

Gene Entrapment Strategy for Functional Genomics Study

林 卿

(Dept. of Microbiology and Immunology, Vanderbilt University
School of Medicine, Nashville, TN 37232, USA)

提纲 (Outline)

- 研究基因功能的方法
- **Gene Trap** 技术在功能基因组研究中的作用
- **Gene Trap** 的基本原理及研究概况
- 本人在 **Dr. Ruley's Lab** 所做的工作
- 胚胎干细胞 **Gene Trap** 技术平台的意义与展望
- 回国研究计划

This proposal containing unpublished data, so please keep the enclosed information confidential until we publish them.

随着人类基因组计划（**Human Genome Project**）以及哺乳动物基因组计划（**Mammalian Genome Project**）的展开，大规模的基因组测序及拼接排序工作都已经接近尾声，我们已经进入一个后基因组时代（**Post-Genomic Era**）。这些计划的完成不但为生命科学研究提供了新的思路、策略和技术方法，而且促进了其他学科，特别是生物医药、化学、信息科学的发展。随着基因组计划的完成，功能基因组计划（**Functional Genomics Study**）就受到了广泛的重视，世界各国纷纷启动了各种功能基因组研究计划，如蛋白质组计划、结构基因组计划和结构蛋白质组计划等。虽然人们在基因组 **DNA** 水平上的认识已经取得了长足的进步，但是最终还是要致力于探索这些基因的结构和功能，只有这样才能有效地利用研究成果来为人类的疾病诊断和防治服务。

➤ 研究基因功能的方法

1. 遗传分析（**Genetic Analysis**）是传统的研究基因功能的方法。通过表型来寻找引起功能异常的突变，但是寻找点突变无异于大海捞针，非常困难。

2. 正向遗传学（**Forward Genetics**）方法 用物理方法或化学诱变剂造成的许多突变，通过表型来研究基因功能，该方法局限在于涉及多点、多个基因的变化，而且很难确定是哪个基因在起作用。

3. 反向遗传学（**Reverse Genetics**）方法 转基因（**Transgene**）技术出现在八十年代初，有效但是稳定性较差，效率不是很高。随着现代分子生物学技术的不断提高和小鼠胚胎干细胞（**Mouse Embryonic Stem Cell**）技术的成熟，八十年代末发展了革命性的基因打靶（**Gene Targeting**）技术。基因打靶技术是一个从分子水平来研究基因功能的有效途径，但是该技术需要对靶基因的结构和功能有相当深的了解，才能设计出可行的同源重组方案。**Gene Trap**（国内有人把它翻译成基因捕获，但是不贴切，所以就用英文）方法几乎是和基因打靶技术同时出现的革新，其基本原理是通过逆转录病毒或其它方法将 **Gene Trap** 载体随机整合到基因组 **DNA** 中，从而使所在位置基因的表达被阻断（见图一），相当于该基因被 **Knockout** 或 **Knockdown**。最近的一些进展如 **Cre-loxp** 和 **Flpe-FRT** 重组系统的运用，在 **Gene Trap** 载体中增加各种易于操作的重组酶介导的盒式交换（**Recombinase Mediated Cassette Exchange, RMCE**），使得蛋白组学研

究 (Proteomics Study) 和 Gene Trap 技术有机地结合在一起, Gene Trap 方法在功能基因组研究中的作用也显得日益突出。

➤ Gene Trap 技术在功能基因组研究中的作用

当前, 人类基因组研究的重心正在由结构向功能转移, 一个以基因组功能研究为主要内容的所谓后基因组时代 (Post-Genomic Era), 也即功能基因组 (Functional Genomics) 时代已经来到。如何获取基因的功能信息, 即与人类重大疾病和重要生理功能相关的基因信息, 就摆在了我们面前。功能基因组计划引起各国政府、科研机构 and 大型制药公司的高度重视, 美国以及欧洲的一些主要国家相继开展一系列的研究工作并取得了极大进展。由于 Gene Trap 技术的实用性和有效性, 并且适合于大规模小鼠胚胎干细胞 Knock-Out 库的构建, 美国以及欧洲的一些主要国家相继建立了 Gene Trap 胚胎干细胞库 (ES Libraries)。Dr. H. Earl Ruley 教授 (我现在所在实验室的老板, Vanderbilt University, Dept. of Microbiology and Immunology) 是这个领域的先驱人物之一, 他在 Gene Trap 方面的研究已经有将近十五年的历史, 从他的实验室培养出 Geoff Hicks (加拿大 University of Manitoba, 现在国际 Gene Trap 协会 (the International Gene Trap Consortium, IGTC) 的发起人之一), Harald von Melchner (德国 Gene Trap 项目的负责人) 等杰出人才。表一列出了目前 IGTC 的成员的信息。

从目前国际上 Gene Trap 的发展动态来看, 美国绝对领先, 美国国立卫生研究院 (NIH) 每年要投入许多经费, 而且大型的制药公司如 Lexicon 也投入巨资建立了世界上最大的胚胎干细胞突变库, 而且号称覆盖了 60% 的小鼠基因。该公司目前提供有偿服务, 每个胚胎干细胞大约为 \$50,000 美元, 还有很多的限制条款。欧洲和加拿大都有政府部门支助的大项目, 亚洲只有日本开始建立 Gene Trap 胚胎干细胞库, 但是日本的数据是不公开的。IGTC 是国际小鼠突变协会 (the International Mouse Mutagenesis Consortium) 的一部分, 有美国, 加拿大等国家的一些骨干实验室组成。IGTC 实现了多个研究机构间胚胎干细胞库的资源共享, 目前总共覆盖了近 30% 的小鼠基因, 目标是覆盖整个小鼠基因组。也许因为技术层次的要求极高, 据我所知, 中国目前没有一家实验室能够建立 Gene

Trap 胚胎干细胞库。

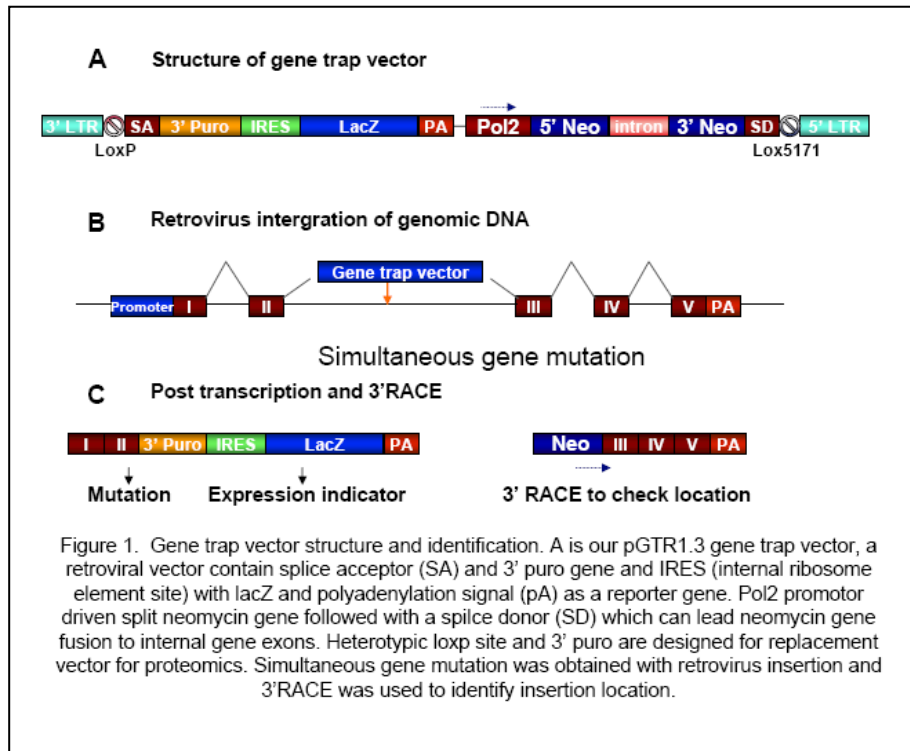
Member	Institute
Bruce R. Conklin	Gladstone Institute of Cardiovascular Disease, University of California San Francisco
Lesley Forrester	John Hughes Bennett Laboratory, Department of Oncology, Western General Hospital
Geoff Hicks	Manitoba Institute of Cell Biology, University of Manitoba
Andras Nagy	Samuel Lunenfeld Research Institute, Mount Sinai Hospital
Barry Rosen	Wellcome Trust Sanger Institute
Patricia Ruiz	Max-Planck Institute for Molecular Genetics, Center for Cardiovascular Research
Earl Ruley	Microbiology and Immunology, School of Medicine, Vanderbilt University
Bill Skarnes	Wellcome Trust Sanger Institute
Philippe Soriano	Program in Developmental Biology, Fred Hutchinson Cancer Research Center
Bill Stanford	Institute of Biomaterials & Biomedical Engineering, Institute of Medical Science, University of Toronto, Samuel Lunenfeld Research Institute, Gene Trap Project, Centre for Modeling Human Disease
Marc Tessier- Lavigne	Senior Vice President, Research Drug Discovery, Genentech, Inc.
Harald von Melchner	Laboratory for Molecular Hematology, University of Frankfurt Medical School
Wolfgang Wurst	GSF Research Center, for Environment and Health, Institute for Developmental Genetics
Ken-Ichi Yamamura	Institute of Molecular Embryology & Genetics Kumamoto University
Stephen G. Young	Gladstone Institute of Cardiovascular Disease, University of California San Francisco

Table 1: Members of IGTC (the International Gene Trap Consortium)

➤ Gene Trap 的基本原理及研究概况

下面简单介绍一下 Gene Trap 的基本原理及目前研究状况。通常来讲，真核生物的基因结构主要有启动子(Promoter)，外显子(Exon)，内含子(Intron)以及 Poly(A) 尾

巴(如图一)。所谓的 Gene Trap，就是通过逆转录病毒将载体随机地插入到基因组 DNA 中，然后经过特定的药物筛选并挑取存活的胚胎干



细胞。一份冻存到液氮里，另一份胚胎干细胞抽提 RNA，然后再利用 3'RACE 或其他方法确定插入位点。Gene Trap 的技术平台到此基本完成，接下去的工作是将这些克隆得到的测序片段进行 Blast Search，整理 Gene Trap 的数据库并且和液氮罐冻存的 ES 细胞互相对应归类。进一步的工作就可以在这个 GeneTrap 胚胎干细胞库的基础上开展。具体流程见图二。

我老板 Dr. H. Earl Ruley 教授 (Vanderbilt University) 在这个领域已经有十五年的研究经历，也培养出国际 Gene Trap 协会的另两个重要成员。本实验室的 Gene Trap 载体经历了几个版本的重大改进，最早的工作有：

1: U3Neo (By Dr. Geoff Hicks) 是比较成功的系列。这种载体属于 Promoter Trap，只能筛选到胚胎干细胞 (ES) 中表达的基因，而且不能进行基因组 DNA 水平的操作，现在已经很少再用它。

2: LANPT1 (By Dr. Anna Ospovich) 是个 Poly(A) Trap，但是这个 ES 突变

库比较小，而且没有一个 ES 细胞能进 Germline Transmission，其中问题也比较多。

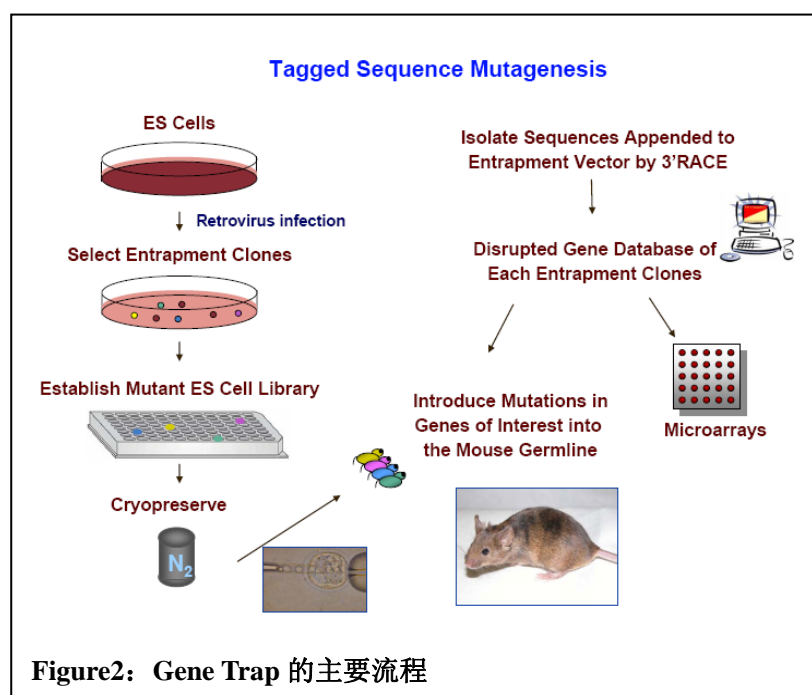
3: pGTR1.x (By Dr. 林卿) 系列载体，目前最新最成功的载体系列。

➤ 本人在 Dr. Ruley's Lab 所做的工作

pGTR1.x

(By Dr. 林卿)

系列载体，在原有失败和成功的经验教训的基础上，加入了大量新的用途和特色（下面会有介绍，主要是用于蛋白组研究）。理论上可以 Trap 所有的基因（包括



ES 细胞中不表达的基因)。是目前最先进的 Gene Trap 载体之一。作为该系列载体主要的设计者和完成者，我在此项研究中倾注了大量的心血。该载体已经申请专利，我也是专利持有人之一。目前有两篇文章（一篇在 *Science*，另一篇在 *Nature Methods*）正在审稿中，希望有好结果。至今我们已经拥有 1000 多个确定的 ES 克隆。我们从中挑选了 10 种不同基因的 ES 细胞突变体注射到小鼠 Blastocyst 中，成功地获得了所有的嵌合体 (Chimeric mice)，F1 子代的 Germline Transmission 为 100%! 日前这十个 Knockout 老鼠系列已经有五个发现有明显的表型，我正对其中的三个老鼠 (Cradd, IL-8Ra and Spink4) 进行详细研究，其结果非常振奋人心，预计会有更多的好文章。Gene trap 项目是一个很大的项目，需要大量的投资，而且胚胎干细胞库的数量也是越大越好。美国 NIH 每年的资助十分可观，Dr. H. Earl Ruley 教授就有一个为期五年近千万美元的项目，其它国家如加拿大、德国、法国等都有政府机构的大量投资。中国这方面的研究几

乎等于零，国内的 **Gene Knockout** 也是尚属起步阶段，且费用很高。如果浙江大学生命科学学院能及时建立胚胎干细胞 **Gene Trap** 技术平台，就可以从战略和技术上占领先机，抓住这一国际前沿领域的发展机遇，及时发展和提高功能基因组研究技术和疾病相关基因研究的整体水平和综合实力，促进我国功能基因组和疾病相关基因研究的跨越式发展。

ES Clone	Gene	Description	Effect on Expression	Phenotype
B3PP3G6	CRADD	CASP2 and RIPk1 domain containing adaptor with death domain	Apparent Null	Yes
B3P3H4	G3BP2	Ras GAP SH3 domain binding protein/contains RNA binding motif (RRM)	ND	ND
B3P3G11	5330402421RIK	Similar to tetratricopeptide repeat domain 18	ND	ND
B3P1H1	Dymeclin	Mouse ortholog of Dymeclin--gene responsible for Dyggve-Melchior-Clausen syndrome	Apparent Null	Skeletal defects similar to human disease syndrome
B3P1B12	V-preb3	Pre B lymphocyte gene 3, Ig domain similar to lambda light chain VLJ region	ND	ND
B3P3D4	IL-8 R like gene	similar to CXC cytokine receptor-1; similar to (63%) and located adjacent to the IL-8 type B receptor	Apparent Null	Yes
B2P1B10	1700022N24Rik	Domains:RING-finger-containing E3 ubiquitin ligase	ND	ND
B3P1A6	Pfdn1	Prefoldin 1, Domains: KE2 family protein.	Apparent Null	Post natal lethal
B2P1B9	Spink-4	Serine protease inhibitor, Kazal type 4	ND	ND
B3P3C8	Hesx-1	ES cell specific homeobox protein	Apparent Null	Forebrain defects and anophthalmia

Table 2. Germline transmission of genes disrupted by pGTR1.3 Analysis of these mutations is currently in progress. ND = not determined.

我在 Dr. Ruley Lab 所做的另一个主要课题是 Cell-Permeable Cre

Recombinase。目前已经发表三篇文章，分别在 *Nature Biotechnology*, *JCB* 和 *BMC Biotechnology* 上。Cre 重组酶是目前在功能基因组研究上用途最广泛的工具酶，我们研制了可以直接穿过细胞膜的 Cre，并且申请了专利，我也是专利享有人之一，该酶已经可以商品化。有了 Cell-Permeable Cre，我们不必再转染细胞，直接往培养基中加 Cre 就可以达到目的，同时也可以向小鼠直接注射 Cre，同样是有用的，具体数据这里不再详细介绍。Cell-Permeable Cre 在 Post-Entrapment 对基因组操作有极大的实用价值，特别是对重组酶介导的盒式交换（Recombinase Mediated Cassette Exchange, RMCE）中有重大意义，为蛋白组的研究提供极大的便利条件。我另外还有一篇有关 Cre/Loxp 技术及应用的章节（Book Chapter）：“DNA site-specific recombinase Cre mediated recombination and application” (In Press). ***Peptide Mediated Transduction Characterization, Optomization and Application***, Edited by Paul Robbins and Steve Dowdy, Kluwer press.

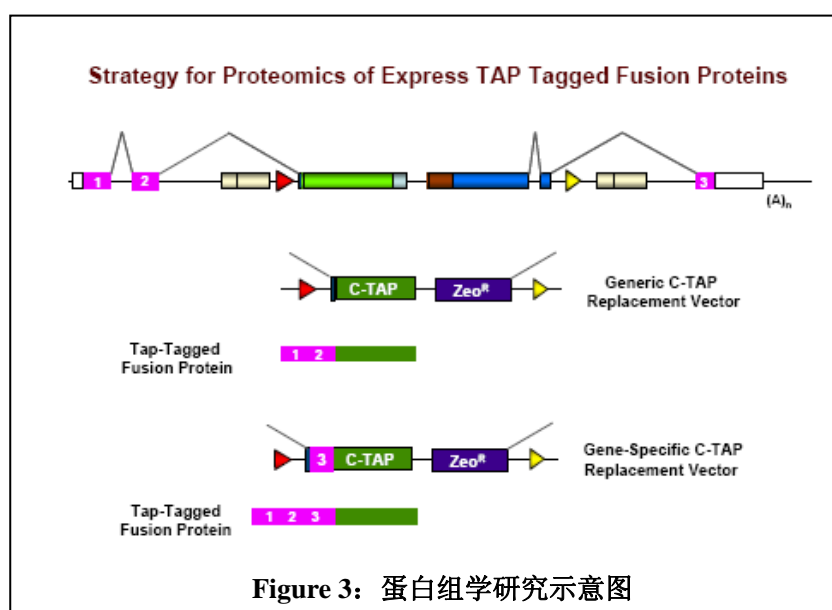
➤ 胚胎干细胞 Gene Trap 技术平台的意义与展望

1. 利用该载体构建 Gene Trap 胚胎干细胞库，建立小鼠胚胎干细胞的 Knock-Out 资源库，并在此基础上提供一个研究基因结构和功能的技术平台。为了有效地利用有限的科研经费，该技术平台可以实现中国小鼠 Knockout 资源库的共享，能够极大地提高国内基因结构和功能研究的水平。而且从费用上来讲也是非常经济的，因为不再需要各个实验室重复构建 Knockout 胚胎干细胞株，而这一过程大概需要 6—18 个月的时间，视库容量的大小而变。根据我的经验，我在大约半年时间中，构建了 1000 多个胚胎干细胞的 Gene Trap 库，耗材和测序的费用大概是 \$50,000 美元。平均每个 Knockout 胚胎干细胞细胞株的费用为 \$50。表三列举了基因打靶和 Gene Trap 的一些比较。

表三：基因打靶和 Gene Trap 一些比较

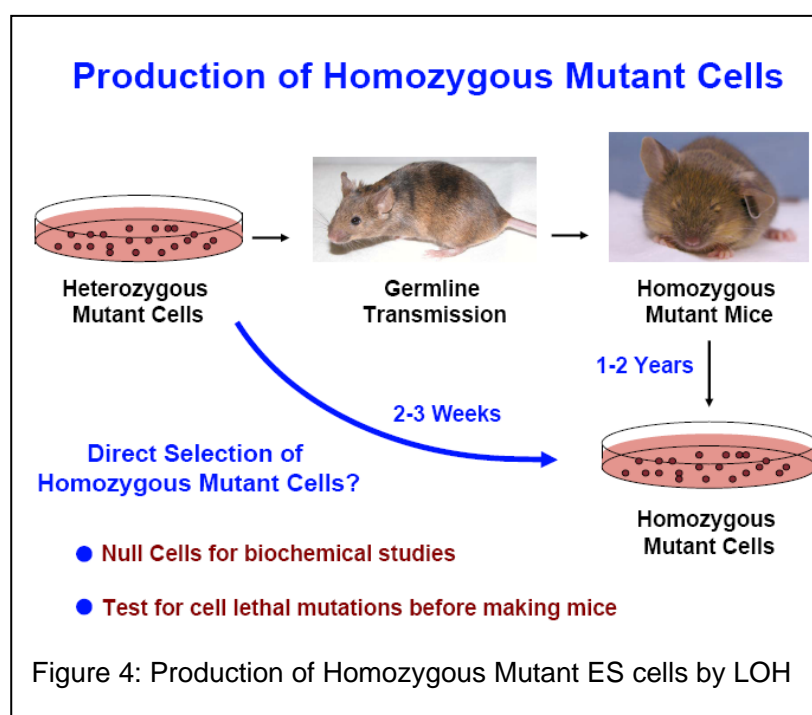
	Gene Trap	Gene Targeting
平均经费	\$50/Per ES Cell	\$3,000-10,000
时间	Large scale, 1,000-3,000/ 6months	6-18 months
成功率	70-100%	60-100%
纯合体小鼠	6-12 months	6-12months

2. 蛋白组学的研究：PGTR1.3 系列载体的一个重要功能是在引入了两个不同的 *Loxp* 位点，通过重组酶介导的盒式交换 (Recombination Mediated



Cassette Exchange, RMCE) 我们可以很容易地对 Gene trap 载体附近的 Genomic DNA 进行操作。许多具体的应用我不多列，下面举例说明 Gene trap 载体在蛋白组学研究中的应用。研究基因的功能，最终还是要回到蛋白-蛋白相互作用上，Gene Trap 载体插入到基因的中间，如果我们将 C-TAP 载体置换到那个位置，就可以形成一个蛋白质复合物，利用 Tandem Affinity Purification (TAP) 方法，纯化整个蛋白复合物，通过液相质谱分析，我们就可以研究哪些蛋白质与之相互作用。这个方法的优点是该基因是在本身启动子 (Native Promoter) 引导下的基因表达，与正常生理情况下该基因的表达完全一致，所以得到的结果更能体现体内环境下蛋白-蛋白相互作用的情况。

3. 基础医学研究: 我们知道动物模型最终是要服务人类, 即为探索人类疾病的发生机制及防治措施做贡献。目前的 Gene Trap 都是基于小鼠胚胎



干细胞的, 但是我们实验室所用的逆转录病毒载体的包装系统是可以用于其它的哺乳动物的, 也就是说适用于人类的胚胎干细胞。目前对人的胚胎干细胞进行 Knockout, 只能剔除其中的一个等位基因, Gene Trap 也有同样的问题。pGTR1.3 的一个非常重要的功能是能够形成同源突变的胚胎干细胞。其原理是利用杂合性缺失 (Loss of Heterozygosity, LOH), 在进行高浓度的药物筛选时可以形成同源突变, 也就是可以剔除两个等位基因。对大多数基因, 我们并非一定要有 Knockout 老鼠, 其实只要有双剔除 (Double Knockout) 的细胞株 (Cell Line) 就可以进行生物化学的研究。而通常情况下, 获得 Double Knockout Cell Line 的途径一般要通过 Double Knockout 小鼠, 从 ES 细胞株到 Double Knockout 小鼠至少需要一到两年, 费时、费力、费钱, 而通过我们的胚胎干细胞库, 只需要两到三周就可以解决问题。胚胎干细胞 Gene Trap 技术平台完全可以满足这些要求, 这是一条非常简捷经济的途径。利用这一个技术平台, 将为我国临床医学研究, 特别是疾病相关基因机理的研究提供有用的资源宝库。

4. Post-Entrapment 基因功能的研究, 包括 Microarray 基础上的发现新基因 (Gene Discovery)、基因表达的研究; Conditional Knockout; 染色体的剔

除、置换和加工等；这里不再详细叙述。

5. 肿瘤研究上的应用：在癌症研究中，杂合性缺失（Loss of heterozygosity, LOH）是一种常用的等位基因缺失检测方法，广泛用于候选抑癌基因的筛选及已知抑癌基因失活机制的阐明等方面。等位基因失衡（Allelic Imbalance）以及杂合性缺失(LOH)是肿瘤细胞染色体不稳定的主要表现形式，研究数据表明恶性肿瘤细胞中有 10,000 多个 LOH 多发区域。等位基因异常在多种肿瘤中被证明是肿瘤发生早期事件，因此杂合性缺失可能成为肿瘤筛查和早期诊断的重要工具。食管癌、结肠癌是世界上最常见的恶性肿瘤之一，近年来对食管癌、结肠癌分子遗传基础研究取得了相当大的进展，但其发生发展的分子机制包括染色体变更(Chromosome Alteration)仍不十分清楚。有关结肠癌研究的数据显示，相比较于已发生浸润的肿瘤组织而言，大量的 LOH 发生于癌前病变组织，这表明结肠癌的病程进展并不需要 LOH。由于癌变早期实际上是肿瘤前体干细胞的扩张，那么这些前体干细胞可能接触到一些致癌物，从而引起遗传损伤的逐渐加深。那么这些致癌物究竟起到什么样的作用，究竟有多大的能力引起 LOH，我们可以利用 Gene Trap 胚胎干细胞库来技术平台来进行研究。我们最近的结果提示，诱变剂例如 Methylnitrosourea (MNU), Diepoxybutane, Mitomycin C 和 Hydroxyurea 所引起的 LOH 的频率大概为 $1-8 \times 10^{-3}$ /cell，远远高于这些诱变引起的点突变的频率。这表明象 MNU 之类的诱变剂除了引起点突变外，还会引起杂合性缺失，从而更快地显出表型。胚胎干细胞库 Gene Trap 技术平台一方面可以帮助研究化合物是否具有在整个基因组范围内诱发 LOH 的能力，另一方面还可以研究染色体上致癌物诱发 LOH 的热点，这对于人类癌变相关基因的研究及防治上具有重大意义。

➤ 回国研究计划

我国于九十年代就积极参与了人类基因组研究计划，并取得了很大的成果。国家已经将功能基因组研究，特别是发现具有疾病相关的功能基因和农业应用前景的功能基因及其调控网络，确定为我国战略需求和重要研究方向。发展用于基

因功能研究的新技术、建立技术平台是功能基因组研究的重要任务，为此国家自然科学基金委和科技部等单位已经启动了许多功能基因组研究计划，如蛋白质组计划和结构基因组计划。

我希望回国为浙江大学服务，建立小鼠胚胎干细胞 **Gene Trap** 技术平台，并发展相关的蛋白组学研究技术。为我国功能基因的遗传学研究的跨越式发展奠定坚实的基础，而且还可以给 **IGTC** 有益的补充。此举对于国内的基因功能研究具有重大的意义，因为这个 **Gene Trap** 小鼠胚胎干细胞技术平台，可以给国内外同行提供各种 **Knockout** 胚胎干细胞，直接注射到小鼠里就可以获得 **Knockout** 小鼠，而不必花费极大的人力物力去构建 **Knockout** 胚胎干细胞，国内基因功能的研究就可以上一个新台阶。基于中国人类胚胎干细胞的研究状况，我们还可以着手构建中国人类胚胎干细胞 **Gene Trap** 库，就可以更加紧密地同基础医学以及疾病相关基因的功能研究联系在一起。**Knockout** 人类胚胎干细胞对于疾病的发生机制及其相应的防治和机理研究都具有重大的意义和具有广泛和深远的临床应用前景。这是领先于其他国家的一个大好机会，我们必须抓住这一国际前沿领域的发展机遇。