during stroke also evidenced significant variation, but with a strain-specific pattern significantly different from the pattern associating infarct volume with collateral extent[12]. This suggests that different pathways are responsible for collateral formation versus collateral remodeling, and that genetic variation in them is likely to be mediated by different loci.

Nothing is known about the genetic loci responsible for the remarkably wide variation in native collateral extent and collateral remodeling described above. However, studies employing gene targeting methods have identified two genes, *Vegfa* and *Clic4*¸ whose expression positively regulates collateral formation in the embryo and thus collateral extent in the adult; one of these (*Vegfa*) also participates in collateral remodeling[34, 37]. Hence, polymorphisms in these genes are *a priori* candidates for causing variation in collateral circulatory function. Identification of the alleles underlying individual variation in collateral extent and remodeling is important not only to help define the responsible signaling pathways for these two processes, but also to allow assessment of risk-severity if occlusive vascular disease develops and to permit stratification of individuals enrolled in clinical trials testing new collaterogenic therapies. Therefore, in the present study we sought to identify genetic loci governing variation in collateral extent and remodeling. To achieve this goal we created 243 F2 mice intercrossed between the C57BL/6 and BALB/c strains and performed linkage analysis to identify quantitative trait loci (QTL). We then used association mapping among 15 inbred strains to identify candidate genes responsible for variation in collateral function.

C. Materials and Methods

Animals

F1 progeny obtained from reciprocal matings of C57BL/6J (B6) and BALB/cByJ (BALB/c, Bc) were mated to produce an F2 population. Chromosome substitution strains were C57BL/6J-Chr7A/J/NaJ (CSS7) and C57BL/6J-Chr17A/J/NaJ (CSS17).

Phenotyping

Mice were subjected to right-side middle cerebral artery occlusion (MCAO) and collateral number and diameter measured 6 days later[12]. To image the pial circulation, the rostral vasculature was maximally dilated, filled with yellow Microfil™ with viscosity adjusted to prevent capillary filling, and fixed (paraformaldehyde). Collaterals connecting the anterior cerebral (ACA) and middle cerebral (MCA) artery trees in both hemispheres were digitally imaged. Cortical territories supplied by the MCA, ACA and PCA trees were determined[12]. ANOVA or t-tests were used together with Bonferroni correction where appropriate.

DNA isolation and genotyping

Tail genomic DNA was genotyped using the 377-SNP GoldenGate genotyping array (Illumina, San Diego, CA). Manual genotyping was done for one marker (rs32420445). SNP positions were obtained from Build 37.1 of the NCBI SNP database.

Linkage analysis

Collateral traits are defined as collateral number, diameter and "conductance" (number-x-diameter) measured in the left (non-occluded) hemisphere. The remodeling trait is the average fold change in collateral diameter between right and left hemisphere 6 days after MCAO. These traits in the F2 population were subjected to linkage analysis

using single and multiple QTL models in the R statistical program. Thresholds for significant and suggestive QTL were defined as *P*=0.05 and *P*=0.63, respectively.

Association mapping

The efficient mixed model association algorithm (EMMA; 21) was applied in R to collateral number measurements on 15 inbred strains, 8-10 individuals each, obtained previously[12, 79] (see **Supplemental figure 3.1**).

An expanded Materials and Methods section is available in the Online Data Supplement at http://circres.ahajournals.org.

D. Results

Native collateral trait differences between B6 and Bc strains are heritable

We previously showed in multiple tissues that Bc mice have smaller collateral number and diameters than B6 mice[16]. This is striking in the pial circulation, where Bc mice average less than one collateral per hemisphere, compared to 9 in B6. We further showed that among 15 strains Bc had the lowest pial collateral number and near-lowest diameter (13μm)[12]. Among those strains, B6 has the largest collateral diameter (23μm), and nearly the largest number of pial collaterals. Therefore, B6 and Bc were chosen for identification of QTL for native pial collateral number and diameter, for a quasi-index of total collateral "conductance" (defined as number-x-diameter) and for increase in collateral diameter (remodeling) induced by MCAO. MCAO in one hemisphere does not alter collateral number or diameter in the other, non-operated hemisphere[12]. After exclusion of 22 individuals for poor filling or other technical problems, 221 F2 mice were analyzed for native collateral properties and 190 for remodeling. The mice were genotyped at 228 informative SNPs across the genome (see Methods).

Cerebral arteriograms from non-operated B6 and Bc mice are shown in **Figure 3.1A**. The entire population of MCA-ACA collaterals is confined to the pial surface, aiding direct quantification[12]. The Bc brain shown, with no pial collaterals, is typical of most Bc mice. Previously reported values for collateral diameter and number in B6 and Bc[12, 16] were confirmed here: 9 and 22.5um in B6, 0.5 and 12.5um in Bc) (Figure 3.1B).

All 3 collateral parameters were intermediate in the F1 population between those for B6 and Bc (Figure 3.1B) although closer to those for B6. Collateral diameter and conductance in F1 mice were significantly lower than in B6 ($p<0.05$) but were still larger than the average of B6 and Bc values. For collateral number, the trend toward fewer than in B6 was not significant. Thus, collateral traits in B6 are semi-dominant over Bc.

Genome-wide single QTL mapping detects a major effect on chromosome 7 (*Canq1*) for native collateral traits

In the F2 population, all 3 collateral traits were approximately normally distributed, indicative of polygenic traits, with the distributions containing suggestions that they may be sums of multiple normal distributions (Figure 3.1C). These traits were independent of sex, parental origin, or body weight (**Supplemental figures 3.2, 3.3**). Variation in collateral traits was also independent of body weight among 15 inbred strains[12]. Collateral number and diameter were minimally correlated ($r^2=0.13$), confirming previous results[12] and suggesting that number and diameter are largely independent traits (**Supplemental figure 3.4**).

The phenotype-genotype data were subjected to genome-wide LOD score profiling using the single QTL model in R/QTL. This identified a single, highly significant locus on distal chromosome 7 for all 3 collateral traits (LOD scores of 21, 17, and 27 for number, diameter, and conductance, respectively; genetic (heritable) effect sizes$\geq$30%

for all three) (**Figure 3.2A; Supplemental table 3.1**). The peak is located 0.5cM telomeric to marker rs13479513 (134.25Mb; Figure 2A inset). The 95% confidence level is 62-67cM, which is estimated to correspond to 132.4-135.8Mb in physical location. The single QTL model also identified a significant QTL (genome-wide 0.05 significance level) for collateral number on chromosome 1, two suggestive QTL (genome-wide 0.63 significance level) on proximal chromosome 6 and chromosome 10 for number, and two suggestive QTL on chromosomes 1 and 10 for conductance (Figure 3.2A).

Additional QTL and epistasis identified by genome-wide multiple QTL mapping

Experiments using haplo-insufficient and null gene targeted mice have shown that *Vegfa*, *Flt1* (VEGFR1) and *Clic4* expression levels are strong positive determinants of collateral number[34, 37, 39]. However, our single QTL analysis revealed no QTL in the regions of these genes (Figure 3.2A), possibly due to the predominant major effect on chromosome 7. Moreover, the multiple suggestive QTL identified using the single QTL model suggest a complex genetic architecture may underlie variation in native collateral extent. Also, a single QTL model cannot detect epistasis. To address these issues, we subjected the data to genome-wide multiple QTL profiling with increased statistical power using the Stepwiseqtl function in R/QTL.

As shown in Figure 3.2B, this analysis confirmed and strengthened the QTL on chromosome 7 for collateral number, the LOD score improving from 21 to 29 and the effect size increasing from 32% to 37%. It also gave a strong QTL (LOD score 13) at 45cM (90Mb) on chromosome 1, which presumably represents the same genetic influence as the weaker QTL on chromosome 1 (LOD score 3.8, 49cM) in the single QTL analysis. Multiple QTL mapping also identified significant loci on chromosomes 3 and 8 for collateral number (LOD scores of 7 and 5, resp.), and interaction between the QTL on chromosomes 1 and 3 (LOD score 7). Suggestive QTL on chromosomes 6 and 10

were not confirmed. These results are summarized in Figure 3.2B; the inset shows the final model for collateral number. We have named the QTL according to convention by descending order of LOD score as *Canq1-4* (collateral artery number QTL), *Cadq1* (diameter), and *Cacq1-2* ("conductance"). **Table 3.1** summarizes effect sizes, the modes of inheritance (additive or dominant) and their degree, and the statistical significance. We also detected slight segregation distortion between proximal chromosome 10 and distal chromosome 6 ($p<0.03$ by Chi-square test). One possible explanation for these findings is that linkage disequilibrium exists between the loci on chromosomes 6 and 10 and the major effect on chromosome 7, a condition controlled for by multiple QTL analysis.

Like the single QTL analysis, multiple QTL analysis failed to find any effect of *Vegfa*, *Fllt1*, or *Clic4*. To search exhaustively for potential epistases with these genes, we fixed all main effects and the interaction, then searched for interaction between *Canq1* and any other locus. No significant interaction was found (**Supplemental figure 3.5**).

To evaluate the genetic effects of the QTL on chromosome 7 for the 3 collateral traits, the average phenotype values were plotted against genotype for mice grouped by genotype at the marker nearest the peak, rs13479513. Mice homozygous for the B6 allele or heterozygous (B6-Bc) did not differ significantly from B6 or the F1 population, respectively, in any of the traits (**Figure 3.3A, B**). The Bc-Bc group also did not differ from the Bc parental in diameter. Thus, these phenotypes are largely determined by the single B6 allele when present. For collateral number, however, Bc-Bc F2 mice had significantly more pial collaterals than the parent Bc (3.9 vs. ≤0.5, $p<0.001$). The same is true of the derived parameter, "conductance" (56 conductance units vs. 12 for Bc, $p<0.05$). In addition, in our study of 15 inbred strains (18), Bc was significantly lower in pial collateral number than the two closest strains (Bc, 0.2 collaterals per hemisphere;

SWR, 1.3, p=0.0004 vs. Bc; AKR, 1.7, p<0.0001 vs. Bc; **Supplemental figure 3.1**). These results suggest that additional loci, possibly unique to Bc among these strains, are responsible for the very low numbers of collaterals in Bc mice and are consistent with our finding of 3 additional QTL for collateral number but not diameter.

To see phenotypic effects, the genotype at a hypothetical marker located exactly at the peak of each of the QTL on chromosomes 1, 3 and 8 was imputed and the mice grouped according to these genotypes (Figure 3.3B). This shows that the QTL on chromosome 1 acts in the same direction as that on chromosome 7 (ie, to favor greater number and conductance in B6; first and third panels in Figure 3.3B), while those on chromosomes 3 and 8 act in the other direction, albeit weakly (first panel).

The major effect of chromosome 7 is confirmed with chromosome substitution strains (CSS)

The inbred strain A/J is similar to Bc in haplotype structure under *Canq1*. It is also similar in hindlimb necrosis and recovery of hindlimb perfusion following femoral artery ligation[17] and infarct volume after MCAO[18] (see Discussion). These phenotypic characteristics, which gave a QTL in the same position as *Canq1*, were recapitulated in both of these studies in the strain C57BL/6J-Chr $7^{A/J}$/NaJ (CSS7), in which chromosome 7 in the B6 stain has been replaced by chromosome 7 from A/J. We therefore measured native collateral traits for CSS7. New data on A/J mice (n=8) in this study confirmed our previous results[12]; A/J has 65% fewer pial collaterals than B6, and CSS7 is not statistically different from A/J (**Figure 3.4A**). CSS7 has slightly larger collateral diameter than Bc (12.6um vs. 8.6um for A/J, p<0.003), but diameter is substantially smaller than in B6 (12.6 vs. 24.6um for B6, p<0.001; Figure 3.4B). Similarly, CSS7 has 70% lower conductance than B6, compared to 86% lower for A/J and 98% lower for Bc (Figure 3.4C). Hence, CSS7 largely mimics Bc in all 3 native collateral traits, as it does for

necrosis in hindlimb ischemia and cerebral infarct volume following MCAO[17, 18]. However, the significant difference between CSS7 and A/J for diameter and conductance and trend toward significant difference for number suggest that loci on additional A/J chromosomes influence collateral traits. Failure to find such evidence in previous studies using CSS7 analysis for recovery of hindlimb perfusion and necrosis score[17] and infarct volume[18] may arise because these downstream traits are removed from the underlying physiological mechanisms—that are dominated by native collateral extent and collateral remodeling.

Different amounts of cortical territory supplied by each of the main arteries (MCA, ACA, and PCA) are determinants of infarct volume after MCAO[12] and could, in principle, be determinants of pial collateral extent. And A/J mice have a larger percentage MCA territory that could potentially confound interpretation of the A/J and CSS7 data[12]. Therefore, we also examined tree territories. The percentage MCA territory in CSS7 mice was not different from that of B6 and Bc mice (Figure 3.4D). In contrast, chromosome 17 (see below) transferred a significant part of the enlarged MCA territory phenotype of A/J to B6, consistent with the notion that loci on chromosomes other than 7 are responsible for the enlarged MCA tree in A/J mice. These data are consistent with our previous finding that collateral number and diameter show essentially no correlation with tree territory (r^2=0.05 and 0.01, resp)[12].

Since no QTL was found on chromosome 17 for collateral traits, cerebral infarct volume[18], or necrosis score and recovery of perfusion in a hindlimb ischemia model[17], CSS17 was also examined for pial collateral traits as a negative “control” to validate the CSS approach. CSS17 was not different from B6 in collateral number and conductance (Figure 3.4A, C). However, collateral diameter was significantly lower in CSS17 than in B6 (20um vs. 24.6um, $p<0.0002$), but it was still larger than in A/J by a substantial amount (20um vs. 8.6um, $p<0.0001$). These data suggest that the

preparation of the CSS strains did not, per se, cause the Bc-like low values for native collateral traits found in CSS7, thereby reinforcing the conclusion that the main effect for low collateral number, diameter, and conductance resides on chromosome 7. The significant reduction in collateral diameter in CSS17 indicates that chromosome 17 of A/J mice harbors a genetic factor(s) that negatively affects this trait when introgressed onto the B6 background.

Importantly, the above findings offer a physiological basis for the previously identified QTL on chromosome 7 linked to hindlimb ischemia and cerebral infarct volume[17, 18], ie, a gene variant(s) that confers variation in native collateral extent.

Cell-cell signaling and immune response pathway genes are enriched in *Canq1*

We subjected all genes in the Entrez database within the 95% confidence interval of *Canq1* (132.3-135.8) to Ingenuity pathway enrichment analysis (Ingenuity Systems, Redwood, CA). After adjustment for multiple-testing errors (Benjamini-Hochberg adjusted p value <0.05), enrichment was found for cell-cell signaling and immune response genes (**Supplemental figure 3.6**). These findings are supported by potential candidate genes responsible for variation in collateral extent revealed by association mapping (below).

Association mapping links pial collateral traits to a narrow region within *Canq1*

Given the strength of the QTL on chromosome 7, one can speculate that inbred strains that share haplotype identity with B6 will have a B6 collateral phenotype, and similarly for Bc. Thus, by comparing the structure within the region of the QTL across multiple inbred strains, it should be possible to narrow the region harboring the genetic element(s) responsible for the QTL. However, as extensively discussed by Kang et al.[79], the population structure and relatedness among inbred strains lead to high false

positive rates and low statistical power. To circumvent these problems, we applied the efficient mixed model association (EMMA) algorithm designed by Kang et al. to the 134 individuals (8-10 per strain) of the 15 inbred strain set we studied previously (Supplemental figure 3.1[12]). The relatedness among inbred strains was modeled as a random effect. The kinship matrix (covariance matrix) was calculated using densely imputed SNPs with confidence levels of ≥0.9 obtained from the Jackson Laboratory web site (http://phenome.jax.org/pub-gi/phenome/mpdcgi?rtn-snps/download). Upon interrogating each of the 41,000 SNPs in the 127-142Mb region on chromosome 7, a tight group of 18 SNPs at 132.356 to 132.529Mb was found to have the greatest significance (p=2.2x10E-5; **Figure 3.5**, group A). This narrow region overlaps the proximal boundary (132.4Mb) of the previously defined 95% confidence interval. This concordance strongly suggests that the 172 kb region harbors genetic variation that determines native collateral extent. Two predicted genes fall in this region (**Table 3.2**). Inclusion of the next most significant set of SNPs (p<4.2x10E-4; Figure 3.5, group B), expands the region telomerically to 132.817, to include a total of 9 genes (Table 3.2).

LOD score profiling for collateral remodeling identifies a QTL on chromosome 11

Obstruction of the trunk of a main artery tree cross-connected to an adjacent tree by collaterals, as in MCAO, induces sustained unidirectional flow and shear stress in the collaterals[38]. This induces the collateral vessels to increase their lumen diameter and wall thickness, a process called outward remodeling or arteriogenesis that can be quantified as a fold change in diameter in the operated hemisphere relative to the non-operated hemisphere[113]. We showed that in mice this process occurs during the first few days after MCAO, and that the amount of remodeling is subject to strong genetic influence[12]. Herein, we examined remodeling 6 days after MCAO because at this time remodeling had reached a maximum in B6 mice but was considerably less in Bc[12]. A

genome-wide scan using the single-QTL model for fold change revealed a significant QTL identical to *Canq1* and a weaker QTL on chromosome 10 (**Figure 3.6**) The relationship between velocity and cross-sectional area proscribes that larger diameter collaterals will have lower fluid shear stress after ligation (eg MCAO), leading to less shear-induced remodeling. In support of this, a highly significant inverse correlation was obtained for collateral remodeling and baseline diameter (Figure 3.6), in agreement with previous results[12]. Hence, diameter is a covariate for remodeling. When set as such in the statistical model, the QTL on chromosomes 7 and 10 were lost, and a significant QTL on chromosome 11 (LOD score 3.5) was obtained (Figure 3.6, blue curve; Table 3.1). These results are consistent with the expectation that genes regulating collateral formation (ie, native number and diameter) in the embryo[38] are likely to differ from those regulating remodeling in the adult[1].

E. Discussion

This study sought to identify the genetic determinants responsible for the wide variation in the number and diameter of native (pre-existing) collaterals in healthy tissue of inbred mice[1, 12, 16, 34, 37, 38, 118, 119] and potentially in humans[27-30, 116, 117], and collateral remodeling after arterial occlusion[16, 38, 118]. We recently reported that these traits in the brain vary widely with genetic background, with C57BL/6 (B6) and BALB/c (Bc) mice being at the two extremes, and that this variation strongly correlates with infarct volume[12]. In the present study, we performed genome-wide linkage analysis in an F2 cross between B6 and Bc mice using collateral number, diameter and conductance as quantitative traits and identified a major locus on chromosome 7 that modulates collateral extent. Further evidence for the involvement of this locus was supported by use of chromosome substitution strains. We also identified a locus on chromosome 11

for collateral remodeling after setting the diameter at baseline as a covariate. High density association mapping was then used to narrow down the chromosome 7 QTL to a 172 kb region containing 9 candidate genes. Importantly, the major QTL on chromosome 7 is at the same locus as a previously reported QTL linked to hindlimb ischemia[17] and cerebral infarct volume[18]. This identification of a common QTL on chromosome 7 is very robust, having been made for physiologically linked but quite distinct phenomena (native collateral extent, cerebral infarct volume and hindlimb necrosis/perfusion recovery) in multiple tissues (cerebral pial circulation, cerebral cortex and skeletal muscle). In the previous papers, the underlying physiological basis was not identified. The present findings strongly suggest that that physiological basis is genetically determined variation in native collateral extent. As such, our results focus attention on the possibility that a major component of the amount of tissue injury in response to arterial obstruction in the brain, heart and peripheral limbs may be a person's genetic predisposition influencing the extent of the collateral circulation.

In the study by Dokun et al[17], the magnitudes of necrosis score and recovery of perfusion assessed 21 days after femoral artery ligation depended in large part on both collateral extent and collateral remodeling, although additional mechanisms could contribute[1, 7, 11, 12, 16, 34, 37, 113-115, 119]. However, infarct volume assessed 24 hours after MCAO[18], when only minimal remodeling of pial collaterals could have occurred, depends to a larger degree on native collateral extent. Given the dependence of shear stress-induced remodeling on collateral diameter, our results argue strongly that variation in native collateral extent exerted by the chromosome 7 locus is the major biological substrate underlying both previously reported QTL[17, 18]. This also provides a physiological mechanism—collateral formation during the perinatal period—on which to prioritize candidate genes and genetic elements for known or suspected involvement. Furthermore, since B6 and Bc strains have high and low native collateral extent,

respectively, in pia, skeletal muscle, and intestine[16], these results suggest that the same genetic determinants will apply to variation in collateral function in other tissues.

Infarct volume is inversely related to the prevailing arterial pressure after cerebral artery occlusion (since arterial pressure is the driving force for collateral perfusion), which could potentially impact interpretation of our results. As previously noted by Keum and Marchuk for Civq1[18], the centromeric tail of Canq1 overlaps the telomeric end of a syntenic rat chromosome 1 locus for blood pressure and infarct volume[120]. However, Yao et al[120] found that stroke-prone, spontaneously hypertensive rats (SPSHR) made congenic for this locus from the normotensive WKY rat showed both lower blood pressure and smaller cerebral infarct volumes after MCAO. The latter effect is *opposite* to what is expected, suggesting that secondary changes in vascular structure, smooth muscle tone or pressure-independent genetic factors also linked to the rat locus might instead be involved[120]. In addition, the location of the telomeric end of the introgressed piece corresponds to 128 Mb in mouse chromosome 7, which is well upstream of the peaks suggested by EMMA and the multiple QTL analysis. Finally, arterial pressures in Bc and B6 mice are virtually identical, whether measured in the anesthetized or conscious state[121-124]. Thus, these findings do not support arterial pressure differences as a physiological mechanism underlying the common QTL identified on chromosome 7 in the present or previous studies[17, 18].

Majid et al.[33] found no significant difference in blood flow velocity between the B6 and Bc strains measured with a laser Doppler probe in an estimated $1mm^3$ of tissue in the “core of the MCA ischemic territory” 10 minutes after MCAO, and concluded that non-hemodynamic variables (ie, including no differences in collateral extent) are responsible for the different infarct volumes in the two strains. However, several considerations call this conclusion into question: Values trended lower for the Bc mice. The Doppler probe that was used measures velocity rather than flow; velocity can

remain unchanged despite reduced flow if diameter also declines (eg, by passive collapse after MCAO), measurement of overall cerebral blood flow velocity in the MCA territory from a single-point probe is unreliable (eg, 24). Lastly, the Doppler signal includes blood flow in the dura mater, which is not supplied by the MCA.

Canq1, *LSq-1* (22) and *Civq1*[18] are broad loci that contain many genes. The gene list under *LSq-1* was reduced by interval-specific haplotype association mapping based on the criterion that haplotypes under the QTL would be identical in Bc and A/J mice and different from B6 (22). For cerebral infarct volume (*Civq1*), the criterion was expanded to include strain SWR, further reducing the list of candidate genes to only 12 in a 10 Mb region, from 132.47 (49334400M02RIK) to 142.37 Mb (Dock1)[18]. These analyses also apply to *Canq1*. However, to further refine this region we applied the recently described efficient mixed model association method (EMMA; 21) to 41,000 known, high quality SNPs in the 15 Mb region. Low statistical power and high false-positive rate can arise in haplotype mapping in mice due to genome-wide application, limited sample size, and lack of control for population structure[79]. Therefore, we confined our association mapping to 127-142 Mb (*Canq1*) and native collateral number from 15 inbred strains (n=8-10 of each strain[12]). We also used dense markers to infer the kinship matrix for the 15 strains. Finally, we applied the resultant kinship matrix to the EMMA algorithm to model population structure as a random effect to account for the population structure problem. This analysis pointed to a narrow region encompassing 9 genes, of which 7 failed to satisfy the haplotype criterion applied by Dokun et al. and by Keum and Marchuk. Thus, only one gene (4933440M02RIK, a predicted gene with EST support but no annotation) is common to the list of Keum and Marchuk and the list suggested by EMMA. There are no known miRNAs in the high-probability 9-gene EMMA region, nor is it known whether other classes of regulatory DNA or RNA elements (eg, long non-coding RNAs) are present. Future work will be needed to identify the genetic element(s)

underlying the major QTL on chromosome 7 and to define the molecular pathways controlling collateral formation in the embryo and thus the wide variation in the adult. Several genes with known influences on native collateral extent or collateral remodeling were not detected in our analysis. Experiments using mice haplo-insufficient or null for *Vegfa*, *Flt1* (VEGFR1) and *Clic4,* have shown that these molecules are strong positive determinants of perinatal collateral formation and maturation and thus native collateral number and diameter in the adult[34, 37, 39]. Furthermore, induction of VEGF-A expression is significantly lower in Bc compared to B6[16, 34, 37, 38, 125, 126] (expression of Flt1 and CLIC4 have not been examined in Bc versus B6). However, we found no QTL in the regions of these genes in either single or multiple QTL analysis. In addition, no interaction between *Canq1* and any other locus was found when all main effects and the epistasis between *Canq2* and *Canq3* were fixed. This outcome was expected for *Vegfa* and *Clic4*, since these genes and their flanking sequences are essentially identical in B6 and Bc. The Flt1 gene and flanking sequences are different between B6 and Bc, and are identical in Bc and A/J. However, whether this has functional consequences in the context of the B6 and Bc backgrounds is not known. One or more of the genetic elements underlying *Canq1-Canq4* may act in a weak *trans* manner to influence expression of VEGF-A, Flt1 or CLIC4, or other molecules involved in collateral formation may compensate for deficiencies in this pathway in Bc mice. Additional studies will be required to determine the role(s) of variants underlying the QTL identified in the present study.

Ingenuity pathway analysis within the 95% confidence interval of *Canq1* found enrichment for immune response genes, and IL4 receptor alpha and IL21 receptor were among the 9 genes in the high significance region identified by EMMA association mapping. While collateral remodeling is known to be highly dependent on the immune system[7, 11, 113-115], no studies have examined whether collateral formation and maturation during the perinatal period, which determine native collateral number in the adult, involve

immune cell function. Van Weel et al.[127] found evidence suggesting involvement of the known difference in natural killer cell (NKC) function in B6 versus Bc mice—associated with a different haplotype on distal chromosome 6 involving the NKC gene complex locus[128]—in the difference in recovery of blood flow after femoral artery ligation in these strains (which is dependent, in large part, on native collateral extent and collateral remodeling). However, this group recently reported, using congenic mice with this locus introgressed into the opposite strain, that the difference in native collateral number does not depend on the NKC locus[129]. This is consistent with our finding of no QTL on distal chromosome 6. It also is consistent with lack of agreement of this chromosome 6 haplotype in 8 inbred mouse strains with native collateral number in these same strains[12, 129]. Our results linking variation in remodeling of pial collaterals to its dependence on initial collateral diameter and thus the chromosome 7 locus, plus the diameter-independent locus on chromosome 11, are not congruent with a role for chromosome 6 in collateral remodeling. Van Weel et al did not examine if mice congenic for the chromosome 6 NKC locus show transference of differences in the remodeling phenotype.

We identified different genetic architectures for variation in collateral number, diameter and collateral remodeling. Native collateral number and diameter in young adult mice (eg, 10 week-old as studied herein) are determined by two processes—collateral formation which occurs late embryonically after the general circulation has formed—and collateral maturation which occurs during the first three postnatal weeks[38]. The former involves sprouting of a unique plexus of arterial-fated endothelial cell tubes, while the latter involves pruning away of a portion of these nascent collaterals, followed by lumen enlargement and smooth muscle cell investment of those that are retained[1, 38]. Collateral remodeling, which occurs in obstructive disease in the adult, involves a complex signaling and restructuring process[7, 11, 113-115]. Although much is known about the molecular signaling pathways directing collateral remodeling, almost nothing is

known about those directing formation and maturation[1]. The very different mechanisms and times of occurrence of the three processes suggest they are guided, at least in part, by genes unique to each. In support of this, collateral number and diameter shared a common QTL (*Canq1*), while number was linked to three additional QTL, and a QTL for remodeling was found on chromosome 11. In addition, collateral number, diameter and amount of remodeling have unique rank-orders among 15 inbred strains[12]. Importantly, all three traits are critical determinants of the severity of stroke and ischemic disease of the heart and other tissues[1].

In conclusion, our findings show that the extent and remodeling of the native collateral circulation is subject to wide variation that is due, primarily, to a remarkably strong polymorphism on chromosome 7, with smaller contributions from several additional QTL. These findings provide the underlying physiological basis for a recently reported QTL at the same locus on chromosome 7 linked to severity of hindlimb ischemia and cerebrovascular stroke[17, 18]. Variation in collateral traits reflect yet-to-be identified genetic and environmental factors that impact formation and maturation of collaterals, maintenance of them during natural growth to adulthood and subsequent aging, and collateral remodeling in obstructive disease. Our findings establish a foundation for future studies to identify the alleles and molecular pathways that direct these processes and account for their wide variability among healthy individuals, with obvious implications for patient evaluation and management.

Table 3.1. Chromosome location, LOD score, confidence interval and genetic descriptors of collateral QTL identified using multiple QTL model.

Trait	QTL	Chr	Location*	LOD	95% CI†	Effect size‡	Additive ± SE§	Dominance ± SE§	P value
Number	Canq2	1	45.0 (90)	12.6	38-49	13.3	-1.09 ± 0.22	0.34 ± 0.32	6.3E-10
	Canq3	3	5.0 (18)	6.6	0-28	6.3	0.09 ± 0.22	-0.15 ± 0.31	1.1E-04
	Canq1	7	63.6 (134)	28.7	62-65	37.2	-2.67 ± 0.22	1.42 ± 0.30	< 2.0E-16
	Canq4	8	16.8 (47)	5.0	6-24	4.9	0.85 ± 0.21	-0.67 ± 0.31	2.2E-05
	1:3			6.6		6.2			1.8E-05
Diameter	Cadq1	7	64.0 (134)	17.2	61-66	30.0	-3.25 ± 0.35	1.68 ± 0.49	1.8E-04
Conductance	Cacq2	1	47.3 (106)	6.0	31-81	6.6	-27.27 ± 5.05	0.18 ± 7.07	1.1E-06
	Cacq1	7	64.0 (134)	30.5	62-66	44.0	-67.90 ± 5.06	28.31 ± 7.07	< 2.0E-16
Remodeling	Carq1	11	55.6 (93)	3.5	41-72	8.1	0.15 ± 0.01	-0.03 ± 0.01	<0.05

*Location; cM (Mb); †Confidence interval, cM; ‡Percent of the total phenotypic variation explained by locus.

§Additive or dominance coefficients; SE, standard error of the mean.

Table 3.2. Candidate genes resulting from EMMA analysis of chromosome 7 QTL locus.

Start (bp)	Symbol	Orient	E	Description
132362957	LOC670828	-	protein	similar to 40S ribosomal protein S7 (S8)
132476076	4933440M02Rik	-	mRNA	RIKEN cDNA 4933440M02 gene
132554205	LOC100043014	-	protein	predicted gene 4171
132588190	Jmjd5	+	best RefSeq	jumonji domain containing 5
132611162	Nsmce1	-	best RefSeq	non-SMC element 1 homolog (S. cerevisiae)
132690783	EG244214	+	mRNA	predicted gene, EG244214
132695796	Il4ra	+	best RefSeq	interleukin 4 receptor, alpha
132746991	Il21r	+	mRNA	interleukin 21 receptor
132784468	Gtf3c1	-	best RefSeq	general transcription factor III C 1

Genes in grey area have $p < 2.2 \times 10^{-5}$: genes in white area have $p < 4.17 \times 10^{-4}$. Orient, orientation of transcription

Supplemental Table 3.1. Chromosome location, LOD score, confidence interval and genetic descriptors of collateral QTL identified using single QTL model.

Trait	Chr	Location*	LOD	95% CI†	Effect size‡	Additive ± SE§		Dominance ± SE§		P value
Number	1	49	3.84	33-78	5.90	-1.15	0.23	0.09	0.33	9.96E-06
	6	5	2.45	0-81	2.00	-0.65	0.23	0.01	0.36	1.41E-02
	7	64	21.24	62-67	32.20	-2.50	0.23	1.30	0.32	< 2E-16
	10	53	3.17	38-69	1.20	-0.34	0.22	0.47	0.34	8.68E-02
Diameter	7	64	17.16	61-66	30.00	-3.25	0.35	1.68	0.49	1.84E-04
Collateral Conductance	1	49	3.42	31-81	5.42	-25.61	5.25	-1.16	7.28	9.63E-06
	7	64	26.84	62-66	41.00	-66.24	5.10	26.45	7.07	< 2E-16
	10	53	3.47	62-66	1.73	-10.01	4.84	13.65	7.47	2.18E-02

*Location (cM); †Confidence interval (cM); ‡Percent of the total phenotypic variation explained by locus.
§Additive or dominance coefficients; SE, standard error of the mean.

Figure 3.1. Phenotypes of the native cerebral cortical pial collateral circulation in C57BL/6 (B6), BALB/c (Bc), F1 and F2 mice. A, Mouse pial arterial circulation in B6 (left) and Bc (right) after clearing, dilating, and filling with yellow Microfil^{TM} restricted from capillary transit. Collaterals cross-connecting the ACA and MCA trees in B6 are indicated (*). The Bc brain that is shown has no MCA-ACA collaterals. **B,** Number, diameter and "conductance" (number x diameter) of collaterals compared for B6, Bc and F1 mice. Data are given as mean ± SEM, with (n) = number of mice studied in this and subsequent figures unless indicated otherwise. The small n-size for Bc diameter and conductance reflects the large number of individuals in this strain with no collaterals. **C,** Distributions and density curves for number per hemisphere, diameter, and conductance of collaterals of the F2 cohort (n=221).

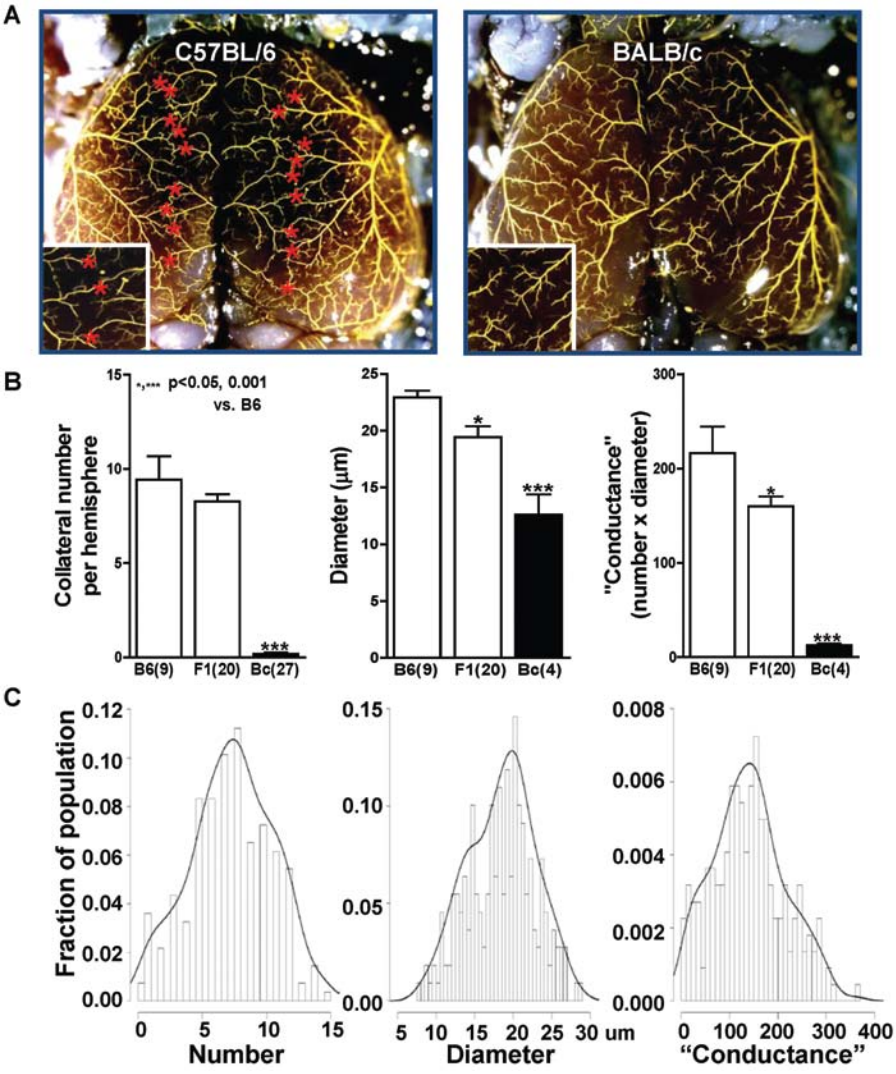


Figure 3.2. Genome-wide mapping of collateral traits in 221 F2 mice. A, LOD profiling using the single QTL model. Locations of genotyping SNPs are shown as ticks on the abscissa. 95% confidence level (dashed line) was estimated using 1000 permutations. Inset, higher resolution genetic map of the peak on chromosome 7. The location of the marker nearest the peak is indicated (triangle). **B,** LOD profiling using the multiple QTL model. Significant QTLs were found on chromosomes 1, 3, 7, and 8 for number (black line), on chromosome 7 for diameter (dashed line), and on chromosomes 1 and 7 for conductance (dot-dash line). Inset, schematic of the multiple QTL model for collateral number having the highest LOD. The solid line denotes interaction between the QTLs on chromosomes 1 and 3 (LOD score 7).

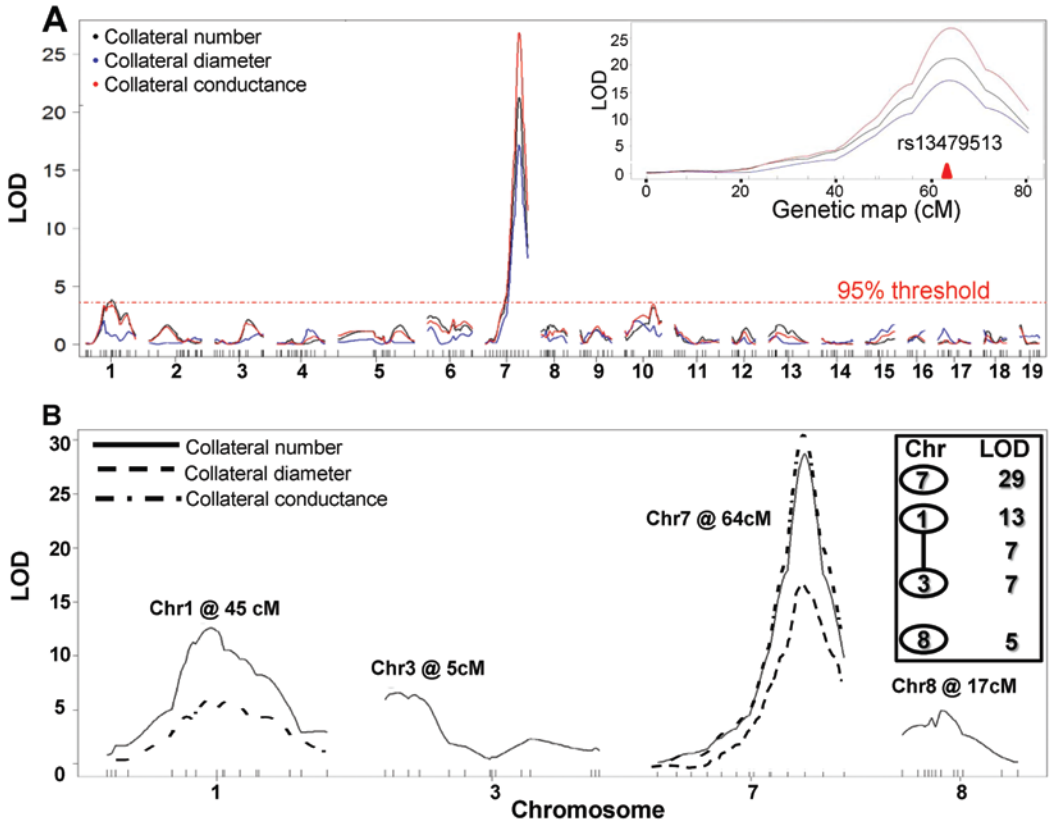


Figure 3.3. Influence of identified QTL on collateral traits. A, Phenotypic effects of the chromosome 7 QTL. Collateral number, diameter and conductance of parental, F1 and F2 mice grouped according to genotype at marker rs13479513 (see Fig. 3.1 inset). F2-B6/B6 mice and F2-B6/Bc mice are similar to parental B6 and F1 mice, resp., in all 3 traits. F2-Bc/Bc mice are similar to Bc in diameter (middle panel). These suggest that the chromosome 7 QTL accounts for the major difference in collateral number between B6 and Bc. However, F2-Bc/Bc mice are significantly different from Bc in collateral number and conductance, suggesting that additional loci must act with the chromosome 7 QTL to cause the very low collateral number in Bc mice. **B,** Phenotypic effects of the QTL identified by multiple QTL mapping. For each trait and QTL, F2 mice were grouped by genotype at the imputed peak SNP (see Supplemental Methods (3)) and the average for each group plotted vs. genotype (see text (3)).

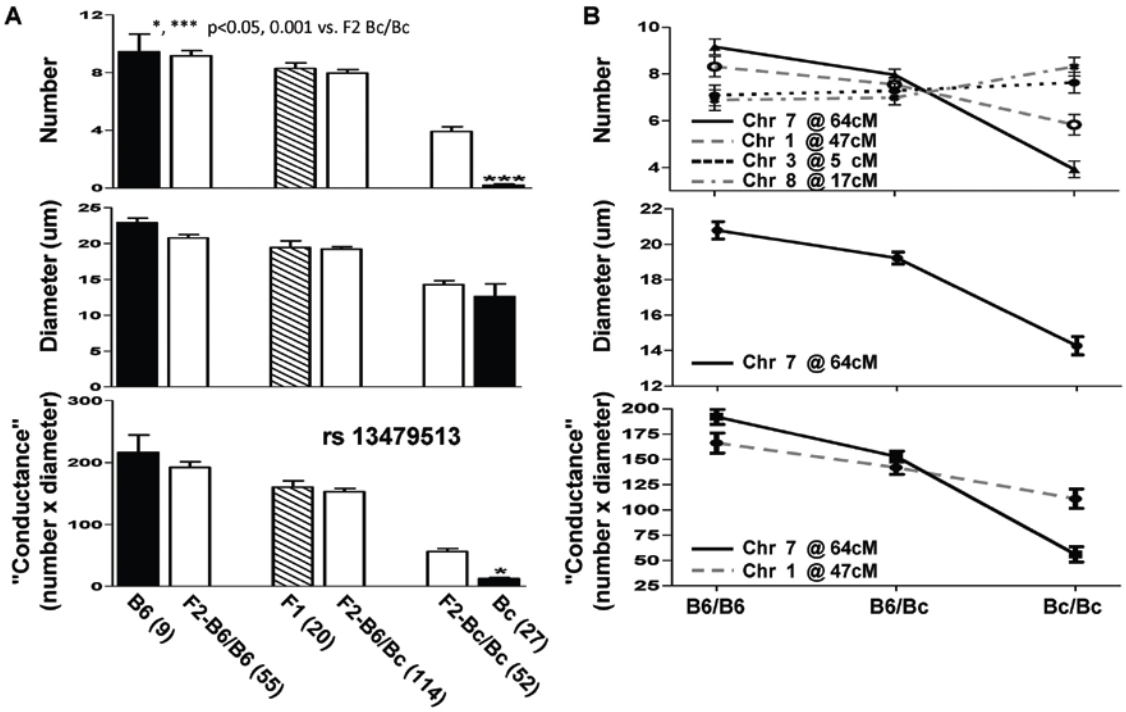


Figure 3.4. Chromosome substitution strains confirm a major effect of chromosome 7. A-C, Collateral extent traits in B6, Bc, A/J, CSS7 and CSS17 mice. A/J and Bc are similarly different from B6 in number and diameter ($p<0.001$; confirms Zhang et al. (12)) and in conductance. CSS7 mice (chromosome 7 in B6 replaced by chromosome 7 from A/J) are not different from A/J in number but are significantly different ($p<0.003$) from A/J in diameter and "conductance". Nonetheless, the replacement transfers most of the B6-A/J difference in each trait. In CSS17, intended as a negative control for validation of the CSS7 approach, collateral number is unchanged from B6 but diameter is modestly reduced. **D**, Fractions of the cerebral cortex area (territories) supplied by the ACA, MCA, and PCA. B6 and Bc have identical territories (confirms Zhang et al. (12)), while in A/J the MCA is slightly larger and the PCA correspondingly smaller than in B6. This difference is not transferred to CSS7, confirming the dissociation of collateral and cerebral artery tree territory phenotypes among inbred mouse strains (12).

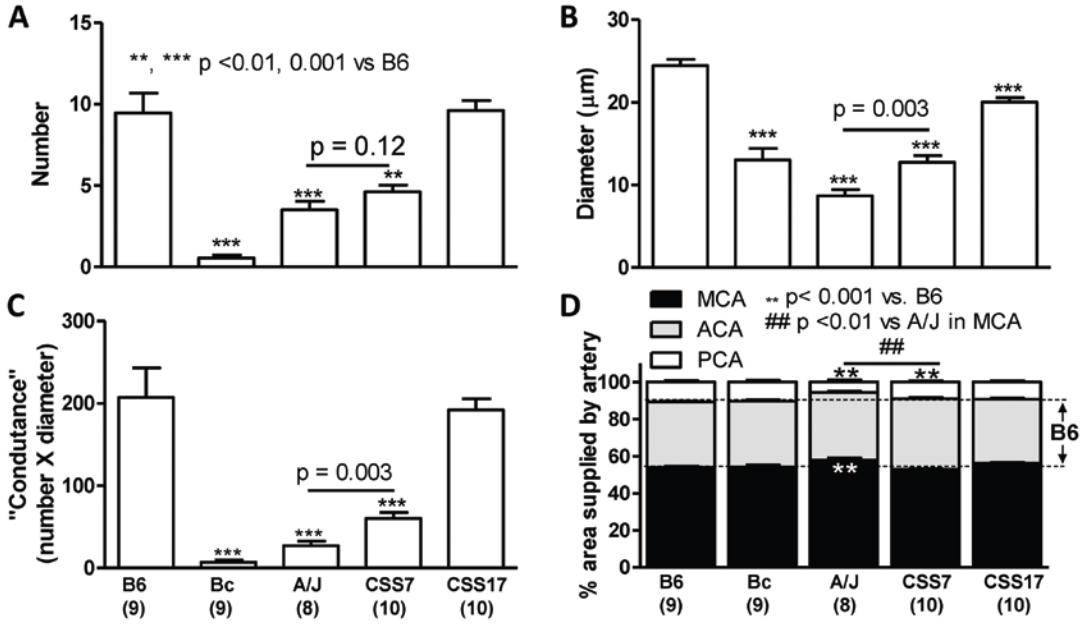


Figure 3.5. Highly significant region within *Canq1* identified by association mapping for native collateral number. The EMMA algorithm (79) was applied to collateral trait values in the 134 individuals of the 15 strain data of Zhang et al. (12) and the 41,000 high quality imputed SNPs under the *Canq1* peak (127 to 143 Mb; cM shown in the inset). A highly significant group of 18 SNPs (p=2.2E-5, Group A) slightly proximal to the QTL peak resulted (inset, black triangle). The second most significant group (155 SNPs, p=4.17E-4; Group B) spans a 290kb region. The genes in these groups are shown in Table 3.2. The vertical dashed lines in the insert delimit the 95% confidence interval of *Canq1*.

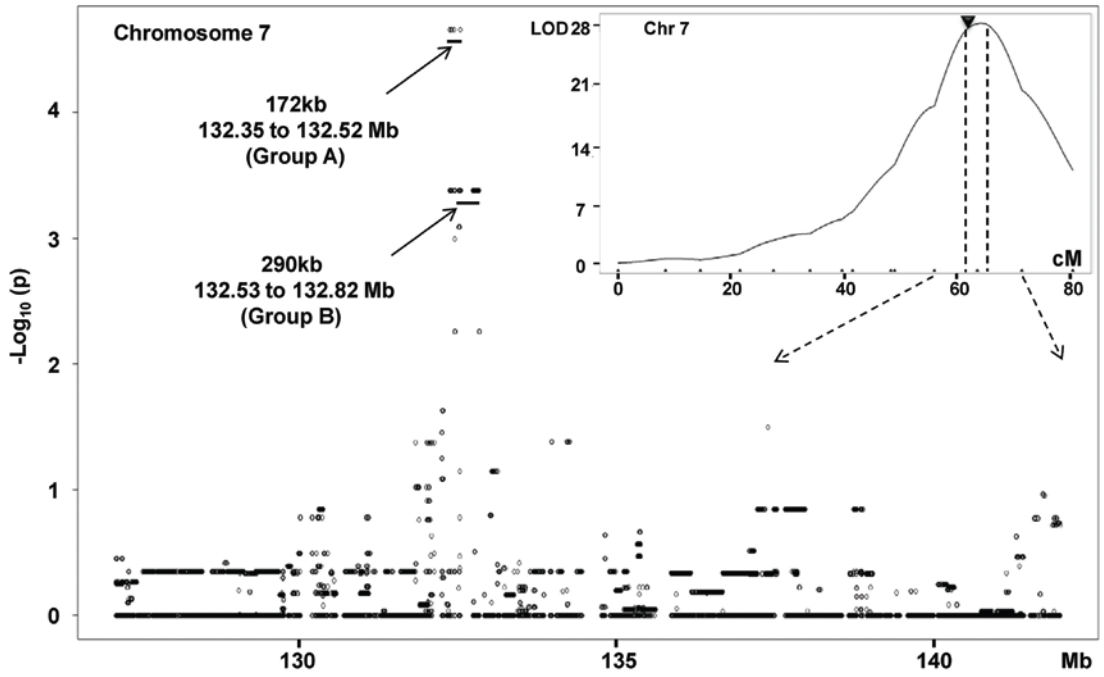
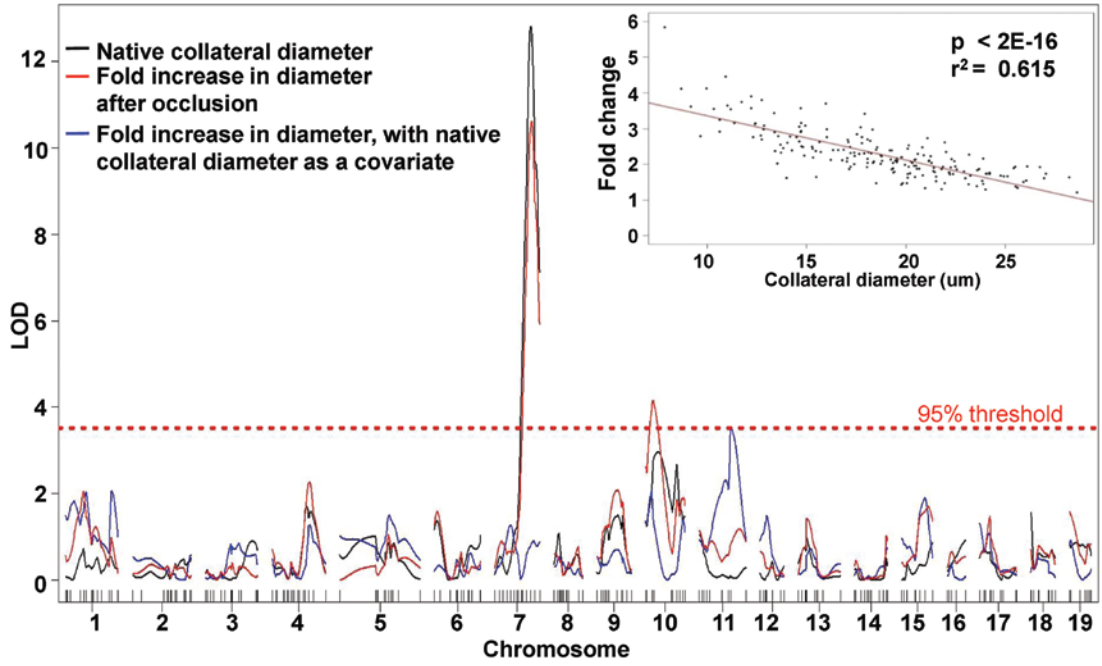
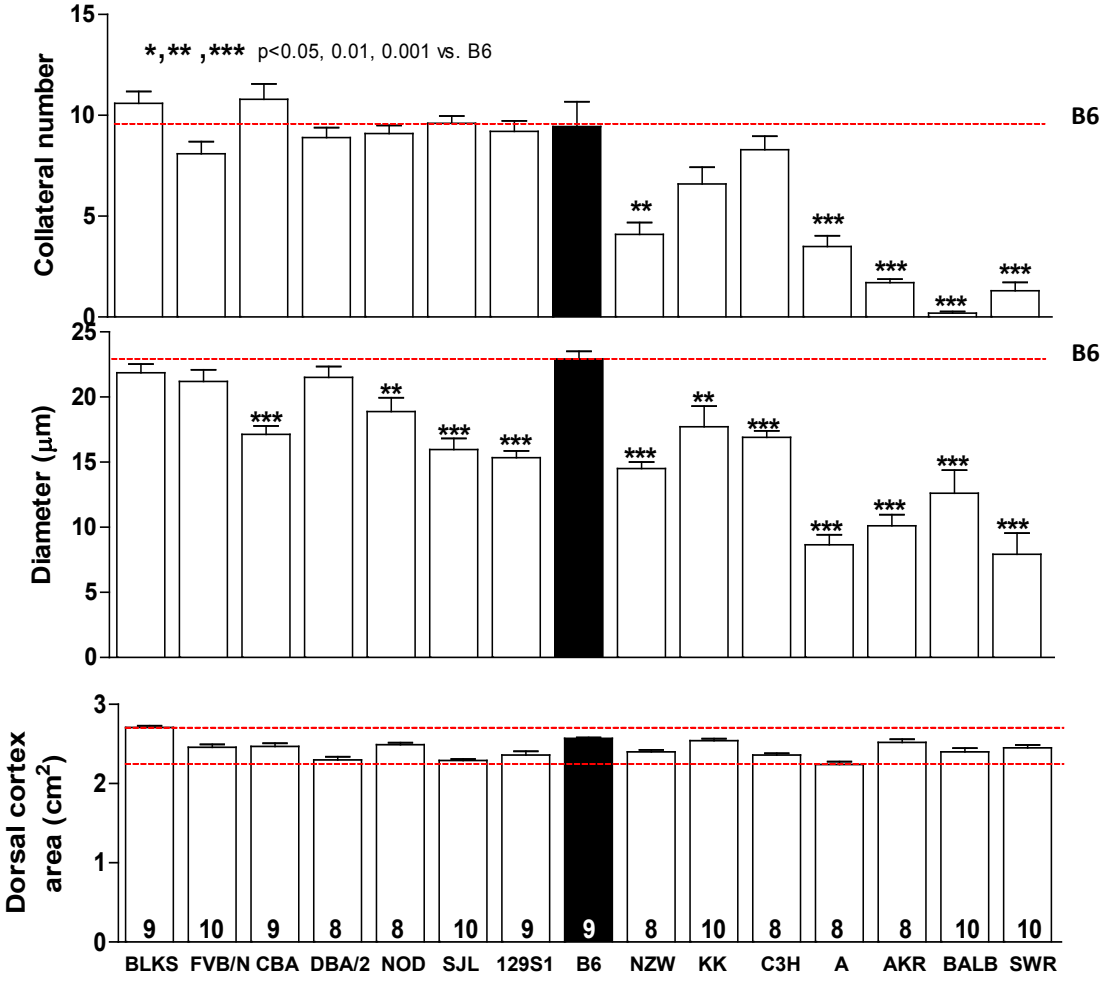


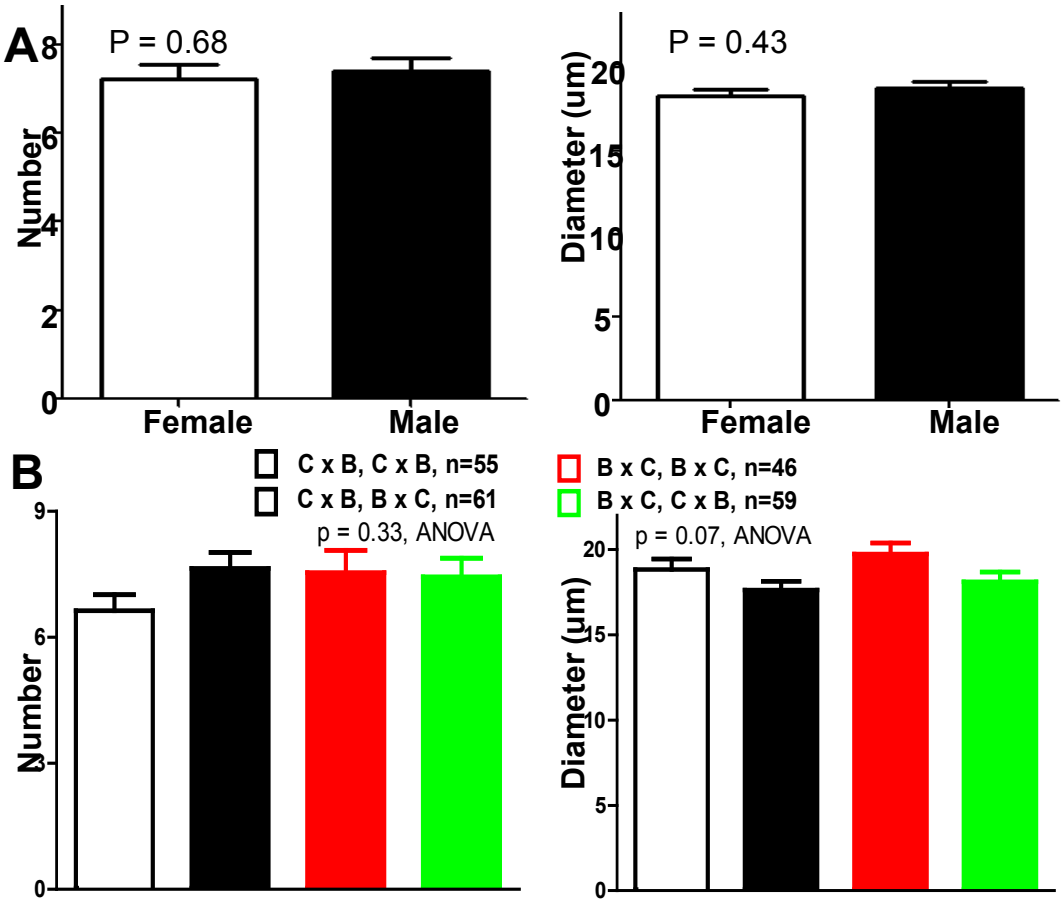
Figure 3.6. Collateral remodeling after MCA occlusion: Genome-wide mapping identifies a QTL on chromosome 11. Diameter measurements made in 190 F2 mice 6 days after right-side MCA occlusion were subjected to single QTL mapping. Black curve, native collateral diameter (non-occluded side); the LOD score is lower than in Figure 3.1 because there were fewer mice (See Supplemental Methods (3)). Red curve, fold increase in collateral diameter after remodeling (occluded over non-occluded). Blue curve, fold increase in diameter after rescanning with native collateral diameter as a covariate (see inset for close correlation between fold increase in diameter in the right hemisphere and native collateral diameter). The chromosome 7 QTL for remodeling (red curve) was lost (blue curve), but a QTL for remodeling on chromosome 11 achieved significance (*Carq1*, Table 3.1; $p<0.05$).



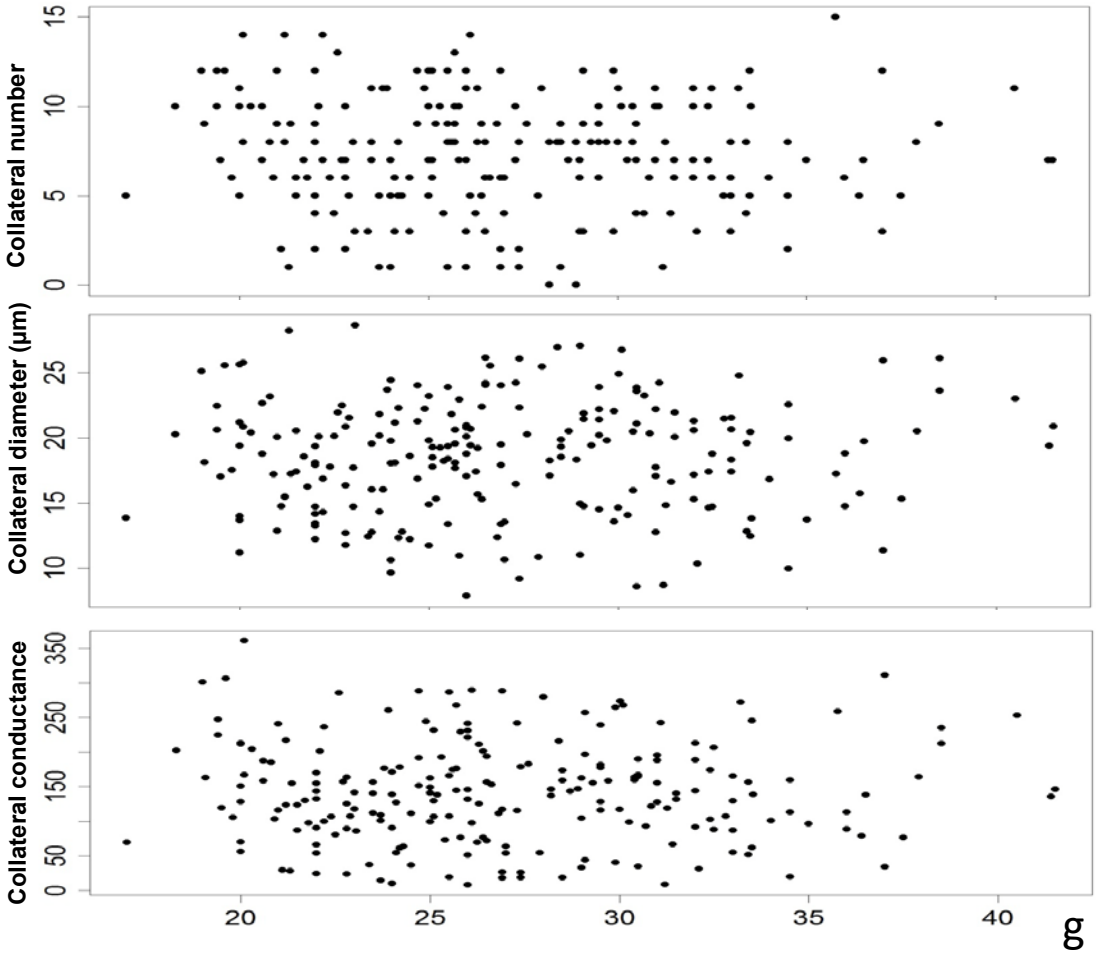
Supplemental Figure 3.1. Pial collateral number and diameter and cortical areas for 15 inbred strains used in the EMMA analysis (reprinted from ref 12). Variation in collateral number (per hemisphere) and diameter is unrelated to small variations in cortex area. Number of animals is given at the base of each column. Dashed lines on the dorsal cortex area panel delimit the 21% range of variation among the 15 strains.



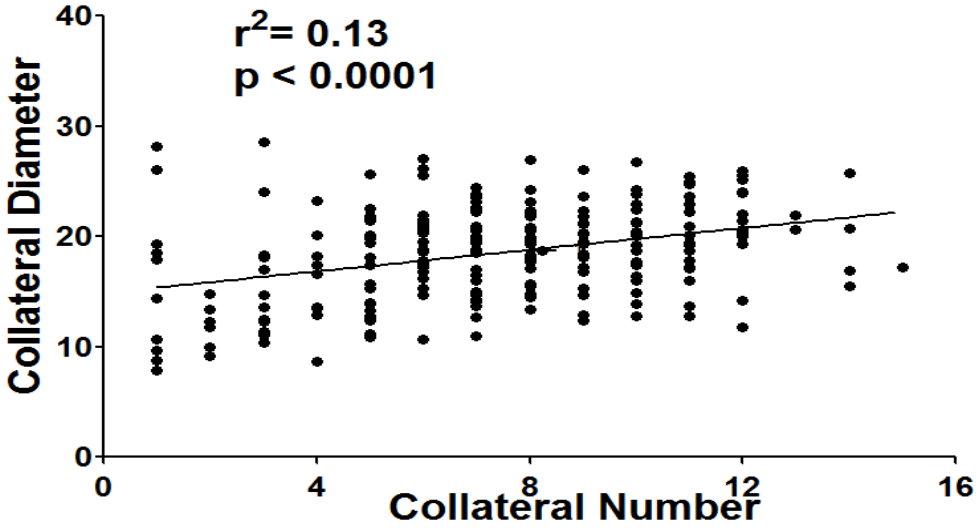
Supplemental Figure 3.2. Gender and parental origin have no effect on native collateral dimensions. **A,** Native collateral number (left panel) and diameter of female and male F2 individuals (n = 221) were not different. **B.** Average collateral number (left panel) and diameter for the 221 F2 mice grouped according to parental origins in reciprocal crosses. In the open bar group (CxB, CxB; F1 female, F1 male), both F1 parents resulted from mating of a BALB/c dam (C) and a C57BL/6 sire (B). Corresponding notation holds for the other 3 groups.



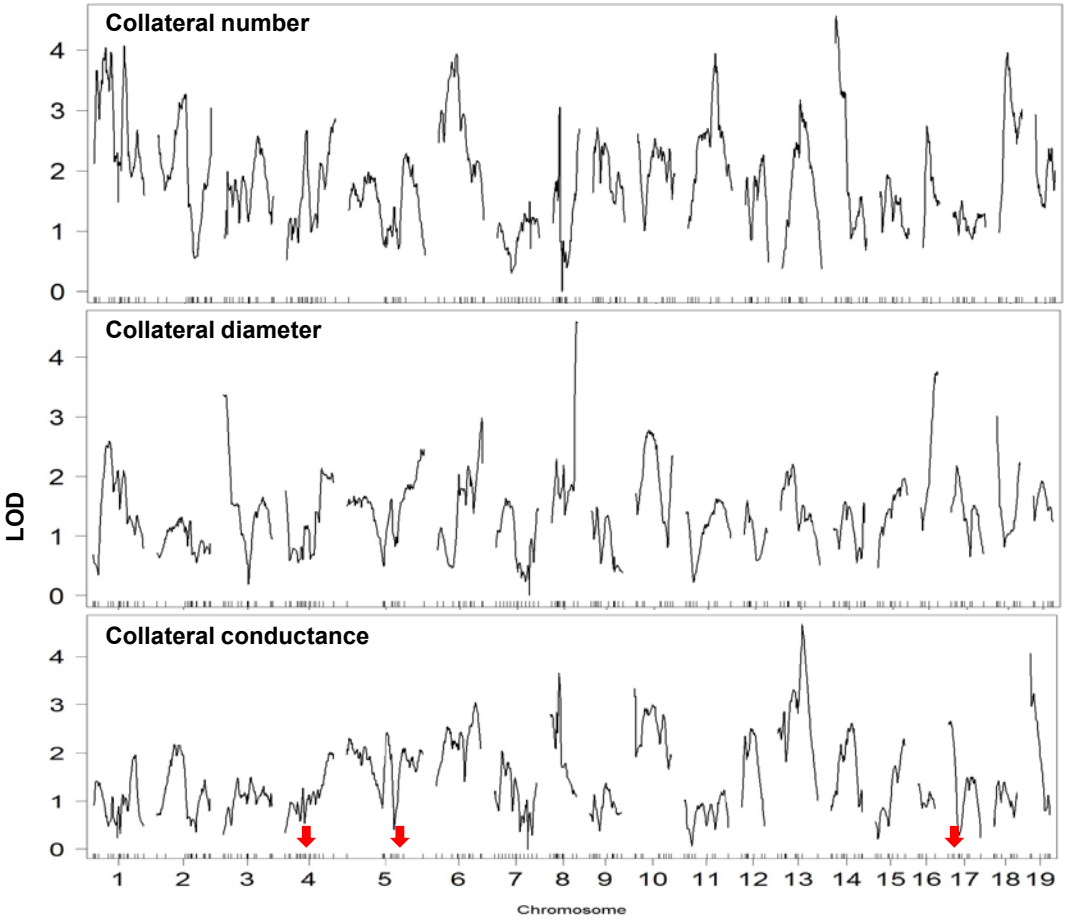
Supplemental Figure 3.3. Native collateral traits have no relationship to body weight.



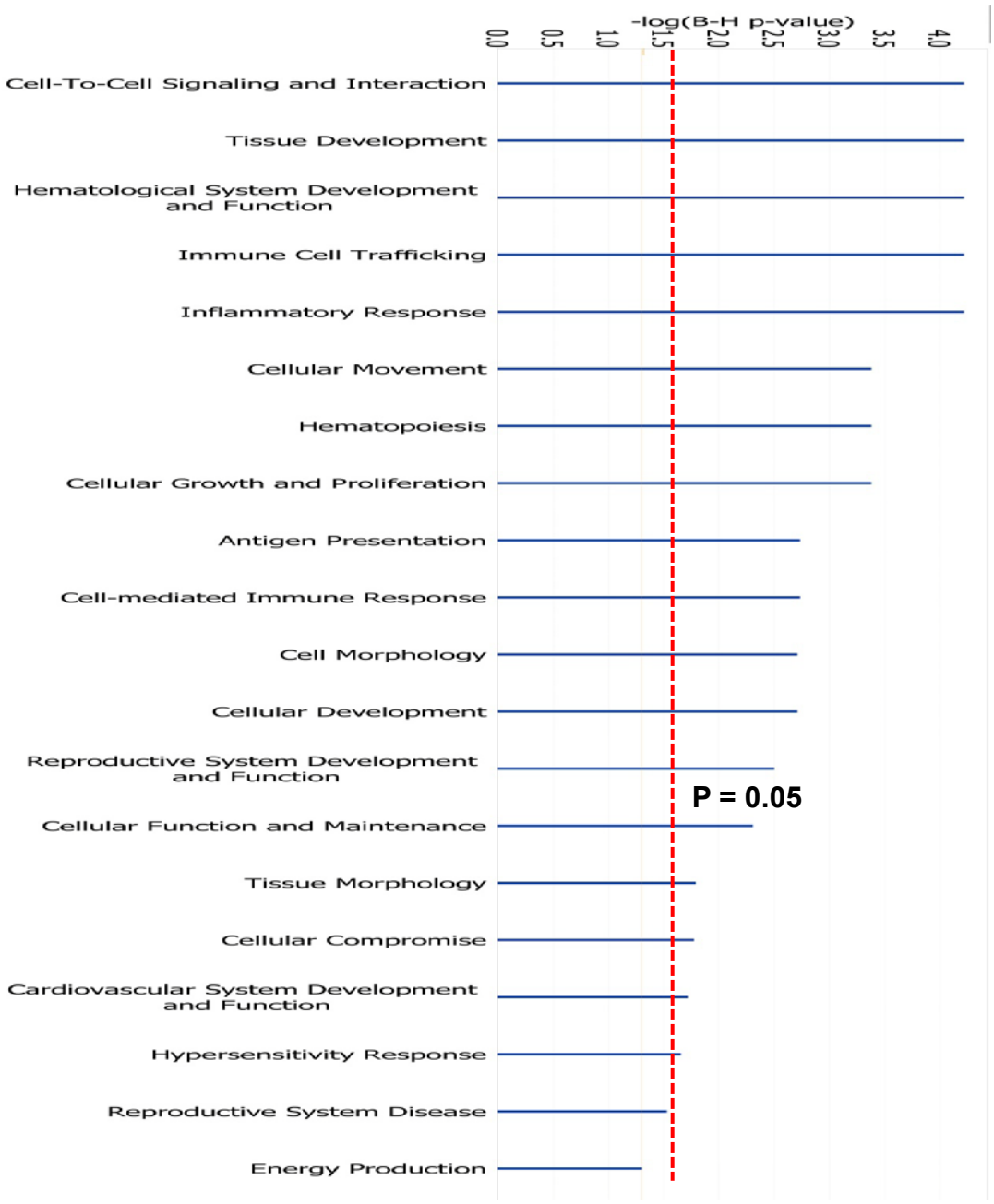
Supplemental Figure 3.4. Collateral number and diameter in the F2 population are largely distinct traits, based on the low r^2 from linear regression of the two traits.



Supplemental Figure 3.5. Genome-wide scan finds no significant interactions between chromosome 7 QTL and other loci known to influence collateral traits. Interactions were sought using likelihood ratio tests between full and additive models, with all identified QTL fixed in the linear model (11). From left to right, arrows indicate approximate positions of *Clic4, Flt1*, and *Vegfa,* genes whose expression levels are positive determinants of native collateral number and diameter (114; 39; 113). Empirical p-value threshold for significant interaction is LOD 6.3 (1).



Supplemental Figure 3.6. Cell-cell signaling and immune response pathways are highly enriched in chromosome 7 QTL region. Pathway enrichment analysis was conducted for all genes between 95% CI within Chr 7 QTL (132.3 Mb - 135.8 Mb) for all known pathways in the Ingenuity database. P values were adjusted for multiple comparisons (Benjamini-Hochberg; B-H p-value) and negative-logarithm-transformed.



CHAPTER IV

CANDIDATE GENES GOVERNING VARIATION IN EXTENT OF THE COLLATERAL CIRCULATION

A. Abstract

Rationale. Native (pre-existing) collaterals, which are arteriole-to-arteriole anastomoses that connect adjacent arterial trees and are critical in ischemic disease, vary widely in extent (ie, number and diameter) among inbred mouse strains. We previously identified a major QTL on chromosome 7 (*Canq1*; LOD=29) predominantly responsible for variation in collateral extent in a C57BL/6 (B6) X BALB/c (Bc) F2 population.

Objective. Herein, we sought to identify candidate genes responsible for collateral variation.

Methods and Results. Collateral extent in the pial cerebral circulation was intermediate to the Bc and B6 parental strains in a CXB recombinant inbred line that splits *Canq1* between the two strains. Phenotyping and mapping of an expanded panel of 21 informative inbred strains narrowed the locus from the previously reported 462 kb interval to 1.7 kb. Collateral extent, infarct volume after middle cerebral artery occlusion, bleeding time and rebleeding time did not differ in *Itgal*$^{-/-}$, *IL4*$^{-/-}$ and *IL4*-receptor-$\alpha^{-/-}$ mice relative to their respective host strains. mRNA expression of 120 genes within the 95% CI of *Cang1*, plus 30 selected angiogenic- and proliferation/aging-related genes, in the pial vasculature of B6 and Bc embryos at E14.5, E16.5 and E18.5 when the collateral circulation forms, identified 19 differentially expressed genes within the *Canq1* locus, plus *Vegfa* shown previously to impact collateral formation.

Conclusions. These findings narrow the *Canq1* locus by additional association mapping, argue against involvement of Itgal and IL4-receptor-α, and identify several genes within the locus as high-priority candidates important in collateral formation and its wide genetic variation.

Key Words: Collateral vessels, genetics, quantitative trait loci, cerebral circulation, arteriogenesis

B. Introduction

Occlusive vascular disease, including coronary artery disease (CAD), stroke, and peripheral artery disease (PAD), is a leading cause of morbidity and mortality[1]. When arterial obstruction occurs many physiological adaptive mechanisms are recruited to lessen the resultant ischemic tissue injury. Among several vascular protective mechanisms, the contribution of the native (pre-existing) collateral circulation and variation in its extent (ie, number and diameter of collateral vessels) have become increasingly appreciated as critical determinants of the large variation in tissue injury among individuals[1, 11, 12]. Collaterals are arteriole-to-arteriole anastomoses present in most healthy tissues that cross-connect occasional distal-most arterioles of adjacent artery trees. They serve as endogenous bypass vessels in the face of arterial obstruction and provide the initial collateral network whose conductance is subsequently increased over days-to-weeks by enlargement/remodeling of the native collaterals (arteriogenesis). Recent findings suggest that differences in extent are major determinants, rather than collateral remodeling, in the severity of tissue injury after arterial occlusion[16].

The severity of ischemic tissue injury has long been known to vary widely among species and individuals. Findings suggest that variation in injury may largely depend on variation in the extent of native collateral circulation. Meier et al. found that coronary collateral conductance measured by coronary flow index (CFI_p) varied by approximately 10-fold in 106 individuals without coronary artery or other heart diseases[24].

They also followed 739 patients with chronic stable coronary artery disease over a 10-year period, and found that CFI_p is a good predictor of 10-year survival rate, with good collateral conductance associated with a 75 percent reduction in mortality. In the cerebral circulation, Bang et al., using dynamic angiography, found evidence for a similar wide variability in native collateral conductance in patients with acute obstruction of the middle cerebral artery (MCA)[27]. In mice where direct measurement of collateral density and diameter can be obtained, native collaterals evidenced large differences in both parameters in multiple tissues in adult C57BL/6 (B6) and BALB/c (Bc) strains[16]. Chalothorn et al. further found, using the cerebral pial collateral circulation as a model, that collaterals begin forming at ~ embryonic day 14.5 (E14.5), with collateral density peaking at ~ embryonic day 18.5 (E18.5) in both B6 and Bc[38]. After E18.5, both strains evidence similar amounts of collateral pruning and maturation after birth, yielding by P21 the collateral number and diameter present in the adult. Thus, differences in extent of the collateral circulation, secondary to natural differences in genetic background, are primarily determined during collateral formation in the embryo.

We recently reported that collateral extent varied widely among 15 inbred mouse strains[12]. Moreover, the same strains had an equally wide variation in infarct volume measured 24 hours after MCA occlusion that was tightly and inversely correlated with collateral number and diameter. These findings establish a major role for genetic variation in the extent of the collateral circulation in determining variation in severity of ischemic tissue injury.

Previously, we performed a genome-wide linkage analysis in an F2 cross between B6 and Bc strains using collateral number, diameter and conductance—the product of number and diameter—as quantitative traits[3]. We identified a major QTL on Chr 7 (*Canq1*) linked to all three traits. This locus accounts for 37% of the variation in

collateral number (LOD = 29) in a B6 X Bc F2 population; three other loci and one interaction were responsible for an additional 35% of the variation. Importantly, this major QTL on Chr 7 is at the same locus as a previously reported QTL linked to hindlimb ischemia and cerebral infarct volume[17, 18]. Moreover, our findings identify the physiological substrate—variation in collateral extent—underlying this previously reported locus[16]. We also used the efficient mixed model association mapping (EMMA) algorithm together with high-density SNPs for 15 inbred strains to narrow the Chr 7 QTL to a 172 kb region containing 2candidate genes (the "EMMA region")[3].

Given the physio- and pathophysiological significance of this locus, it is important to explore the underlying candidate genes for involvement in regulating collateral formation. Identification of the alleles responsible for individual variation in collateral extent could help define the signaling pathways controlling collateral formation, permit stratification of patients, and aid the development of collaterogenic therapies[1]. Therefore, the goals of the present study were to further refine the EMMA region and to identify candidate genes underlying the Chr 7 QTL. To this end, we phenotyped collateral extent in additional inbred strains sharing common SNP patterns, measured collateral extent in mice genetically deficient for several candidate genes, and determined mRNA levels within the pial circulation in mouse embryos during the period of collateral formation.

C. Materials and Methods

Animals

Mouse strains: LP/J, NON/ShiLtJ (NON), LEWES/EiJ, 129X1/SvJ (129X1), CAST/EiJ, C57BL/6J (B6), BALB/cByJ (Bc), Itgal$^{-/-}$ (B6 background), IL4$^{-/-}$ (B6 background), IL4Ra$^{-/-}$ (Bc background) and CXB11 recombinant inbred strain (RIL derived from Bc and B6 progenitors) at ~10 weeks of age were obtained from Jackson Laboratories. PWD at ~10 weeks age were obtained from UNC-CH.

Phenotyping

Mice were phenotyped for collateral number and diameter as described previously[3, 12]. Briefly, the circulation was cleared and the pial circulation exposed. Arterial vessels were then maximally dilating, filled with a casting agent and fixed with paraformaldehyde. Collaterals connecting the anterior cerebral (ACA) and middle cerebral (MCA) artery trees in both hemispheres were measured for number and lumen diameter from digitized images (Leica MZ16FA, Bannockburn, IL). Cortical territories supplied by the MCA, ACA and PCA trees were determined morphometrically.

Bleeding and re-bleeding assays

Bleeding and re-bleeding assays were conducted as previously described[130]. Under 1.2% isoflurane anesthesia, the tail was pre-warmed for 5 min in 50 mL of 37°C saline. A 5 mm tip of the tail was quickly amputated with a fresh sterile #11 scalpel, and the tail was immediately returned to the saline. The vertical distance between bleeding point and mouse body was 10 cm. Bleeding time was measured as the time between start and cessation of bleeding. Re-bleeding time was measured as the time between cessation and the start of the second bleeding.

Middle cerebral artery occlusion (MCAO) and measurement of infarct volume

As detailed previously[12], the MCA was occluded at a midpoint between the zygomatic and temporal bones. Twenty-four hours later, the brain was removed, sliced into 1-mm coronal sections and sections were stained with 2% triphenyltetrazolium chloride (TTC). Infarct volumes were calculated as the sum of the cerebral cortical volume devoid of TTC in each section, and expressed as a percent of total cerebral cortex volume[34].

mRNA quantification

B6 and Bc breeding pairs were paired at 10-12 weeks of age. The presence of a vaginal plug the morning after pairing was designated as 0.5 days post-coitus (E0.5). Mice were collected at E14.5, E16.5, and E18.5. Embryos were removed under deep anesthesia (ketamine+xylazine, 100+15 mg/kg) and staged according to crown-to-rump length. Brains were quickly removed from embryos into RNAlater® (Sigma-Aldrich Corp., St. Louis, MO) and stored at -20ºC. At least 24 hours later, the embryonic pial circulation was peeled from the underlying dorsal cerebral cortex of both hemisphere using fine forceps under a stereomicroscope and stored in RNAlater® at -20ºC; pia from 8-10 embryos of each strain were pooled for each RNA sample ($\geq$ 2 litters per pool) at three embryonic time-points (E) days: E14.5, E16.5 and E18.5. At each time-point for each strain, three pooled RNA sample were prepared (18 total samples). Samples were thawed in RNAlater and homogenized (TH, Omni International, Marietta, GA) in Trizol Reagent (Invitrogen, Carlsbad, CA). The aqueous phase was extracted and RNA precipitated with ethanol. Total RNA was purified using the RNeasy Micro Kit according to the manufacturer (Qiagen, Valencia, CA). RNA concentration and quality were determined by NanoDrop 1000 (Thermo Scientific, Wilmington, DE) and by Bioanalyzer

2100 (Agilent, Foster City, CA), respectively. mRNA measurement was conducted by the genomics facility at UNC-CH using NanoString customized probes (NanoString, Seattle, WA)[131, 132].

Association mapping

As detailed previously[3, 79], the EMMA algorithm was applied in R to the collateral number trait on 21 inbred strains, representing 6 new strains (8-10 individuals each) in addition to data obtained for 15 strains previously[12]. Briefly, the kinship matrix among 21 inbred strains was calculated using the dense SNP set within the *Canq1* region, where kinship matrix represents the pair-wise relatedness among 21 strains[80]. To account for the relatedness among the strains, the kinship matrix was modeled as random effects. Each SNP within *Canq1* was modeled as a fixed effect. After the mixed model was fitted, an F test was conducted and p value obtained at each SNP location[79].

Statistical analysis

ANOVA, Student's t-tests and Bonferroni correction on pre-planned multiple comparisons were conducted where appropriate. Statistical treatment of association mapping data was as described previously[3].

D. Results

Collateral number in CXB11.HiAJ RIL verifies chromosome 7 QTL, *Canq1*

The CXB recombinant inbred lines (RILs) originated from Bc and B6 parentals [105]. Among them, CXB11.HiAJ (CXB11) is particularly relevant. Given the unique mosaic genetic structure for each RIL created by cross-over, we hypothesized that if *Canq1* in general, and if the EMMA region in particular, are predominant determinants of collateral extent, an RIL that inherits these loci from either B6 (abundant large diameter collaterals) or Bc (sparse small diameter collaterals) will exhibit a collateral extent dominated by the parental phenotype[3]. CXB11 splits the *Canq1* locus such that both the EMMA region and the proximal half of the locus are obtained from B6 (**Figure 4.1A**).

CXB11 averaged 5.2 collaterals per hemisphere, compared to 9.2 in B6 and 0.2 in Bc. Lumen diameter averaged 21um for CXB11, close to that of B6 (23um) and different from Bc (13um) (Figure 4.1B). These results confirm our previous finding[3] that *Canq1* harbors one or more variants governing collateral extent, probably in the proximal half. However, it is not possible to differentiate between an effect of mosaicism elsewhere in the CXB11 genome versus a contribution of a variant (s) within *Canq1* but distal to Itgal to explain the intermediate phenotype for collateral number. To differentiate these two confounding effects, a congenic strain which contains only half of *Canq1* may be very useful in the future. Because varying amounts of cortical territory supplied by each of the cerebral arteries (MCA, ACA, and PCA) (ie, “tree size”) could, in principle, be determinants of pial collateral extent12, we measured the percentage of territory supplied by each tree. These data (Figure 4.1c) fall within the range of tree size data previously shown to have no correlation with collateral extent and so do not explain the CXB11 results.[12].

Additional inbred strains strengthen and refine the EMMA region

EMMA mapping is more efficient computationally and better accounts for population structure embedded in inbred strains, compared to traditional association mapping algorithms[79]. We previously applied this algorithm to 15 inbred strains and obtained a most significant 172kb region ($p = 2.2 \times 10^{-5}$) and a second-most significant region (290kb, $p = 4.2 \times 10^{-4}$)[3]. To facilitate analysis, we generated a heatmap of the imputed SNPs within the 172 kb region for 74 inbred strains (**Figure 4.2**). The lowest 15 strains on the y-axis were those used previously. The lowest 5 of these have significantly fewer collaterals than B6. Among these 15 strains, 9 out of 10 with high collateral number have a blue DNA haplotype-like block, whereas all 5 strains with low collateral number have a largely green block, i.e. mostly green, not much blue. An exception is SJL/J, which has a green block but averages 10 pial collaterals per hemisphere. To verify the robustness of the EMMA region for variation in collateral number in mice, we chose 6 more informative inbred strains from the 74 lines: 3 strains with haplotype structures within the EMMA region predicting high collateral number (blue; 129X1, LP, NON), and 3 strains with haplotypes predicting low collateral number (green; CAST/Ei, PWD, LEWES/Ei). Average collateral numbers per hemisphere for 129X1, LP and NON were 10.4, 6.0, and 8.8, whereas CAST/Ei, PWD, LEWES/Ei were 8.6, 10.1 and 3.0 (**Figure 4.3**, **Supplemental figure 4.1**). Thus, collateral numbers in 4/6 strains fit the phenotypes predicted by their haplotype structures. The two strains that do not are wild-derived strains, which have more complex haplotype structures.

An important consideration is whether the EMMA peak shifts position or the previous significant SNPs change their p values after EMMA mapping of the above additional 6 strains added to the previous 15 stains. The results show that the most significant region did not change position, that this previous EMMA peak region was narrowed from 172kb to 2kb, and that the new peak evidenced increased significance (p

= 9 X 10^{-6}, originally p = 2.2 X 10^{-5}) (**Figure 4.4**). Of note, we also eliminated possible errors from imputations by replacing 2,859 imputed SNPs with new genotyped data downloaded from http://compgen.unc.edu/wp/.

The kinship matrix used to account for population structure can vary with use of different SNP datasets[80]. Furthermore, we previously didn't allow the missing genotypes with consideration of the decrease of statistical power[3]. We thus examined the effects of several parameters on our new EMMA mapping results by doing or not doing the following: allowing missing genotypes in less than 3 strains; using only informative SNPs; excluding wild-derived inbred strains. Irrespective of these mapping parameters, 4 SNPs always had the most significant p values (rs32978185, rs32978627, rs32973294, and rs32973297; 132.502-132.504 Mb; **Supplemental figures 4.2-4.6**). However, there is nothing annotated in this region.

Collateral extent, hemostasis and infarct volume in IL4Ra$^{-/-}$, IL4$^{-/-}$ and Itgal$^{-/-}$ do not support these genes as candidates for *Canq1*

We previously identified IL4Ra (interleukin 4 receptor alpha chain) in the second-most significant region defined by EMMA mapping[3]. Homodimers or heterodimers of IL4Ra and Yc function as receptors for IL4 and IL13 cytokines[133-135]. IL4 is a pleiotropic cytokine that mediates a Th2 inflammatory response in addition to inducing production of IgE in B cells. Evidence indicates that IL-4 may promote an inflammatory vascular endothelial phenotype including up-regulation of VCAM-1, MCP-1, and vascular endothelial growth factor (VEGF)[136, 137]. Yamaji-Kegan et al. recently demonstrated that hypoxia-induced VEGF is significantly diminished in IL4-/- mice (B6 background). Furthermore, hypoxia-induced pulmonary angiogenesis and proliferation in airways and pulmonary artery were suppressed in IL4 deficient mice[137]. In addition,

VEGF is an important determinant of collateral formation in the embryo[34]. Despite these rationales, IL4$^{-/-}$ mice had no deficit in collateral extent (9.5 collaterals/hemisphere, 21um diameter) compared to B6 (9.3 and 23um, respectively) (**Figure 4.5**). Bc is well known to mount a Th2-dominant response, compared with B6 which is Th1-dominant[135, 138]. Accordingly, T cells from Bc produce greater amounts of IL4, IL4Ra, and IL13. We thus examined IL4Ra-/- (only available on Bc background). Collateral extent in IL4Ra deficient mice was similar to Bc (0.2 collaterals, 17um diameter versus Bc, 0.2 and 13um, respectively). The above results from two knockout strains do not support involvement of IL4Ra in variation of collateral extent.

Itgal is located at 134Mb near the peak of *Canq1* and encodes an integrin alpha L chain. This polypeptide binds an integrin beta-2 chain to form lymphocyte function-associated antigen-1 (LFA-1), which is expressed on the surface of leukocytes[139-141]. LFA-1 plays a central role in leukocyte trafficking through interactions with its intercellular adhesion molecule (ICAM) ligands expressed on endothelial cells. Several signaling pathways are activated upon binding of ICAM by LFA-1 to effect infiltration of lymphocytes. Many lines of evidence suggest that Itgal is involved in immune response, inflammation and ischemic injury[140]. Keum and Marchuk identified *Itgal* as a candidate gene for infarct volume measured 24 hours after MCAO in a B6 X Bc F2 population, using genome wide linkage analysis and interval-specific SNP haplotype association mapping[18]. Furthermore, Arumugam et al., using a 1 hour transient MCAO model, found that LFA-1$^{-/-}$ mice had smaller infarct volumes measured 24 hours after reperfusion, compared to B6[142]. Although, many factors contribute to the severity of ischemic injury besides collateral extent, including types of ischemic models, the no-reflow effect of leukocyte plugging, and hemostasis[143-147], these findings led us to

examine whether *Itgal* modulates collateral abundance. However, *Itgal*$^{-/-}$ and B6 controls had similar collateral extent (Figure 4.5).

We also determined infarct volume 24h after permanent MCAO, since collateral extent is the major determinant of strain-dependent variation in this model of stroke[12] and the above mentioned previous studies suggest involvement of Itgal[18, 142]. Consistent with our finding of no difference in collateral extent, infarct volume was not different in *Itgal*$^{-/-}$ mice, compared to B6 controls and as referenced to Bc that exhibit larger infarctions (**Figure 4.6**). Neither was it affected in *IL4*$^{-/-}$ mice (B6 background), in agreement with collateral extent (Figure 4.5). These findings do not support *Itgal* (or *IL4 or IL4ra*) as a candidate gene for *Canq1*^{3} or the previously reported *Civq1*[18].

Given that MCA tree size is also a determinant of infarct volume after permanent[12] or transient MCAO, we also examined MCA tree size in these strains. Tree size was smaller in *Itgal*$^{-/-}$ mice (Figure 4.5), which may explain the smaller infarct volume observed after transient MCAO in *Itgal*$^{-/-}$ mice[142]. We did not observe a difference with permanent MCAO (Figure 4.6), which may result from an effect on re-bleeding time (see below).

Leukocyte/platelet adhesion and hemostasis have complex effects on infarct volume, especially secondary to transient occlusion, through effects on thrombosis/thrombolysis, hemostasis and inflammatory mechanisms[148]. Moreover, thrombosis and thrombolysis have been reported to vary widely dependent on genetic backgrounds[130]. Because these mechanisms could be affected in *Itgal*$^{-/-}$ mice, and possibly also in *IL4*$^{-/-}$ mice, we measured bleeding and re-bleeding time to assay thrombosis and thrombolysis, respectively. No differences were observed among knockouts, B6 controls or the Bc strain (Figure 4.6), with the exception of shorter time to re-bleeding in *Itgal*$^{-/-}$ mice. Depending on models and severity of stroke, increased

thrombolysis can either reduce no-reflow and lessen infarct volume, or increase petechial hemorrhage and increase infarct volume[148].

Expression of genes within *Canq1*

The highly significant SNPs within the EMMA peak (Figure 4.4), together with the strength of the *Canq1* locus (LOD 29, heritability 37%)[3], strongly suggests that a gene(s) or genetic element(s) within the EMMA region or the much wider *Canq1* locus may be polymorphic, resulting in differences in expression of the given or an adjacent gene(s) that governs variation in collateral formation. To explore this possibility, we measured mRNA levels of genes within the 95% CI of *Canq1* in the B6 and Bc parental strains of the F2 population from which the locus was mapped. In the pial circulation collateral formation begins with the appearance of a plexus of collateral vessels that starts to form between the crowns of the cerebral artery trees and between branches within the trees (intra-tree anastomoses, ITAs) at E14.5 and reaches its peak density by E18.5 in both B6 and Bc[38]. 60-70% fewer collaterals and ITAs form in the Bc strain. Between E18.5 and postnatal day 21, an equal number of collaterals and virtually all of the ITAs are pruned away in both strains, yielding the number and diameter of collaterals present in the adult (eg, Figure 4.5)[16, 38]. We thus examined expression in both strains at E14.5, E16.5 and E18.5 for three groups of genes: genes with SNP variation between B6 and Bc strains in coding or regulatory (ie, the 5' and 3' 2kb non-coding) regions within the 95% CI of *Canq1* including the 172 kb EMMA region (132.356 - 132.528 Mb); selected angiogenic-related genes, and selected genes related to proliferation and aging. Among 150 genes, 19 genes located within the 95% CI of *Canq1* have adjusted p values less than 0.05 for a difference between strains (Bonferroni correction, **Table 4.1**, **Figure 4.7**). Nine of these are more than 2-fold higher in Bc than in B6 at all time-points, while

one gene, Pycard, is at least 60% higher in B6. Nsmce1 (132.6 Mb), located in the second-most significant EMMA region (132.53 - 132.82 Mb), has three splicing isoforms, with two expressing 2-to-7-fold higher levels in Bc (Table 4.1, **Supplemental table 4.1**). On the other hand, expression of 6 genes significantly changed over the 3 time-points, including two angiogenic-related genes and four genes in *Canq1*.

E. Discussion

In present study we sought to confirm *Canq1* and the EMMA region reported previously[3]. The collateral phenotype of CXB CXB11, in which the proximal half of *Canq1* derives from B6, was intermediate for collateral number between B6 and Bc, and not different from B6 for collateral diameter. With *Canq1* alone accounting for a large fraction of heritable variation in collateral extent[3], this result confirms our previous findings *that Canq1* harbors one or several important variants governing collateral extent. Our previous linkage analysis showed that *Canq1* alone accounts for 37% of the phenotypic variation, with three other loci plus an interaction explaining 35% additional variation. However, because of the small effects (LOD < 6) and broad ranges (> 5Mb) of these other QTL on Chr 1, 3, and 8, CXB11 has mosaic structures within these loci. Thus, it is impossible to verify whether the EMMA region harbors the causal variation of *Canq1* by phenotyping CXB11 alone. To verify and possibly narrow down this region, we phenotyped 6 more inbred strains chosen according to their haplotype structures within the region. Four of 6 inbred strains had the predicted collateral extent. We then subjected this panel of now 21 strains to EMMA mapping, narrowing the region considerably. Together, these three approaches verify, strengthen and better define the importance of *Canq1* and the EMMA region within it.

By examining mRNA expression of candidate genes within both regions, we were able to identify for future study a number of genes as high-priority candidates potentially regulating collateral formation. As a first step in this effort, we examined knock-out mice for two genes within *Canq1* (*Itgal, IL4Ra*) and one acting on a gene within the locus (IL4) that may be involved in [149], and found no supportive evidence. While conclusions based on gene knockout mice have limitations (eg, compensations from global deletion, lack of ability to test for variants producing gain-of-function or increased expression), this conclusion is strengthened by the absence of a significant difference in expression of *Itgal* and *IL4ra* between B6 and BALB/c during the time when collateral formation is occurring (adjusted p >0.76).

Two of the 6 strains chosen, Cast/Ei and PWD, for expanded EMMA mapping were outlier (Figure 4.3). The unique haplotype structures of these wild-derived strains have discouraged their use in previous mapping studies[150, 151], a finding which our results confirm. However, LEWES—also a wild-derived strain—fit the predicted collateral number. Several studies have utilized dense SNPs from the recently re-sequenced 15 strains (NIEHS and Perlegen) to infer ancestral haplotype origins for the classical inbred strains[151, 152]. By using Cast/Ei, PWD, and WSB as references for *Mus Musculus* subspecies *castaneus, musculus*, and *domesticus,* together with the hidden Markov Model algorithm, Frazer et al. found that the classical 11 inbred strains shared 68% haplotype origins with *domesticus* (WSB), 6% with *musculus* (PWD), and 3% with *castaneus* (Cast/Ei)[152]. In another study, Yang et al. found that wild-derived strains, which are presumed to represent different mouse subspecies, show substantial intersubspecific introgression[151]. Using the 72% of the autosomal and X-chromosomal 100 kb genomic intervals for which the ancestry of the reference strains is unambiguous, they estimated that the classical 11 inbred strains inherited, on average, 92% of their

genomes from *Mus domesticus* origin, 3-12% from *musculus* origin, 1-2% from *castaneus* origin. In a recent study, Kirby et al. densely genotyped 121k SNPs for 94 inbred strains[150]. Based on SNP analysis, LEWES has close ancestral origins with the WSB strain. We thus postulate that because Cast/Ei and PWD share minimal ancestral haplotype origins with the classical inbred strains, these two wild-derived strains were outlier in our study, while LEWES, which shares major ancestry origin with the classical inbred strains, evidenced the predicted collateral number (Figure 4.3). However, regardless of whether the three wild-derived strains were included or excluded in EMMA mapping, the same significant region was obtained (132.502 to 132.504 Mb) with invariant p values ($<10^{-6}$) (Supplemental Figures 4.2-4.6). Moreover, the diverse genetic background used in our EMMA mapping was corrected by applying the kinship matrix. For example, the average identical by descent (IBD) is 0.45 between Cast/Ei and the other 20 stains, whereas IBD averages 0.82 between B6 and the other 20 strains.

The most significant region ($p < 9 \times 10^{-6}$, 132.502 to 132.504Mb) and its adjacent extended region ($p < 3.8 \times 10^{-4}$, 132.495 to 132.513Mb) identified by EMMA mapping with 21 inbred strains were examined for mRNA annotation. No annotation was found in this region (http://uswest.ensembl.org/Mus_musculus/Location/View?r=7:132495610-132513000). The examination in UCSC genome browser database confirmed this result. The gene 4933440M02RIK is located 2.3 kb at the upstream of this region while Jmjd5 positions 6.5 kb downstream. MicroRNA and ncRNA were also absent in this region (http://www.ncrna.org/). We further examined this extended region for transcription binding factors using Transcription Element Search System (TESS) (http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME)[153, 154]. No mammalian transcription factor binding sites were identified.

SJL, which has high pial collateral density yet possesses a haplotype structure similar to that of SWR, is also outlier (Figure 4.2 and Figure 4.3). Among approximately one thousand SNPs within the most significant EMMA region, 400 SNPs imputed with high confidence (>0.9) are identical between SWR and SJL, and the genotypes of the remaining SNPs imputed with lower confidence are also the same. If no variation exists within the EMMA region between SWR and SJL, what genetic components regulate the variation in native collateral extent between these two strains? Effort to identify the novel genetic component(s) using an SWR X SJL F2 cross is currently under way.

In conclusion, the present results significantly reduce the priority of several candidate genes previously identified underlying the Chr 7 QTL[3, 17, 18]. On the other hand, locus-wide expression analysis suggests a number of interesting new candidates, eg, Nsmce1. Use of additional inbred strains has verified and narrowed down the previous EMMA region. Although this region (132.502 to 132.504 Mb) contains no genes, variation within this region is highly associated with variation in native collateral extent. Thus, complete sequencing of this region may lead to clues regarding this important source of collateral variation. Additional insight may be obtained if more inbred mouse strains with predicted low pial collateral density can be identified as new information on ancestral haplotype sharing patterns becomes available.

Table 4.1. Genes having differential expression either between strains or among different embryonic days

Gene* number	Name	Description	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
111	Pycard	PYD and CARD domain containing Gene	II	135135617	-2.29	-1.67	-1.79	0.0003	0.0595
124	Inpp5f	inositol polyphosphate-5-phosphatase F	II	135754842	1.90	1.69	2.30	0.0036	1.0000
47	Tbx6	T-box 6	II	133924997	2.38	2.09	2.89	0.0055	1.0000
16	D430042O09Rik-204	RIKEN cDNA D430042O09 gene	II	132851390	3.94	2.61	3.65	0.0073	1.0000
9	Nsmce1-003	non-SMC element 1 homolog	I	132611154	5.54	3.58	7.09	0.0084	1.0000
150	telomerase	NM_009354.1	IV	73764438	3.99	2.14	3.87	0.0085	1.0000
8	Nsmce1-002	non-SMC element 1 homolog	I	132611154	2.52	2.05	3.57	0.0122	1.0000
26	Nfatc2ip	nuclear factor calcineurin-dependent 2 interacting protein	II	133526368	1.68	1.69	2.05	0.0125	1.0000
55	Taok2	TAO kinase 2	II	134009192	1.65	1.38	1.97	0.0143	1.0000
110	B230325K18Rik	RIKEN cDNA B230325K18	II	135126593	9.69	4.68	7.75	0.0165	1.0000
91	1700120K04Rik	RIKEN cDNA 1700120K04 gene	II	134747592	3.37	1.99	2.35	0.0177	1.0000
80	9130019O22Rik	RIKEN cDNA E430018J23 gene	II	134525774	3.63	2.28	3.78	0.0190	1.0000
87	1700008J07Rik	non-coding RNA	II	134655683	2.64	1.87	3.05	0.0202	1.0000
120	BC017158	UPF0420 protein C16orf58 homolog	II	135414893	1.56	1.43	1.67	0.0226	1.0000
140	Vegfa188	vascular endothelial growth factor A	III	46153942	1.88	1.45	2.05	0.0236	1.0000
88	Phkg2	phosphorylase kinase, gamma 2	II	134716854	1.70	1.34	2.19	0.0250	1.0000
19	D430042O09Rik-209	RIKEN cDNA D430042O09 gene	II	132851390	2.04	2.09	5.63	0.0284	1.0000
41	Giyd2	GIY-YIG domain containing 2	II	133832982	1.57	1.63	2.40	0.0291	1.0000
54	Ino80e	INO80 complex subunit E	II	133995094	4.64	2.71	3.79	0.0343	1.0000
127	Dock1	dedicator of cytokinesis 1	II	141862370	1.87	2.15	2.47	0.0384	1.0000

Table 4.1 (continued)

Gene* number	Name	Description	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
122	Tial1	Tia1 cytotoxic granule-associated RNA binding protein-like 1	II	135583291	2.02	1.54	2.12	0.0418	1.0000
109	Fus	malignant liposarcoma gene	II	135110971	1.47	1.49	1.72	0.0428	0.6911
40	Sult1a1	sulfotransferase family 1A, phenol-preferring, member 1	II	133816379	-2.29	1.01	1.76	1.0000	0.0026
129	Angiopoietin 2	angiopoietin 2	III	18690263	-1.15	1.14	1.61	1.0000	0.0077
139	TGFb	growth factor beta binding protein 1	III	75404869	1.30	1.18	1.32	0.1245	0.0087
37	Nupr1	nuclear protein 1	II	133766763	-1.72	1.21	1.70	1.0000	0.0096
65	Qprt-001	quinolinate phosphoribosyltransferase	II	134250628	-1.55	-1.19	1.10	1.0000	0.0377
20	AC150648.2	GSG1-like	II	133024217	-1.59	-1.28	-1.27	1.0000	0.0448

Expression assay by Nanostring nCounter. 3 RNA samples of pia at each time point for each strain, with each sample composing of ≥ 8 embryos from ≥ 2 litters.

***** Gene number same as in Figure 4.7 and Supplemental Table 4.1.

† **II**, gene in 95% CI of Chr 7 QTL; **I**, genes within the EMMA region; **IV**, proliferation/aging-related genes; **III**, angiogentic-related genes.

‡ Fold change for BALB/c vs B6 if positive and B6 vs BALB/c if negative.

§ Bonferroni-adjusted p values from two-way ANOVA for strain and embryonic day.

Supplemental Table 4.1. Expression for all genes as assayed by Nanostring nCounter

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
1	EG545999	XM_620562.3	I	132218849	-1.02	1.3	1.33	1	1
2	LOC670828	LOC670828.1	I	132362963	1.44	1.12	1.84	1	1
3	4930533L02Rik	AK015957.1	I	132461873	-1.33	1.13	1.41	1	1
4	4933440M02Rik	XR_035453.1	I	132473157	1.48	-1.16	1.25	1	1
5	Jmjd5-001	ENSMUST00000033010.1	I	132588190	1.19	1.35	1.3	1	1
6	Jmjd5-002	ENSMUST00000135129.1	I	132588190	-1.62	1.07	1.31	1	1
7	Nsmce1-001	ENSMUST00000033006.1	I	132611154	1.36	1.36	1.66	0.0852	1
8	Nsmce1-002	ENSMUST00000138616.1	I	132611154	2.52	2.05	3.57	0.0122	1
9	Nsmce1-003	ENSMUST00000149289.1	I	132611154	5.54	3.58	7.09	0.0084	1
10	EG244214	XR_002216.2	I	132690783	1.26	1.01	1.69	1	1
11	Il4ra-201	ENSMUST00000033004.1	I	132695785	-1.92	-1.93	-1.02	0.7669	1
12	Il21r-201	ENSMUST00000033000.1	I	132746983	-1.19	-1.1	-1.06	1	1
13	Il21r-202	ENSMUST00000084605.1	I	132746983	1.07	1.17	-1.19	1	1
14	Gtf3c1-201	ENSMUST00000055506.1	I	132784469	1.25	1.19	1.47	0.1134	0.1431
15	Gtf3c1-202	ENSMUST00000106423.1	I	132784469	1.04	-1.09	1.09	1	1
16	D430042O09Rik-204	ENSMUST00000122337.1	II	132851390	3.94	2.61	3.65	0.0073	1
17	D430042O09Rik-205	ENSMUST00000124223.1	II	132851390	1.13	-1.06	-1.14	1	1
18	D430042O09Rik-206	ENSMUST00000132204.1	II	132851390	1.24	-1.04	1.11	1	1
19	D430042O09Rik-209	ENSMUST00000155059.1	II	132851390	2.04	2.09	5.63	0.0284	1
20	AC150648.2	NM_001101488.1	II	133024217	-1.59	-1.28	-1.27	1	0.0448

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
21	Xpo6	NM_028816.2	II	133245237	1.49	1.25	1.63	0.0861	1
22	Sbk1	NM_145587.2	II	133416133	1.08	-1.11	-1.09	1	1
23	AC135809.1	XM_001479519.1	II	133442948	1.09	1.11	1.55	1	1
24	Lat	NM_010689.2	II	133507346	-1.34	1.34	1.09	1	1
25	Spns1	NM_023712.2	II	133513574	-1.25	1.11	1.23	1	1
26	Nfatc2ip	NM_010900.2	II	133526368	1.68	1.69	2.05	0.0125	1
27	Cd19	NM_009844.2	II	133552519	1.38	1.63	1.51	1	0.9776
28	Rabep2	NM_030566.2	II	133572273	1.31	1.41	2.22	0.2858	1
29	Atp2a1	NM_007504.2	II	133589372	1.38	1.15	1.34	1	1
30	Sh2b1	NM_001081459.1	II	133610508	1.18	1.09	1.36	1	1
31	Tufm	NM_172745.3	II	133630892	1.18	1.17	1.61	1	1
32	Atxn2l	NM_183020.1	II	133635224	1.46	1.21	1.52	0.1097	1
33	Eif3c	NM_146200.1	II	133690426	1.34	1.3	1.63	0.1887	1
34	Cln3	NM_001146311.1	II	133714721	1.6	1.28	2.23	0.0697	1
35	Apob48r	NM_138310.1	II	133728456	-1.07	1.12	1.38	1	1
36	Il27	NM_145636.1	II	133732524	1.44	-1.15	1.17	1	1
37	Nupr1	NM_019738.1	II	133766763	-1.72	1.21	1.7	1	0.0096
38	2510046G10Rik	ENSMUST00000084587.1	II	133792737	1.26	1.26	1.75	0.6995	1
39	Ccdc101	NM_029339.2	II	133792823	1.5	1.5	2.21	0.2086	1
40	Sult1a1	NM_133670.1	II	133816379	-2.29	1.01	1.76	1	0.0026

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
41	Giyd2	NM_029420.3	II	133832982	1.57	1.63	2.4	0.0291	1
42	Bola2	NM_175103.3	II	133838915	1.35	1.26	1.75	0.4726	1
43	Coro1a	NM_009898.2	II	133843287	1.15	1.05	1.42	1	0.4048
44	Mapk3	NM_011952.2	II	133903115	1.05	1.18	1.41	1	1
45	Gdpd3	NM_024228.2	II	133909928	1.74	1.35	2	0.8226	1
46	Ypel3	NM_026875.2	II	133920469	1.19	1.15	1.66	0.3037	1
47	Tbx6	NM_011538.2	II	133924997	2.38	2.09	2.89	0.0055	1
48	Ppp4c	NM_019674.3	II	133929421	1.16	1.18	1.33	1	1
49	Aldoa	NM_007438.3	II	133938748	1.25	1.04	1.58	1	1
50	Fam57b-201	ENSMUST00000079423.1	II	133963344	1.79	-1.08	1.55	1	1
51	Fam57b-202	ENSMUST00000098032.1	II	133963344	1.23	1.36	1.63	1	1
52	4930451I11Rik	NM_183131.2	II	133973988	-1.22	-1.3	1.19	1	1
53	Doc2a	NM_010069.1	II	133990930	3.7	-1.27	1.94	1	1
54	Ino80e	NM_153580.1	II	133995094	4.64	2.71	3.79	0.0343	1
55	Taok2	NM_001163774.1	II	134009192	1.65	1.38	1.97	0.0143	1
56	Tmem219	NM_028389.1	II	134029685	1.48	1.39	1.99	0.1386	1
57	Kctd13	NM_172747.2	II	134072393	1.26	-1.02	1.28	1	1
58	Sez6l2	NM_144926.4	II	134085658	1.36	-1.23	1.24	1	1
59	Asphd1	NM_001039645.1	II	134089081	1.7	1.08	1.61	1	1
60	Cdipt	NM_138754.3	II	134119902	1.06	1.28	1.5	1	1

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
61	Mvp	NM_080638.2	II	134130382	-1.26	1.13	1.21	1	1
62	2900092E17Rik	NM_030240.1	II	134144996	1.71	1.43	2.42	0.1198	1
63	Kif22	NM_145588.1	II	134171246	1.37	1.45	1.71	1	1
64	AI467606	NM_178901.3	II	134234873	-1.01	1.35	1.62	1	1
65	Qprt-001	ENSMUST00000032912.1	II	134250628	-1.55	-1.19	1.1	1	0.0377
66	Qprt-002	ENSMUST00000129332.1	II	134250628	1.35	1.72	1.81	0.2412	0.1273
67	Qprt-003	ENSMUST00000142403.1	II	134250628	1.61	2.19	4.1	1	1
68	Cd2bp2	NM_027353.3	II	134337200	1.32	1.23	1.46	0.0554	1
69	Tbc1d10b	NM_144522.5	II	134340979	1.01	-1.02	1.18	1	1
70	Mylpf	NM_016754.5	II	134355122	1.17	1.14	1.62	1	1
71	Sept1	NM_017461.2	II	134357961	1.33	1.68	1.66	0.2131	1
72	AC133494.1	NR_030674.1	II	134376338	3.03	1.9	3.23	0.0625	1
73	Zfp553	NM_146201.1	II	134376575	1.02	1.05	1.06	1	0.8373
74	Zfp771	NM_177362.3	II	134388040	1.16	1.05	1.31	1	1
75	Dctpp1	NM_023203.1	II	134400473	1.38	1.11	1.48	1	1
76	Sephs2	NM_009266.3	II	134416075	1.14	1.12	1.27	1	1
77	Itgal-003	ENSMUST00000120857.1	II	134439774	1.82	2.09	1.76	1	1
78	Zfp768	NM_146202.1	II	134486312	1.22	-1.02	1.65	1	1
79	Zfp747	NM_175560.3	II	134516078	1.49	1.42	1.98	0.0991	1
80	9130019O22Rik	NM_030226.2	II	134525774	3.63	2.28	3.78	0.019	1

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
81	AC133494.2	NM_198011.1	II	134527268	1.35	1.15	1.65	0.8822	1
82	Zfp764	NM_146203.3	II	134547182	1.4	1.44	2.1	0.1708	1
83	Zfp689	NM_175163.3	II	134587266	1.31	1.44	1.9	0.2063	1
84	Prr14	NM_145589.2	II	134603125	1.26	1.12	1.64	0.7564	1
85	Fbrs	NM_010183.1	II	134629170	1.22	1.14	1.46	0.2593	1
86	Srcap	XM_001480403.1	II	134655516	1.33	1.15	1.53	0.255	1
87	1700008J07Rik	NR_024331.1	II	134655683	2.64	1.87	3.05	0.0202	1
88	Phkg2	NM_026888.3	II	134716854	1.7	1.34	2.19	0.025	1
89	Gm166	NM_001033040.2	II	134726602	2.31	1.63	2.26	0.2001	1
90	Rnf40	NM_172281.1	II	134732281	1.18	1.25	1.42	0.1315	1
91	1700120K04Rik	NR_027915.1	II	134747592	3.37	1.99	2.35	0.0177	1
92	Zfp629	NM_177226.5	II	134750545	1.24	1.08	1.62	1	1
93	Bcl7c	NM_009746.2	II	134806867	-1.05	-1.01	1	1	1
94	Ctf1	NM_007795.1	II	134856258	1.53	1.1	1.54	0.2593	0.1427
95	Ctf2	NM_198858.1	II	134861620	1.32	1.05	1.15	1	1
96	Fbxl19	NM_172748.2	II	134890289	1.47	1.07	1.5	1	0.8551
97	Orai3	NM_198424.3	II	134913329	2	1.59	2.73	0.1136	1
98	Setd1a	NM_178029.3	II	134920184	1.39	1.3	1.58	0.1413	1
99	Hsd3b7	NM_133943.2	II	134929122	1.06	1.27	1.53	1	0.2238
100	Stx1b	NM_024414.2	II	134947414	1.93	1.03	1.76	1	1

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
101	Stx4a	NM_009294.3	II	134967808	1.34	1.41	1.88	0.7125	1
102	Zfp668	NM_146259.3	II	135008684	1.7	1.21	1.82	0.1367	1
103	Zfp646	NM_172749.4	II	135020310	1.35	1.31	1.59	0.076	1
104	BC039632	NM_001081268.1	II	135029355	1.88	1.29	2.15	1	1
105	Vkorc1	NM_178600.2	II	135029741	1.07	1.27	1.68	1	1
106	Myst1	NM_026370.1	II	135056031	1.56	1.33	2.1	0.0889	1
107	Prss8	NM_133351.2	II	135069232	-1.26	1.5	1.71	1	1
108	Prss36	NM_001081374.1	II	135076152	1.15	1.25	1.65	1	1
109	Fus	NM_139149.2	II	135110971	1.47	1.49	1.72	0.0428	0.6911
110	B230325K18Rik	NM_176936.2	II	135126593	9.69	4.68	7.75	0.0165	1
111	Pycard	NM_023258.4	II	135135617	-2.29	-1.67	-1.79	0.0003	0.0595
112	Trim72	NM_001079932.2	II	135147503	2.19	2.07	2.27	0.7767	1
113	Itgax	NM_021334.2	II	135273061	1.25	1.75	1.36	1	1
114	Itgad	NM_001029872.1	II	135298292	1.65	-1.07	1.06	1	1
115	Cox6a2	NM_009943.2	II	135349128	1.05	1.18	1.76	1	1
116	9130023H24Rik	NM_177001.3	II	135379920	1.21	1.39	1.74	0.6463	1
117	Armc5	NM_146205.2	II	135381630	-1.15	1.03	1.07	1	1
118	Tgfb1i1	NM_009365.2	II	135390385	5.01	4.35	5.85	0.0634	1
119	Slc5a2	NM_133254.3	II	135409171	1.14	1.29	1.5	1	1
120	BC017158	NM_145590.1	II	135414893	1.56	1.43	1.67	0.0226	1

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
121	Rgs10	NM_026418.2	II	135517135	1.1	-1.22	1.31	1	1
122	Tial1	NM_009383.2	II	135583291	2.02	1.54	2.12	0.0418	1
123	Bag3	NM_013863.4	II	135667130	-1.42	-1.16	-1.15	1	1
124	Inpp5f	NM_178641.5	II	135754842	1.9	1.69	2.3	0.0036	1
125	Ctbp2	NM_001170744.1	II	140179304	1.22	1.15	1.33	1	1
126	LOC100043248	XM_001479821.1	II	141616800	2.49	-1.04	1.31	1	1
127	Dock1	NM_001033420.2	II	141862370	1.87	2.15	2.47	0.0384	1
128	Angiopoietin 1	NM_009640.3	III	42256273	1.05	1.01	1.66	1	1
129	Angiopoietin 2	NM_007426.3	III	18690263	-1.15	1.14	1.61	1	0.0077
130	CLIC4	NM_013885.2	III	134769884	-1.17	1.08	1.25	1	1
131	Dll4	NM_019454.2	III	119151520	1.09	1.43	1.6	1	1
132	Ephrin B2	NM_010111.5	III	8617434	1.24	1.02	1.05	1	1
133	Flk1	NM_010612.2	III	76328852	1.13	1.46	1.61	1	1
134	Flt1	NM_010228.3	III	148373180	1.15	1.86	2.15	0.2966	1
135	KLF2	NM_008452.2	III	74842932	-1.66	1.17	-1.31	1	0.8971
136	KLF4	NM_010637.3	III	55540015	-1.43	1.18	1.29	1	0.0591
137	PDGF-A	NM_008808.3	III	139451968	1.43	1.11	1.29	1	1
138	PDGF-B	NM_011057.3	III	79826330	-1.18	1.32	1.12	1	1
139	TGFb	NM_019919.2	III	75404869	1.3	1.18	1.32	0.1245	0.0087
140	Vegfa188	NM_001025250.3	III	46153942	1.88	1.45	2.05	0.0236	1

Supplemental Table 4.1 (continued)

Gene* number	Name	Accession number	Gene† Group	Start	Fold‡ E14.5	Fold‡ E16.5	Fold‡ E18.5	Bonadj§ strain	Bonadj§ Eday
141	AMPK	NM_001013367.3	IV	5093861	1.35	1.3	1.69	0.4531	1
142	Ki67	NM_001081117.2	IV	142881772	1.37	1.65	1.55	0.9798	1
143	LKB1	NM_011492.3	IV	79578548	1.15	1.11	1.26	1	1
144	p16INK4a	NM_009877.2	IV	88920377	1.44	1.39	1.3	1	1
145	p21	NM_007669.4	IV	29227924	1.33	1.13	1.32	1	1
146	p27	NM_009875.4	IV	134870419	1.4	1.22	1.68	0.0893	1
147	p53	NM_011640.1	IV	69393861	1.22	1.39	1.6	0.2812	1
148	PCNA	NM_011045.2	IV	132074898	1.31	1.11	1.33	1	1
149	SIRT 1	NM_019812.1	IV	62781753	1.48	1.14	1.57	1	1
150	telomerase	NM_009354.1	IV	73764438	3.99	2.14	3.87	0.0085	1

Expression assay by Nanostring nCounter. 3 RNA samples of pia at each time point for each strain, with each sample composing of ≥ 8 embryos from ≥ 2 litters.

***** Gene number same as in Figure 4.7 and Table 4.1.

† **I**, genes within the EMMA region; **II**, gene in 95% CI of Chr 7 QTL; **III**, angiogentic-related genes; **IV**, proliferation/aging-related genes

‡ Fold change for BALB/c vs B6 if positive and B6 vs BALB/c if negative.

§ Bonferroni-adjusted p values from two-way ANOVA for strain and embryonic day.

Figure 4.1. Collateral extent in CXB11.HiAJ (CXB11) confirms importance of Chr 7 QTL. A, Inheritance patterns of CXB11 (BALB/c X C57BL/6) recombinant inbred line (RIL) within Chr 7 QTL. Y axis, SNP positions on Chr 7 in Build 37. Genotypes of informative SNPs inherited from BALB/c (Bc) denoted in black and C57BL/6 denoted in grey. CXB11 inherits the EMMA region through Itgal gene from B6 (arrows). **B**, Collateral number is intermediate between B6 and Bc in CXB11. **C**, Fractions of the cerebral cortex area (territories) supplied by the ACA, MCA, and PCA. B6 and Bc have identical territories (confirms Zhang et al. (12)), while CXB11 has slightly smaller MCA, larger ACA and smaller PCA territories than B6. Differences in those territories do not correlated with variation in collateral number or diameter in these strains. Numbers of mice given in parentheses.

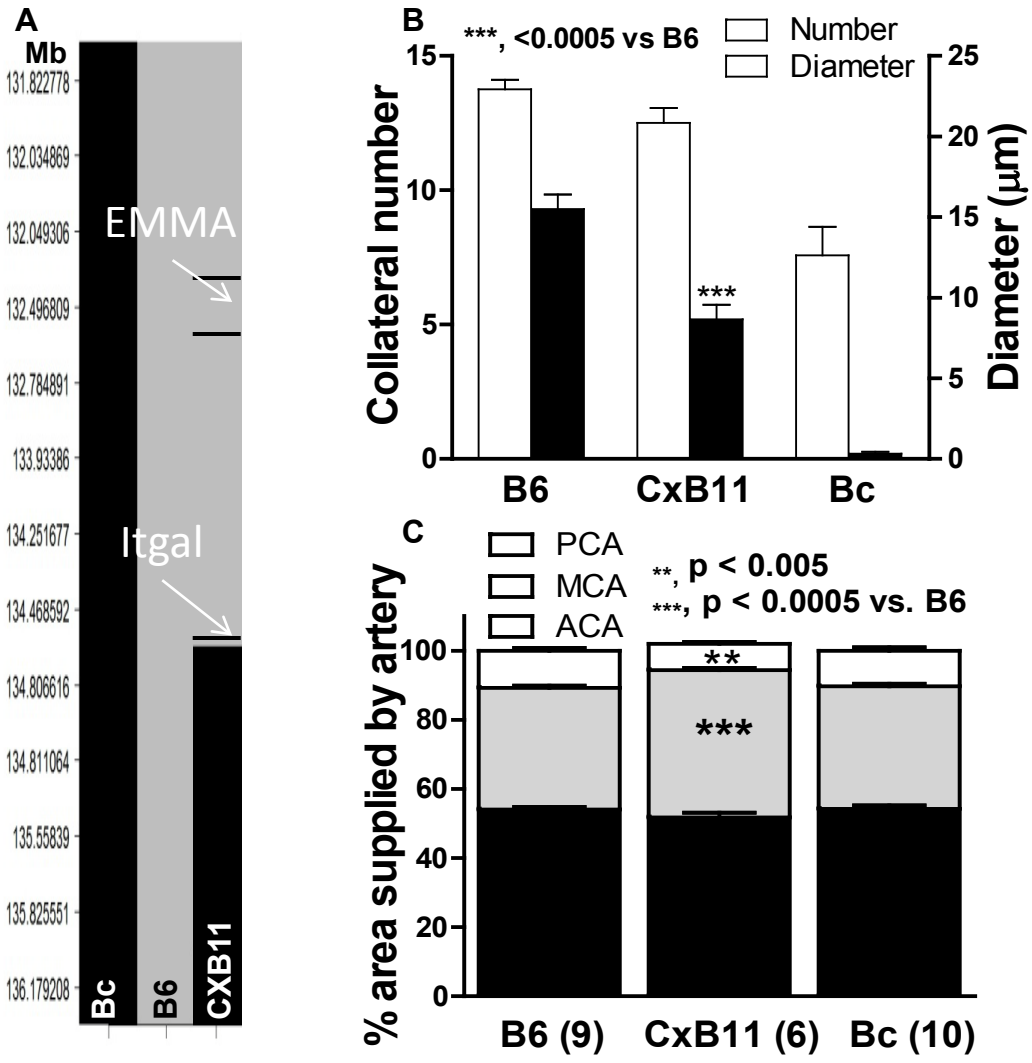


Figure 4.2. Heatmap of the imputed SNPs for 74 strains within the EMMA region. X axis, the EMMA region ($p = 2.2X10^{-5}$). Abscissa, physical locations of selective SNPs; Ordinate, 74 inbred strains. Each vertical line represents a SNP (Blue, SNP genotype is same as in B6; Green, different from B6; Blank, genotype unknown).

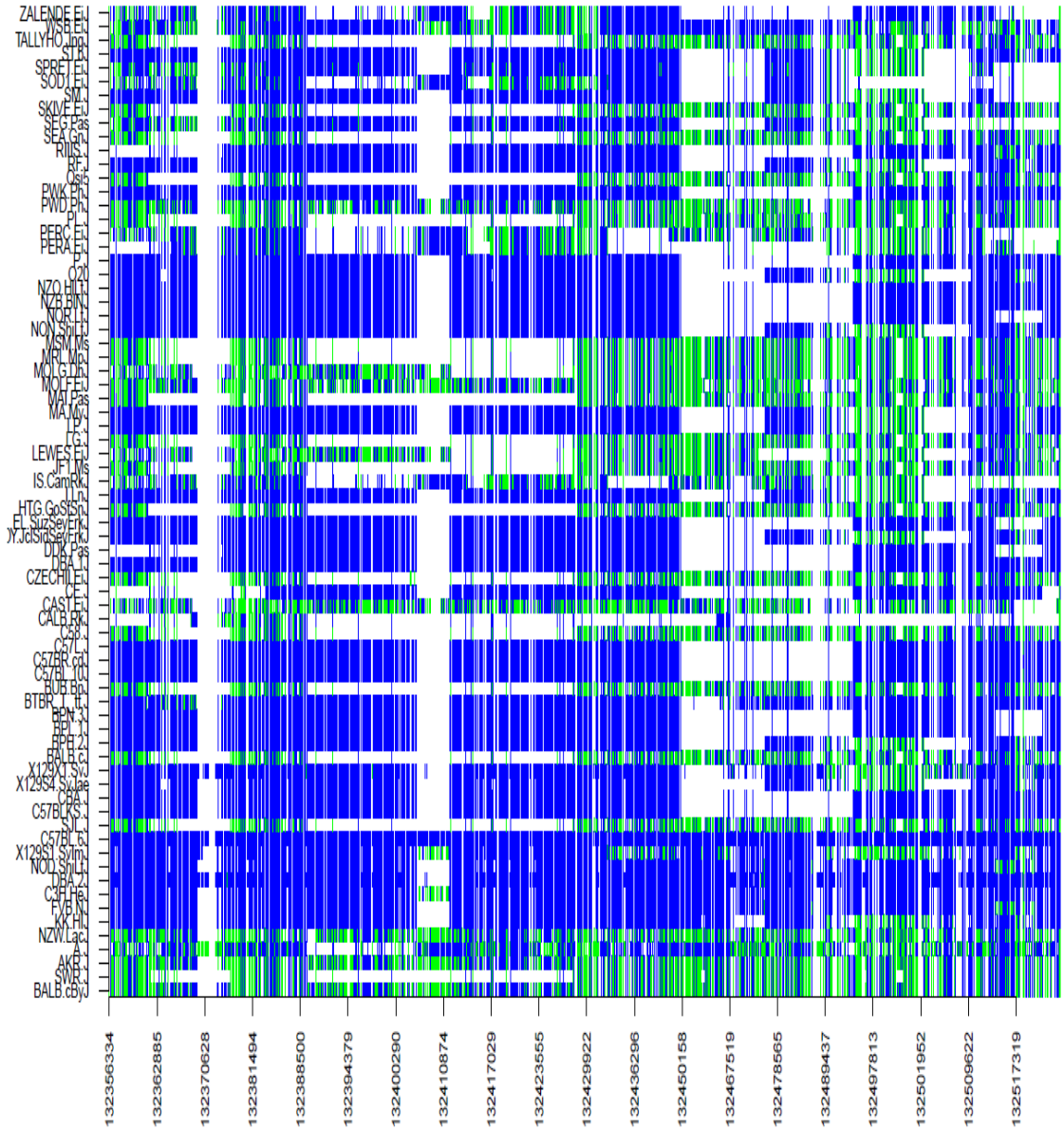


Figure 4.3. Collateral number for 21 inbred strains and heatmap of their SNPs within the EMMA region. Left, Heatmap of known and imputed SNPs for 21 strains within the EMMA region (color scheme, same as Figure 4.2). **Right**, collateral number for 21 strains. Strain names in red denote newly phenotyped strains. Data for other strains from Zhang et al. N = 8-10/bar. Dashed line, reference to B6.

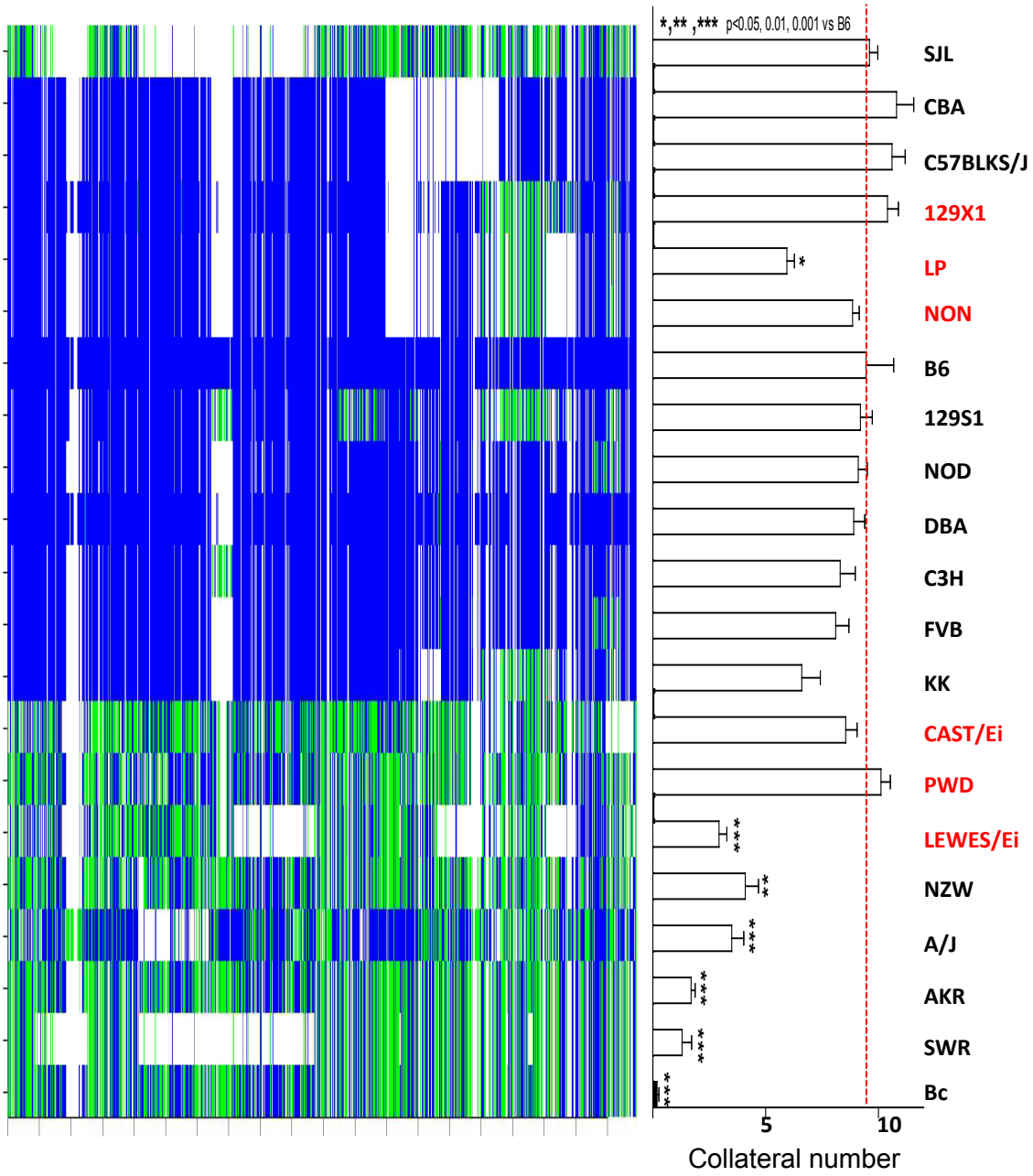


Figure 4.4. Six more strains (red strains in Figure 4.3) strengthen and narrow the EMMA region. The EMMA algorithm was applied to collateral number for 21 strain data (6 new strains plus15 from Zhang et al. (12)). ~49,000 high quality imputed SNPs under the *Canq1* peak were tested. The previous highly significant region was strengthened and narrowed down to 1.7Kb ($p < 9 \times 10^{-6}$, 4 SNPs). Black and grey bars denote the most significant and the second most significant regions in the previous study (3). Mapping allows fewer than three SNPs with missing genotypes.

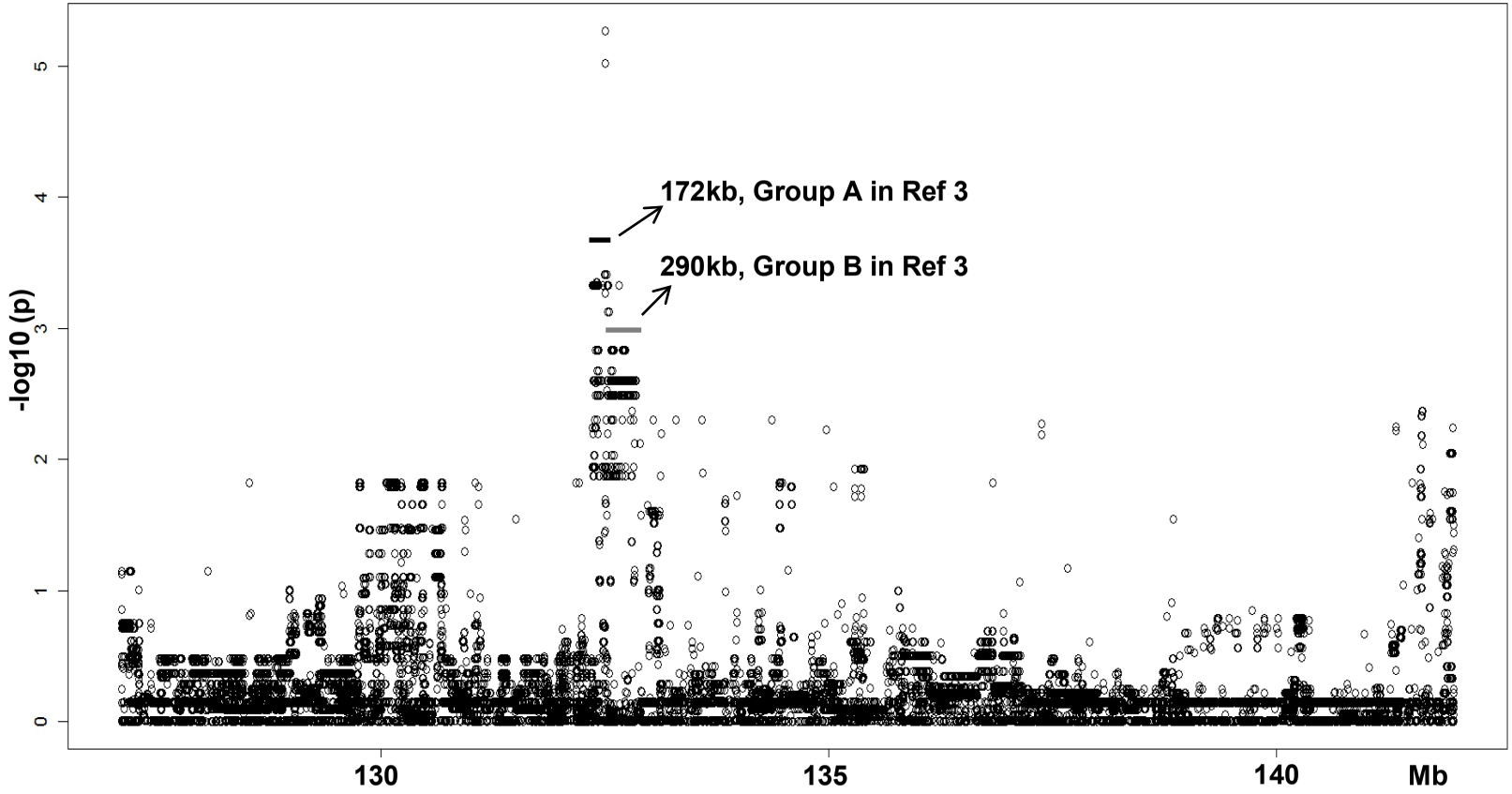


Figure 4.5. Collateral extent not altered in Itgal, IL4 or IL4 receptor alpha (Ra) deficient mice relative to their host strains. **A,** Itgal $^{-/-}$ and IL4 $^{-/-}$ mice are B6 background; IL4Ra $^{-/-}$ are Bc background. Comparisons in both number and diameter between Itgal $^{-/-}$, IL4 $^{-/-}$ and B6 are not significantly different, nor are those between IL4Ra$^{-/-}$ and Bc (t tests). **B,** Fractions of the cerebral cortex area (territories) supplied by the ACA, MCA, and PCA. B6 and Bc are identical, while those in Itgal$^{-/-}$ and IL4Ra$^{-/-}$ differ, confirming dissociation of collateral and cerebral artery tree territory phenotypes among inbred mouse strains (Ref. 12). Number of mice given in parentheses.

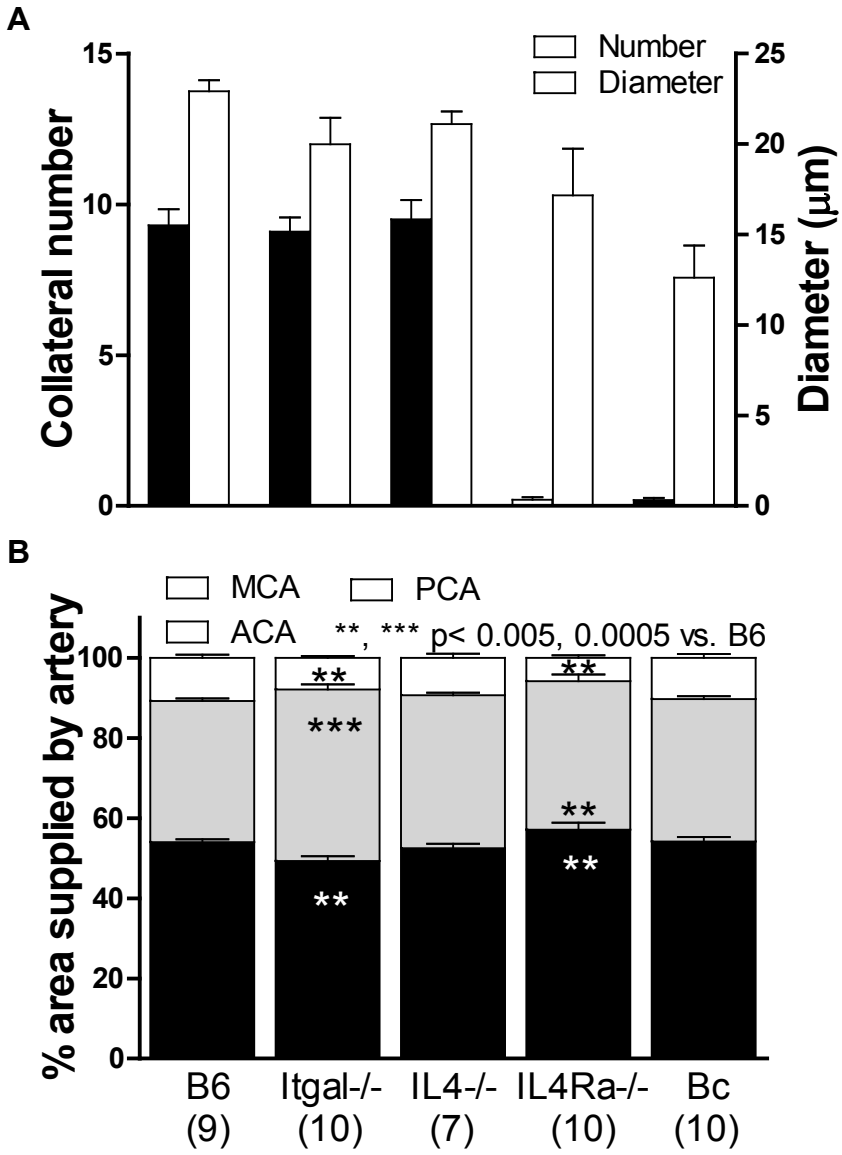


Figure 4.6. Tail bleeding assays and MCA occlusion-induced cerebral infarct volume. Itgal$^{-/-}$ and IL4$^{-/-}$ mice are B6 background. Bleeding time and infarct volume on knockouts not different from B6; Rebleeding time in IL4$^{-/-}$ not different from B6, but shorter in Itgal$^{-/-}$ (t-tests). Number of mice given in parentheses.

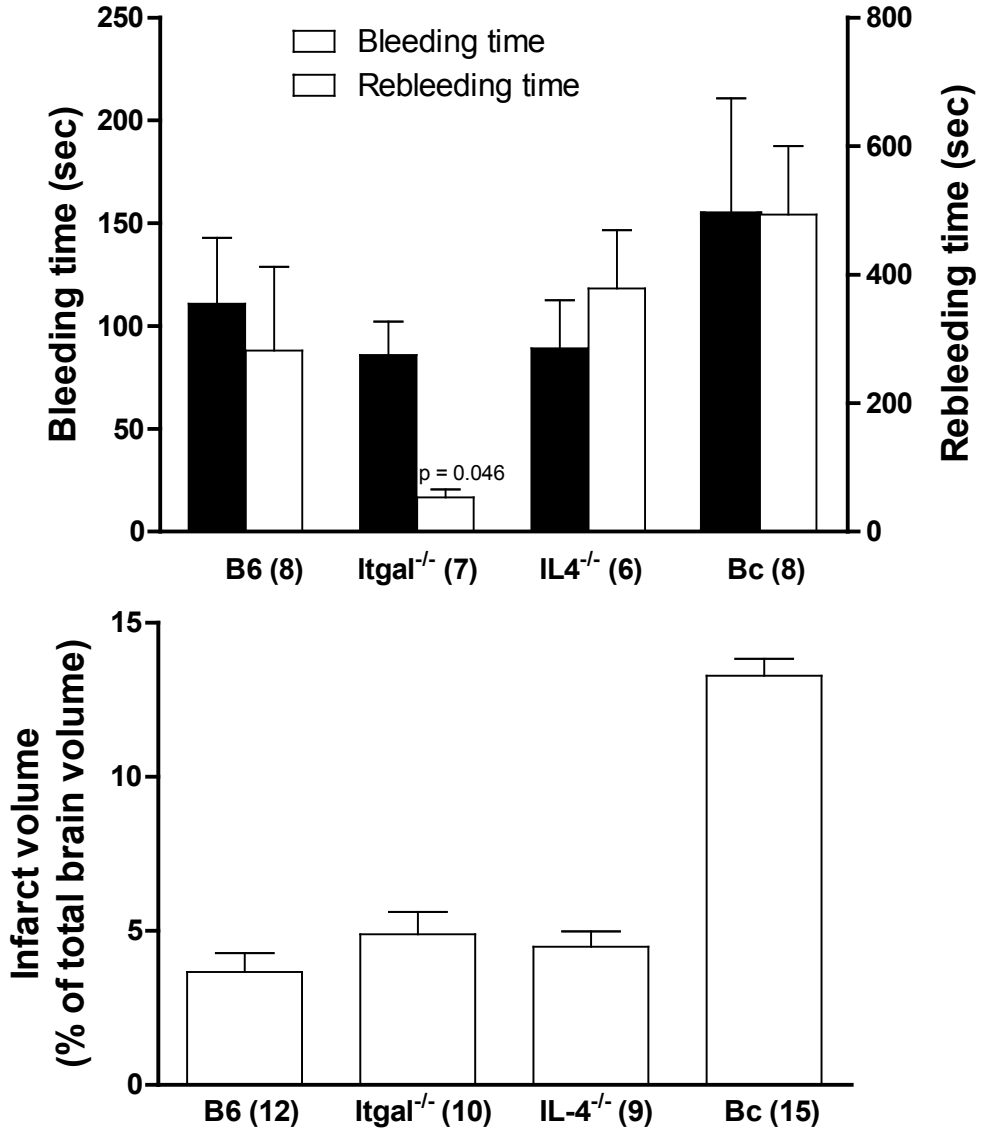
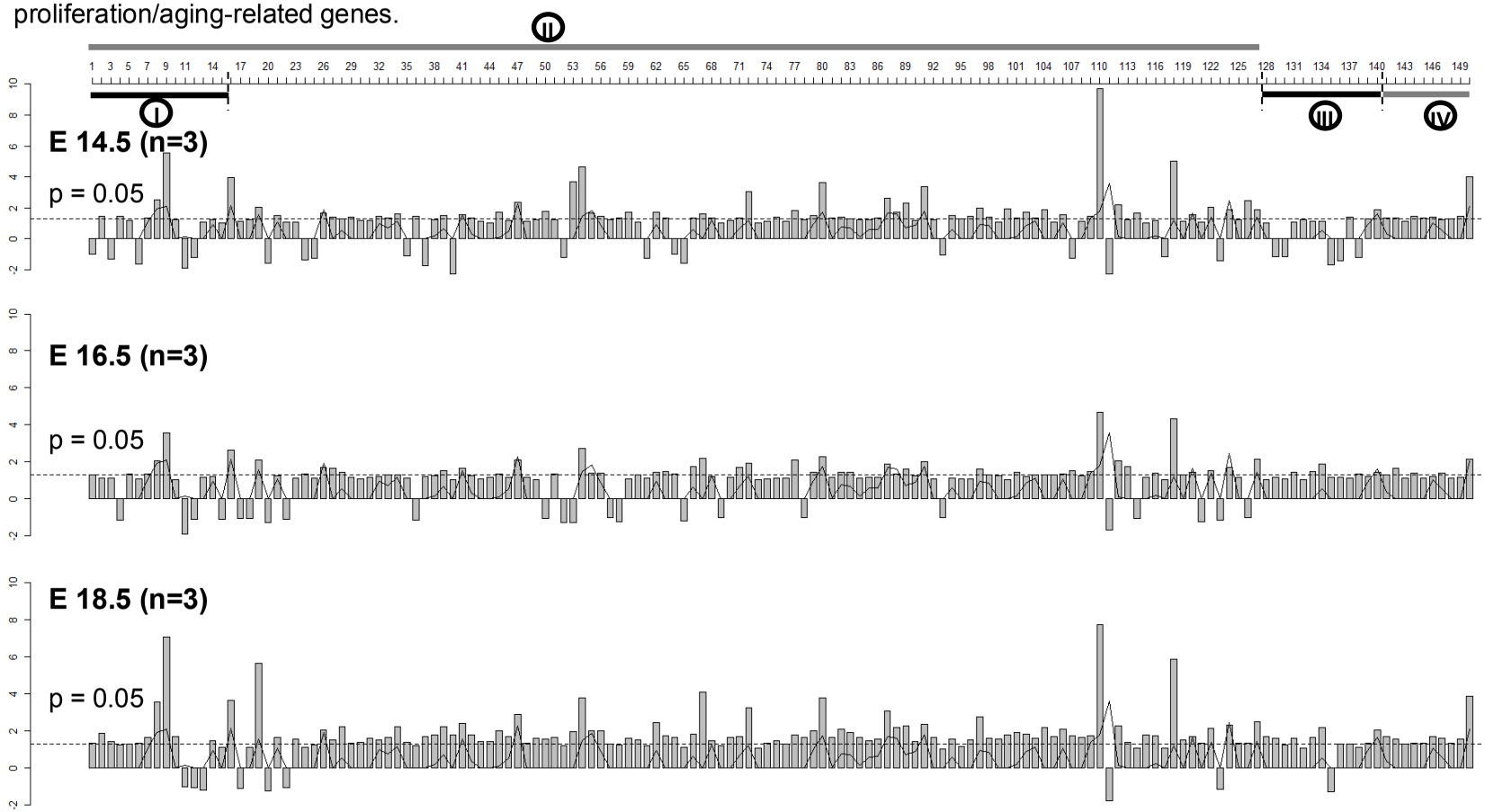
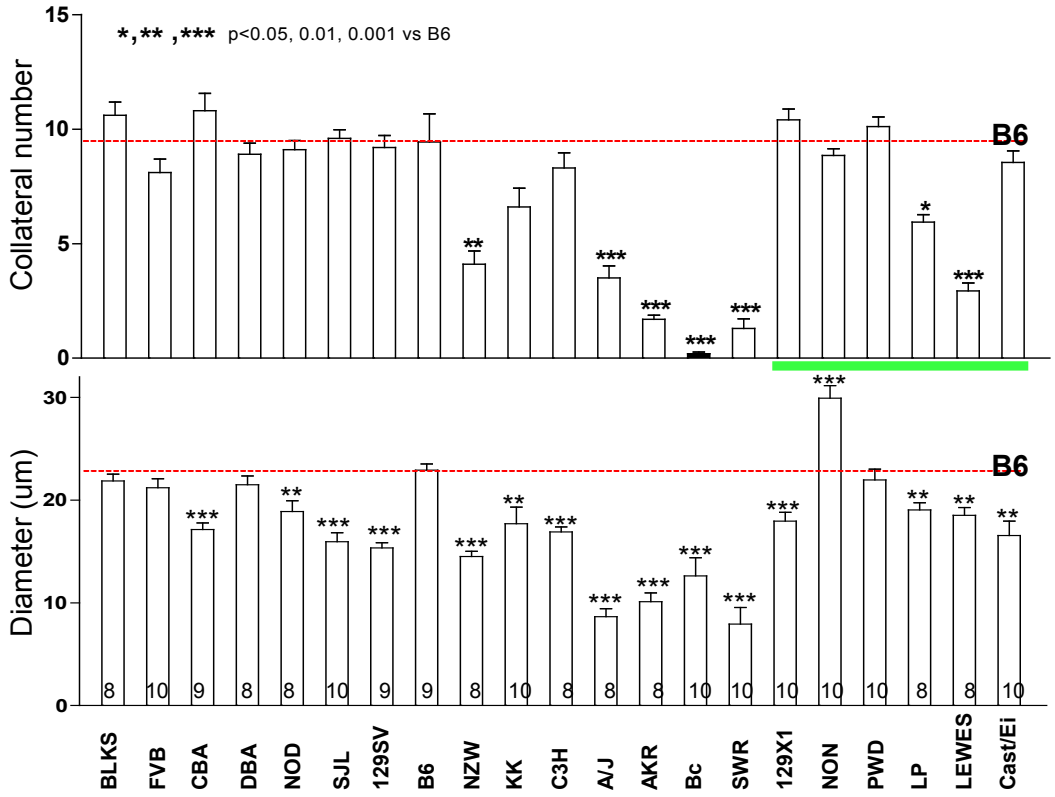


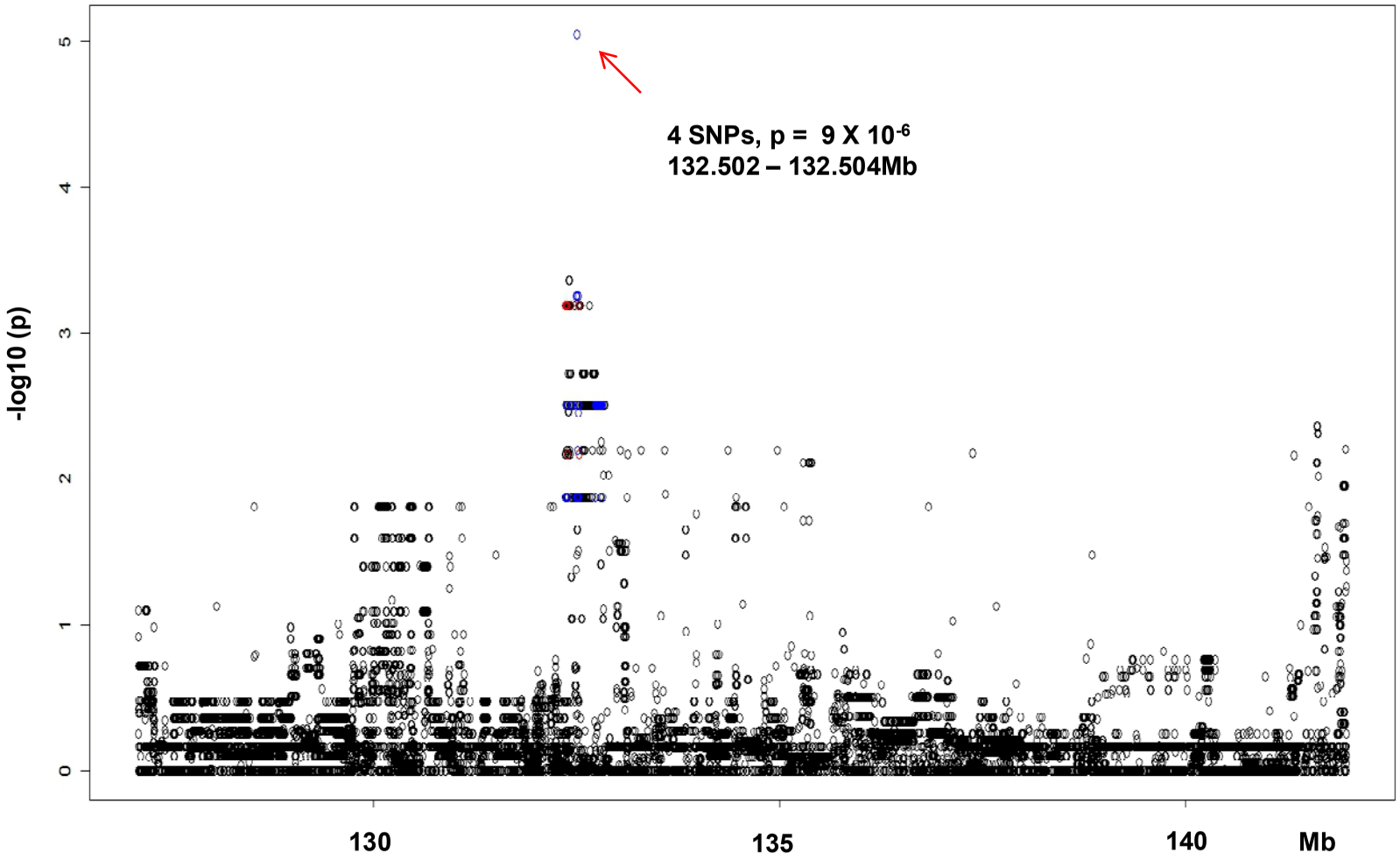
Figure 4.7. Fold changes between BALB/c and C57BL/6 in mRNA levels of genes within Chr 7 QTL and selected angiogenic- and proliferation/aging- related genes at embryonic day E14.5, E16.5 and E18.5. The numbers 1-150 at top denote gene numbers in Supplement Table 4.1. Y-axis, Fold changes BALB/c vs B6 if positive; B6 vs BALB/c, if negative. Black curve, negative logarithm adjusted p values for strain effect in two-way ANOVA (Bonferroni correction). Genes 1-128 located in 95% confidence interval of the Chr 7 QTL are arranged in order of physical locations. **I**, genes within the EMMA region (132.356 to 132.528 Mb); **II**, genes in 95% CI of Chr 7 QTL; **III**, angiogenic-related genes; **IV**, proliferation/aging-related genes.



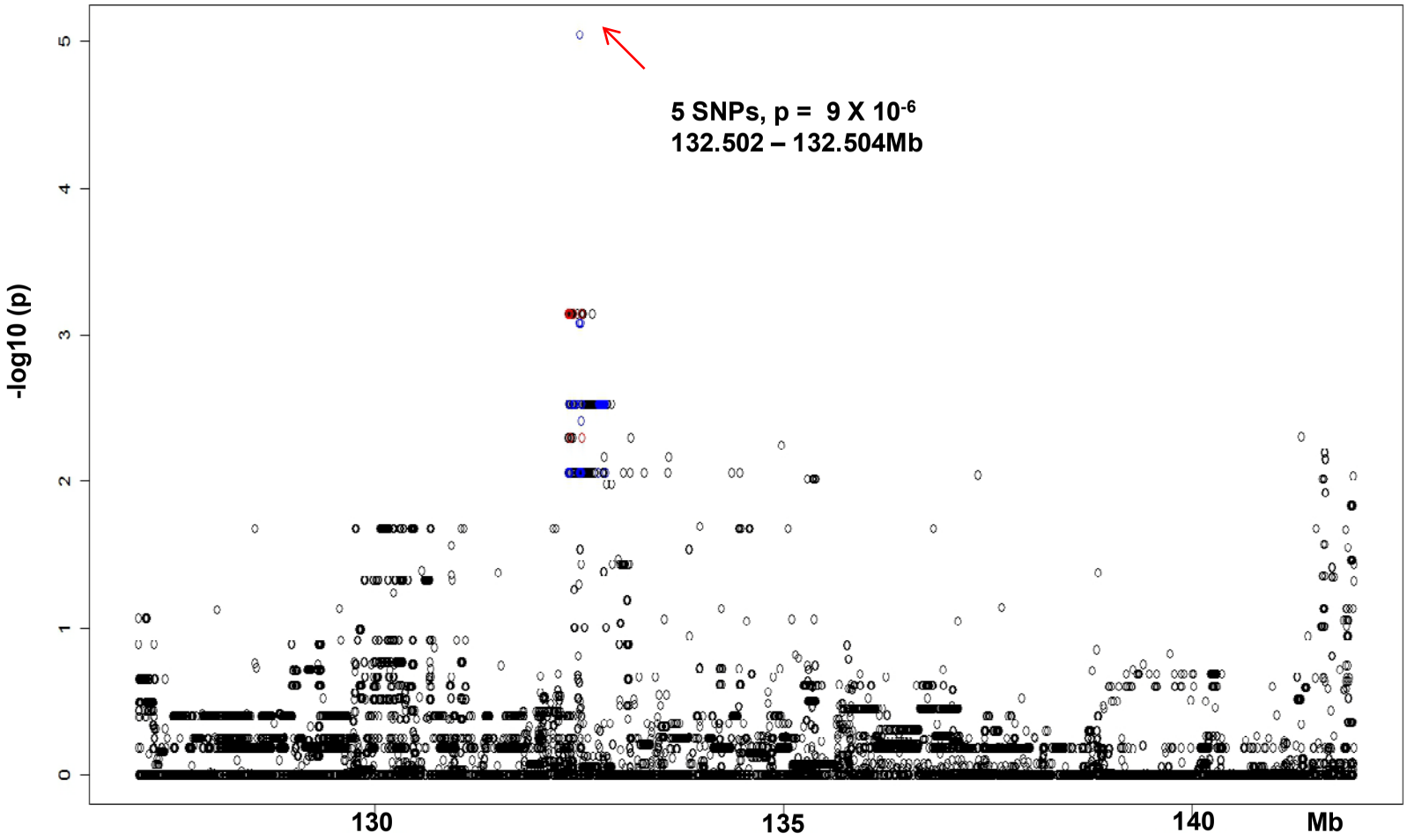
Supplemental Figure 4.1. Pial collateral number per hemisphere and diameter for 21 inbred strains, including 15 strains published before (ref 12). Green bar denotes six newly phenotyped strains. Number of animals is given at the base of each column.



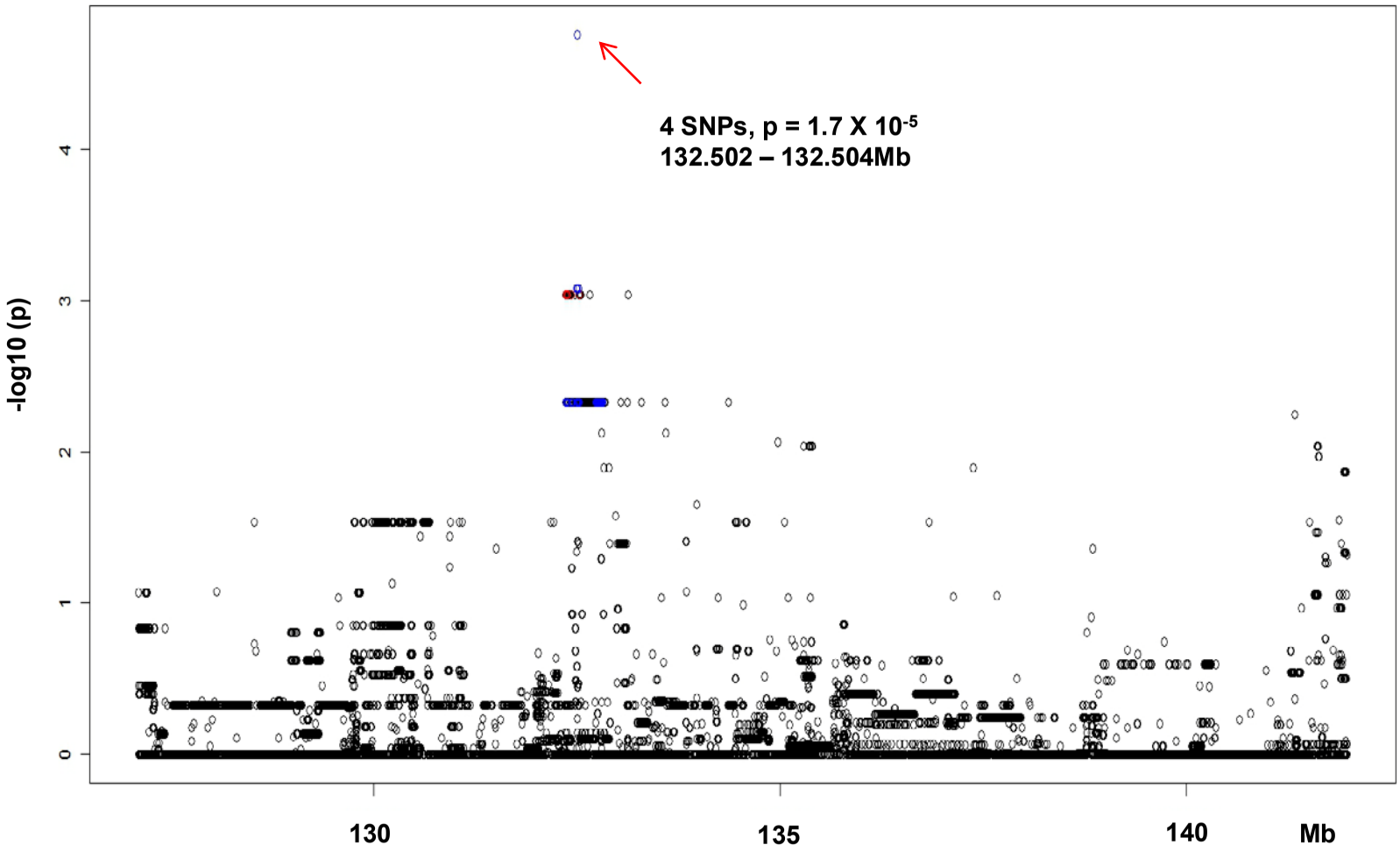
Supplemental Figure 4.2. EMMA mapping using 20 strains excluding Cast/Ei. Red and blue dots, the most significant and the second significant SNPs in the previous EMMA mapping (Ref 3), the same color scheme in the following EMMA mappings. Allows fewer than 3 SNPs with missing genotypes



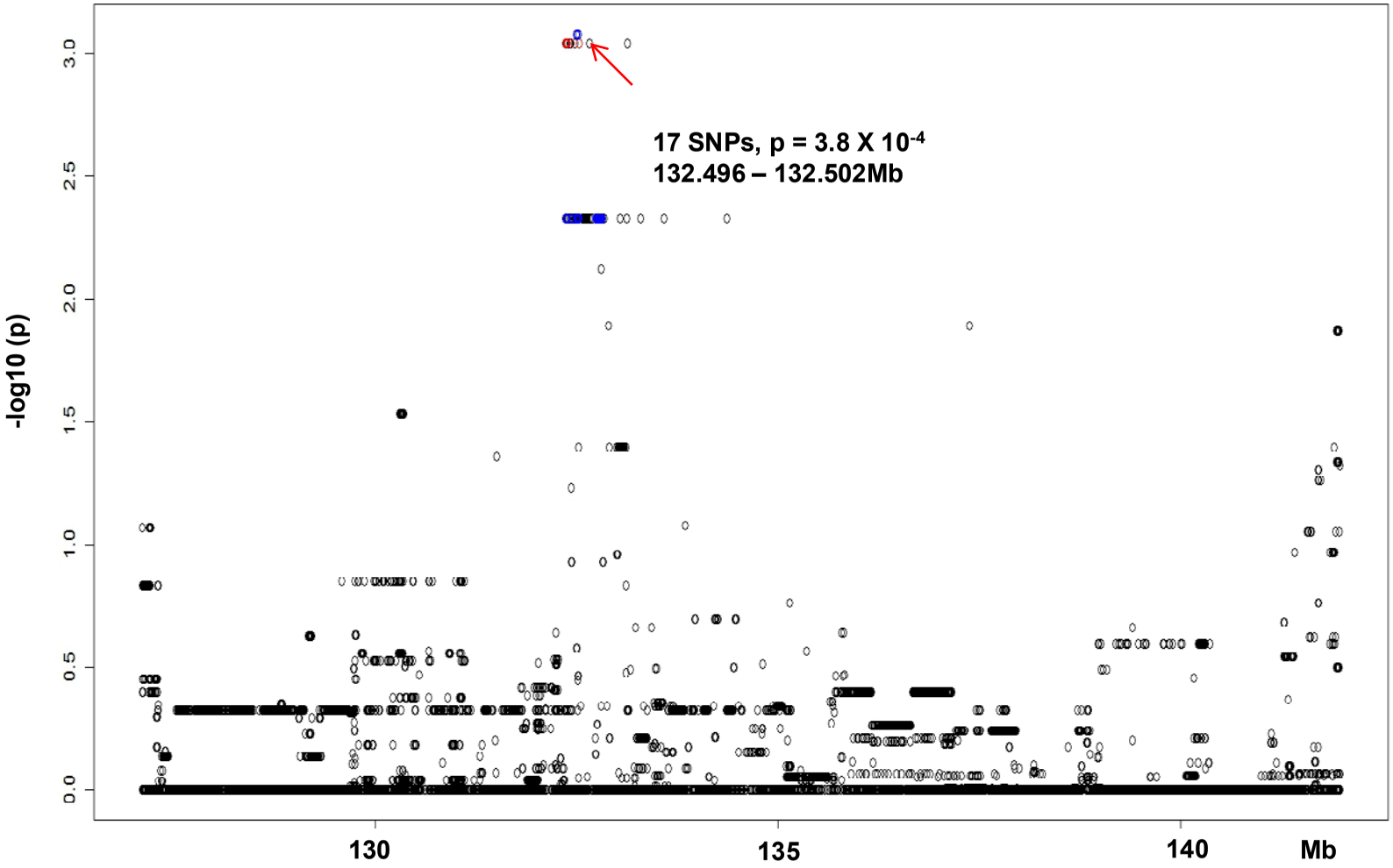
Supplemental Figure 4.3. EMMA mapping using 19 strains excluding Cast/Ei and PWD. Allows fewer than 3 SNPs with missing genotypes



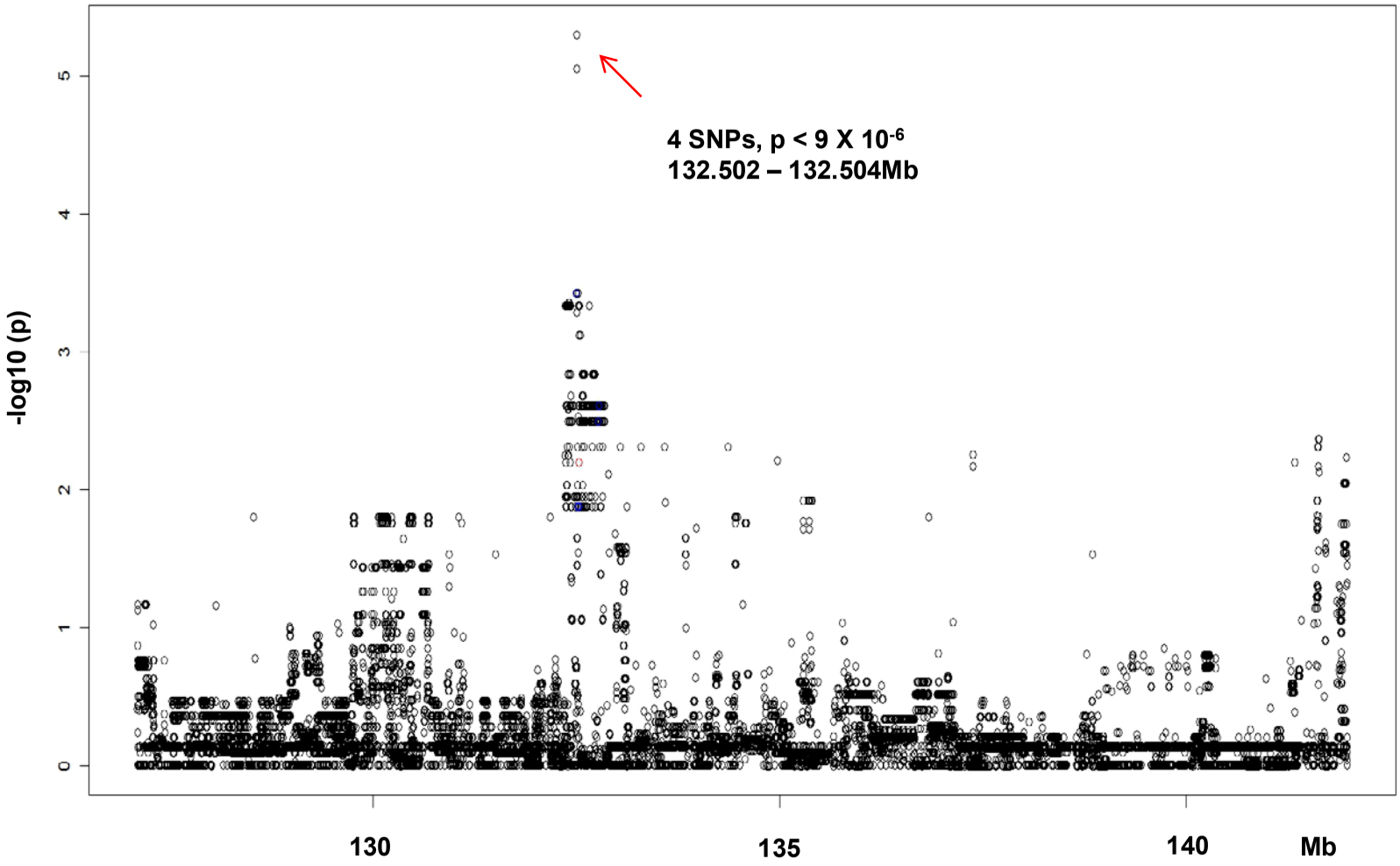
Supplemental Figure 4.4. EMMA mapping using 18 strains excluding Cast/Ei, PWD, and LEWES. Allows fewer than 3 SNPs with missing genotypes



Supplemental Figure 4.5. EMMA mapping using 21 strains. No SNPs with missing genotypes allowed.



Supplemental Figure 4.6: EMMA mapping using 21 strains with only informative SNPs. Allows fewer than 3 SNPs with missing genotypes



CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Summary

Occlusive vascular disease, including myocardial infarction, stroke, and chronic ischemic disease of the heart, brain and lower extremities, is the leading cause of morbidity and mortality in the developed countries [3, 16]. Two vascular protective mechanisms during ischemia have drawn much attention: the extent of native collaterals—the number and diameter of arteriole-to-arteriole anastomoses present in the tissue before the onset of disease—has become increasingly appreciated, as well as their ability to remodel which has been studied for a number of years[1]. These native collaterals not only supply blood flow immediately after occlusion, but their baseline number and diameter in a tissue sets the initial capacity for their remodeling/growth in diameter (arteriogenesis) which improves recovery of flow over the days-to-weeks after arterial occlusion. Arteriogenesis describes the process of a small collateral (~30 microns diameter in healthy tissue) increasing its anatomic diameter in 2-10 fold, thus growing into an artery[5-8]. Since conductance is related to diameter to the fourth power, the resultant increases in capacity to provide flow to ischemic tissue is considerable[1, 113, 155, 156]. The degree of ischemic injury is thus primarily determined by the native collateral extent and the amount of remodeling.

Many lines of evidence suggest that native collateral extent and collateral remodeling are subjected to the wide variation among individuals in both humans and inbred mice[12, 24, 26, 30, 31]. Previous studies have shown that C57BL/6 (B6) and BALB/c (Bc) have dramatically different collateral extent and remodeling[16, 34]. To dissect the genetic components underlying these variations, we created 243 F2s by reciprocally crossing B6 and Bc. Around 10 weeks of age, a permanent MCAO was conducted in one side of brain in each F2 mouse. 6 days after MCAO, collateral number and diameter were measured in both hemispheres and DNA was prepared. 228 informative SNPs which evenly span the whole genome were genotyped. Four phenotypes

obtained: native collateral number, average diameter, and collateral conductance (number X diameter) in the normal non-ligated side of brain, and fold-increase in diameter 6 days after MCAO were mapped to the genome, in order to identify QTL[3].

A. Chr 7 QTL predominantly determines native collateral extent

The collateral extent in F2s was not affected by differences in sex, parental origin, age, or body weight. Mapping using both genome-wide single and multiple QTL models detected a major QTL on Chr 7 for three native collateral traits (number, diameter, and the product—an index of conductance). In addition to Chr 7 QTL, multiple QTL model mapping also identified additional QTL on Chr's 1, 3, 8 and an interaction between the QTL on Chr 1 and Chr 3 (LOD > 5 for all loci and the interaction). The Chr 7 QTL for native collateral number (designated as *Canq1*) alone accounts for 37% phenotypic variation, whereas the rest of genetic components are responsible for 35% of the variation. This result establishes a predominant role of the Chr 7 QTL in determining native collateral number. The peak of Chr 7 QTL is located 0.5 cM telomeric to marker rs13479513, with a high level of significance (134.25 Mb, $p < 2X10^{-16}$). The allele inherited from Bc at the position right underneath the Chr 7 QTL peak was predicted to reduce collateral number relative to the B6 allele. Furthermore, I used the chromosome substitution strain-7, i.e., C57BL/6 mice with Chr 7 replaced by Chr 7 from the A/J strain, a distinct low collateral density strain, and verified the direction and effect size of the Chr 7 QTL. In this strain, the substitution caused most of the native collateral phenotype from B6-like to become A/J-like, thus confirming the predominant effect of this QTL. Interaction between the Chr 7 QTL and *Vegfa*, *Clic4*, or *Dll4* genes, which were previously found to affect collateral extent[1, 34, 37] was not detected by likelihood ratio tests[3]. Importantly, the Chr 7 QTL-*Canq1* is at the same locus as two previously

reported QTL: *LSq-1* and *Civq1*. *LSq-1* was identified for both hindlimb necrosis and perfusion recovery 21 days after femoral artery ligation[17], whereas *Civq1* was detected for cerebral infarct volume 24 hours after MCAO[18]. Both *LSq-1* and *Civq1* were identified in B6 and Bc crosses. However, the underlying physiological mechanisms were not identified in these studies, owing to the complexity of the traits measured. Along with recent findings by Chalothorn et al.[36], my results strongly suggest that physiological basis for these QTL is that this locus is critical in determining variation in native collateral number and diameter.

B. A QTL on Chr 11 links to collateral remodeling

Arteriogenesis, initiated by hemodynamic changes upon occlusion, is also subject to variation among mouse strains[12]. Although much is known about molecular pathways for collateral remodeling[7, 8, 113], no unbiased genetic study has been conducted to identify underlying risk factors. Fold increase in collateral diameter 6 days after MCAO, a parameter indicating collateral outward growth, was also mapped in B6 X Bc F2s. As expected, a QTL on Chr 7 again was identified to be associated with collateral remodeling. As noted before, the remodeling process is stimulated by shear stress. And shear stress is negatively correlated to native collateral diameter[157]. Thus, it is not surprise that remodeling has a same QTL on Chr 7 as native collateral diameter does. When native collateral diameter was set as a covariate in the subsequent mapping, Chr 7 QTL disappeared. Instead, a new significant QTL on Chr 11 was obtained: *Carq1*. Although the effect size for this QTL is small, this finding is the first time to directly show a genetic component underlying the remodeling[3]. Interestingly, the studies cited above only found the Chr7 QTL for response to ischemic injury, but failed to identify *Carq1*[17, 158].

C. Association mapping narrows down the candidate genetic elements in the Chr 7 QTL

Association mapping in inbred mice, previously termed *in silico* mapping, was first proposed by Grupe et al[76]. However, along with haplotype association mapping, this approach has been criticized primarily because of high false-positives due to limited sample sizes and the relatedness embedded among inbred strains[74, 75, 77, 79]. Two previous studies related to ischemic injury employed interval-specific SNP haplotype association mapping in which only 3 or 4 strains were used[17, 18]. Here, I confined association mapping only within the Chr 7 QTL and applied the efficient mixed model association (EMMA) algorithm in which the relatedness among 15 classical inbred strains was accounted for by a kinship matrix[3, 79]. The most significant 172kb region ($p = 2.2 \times 10^{-5}$), designated as the EMMA region hereafter, and a second most significant region ($p = 4.2 \times 10^{-4}$) adjacent to the first, were obtained by EMMA mapping using data that was previously obtained for native collateral number for the 15 strains[3, 12].

To verify and possibly narrow down the EMMA region, I employed a novel strategy: Using a heat map approach, I visualized SNP variation within the most significant EMMA region for 74 inbred strains, employing known and high quality imputed genotypes. Using this visualization, I chose 6 inbred strains judged to be informative by their similarities in haplotype structure within the EMMA region to that in B6-high collateral density strain or that in Bc–low density strain. Four of six strains fit the expectation by their haplotype similarities: 129X1/Sv, NON/ShiLt, LP, and LEWES/Ei (Chapter IV). Two strains - Cast/Ei and PWD that did not fit the expectation are wild-derived strains. Recent studies suggested that these two strains share < 3% haplotype ancestral origins with classical inbred strains[151, 152, 159]. On the other hand, LEWES, also a wild-derived strain, has few collaterals—as predicted by its haplotype structure.

Results from a study that densely genotyped SNPs for 94 inbred strains suggest that LEWES shares more haplotype ancestral origins with classical inbred strains than do Cast/Ei or PWD[150]. Further EMMA mapping for collateral number using these six strains along with the previous 15 strains refined the previous most significant region and narrowed it to several SNPs (Chapter IV).

D. Collateral extent in knockout mice argues against a role for Itgal or IL4-receptor-α

To identify which genes govern native collateral extent, we examined gene-deleted mice for three candidate genes, Itgal$^{-/-}$ and IL4$^{-/-}$ on the B6 background and IL4ra$^{-/-}$ on the Bc background. All three had similar pial collateral extents as their respective host strains. This finding argues against the notation that these three genes harbor causal alleles for collateral variation. We further examined infarct volume after MCAO-induced ischemic injury in Itgal$^{-/-}$ and IL4$^{-/-}$ mice. The results indicated that these gene-deletion mice also have similar infarct volume as their host strains. This result is consistent with our findings of collateral extent and argues against a regulatory role of these two genes (Chapter IV). However, knockout strategies have limitations. For instance, knockout strains are usually made in X129 embryonic stem cells. The resulting mice are then bred back to the desired strain for 6-10 generations. The backcross leaves an island of X129 DNA at the site of the mutated gene that, theoretically, can have confounding effects from incorporated neighboring X129 genes. Limitations in conclusions can also arise from potential compensation due to deletion. Moreover, gene deletion does not test for an allele change that causes a new function on one of the respective backgrounds. Nevertheless, gene-deleted mice remain an important approach to examine candidate genes[160].

E. mRNA quantification of genes within the 95% CI of *Canq1* identifies high priority genes

Using the cerebral pial collateral circulation as a model, Chalothorn et al. discovered that collaterals begin forming at ~E14.5 and peak in number at ~ E18.5 in both B6 and Bc[38]. Furthermore, he found that differences in extent of the collateral circulation in adults are primarily determined during collateral formation in the embryo[38], although processes governing stabilization of nascent collaterals in the first few weeks of life and maintenance of them during growth to adulthood also contribute[161]. Therefore, I dissected the pial membrane containing the dorsal cortical pial circulation from B6 and Bc at E14.5, E16.5 and E18.5 when the collateral circulation is forming. At each time point for each strain, three RNA samples were prepared with each from at least 8 embryos derived from at least 2 litters. I then measured expression of 150 genes comprised of 120 genes within 95% CI of *Canq1,* plus angiogenic-related, proliferation/aging-related genes. Among the 120 genes, 19 were expressed differentially between B6 and Bc, with 9 consistently more than 2-fold higher in Bc than in B6 at each time-point. One gene, Pycard, was expressed at least 60% higher in B6 at all time-points. Nsmce1 (132.6 Mb), located in the second significant EMMA region (132.53 - 132.82 Mb), has three splicing isoforms, two of which express at least 2-fold higher in Bc (Chapter IV).

Future Directions

F. A new F2 cross—SWR X SJL—to identify genetic components regulating collateral extent

My EMMA mapping identified the SJL strain as an outlier: SJL and SWR share the same genotypes for the most significant SNPs within the EMMA region. However, SJL averages 9.6 pial collaterals while SWR has 1.3 (Figure 4.3). Thus, what genetic components are responsible for the variation in collateral number between these two strains? We hypothesized that the genetic architecture underlying variation in collateral number between SJL and SWR is different from that between B6 and Bc. To test this hypothesis, I created 125 F2 crossed between SJL X SWR. For each mouse at ~10 weeks of age, we filled the pial circulation with Microfil and counted pial collateral number microscopically following the procedures described before[3, 12]. The genotypes of 1449 SNPs across the genome for each mouse were obtained using the GoldenGate middle density linkage array. Linkage analysis and association mapping to identify underlying novel genetic components is currently underway.

G. advanced search for genetic sources of variation in collateral formation in the mouse as a model for man

The Collaborative Cross (CC) is a large panel of recombinant inbred lines (RILs) derived from a genetically diverse set of founders [162-164] that could be applied to future extension of my findings. The CC, which is being completed through efforts at several centers around the world including UNC, will provide a unique and powerful resource for functional studies of biological networks that will be essential for understanding the mechanisms of many diseases[163]. The eight founders consist of five classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/H1LtJ) and three wild-derived strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ). These strains are predicted to

capture on average 89% of the known allelic diversity in the mouse genome [108]. We have already phenotyped 4/5 classical and 1/3 wild-derived strains for collateral extent, with a good distribution of differences in collateral number and diameter that the three additional strains will likely enhance.

As reported by the latest review for the CC, ~500 eight-way RI lines are extant, with the goal of eventually acquiring ~1000 lines[162]. Compared to other genetic sources such as traditional recombinant inbred lines from two progenitors, congenic strains, chromosome substitution strains, heterogeneous stocks, and the mouse phenotype diversity panel, the CC has multiple advantages[108]: (1) The CC incorporates a large amount of variation, with the distribution of minor allele frequency approximating that in human. For example, CXB RILs only capture 14% variation while the CC captures 89% of the variation in the mouse genome[108]. Conducting association mapping for collateral extent in the CC will likely allow us to identify many novel genetic components because of the diverse genetic variation; (2) The CC provides association mapping with a larger sample size and better precision. While association mapping for collateral extent in the CC can identify loci more relevant to human population, the sufficient sample size (up to 1000 RILs) can also achieve better statistical power: 500 RILs can detect a QTL having 5% genetic effect with power of 0.67 and precision of 0.96 cM [165, 166]. With such a precision and power, one can predict that the Chr 7 QTL, with its large effect, will be narrowed down to the 1-2 gene level. And many QTL with small effects can be detected with better resolution; (3) In addition, the CC offers reproducibility and cost efficiency. Unlike F2s, unlimited mice with identical genetic background for each line are available in the CC[163]. This feature is especially important in searching for genes governing collateral extent. For example, it is not possible to extract enough RNA from collaterals in the pial circulation of a single mouse for expression analysis. Moreover, filling the pial collaterals with yellow Microfil, required for

phenotyping, is not compatible with RNA extraction in the same mouse. This technical difficulty greatly limits our capability to identify candidate genes: It eliminates genome-wide expression-QTL (eQTL), systems genetics and causality analyses. With the CC, we could dissect individual pial collaterals (10-40um diameter X 100-300um length) in multiple mice from each line, as well as phenotype multiple mice in a second group of the same line. This would allow us to perform, for example, expression QTL mapping. Here, *cis* expression QTL (eQTL), defined as cis-eQTL with mapping position nearby and proximal to a given gene's location, can be used to prioritize the candidate genes for a QTL in linkage analysis. Combining gene expression data and collateral extent with genotypic data would permit performance of causal inference analysis to identify the underlying genes responsible for variation in collateral extent [110, 167, 168], i.e., those genes that control the process of collateral formation. Furthermore, the CC is cost-efficient because the genotypes of the dense SNPs for all RILs in the cross are already known[162].

H. Detailed mapping of the genetic architecture of collateral remodeling

In their study, Dokun et al. reported an identical QTL on chromosome 7 for recovery of flow (LOD= 3.71, p= 0.05) and tissue necrosis score (LOD = 7.96, p = 0.001) after femoral artery ligation (FAL) in a B6 X Bc mapping population. The recovery of flow and tissue necrosis score at day 7, 14, and 21 after FAL are largely dependent on collateral number, diameter, and outwards remodeling. Since collateral extent and remodeling were not quantified in this study, the underlying physiological mechanisms were not revealed. Wang et al. identified a highly significant QTL at the same locus for pial collateral traits: number, diameter, and conductance in 243 B6 X Bc F2s. Subsequent multi-QTL mapping showed that native (baseline) diameter is a major determinant of collateral remodeling. This is in agreement with the well-known fact that

shear-stress varies by diameter-cubed, and that the shear stress applied to collateral endothelial cells after artery obstruction is the initiating stimulus and major subsequent driving stimulus for the biological program governing collateral remodeling. This implies that loci governing variation in baseline collateral diameter (I found one with the same locus as for number on chromosome 7, i.e., *Cadq1)* will be, indirectly, major determinants of variation in collateral remodeling. However, multi-QTL mapping with baseline diameter treated as a co-variate, identified a barely significant QTL on Chr 11 for collateral remodeling independent of initial diameter[3]. This finding is particularly promising, because it was derived from a cross that other work in the lab has now shown that the two strains are not optimal parentals to map remodeling trait. Zhang et al (2010) showed that remodeling of pial collaterals is only slightly weaker in Bc compared to B6. In contrast, A/J are 3-fold stronger than B6, a finding recently confirmed in hindlimb collaterals by Chalothorn et al[36]. Taken together, these findings indicate that variation in collateral remodeling is not only governed by baseline diameter, but also by genetic elements distinct from those that specify native diameter. Zhang et al. also measured pial collateral outward remodeling in 12 additional strains. While they found that B6 and Bc have similar fold-change in diameter (~1.8), remodeling in SWR is ~5-fold stronger than that in Bc and the strongest among all 15 strains studied. Indeed, the rank order of collateral number and diameter are distinct from that for remodeling among the 15 strains. These results partially explain why we found a barely significant but distinct QTL on Chr 11 for remodeling. As suggested by many lines of evidence, collateral remodeling is subject to variation [12, 89, 156]. The phenotypic variation between parentals in a segregating population will greatly affect the power to detect QTL (Churchill, 2007). To map the genetic components underlying collateral remodeling, a mapping population of Bc X SWR would be predicted to be much more likely to confirm, with a higher LOD score, the Chr 11 QTL for remodeling, as well as to identify additional QTL independent

of native collateral diameter, since diameter and number are comparable in these two strains[12].

I. Identify additional low-collateral-density strains

Our ability to identify the genetic source of variation in collateral number, the most important determinant of tissue injury in occlusive diseases, is greatly impaired by few low-collateral-density strains. Among 21 strains phenotyped for collateral extent, only 6 strains are significantly lower than B6 in number (and none is significantly higher). All of these strains originated from the *Mus domesticus* subspecies, allowing us to postulate that the variant we identified on Chr 7 QTL arose as a mutation during the history of generation of the classical inbred strains in the past 100 years[159, 169]. We hasten to point out here, as we discussed in Chapter I, that wide variation in native collateral extent also appears to exist in the healthy human population[24]. The heatmap representation (Chapter IV), which facilitates choosing informative strains, is convenient to construct. However, it lacks strength for haplotype inference and did not prove perfect for predicting collateral phenotype. To identify more strains with low collateral density, ancestral haplotype origins for inbred strains need to be inferred, especially for strains originating from *Mus domesticus*.

Wade et al.[159] observed high SNP-rate (~40 SNPs/10 kb) and low SNP-rate (~0.5 SNPs/10 kb) regions alternatively within the genome of most inbred strains, using strain-strain pair-wise comparisons[169]. The high SNP-rate regions cover one third of the genome on average. An ancestry breakpoint was defined as a location at which any of the pairwise comparisons transited from a low SNP-rate to a high SNP-rate region. Therefore, the ancestral haplotype can be inferred using this approach. Further, they showed that positional cloning of a region identified in a QTL mapping study can be

efficiently obtained by correlating phenotypes with ancestral haplotypes. As an example, genetic components for albinism were quickly narrowed to a 500kb region close to the tyrosinase gene using this method[169]. Two subsequent studies utilized SNP data from the 15 strain re-sequencd data by Perlegen to infer the ancestral haplotype origins for several classical inbred strains using Hidden Markov Model (HMM)[151, 152]. Both studies found that the classical inbred strains inherited the majority of their genomes from *Mus domesticus*. To infer ancestral haplotype origins, the genotypes for dense SNPs are needed for more inbred strains. A recent study by Kirby et al. expanded this effort and genotyped 121,000 SNPs for 94 inbred strains, with SNPs dispersing at 20kb interval on average across the genome[150]. Since the major source of genetic variation governing collateral extent may have arisen during the generation of classical inbred strains, one novel approach going forward would be to select classical inbred strains different from the 21 we've studied and to infer their ancestral haplotype origins based on the dense SNPs within the Chr 7 QTL published by Kirby et al[150]. A proposed study is to choose non-wild-derived inbred lines among the 94 densely genotyped strains. High SNP-rate and low SNP-rate regions can then be inferred in pairwise strain-strain comparisons using the HMM algorithm. Haplotype structure can then be inferred by defining ancestry break points among any of the pairwise comparisons. Based on the inferred haplotype structure, one can then identify predicted low-collateral-density strains to phenotype.

Given the high prevalence of arterial obstructive disease, a therapeutic compound to stimulate new collateral vessel formation or enlargement of the diameter of the native population in adults could have a major impact on morbidity and mortality. Although results from animal studies over the past 10 years have held promise for various agents to stimulate arteriogenesis, not a single drug has proved applicable in clinical practice thus far, either because of lack of efficacy (due to either poor response or high variability in response) or because of undesired side effects[170]. If the causal

gene(s) or genetic element(s) for the Chr 7 QTL can be identified, the orthologous loci in the human could then be tested as both biomarkers for collateral extent and assignment of risk-severity for stroke, myocardial infarction, peripheral artery disease and other obstructive conditions. Such a biomarker(s) would also allow patients to be stratified in trials of new proposed collateralogenic therapies, thus reducing the variation in the test population by assigning patients to low and high collateral extent groups. Moreover, discovering these variants would, by definition, point to the signaling molecules responsible for formation of these unique vessels in the embryo. This information would then permit a rational approach to developing therapies aimed at re-activating this developmental process in adult tissues.

CHAPTER VI

REFERENCES

1. Faber JE DX, Lucitti J. *Genetic and environmental mechanisms controlling formation and maintenance of the native collateral circulation*. Science; 2010.

2. Brozici M, van der Zwan A, Hillen B. Anatomy and functionality of leptomeningeal anastomoses: A review. *Stroke*. 2003;34:2750-2762

3. Wang S, Zhang H, Dai X, Sealock R, Faber JE. Genetic architecture underlying variation in extent and remodeling of the collateral circulation. *Circulation research*. 2010

4. Deindl E, Schaper W. The art of arteriogenesis. *Cell Biochem Biophys*. 2005;43:1-15

5. Busch HJ, Buschmann IR, Mies G, Bode C, Hossmann KA. Arteriogenesis in hypoperfused rat brain. *J Cereb Blood Flow Metab*. 2003;23:621-628

6. Deindl E. Arteriogenesis: A focus on signal transduction cascades and transcription factors. *Thromb Haemost*. 2007;98:940-943

7. Grundmann S, Piek JJ, Pasterkamp G, Hoefer IE. Arteriogenesis: Basic mechanisms and therapeutic stimulation. *Eur J Clin Invest*. 2007;37:755-766

8. Helisch A, Schaper W. Arteriogenesis: The development and growth of collateral arteries. *Microcirculation*. 2003;10:83-97

9. Wei L, Erinjeri JP, Rovainen CM, Woolsey TA. Collateral growth and angiogenesis around cortical stroke. *Stroke*. 2001;32:2179-2184

10. de Groot D, Pasterkamp G, Hoefer IE. Cardiovascular risk factors and collateral artery formation. *Eur J Clin Invest*. 2009;39:1036-1047

11. Kinnaird T, Stabile E, Zbinden S, Burnett MS, Epstein SE. Cardiovascular risk factors impair native collateral development and may impair efficacy of therapeutic interventions. *Cardiovasc Res*. 2008;78:257-264

12. Zhang H, Prabhakar P, Sealock R, Faber JE. Wide genetic variation in the native pial collateral circulation is a major determinant of variation in severity of stroke. *J Cereb Blood Flow Metab*. 2010;30:923-934

13. Harmon KJ, Couper LL, Lindner V. Strain-dependent vascular remodeling phenotypes in inbred mice. *Am J Pathol*. 2000;156:1741-1748

14. Dodd-o JM, Hristopoulos ML, Welsh-Servinsky LE, Tankersley CG, Pearse DB. Strain-specific differences in sensitivity to ischemia-reperfusion lung injury in mice. *J Appl Physiol*. 2006;100:1590-1595

15. Riggs HE, Rupp C. Variation in form of circle of willis. The relation of the variations to collateral circulation: Anatomic analysis. *Arch Neurol*. 1963;8:8-14

16. Chalothorn D, Clayton JA, Zhang H, Pomp D, Faber JE. Collateral density, remodeling, and vegf-a expression differ widely between mouse strains. *Physiological genomics*. 2007;30:179-191

17. Dokun AO, Keum S, Hazarika S, Li Y, Lamonte GM, Wheeler F, Marchuk DA, Annex BH. A quantitative trait locus (lsq-1) on mouse chromosome 7 is linked to the absence of tissue loss after surgical hindlimb ischemia. *Circulation*. 2008;117:1207-1215

18. Keum S, Marchuk DA. A locus mapping to mouse chromosome 7 determines infarct volume in a mouse model of ischemic stroke. *Circulation.Cardiovascular genetics*. 2009;2:591-598

19. Winkler B SS, Binz K, Schaper W. Myocardial blood flow and myocardial infarction in rats, guinea pigs and rabbits. *J Mol Cell Cardiol*. 1984;16

20. Maxwell MP, Hearse DJ, Yellon DM. Species variation in the coronary collateral circulation during regional myocardial ischaemia: A critical determinant of the rate of evolution and extent of myocardial infarction. *Cardiovasc Res*. 1987;21:737-746

21. Kume T. Specification of arterial, venous, and lymphatic endothelial cells during embryonic development. *Histol Histopathol*. 2010;25:637-646

22. Herrick JB. Landmark article (jama 1912). Clinical features of sudden obstruction of the coronary arteries. By james b. Herrick. *JAMA*. 1983;250:1757-1765

23. Schaper W. *Experimental infarcts and the microcirculation.*: Raven Press; 1984.

24. Meier P, Gloekler S, Zbinden R, Beckh S, de Marchi SF, Zbinden S, Wustmann K, Billinger M, Vogel R, Cook S, Wenaweser P, Togni M, Windecker S, Meier B, Seiler C. Beneficial effect of recruitable collaterals: A 10-year follow-up study in patients with stable coronary artery disease undergoing quantitative collateral measurements. *Circulation*. 2007;116:975-983

25. Spaan JA, Piek JJ, Hoffman JI, Siebes M. Physiological basis of clinically used coronary hemodynamic indices. *Circulation*. 2006;113:446-455

26. Seiler C. The human coronary collateral circulation. *Heart*. 2003;89:1352-1357

27. Bang OY, Saver JL, Buck BH, Alger JR, Starkman S, Ovbiagele B, Kim D, Jahan R, Duckwiler GR, Yoon SR, Vinuela F, Liebeskind DS. Impact of collateral flow on tissue fate in acute ischaemic stroke. *J Neurol Neurosurg Psychiatry*. 2008;79:625-629

28. Christoforidis GA, Karakasis C, Mohammad Y, Caragine LP, Yang M, Slivka AP. Predictors of hemorrhage following intra-arterial thrombolysis for acute ischemic stroke: The role of pial collateral formation. *AJNR Am J Neuroradiol*. 2009;30:165-170

29. Miteff F, Levi CR, Bateman GA, Spratt N, McElduff P, Parsons MW. The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke. *Brain*. 2009;132:2231-2238

30. Maas MB, Lev MH, Ay H, Singhal AB, Greer DM, Smith WS, Harris GJ, Halpern E, Kemmling A, Koroshetz WJ, Furie KL. Collateral vessels on ct angiography predict outcome in acute ischemic stroke. *Stroke*. 2009;40:3001-3005

31. Debakey ME, Simeone FA. Battle injuries of the arteries in world war ii : An analysis of 2,471 cases. *Ann Surg*. 1946;123:534-579

32. Barone FC, Knudsen DJ, Nelson AH, Feuerstein GZ, Willette RN. Mouse strain differences in susceptibility to cerebral ischemia are related to cerebral vascular anatomy. *J Cereb Blood Flow Metab*. 1993;13:683-692

33. Majid A, He YY, Gidday JM, Kaplan SS, Gonzales ER, Park TS, Fenstermacher JD, Wei L, Choi DW, Hsu CY. Differences in vulnerability to permanent focal cerebral ischemia among 3 common mouse strains. *Stroke*. 2000;31:2707-2714

34. Clayton JA, Chalothorn D, Faber JE. Vascular endothelial growth factor-a specifies formation of native collaterals and regulates collateral growth in ischemia. *Circulation research*. 2008;103:1027-1036

35. Sen S, Satagopan JM, Broman KW, Churchill GA. R/qtldesign: Inbred line cross experimental design. *Mamm Genome*. 2007;18:87-93

36. Chalothorn D, Faber JE. Strain-dependent variation in collateral circulatory function in mouse hindlimb. *Physiol Genomics*. 2010;42:469-479

37. Chalothorn D, Zhang H, Smith JE, Edwards JC, Faber JE. Chloride intracellular channel-4 is a determinant of native collateral formation in skeletal muscle and brain. *Circ Res*. 2009;105:89-98

38. Chalothorn D, Faber JE. Formation and maturation of the native cerebral collateral circulation. *J Mol Cell Cardiol*. 2010;49:251-259

39. Lucitti J FJ. Vascular endothelial growth factor specifies native collateral formation during embryogenesis. *Keystone Angiogenesis Conference*. 2010

40. Wang DB, Blocher NC, Spence ME, Rovainen CM, Woolsey TA. Development and remodeling of cerebral blood vessels and their flow in postnatal mice observed with in vivo videomicroscopy. *J Cereb Blood Flow Metab*. 1992;12:935-946

41. Zhang L, Zhao Y, Zhou JW. The three-dimensional structure and the relationship between external and internal vascularizations in the brain of rat embryos. *Chin Med J (Engl)*. 2004;117:280-285

42. Pages G, Pouyssegur J. Transcriptional regulation of the vascular endothelial growth factor gene--a concert of activating factors. *Cardiovasc Res*. 2005;65:564-573

43. Vogel J, Gehrig M, Kuschinsky W, Marti HH. Massive inborn angiogenesis in the brain scarcely raises cerebral blood flow. *J Cereb Blood Flow Metab*. 2004;24:849-859

44. Heinzer S, Kuhn G, Krucker T, Meyer E, Ulmann-Schuler A, Stampanoni M, Gassmann M, Marti HH, Muller R, Vogel J. Novel three-dimensional analysis tool for vascular trees indicates complete micro-networks, not single capillaries, as the angiogenic endpoint in mice overexpressing human vegf(165) in the brain. *Neuroimage*. 2008;39:1549-1558

45. Ulmasov B, Bruno J, Gordon N, Hartnett ME, Edwards JC. Chloride intracellular channel protein-4 functions in angiogenesis by supporting acidification of vacuoles along the intracellular tubulogenic pathway. *Am J Pathol*. 2009;174:1084-1096

46. Zou F. Qtl mapping in intercross and backcross populations. *Methods Mol Biol*. 2009;573:157-173

47. Li R, Lyons MA, Wittenburg H, Paigen B, Churchill GA. Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics*. 2005;169:1699-1709

48. Manichaikul A, Broman KW. Binary trait mapping in experimental crosses with selective genotyping. *Genetics*. 2009;182:863-874

49. Da Costa e Silva L, Zeng ZB. Current progress on statistical methods for mapping quantitative trait loci from inbred line crosses. *J Biopharm Stat*. 2010;20:454-481

50. Banerjee S, Yandell BS, Yi N. Bayesian quantitative trait loci mapping for multiple traits. *Genetics*. 2008;179:2275-2289

51. Liu B. *Statistical genomics: Linkage mapping and qtl analysis.*: CRC Press; 1998.

52. Sax K. The association of size differences with seed-coat pattern and pigmentation in phaseolus vulgaris. *Genetics*. 1923;8:552-560

53. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*. 1989;121:185-199

54. Dempster A, Laird, NM, Rubin, DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat. Soc*. 1977:38

55. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*. 1992;69:315-324

56. Knapp SJ, Bridges WC. Using molecular markers to estimate quantitative trait locus parameters: Power and genetic variances for unreplicated and replicated progeny. *Genetics*. 1990;126:769-777

57. Martinez O, Curnow, RN. Estimating the locations and sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 1992;85:9

58. Doerge RW. Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet*. 2002;3:43-52

59. Zeng ZB. Precision mapping of quantitative trait loci. *Genetics*. 1994;136:1457-1468

60. Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008;9:855-867

61. Kao CH, Zeng ZB. General formulas for obtaining the mles and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the em algorithm. *Biometrics*. 1997;53:653-665

62. Sen S, Churchill GA. A statistical framework for quantitative trait mapping. *Genetics*. 2001;159:371-387

63. Broman KW, and T. P. Speed. A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. R. Stat. Soc*. 2002;64:16

64. Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics*. 2009;181:1077-1086

65. George E, McCulloch, RE. Variable selection via gibbs sampling. *J. Am. Stat. Asso.* 1993;88:9

66. Wang H, Zhang YM, Li X, Masinde GL, Mohan S, Baylink DJ, Xu S. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics*. 2005;170:465-480

67. Yi N, Shriner D. Advances in bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity*. 2008;100:240-252

68. Yi N, Xu S. Bayesian lasso for quantitative trait loci mapping. *Genetics*. 2008;179:1045-1055

69. Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics*. 2005;170:1333-1344

70. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics*. 1994;138:963-971

71. Mangin B, Goffinet B, Rebai A. Constructing confidence intervals for qtl location. *Genetics*. 1994;138:1301-1308

72. Manichaikul A, Dupuis J, Sen S, Broman KW. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics*. 2006;174:481-489

73. Tsaih SW, Korstanje R. Haplotype association mapping in mice. *Methods Mol Biol*. 2009;573:213-222

74. McClurg P, Pletcher MT, Wiltshire T, Su AI. Comparative analysis of haplotype association mapping algorithms. *BMC bioinformatics*. 2006;7:61

75. McClurg P, Janes J, Wu C, Delano DL, Walker JR, Batalov S, Takahashi JS, Shimomura K, Kohsaka A, Bass J, Wiltshire T, Su AI. Genomewide association analysis in diverse inbred mice: Power and population structure. *Genetics*. 2007;176:675-683

76. Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G. In silico mapping of complex disease-related traits in mice. *Science (New York, N.Y.)*. 2001;292:1915-1918

77. Chesler EJ, Rodriguez-Zas SL, Mogil JS. In silico mapping of mouse quantitative trait loci. *Science (New York, N.Y.)*. 2001;294:2423

78. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006;38:203-208

79. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics*. 2008;178:1709-1723

80. Lynch M, Ritland K. Estimation of pairwise relatedness with molecular markers. *Genetics*. 1999;152:1753-1766

81. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M. An arabidopsis example of association mapping in structured samples. *PLoS Genet*. 2007;3:e4

82. Liu P, Vikis H, Lu Y, Wang D, You M. Large-scale in silico mapping of complex quantitative traits in inbred mice. *PLoS One*. 2007;2:e651

83. Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang WP, He A, Kayne P, Gargalovic P, Kirchgessner T, Pan C, Castellani LW, Kostem E, Furlotte N, Drake TA, Eskin E, Lusis AJ. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res*. 2010;20:281-290

84. Ghazalpour A, Doss S, Kang H, Farber C, Wen PZ, Brozell A, Castellanos R, Eskin E, Smith DJ, Drake TA, Lusis AJ. High-resolution mapping of gene expression using association in an outbred mouse stock. *PLoS Genet*. 2008;4:e1000149

85. DiPetrillo K, Wang X, Stylianou IM, Paigen B. Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet*. 2005;21:683-692

86. Burgess-Herbert SL, Cox A, Tsaih SW, Paigen B. Practical applications of the bioinformatics toolbox for narrowing quantitative trait loci. *Genetics*. 2008;180:2227-2235

87. Burgess-Herbert SL, Tsaih SW, Stylianou IM, Walsh K, Cox AJ, Paigen B. An experimental assessment of in silico haplotype association mapping in laboratory mice. *BMC Genet*. 2009;10:81

88. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:1572-1577

89. Schaper W, Scholz D. Factors regulating arteriogenesis. *Arterioscler Thromb Vasc Biol*. 2003;23:1143-1151

90. de Koning DJ, Haley CS. Genetical genomics in humans and model organisms. *Trends in genetics : TIG*. 2005;21:377-381

91. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusis AJ. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature genetics*. 2005;37:710-717

92. Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh LV, de Haan G, Su AI, Jansen RC. Genetical genomics: Spotlight on qtl hotspots. *PLoS genetics*. 2008;4:e1000232

93. Mozhui K, Ciobanu DC, Schikorski T, Wang X, Lu L, Williams RW. Dissection of a qtl hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. *PLoS genetics*. 2008;4:e1000260

94. Sun W, Yu T, Li KC. Detection of eqtl modules mediated by activity levels of transcription factors. *Bioinformatics (Oxford, England)*. 2007;23:2290-2297

95. Wu C, Delano DL, Mitro N, Su SV, Janes J, McClurg P, Batalov S, Welch GL, Zhang J, Orth AP, Walker JR, Glynne RJ, Cooke MP, Takahashi JS, Shimomura K, Kohsaka A, Bass J, Saez E, Wiltshire T, Su AI. Gene set enrichment in eqtl data identifies novel annotations and pathway regulators. *PLoS Genet*. 2008;4:e1000070

96. Perez-Enciso M, Quevedo JR, Bahamonde A. Genetical genomics: Use all data. *BMC genomics*. 2007;8:69

97. Allan MF, Eisen EJ, Pomp D. Genomic mapping of direct and correlated responses to long-term selection for rapid growth rate in mice. *Genetics*. 2005;170:1863-1877

98. Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S, Allan MF, Pomp D, Schadt EE. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome biology*. 2009;10:R55

99. Allan MF, Eisen EJ, Pomp D. The m16 mouse: An outbred animal model of early onset polygenic obesity and diabesity. *Obesity research*. 2004;12:1397-1407

100. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature biotechnology*. 2001;19:342-347

101. Haley CS, Knott SA, Elsen JM. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics*. 1994;136:1195-1207

102. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. Genetics of gene expression surveyed in maize, mouse and man. *Nature*. 2003;422:297-302

103. Breiman L. *Bagging predictors [electronic resource]*. 1996.

104. Efron B. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 1979;Volume 7:1

105. Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol*. 2006;4:e395

106. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: Qtl mapping in experimental crosses. *Bioinformatics (Oxford, England)*. 2003;19:889-890

107. de la Fuente A, Bing N, Hoeschele I, Mendes P. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*. 2004;20:3565-3574

108. Roberts A, Pardo-Manuel de Villena F, Wang W, McMillan L, Threadgill DW. The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: Implications for qtl discovery and systems genetics. *Mamm Genome*. 2007;18:473-481

109. Chaibub Neto E, Ferrara CT, Attie AD, Yandell BS. Inferring causal phenotype networks from segregating populations. *Genetics*. 2008;179:1089-1100

110. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS computational biology*. 2007;3:e69

111. Valdar W, Holmes CC, Mott R, Flint J. Mapping in structured populations by resample model averaging. *Genetics*. 2009;182:1263-1277

112. Vera G, Jansen RC, Suppi RL. R/parallel--speeding up bioinformatics analysis with r. *BMC bioinformatics*. 2008;9:390

113. Schaper W. Collateral circulation: Past and present. *Basic research in cardiology*. 2009;104:5-21

114. Schirmer SH, van Nooijen FC, Piek JJ, van Royen N. Stimulation of collateral artery growth: Travelling further down the road to clinical application. *Heart*. 2009;95:191-197

115. Sherman JA, Hall A, Malenka DJ, De Muinck ED, Simons M. Humoral and cellular factors responsible for coronary collateral formation. *Am J Cardiol*. 2006;98:1194-1197

116. Abul-Khoudoud O. Diagnosis and risk assessment of lower extremity peripheral arterial disease. *J Endovasc Ther*. 2006;13 Suppl 2:II10-18

117. Bobek V, Taltynov O, Pinterova D, Kolostova K. Gene therapy of the ischemic lower limb--therapeutic angiogenesis. *Vascul Pharmacol*. 2006;44:395-405

118. Helisch A, Wagner S, Khan N, Drinane M, Wolfram S, Heil M, Ziegelhoeffer T, Brandt U, Pearlman JD, Swartz HM, Schaper W. Impact of mouse strain differences in innate hindlimb collateral vasculature. *Arterioscler Thromb Vasc Biol*. 2006;26:520-526

119. Zbinden S, Clavijo LC, Kantor B, Morsli H, Cortes GA, Andrews JA, Jang GJ, Burnett MS, Epstein SE. Interanimal variability in preexisting collaterals is a major factor determining outcome in experimental angiogenesis trials. *Am J Physiol Heart Circ Physiol*. 2007;292:H1891-1897

120. Yao H, Cui ZH, Masuda J, Nabika T. Congenic removal of a qtl for blood pressure attenuates infarct size produced by middle cerebral artery occlusion in hypertensive rats. *Physiol Genomics*. 2007;30:69-73

121. Campen MJ, Tagaito Y, Jenkins TP, Balbir A, O'Donnell CP. Heart rate variability responses to hypoxic and hypercapnic exposures in different mouse strains. *J Appl Physiol*. 2005;99:807-813

122. Deschepper CF, Olson JL, Otis M, Gallo-Payet N. Characterization of blood pressure and morphological traits in cardiovascular-related organs in 13 different inbred mouse strains. *J Appl Physiol*. 2004;97:369-376

123. Ryan MJ, Didion SP, Davis DR, Faraci FM, Sigmund CD. Endothelial dysfunction and blood pressure variability in selected inbred mouse strains. *Arterioscler Thromb Vasc Biol*. 2002;22:42-48

124. Zuurbier CJ, Emons VM, Ince C. Hemodynamics of anesthetized ventilated mouse models: Aspects of anesthetics, fluid support, and strain. *Am J Physiol Heart Circ Physiol*. 2002;282:H2099-2105

125. Vasquez-Pinto LM, Landgraf RG, Bozza PT, Jancar S. High vascular endothelial growth factor levels in nzw mice do not correlate with collagen deposition in allergic asthma. *Int Arch Allergy Immunol*. 2007;142:19-27

126. Ward NL, Moore E, Noon K, Spassil N, Keenan E, Ivanco TL, LaManna JC. Cerebral angiogenic factors, angiogenesis, and physiological response to chronic hypoxia differ among four commonly used mouse strains. *J Appl Physiol*. 2007;102:1927-1935

127. van Weel V, Toes RE, Seghers L, Deckers MM, de Vries MR, Eilers PH, Sipkens J, Schepers A, Eefting D, van Hinsbergh VW, van Bockel JH, Quax PH. Natural killer cells and cd4+ t-cells modulate collateral artery development. *Arterioscler Thromb Vasc Biol*. 2007;27:2310-2318

128. Brown MG, Scalzo AA. Nk gene complex dynamics and selection for nk cell receptors. *Semin Immunol*. 2008;20:361-368

129. Taherzadeh Z, VanBavel E, de Vos J, Matlung HL, van Montfrans G, Brewster LM, Seghers L, Quax PH, Bakker EN. Strain-dependent susceptibility for hypertension in mice resides in the natural killer gene complex. *Am J Physiol Heart Circ Physiol*. 2010;298:H1273-1282

130. Hoover-Plow J, Shchurin A, Hart E, Sha J, Hill AE, Singer JB, Nadeau JH. Genetic background determines response to hemostasis and thrombosis. *BMC Blood Disord*. 2006;6:6

131. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26:317-325

132. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hossein A, Fare T. Multiplexed measurements of gene signatures in different analytes using the nanostring ncounter assay system. *BMC Res Notes*. 2009;2:80

133. Paul WE, Zhu J. How are t(h)2-type immune responses initiated and amplified? *Nat Rev Immunol*. 2010;10:225-235

134. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL. Two types of murine helper t cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. 1986. *J Immunol*. 2005;175:5-14

135. Barner M, Mohrs M, Brombacher F, Kopf M. Differences between il-4r alpha-deficient and il-4-deficient mice reveal a role for il-13 in the regulation of th2 responses. *Curr Biol*. 1998;8:669-672

136. Faffe DS, Flynt L, Bourgeois K, Panettieri RA, Jr., Shore SA. Interleukin-13 and interleukin-4 induce vascular endothelial growth factor release from airway smooth muscle cells: Role of vascular endothelial growth factor genotype. *Am J Respir Cell Mol Biol*. 2006;34:213-218

137. Yamaji-Kegan K, Su Q, Angelini DJ, Johns RA. Il-4 is proangiogenic in the lung under hypoxic conditions. *J Immunol*. 2009;182:5469-5476

138. Brenner GJ, Cohen N, Moynihan JA. Similar immune response to nonlethal infection with herpes simplex virus-1 in sensitive (balb/c) and resistant (c57bl/6) strains of mice. *Cell Immunol*. 1994;157:510-524

139. Cook G, Dumbar M, Franklin IM. The role of adhesion molecules in multiple myeloma. *Acta Haematol*. 1997;97:81-89

140. Hu X, Wohler JE, Dugger KJ, Barnum SR. Beta2-integrins in demyelinating disease: Not adhering to the paradigm. *J Leukoc Biol*. 2010;87:397-403

141. Mace EM, Monkley SJ, Critchley DR, Takei F. A dual role for talin in nk cell cytotoxicity: Activation of lfa-1-mediated cell adhesion and polarization of nk cells. *J Immunol*. 2009;182:948-956

142. Arumugam TV, Salter JW, Chidlow JH, Ballantyne CM, Kevil CG, Granger DN. Contributions of lfa-1 and mac-1 to brain injury and microvascular dysfunction induced by transient middle cerebral artery occlusion. *Am J Physiol Heart Circ Physiol*. 2004;287:H2555-2560

143. Ding Y, Li J, Rafols JA, Phillis JW, Diaz FG. Prereperfusion saline infusion into ischemic territory reduces inflammatory injury after transient middle cerebral artery occlusion in rats. *Stroke*. 2002;33:2492-2498

144. Gidday JM, Gasche YG, Copin JC, Shah AR, Perez RS, Shapiro SD, Chan PH, Park TS. Leukocyte-derived matrix metalloproteinase-9 mediates blood-brain barrier breakdown and is proinflammatory after transient focal cerebral ischemia. *Am J Physiol Heart Circ Physiol*. 2005;289:H558-568

145. Laragh JH. Renin-angiotensin-aldosterone system for blood pressure and electrolyte homeostasis and its involvement in hypertension, in congestive heart failure and in associated cardiovascular damage (myocardial infarction and stroke). *J Hum Hypertens*. 1995;9:385-390

146. Loftus JP, Johnson PJ, Belknap JK, Pettigrew A, Black SJ. Leukocyte-derived and endogenous matrix metalloproteinases in the lamellae of horses with naturally acquired and experimentally induced laminitis. *Vet Immunol Immunopathol*. 2009;129:221-230

147. Carmichael ST. Rodent models of focal stroke: Size, mechanism, and purpose. *NeuroRx*. 2005;2:396-409

148. Mohr JP. *Stroke : Pathophysiology, diagnosis, and management*. New York: Churchill Livingstone; 2004.

149. Ishida T, Kundu RK, Yang E, Hirata K, Ho YD, Quertermous T. Targeted disruption of endothelial cell-selective adhesion molecule inhibits angiogenic processes in vitro and in vivo. *J Biol Chem*. 2003;278:34598-34604

150. Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, Han B, Furlotte N, Kang EY, Rivas M, Bogue MA, Frazer KA, Johnson FM, Beilharz EJ, Cox DR, Eskin E, Daly MJ. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics*. 2010;185:1081-1095

151. Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. On the subspecific origin of the laboratory mouse. *Nature genetics*. 2007;39:1100-1107

152. Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, Pethiyagoda CL, Stuve LL, Johnson FM, Daly MJ, Wade CM, Cox DR. A sequence-based variation map of 8.27 million snps in inbred mouse strains. *Nature*. 2007;448:1050-1053

153. Palin K, Taipale J, Ukkonen E. Locating potential enhancer elements by comparative genomics using the eel software. *Nat Protoc*. 2006;1:368-374

154. Schug J. Using tess to predict transcription factor binding sites in DNA sequence. *Curr Protoc Bioinformatics*. 2008;Chapter 2:Unit 2 6

155. Cai W, Schaper W. Mechanisms of arteriogenesis. *Acta Biochim Biophys Sin (Shanghai)*. 2008;40:681-692

156. van Royen N, Piek JJ, Schaper W, Fulton WF. A critical review of clinical arteriogenesis research. *J Am Coll Cardiol*. 2009;55:17-25

157. Blatnik JS, Schmid-Schonbein GW, Sung LA. The influence of fluid shear stress on the remodeling of the embryonic primary capillary plexus. *Biomech Model Mechanobiol*. 2005;4:211-220

158. Messina LM. Elucidating the genetic basis of peripheral arterial disease: Identification of a quantitative trait locus that determines the phenotypic response to experimental hindlimb ischemia. *Circulation*. 2008;117:1127-1129

159. Wade CM, Daly MJ. Genetic variation in laboratory mice. *Nature genetics*. 2005;37:1175-1180

160. Tymms M, Kola I. *Gene knockout protocols*. Humana Pr Inc; 2001.

161. Dai X, Faber JE. Endothelial nitric oxide synthase deficiency causes collateral vessel rarefaction and impairs activation of a cell cycle gene network during arteriogenesis. *Circ Res*. 2010;106:1870-1881

162. Chesler EJ, Miller DR, Branstetter LR, Galloway LD, Jackson BL, Philip VM, Voy BH, Culiat CT, Threadgill DW, Williams RW, Churchill GA, Johnson DK, Manly KF. The collaborative cross at oak ridge national laboratory: Developing a powerful resource for systems genetics. *Mamm Genome*. 2008;19:382-389

163. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, et al. The collaborative cross, a community resource for the genetic analysis of complex traits. *Nat Genet*. 2004;36:1133-1137

164. Threadgill DW, Hunter KW, Williams RW. Genetic dissection of complex and quantitative traits: From fantasy to reality via a community effort. *Mamm Genome*. 2002;13:175-178

165. Broman KW. The genomes of recombinant inbred lines. *Genetics*. 2005;169:1133-1146

166. Valdar W, Flint J, Mott R. Simulating the collaborative cross: Power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics*. 2006;172:1783-1797

167. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*. 2008;6:e107

168. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet*. 2009;10:23

169. Wade CM, Kulbokas EJ, 3rd, Kirby AW, Zody MC, Mullikin JC, Lander ES, Lindblad-Toh K, Daly MJ. The mosaic structure of variation in the laboratory mouse genome. *Nature*. 2002;420:574-578

170. Buschmann IR, Busch HJ, Mies G, Hossmann KA. Therapeutic induction of arteriogenesis in hypoperfused rat brain via granulocyte-macrophage colony-stimulating factor. *Circulation*. 2003;108:610-615