INTRODUCTION. SH2 domains are regarded as the classic archetype of a large family of interaction domains that control many aspects of signal transduction and cellular regulation through their ability to bind variously to protein, phospholipids, nucleic acids, or small molecule ligands [1]. SH2 domains form a critical link between protein tyrosine kinases and downstream signaling through their ability to bind to phosphorylated tyrosine residues in the context of peptide sequences that provide specificity [2]. The data from the human genome project provides the opportunity to identify the complete human complement of proteins containing recognizable SH2 domains. We have mined publicly available human & mouse genetic sequence data for sequences that contain SH2 domains.

METHOD. Multiple approaches were combined to develop a comprehensive strategy to capture sequences containing putative SH2 domains. A set of human proteins identified by the simple modular architecture research tool (SMART [3]) within the Swiss-Prot & TrEMBL databases as containing SH2 domains was combined with a set GRAIL-predicted [4] open reading frames identified as having SH2 domains using the Pfam or SMART domain descriptions. Both SMART and Pfam domain descriptions were additionally used to search the non-redundant sequence set at NCBI. The combined set contained a significant degree of sequence duplication and redundancy. Thus, the overpopulated domain set was filtered first for identity, and then for proteins identified as having identical genetic loci or representing sequence polymorphisms.
**DISCUSSION.** After removal of duplicates, splice variants and pseudogenes, this afforded a total of 115 SH2 domains contained in 105 distinct proteins. This represents a nearly comprehensive census and early bioinformatic view of SH2 domain containing proteins in the human genome. Comparative evaluation of the sequence and domain organization of these proteins provides contextual information that enhances understanding of individual family members and a framework for the further characterization of this important class of proteins.

The 115 SH2 domains were organized into 11 different functional categories. Many of the SH2 domain-containing proteins have been genetically disrupted in mice. Several knockout mice led to embryonic lethality as a large number were found with defects in lymphocyte development. In addition, a number of SH2-domain-containing proteins are linked to human diseases. The SH2 domain proteins, binding partners, roles in cell signaling, homeostasis and human disease will be discussed.

**ACKNOWLEDGMENT.**
The work was supported by the University of Chicago Cancer Research Center, Argonne National Laboratory, and the Canadian Institutes for Health Research.

**REFERENCES.**